English Language and Linguistics, 29.1: 1–34.
doi:10.1017/S1360674323000679

'When intuitions (don't) fail': combining syntax and sociolinguistics in the analysis of Scots[1]

E JAMIESON
University of York

JENNIFER SMITH
University of Glasgow

DAVID ADGER
Queen Mary University of London

CAROLINE HEYCOCK
University of Edinburgh

and

GARY THOMS
New York University

(Received 20 December 2022; revised 12 December 2023)

A perennial problem for sociolinguists interested in morphosyntactic variation is that such forms are often low frequency, making quantitative analysis difficult or impossible. However, sociolinguists have been generally reluctant to adopt methodologies from syntax, such as acceptability data gleaned from speaker intuition, due to the belief that these judgments are not necessarily reliable. In this article we present data from the *Scots Syntax Atlas*, which employs sociolinguistic methodologies in spoken data alongside the results of acceptability judgments. We target three morphosyntactic variables and compare and contrast these across the two data types in order to assess the reliability of the judgment data at community level. The results show that reliability is variable-dependent. For some variables, there is clear correlation; with others, it appears that, as Labov (1996) phrased it, 'intuitions fail'. We discuss how factors such as salience, social stigma and local identity combine to govern the reliability of judgment data.

Keywords: acceptability judgments, sociolinguistics, syntax, dialect, Scots

[1] This work was funded by UKRI grant AH/M005550/1. We thank the audiences of BICLCE 8 (Bamberg, 2019) and UKLVC 12 (QMUL, 2019) for their questions and input, as well as two anonymous reviewers for *English Language and Linguistics* for their helpful suggestions and comments. We also thank Frankie Macleod for help with data preparation. For the purposes of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. All data supporting the results of the study are available for download at scotssyntaxatlas.ac.uk

1 Introduction

At the core of the variationist paradigm is the quantitative analysis of spontaneous speech, where variable use of forms is the focus in uncovering structured heterogeneity (Weinreich, Labov & Herzog 1968). While there is a wealth of research on phonetic variation (see e.g. Drager & Ketig 2021 for an overview), work on morphosyntactic variation is comparatively uncommon in variationist sociolinguistics (for discussion, see e.g. Rickford *et al.* 1995: 106; Cheshire 1999: 59; Cornips & Corrigan 2005: 2) despite wide-ranging variation, such as (1a–g).

(1) (a) The car needs washed. (*need* + PAST, Edelstein 2014: 242)
(b) They was a cemetery out on Hazel Mountain. (expletive *they*, Tortora 2006: 292)
(c) The teachers asks them to write something. (agreement, Adger & Smith 2010: 1110)
(d) Go you away! (imperatives, Henry 1995: 50)
(e) Here's you a pizza. (presentational datives, Wood & Zanuttini 2018: 9)
(f) Kim isn't running for office now, but she has done in the past. (*do*-ellipsis, Thoms & Sailor 2018: 145)
(g) She weren't that close to you. (*weren't* levelling, Parrott 2007: 298)

One reason for the lack of studies is frequency: many of these forms would not occur frequently enough in spontaneous speech for the type of quantitative analysis employed in the variationist paradigm. A potential solution is for sociolinguists to employ methodologies used by syntacticians, specifically 'elicited judgments, the intuitions of the native speakers' (Labov 1996: 77), in which speakers are asked to introspect on what they believe themselves to say. However, sociolinguists have remained sceptical of introspective data (Labov 1972; Carden & Dieterich 1981; Bender 2001; Green 2010), even going so far as to suggest it may 'fail' (cf. Labov's (1996) 'When intuitions fail') to capture the linguistic situation, in that speakers may 'agree that a certain form is completely unacceptable, yet use it themselves freely in every-day speech' (Labov 1996: 78). This impasse has meant many morphosyntactic variables remain understudied in sociolinguistic research, leaving a significant gap in our knowledge of variation across all levels of the grammar. This leads Rickford (2019) to suggest that the utility of introspective data should be revisited, and specifically:

> What is needed is a concentrated effort to determine what kind of intuitive judgments are more robust than others, what factors influence their variability, and what methods we might use for calibrating them against observational and other evidence. (Rickford 2019: 102)

In this article we address these issues using data from the *Scots Syntax Atlas*, herein SCOSYA (scotssyntaxatlas.ac.uk) (Smith *et al.* 2019). SCOSYA includes two types of data:

1. Likert-scale responses to a questionnaire of 200 dialect syntax forms given by over 500 speakers ('acceptability judgments')
2. A total of 275 hours of conversation gathered through sociolinguistic interviews with the same speakers ('spontaneous spoken data')

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

With access to both these data types from the same speakers, we can assess the reliability of the judgment data against spoken data. In doing so, we address the questions of where and why speakers' acceptability judgements depart from their use of variables in everyday speech, and more broadly, how judgment tasks, and other introspective methods, can be used in combination with variationist analysis in future sociosyntactic research.

2 Contrasting methods in the analysis of morphosyntactic variation

2.1 *Spontaneous speech*

Vernacular speech is considered the gold standard in variationist sociolinguistics, primarily collected through the SOCIOLINGUISTIC INTERVIEW (Labov 1972), which aims to mitigate the OBSERVER'S PARADOX: how to tap the most unmonitored style of speech when a participant is being monitored. The PRINCIPLE OF ACCOUNTABILITY (Labov 1966: 49) is also key to variationist study, and in particular, the LINGUISTIC VARIABLE, where a particular form is studied in relation to the form(s) it is in variation with. Moreover, the linguistic variable must be 'high in frequency' (Labov 1984: 49) for systematic patterns of variation to emerge. While debates on equivalence are no longer at the forefront of discussion (Lavandera 1978; Buchstaller 2009), the issue of frequency with morphosyntactic variables remains, as they 'often involve special semantic and pragmatic circumstances which may occur rarely or unpredictably in interview' (Rickford *et al.* 1995: 106). For example, Henry's (2005: 1609) work on the *after*-perfect in Belfast English found only three tokens in 720,000+ words, and Coats (2023: 706) finds 2.4 double modals per one million words in a corpus of Scots, despite both forms being robustly attested as part of the respective varieties. Studies on morphosyntactic variation thus often converge on a limited set of variables which appear frequently enough for analysis, including agreement (e.g. Cheshire & Fox 2009; Rupp & Britain 2019), relative pronouns (e.g. Guy & Bayley 1995; Tagliamonte, Smith & Lawrence 2005) and negative concord (e.g. Cheshire 1982; Smith 2001).

2.2 *Acceptability judgments*

Acceptability judgments have been the go-to methodology for theoretical syntacticians since the 1960s (Chomsky 1966),[2] and their reliability for investigating at least standard English has been experimentally demonstrated (e.g. Sprouse & Almeida 2017; Langsford *et al.* 2018; Goodall 2021). The methodology has been extensively employed in the study of non-standard morphosyntactic variation, from early introspective work (Rizzi 1982; Poletto 1993) to recent large-scale crowdsourced projects (e.g. Zanuttini *et al.* 2018). The judgment methodology has been refined for non-standard varieties, highlighting factors such as participant selection, conducting the task in the local dialect, and

[2] We focus on judgments in generative syntax, but note use of judgments in other areas of linguistics, such as language documentation (e.g. Munro 2017), which also often complement corpora.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

oral presentation of stimuli (Cornips & Corrigan 2005; Cornips & Poletto 2005; Barbiers & Bennis 2007; Buchstaller & Corrigan 2011), and some studies have shown that judgment data can usefully complement spoken data. For example, Henry (2005) compares the judgments of nine Belfast English speakers on a set of examples with non-standard agreement patterns, to the agreement patterns found in a 130-hour corpus of Belfast English speech. She finds that the corpus gives useful information about frequency, but the judgments present an opportunity to find out what is *ungrammatical* as opposed to simply infrequent. However, concerns about whether reported judgments of acceptability truly reflect community speech patterns remain (Henry 2005; Green 2010; Eide & Åfarli 2020).

What factors might influence the accuracy of judgments in non-standard varieties? In the next section, we summarise Labov (1996), who proposes reasons as to why speaker intuitions, as expressed through judgment tasks, might 'fail' to match up with speech patterns.

2.3 Why might intuitions fail?

Labov (1996: 85) provides an overview of some linguistic variables where comparison between such intuitions and spontaneous speech is available.[3] One such case is positive *anymore* (2), 'roughly equivalent to *nowadays*':

(2) Cars are sure expensive *anymore*.

Despite widespread use in conversation data, speaker intuitions about the form were 'very erratic indeed' (Labov 1996: 84), with some speakers denying knowledge of the form despite using it in the same interview.

Another variable, *ain't,* has a complex set of sociolinguistic rules which provided a good testing ground for judgment data. With *ain't* for *isn't* (3), there is a direct correlation between intuitions and production: speakers in communities across the US produce it (Labov *et al.* 1968; Feagin 1979; Hazen 1996), and recognise it as grammatical (Labov 1996: 90).

(3) He *ain't* too smart.

However, in African American English (AAE), *ain't* can also be used as a variant of *didn't* (4). In Labov's data, AAE speakers recognise (4) as grammatical to the same extent they do unattested forms, such as (5) in which *ain't* is used for *don't*.[4] Both receive 40–50 per cent acceptance.

[3] Labov uses the term 'intuitions' as a shorthand for the outcomes of the elicitation of introspective judgments. This is not quite the same approach as is taken in generative grammar. Chomsky (1965: 18–20), for example, uses the term 'linguistic intuition' to refer to the system of knowledge that a speaker has, as opposed to data which can inform the linguist about that system. Chomsky uses the term 'introspective reports' for what Labov calls intuitions, and reserves the term 'intuition' for the systematic knowledge the speaker has. Of course, informal use tends to have been much looser.

[4] Fisher (2018: 29–30) reports *ain't* for *don't*, almost entirely with *got/gotta*. *Ain't* for *don't* may have been a real (rare) feature of Labov's speakers' grammars; however, (5) (with *wanna*) would still not be expected.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

(4) He *ain't* see her yesterday.
(5) I *ain't* really wanna do that.

Two issues arise for the use of introspective judgments here: rejection of forms attested in speech (positive *anymore*, *ain't* for *didn't*), and acceptance of forms not found in speech data (*ain't* for *don't*). From these observations, Labov (1996: 98) proposes reasons why speakers' intuitions might 'fail':[5]

I. A 'socially superordinate norm' may take precedence because:
   a. The availability of a non-regionally specific variant may override judgment of a regional form;
   b. There may be social stigma attached to the form.
II. Infrequency: 'judgments that a form is ungrammatical may actually be motivated by the fact that it is rare' (Labov 1996: 99).

Regarding (Ia), as we saw above, intuitions about positive *anymore* were 'erratic' among speakers who used it. Labov states that 'the social bias [against *anymore*] is not at all obvious' (1996: 98), but nevertheless the non-regionally specific *nowadays* overrides speakers' ability to introspect about *anymore*. Regarding (Ib), social stigma is complex, involving factors such as class (e.g. Niedzielski & Preston 2003), race (e.g. Alim, Rickford & Ball 2016), region (e.g. Preston 1989), age (e.g. Cheshire 2005). What is considered prestigious varies across communities (Coupland & Bishop 2007), although the 'standard' variety tends to hold sway across the board. Furthermore, stigma is closely related to social salience (Labov 1994, 2001; Labov *et al.* 2011), and to the divisions between INDICATORS (socially stratified but not identifiable by speakers), MARKERS (controlled depending on speech context) and STEREOTYPES ('the overt topics of social comment') (Labov 1994: 78). While markers and especially stereotypes may be stigmatised, indicators are, by hypothesis, below the level of consciousness.[6] The salience of a feature, and its possibility of being stigmatised, is therefore relevant to understanding why intuitions might fail. The influence of stigma on the results of judgment tasks can be seen in Blanchette's (2017: 2) study of negative concord. Although her participants report negative concord as unacceptable due to 'heavy social stigma', they 'have grammatical knowledge' of it, demonstrated in subtly different Likert-scale acceptability judgments for different types of sentences.

In terms of (II), absolute frequency may not be particularly important for acceptability. For example, *it*-clefts in English appear in 'less than one tenth of a percent of all sentences' (Roland, Dick & Elman 2007: 353) but, given appropriate pragmatic contexts, are judged more acceptable than unclefted sentences (Destruel, Beaver & Coppock 2019). This seems to apply also to non-standard features: for example, Wood *et al.* (2020: 4) find

[5] Labov also lists other factors which may lead to failure of judgments in phonology or semantics.
[6] Indicators may nevertheless be stigmatised. Smith & Holmes-Elliot (2022) show non-temporal *never* is an indicator for speakers in Buckie, who do not alter their usage patterns of *never* when speaking to a community outsider as opposed to an insider – however, it is still subject to prescriptive judgment. Furthermore, Cornips (2022) shows that in Heerlen Dutch, syntactic forms which would be expected to be 'under the radar' were still able to be socially stratified, even at low frequencies.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

that although the datives they investigate (1e) are 'not frequently attested in written forms or even spoken corpora', they are 'robustly accepted' in judgment tasks.

However, evidence exists for correlation between acceptability and the *relative* frequency of variants – although the picture gets more complicated at lower frequencies (Featherston 2005; Kempen & Harbusch 2005; Arppe & Järvikivi 2007; Bresnan 2007; Divjak 2017; Bader & Häussler 2010; Bermel & Knittl 2012). What is considered 'high frequency' varies, from over 50 per cent (Bermel & Knittl 2012) to just 3 per cent (Bader & Häussler 2010), but what is consistent is that relatively frequent variants receive high acceptability ratings. On the other hand, grammatical but rarely produced variants are neither categorically accepted nor rejected. Acceptability rates ranged from 20 to 60 per cent for low-frequency grammatical options in Bader & Häussler's (2010) study of German word order, compared to over 80 per cent acceptability for more frequent grammatical word orders, and < 10 per cent acceptability for ungrammatical word orders. This gives a good starting point for considering how relative frequency might impact the results of judgment tasks, though there may be additional factors at play when considering a non-standard variant alongside standard ones.

2.4 Summary

A range of factors may affect speakers' abilities to introspect about the acceptability of morphosyntactic features, including frequency and social stigma. However, it remains unclear when these factors come into play and how they combine to influence the results of acceptability judgment tasks. Furthermore, there have been considerable adaptations made to judgment tasks to collect data from non-standard varieties. It is unclear to what extent the factors discussed in the context of earlier research continue to influence these adapted judgment tasks.

In the remainder of the article, we address these questions by comparing judgment data with spontaneous spoken data from the same speakers. In section 3, we introduce the social context of Scots and Scottish English in Scotland, SCOSYA, and its data collection methods, before comparing three different phenomena in section 4.

3 The *Scots Syntax Atlas*

3.1 Scots and Scottish English in Scotland

Historically, Scots was a distinct Germanic language within the British Isles, spoken in Lowland Scotland. However, over the centuries following the political union of Scotland and England in the 1600s, the role and prestige of English within Scotland grew. In the present day, Scots – which is distinct from English lexically, morphosyntactically and phonetically – exists on a continuum with (Standard Scottish) English. The existence of the continuum at an interspeaker level intersects with various sociodemographic issues – e.g. broader Scots is often associated with working-class speakers – but on an intraspeaker level, many speakers of Scots style shift along the continuum depending on their interlocutors and the linguistic task at hand. Use of

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

broad Scots in a professional context may be liable for disapproval, but so may use of Standard Scottish English when speaking to members of a speaker's own community. For more information on the continuum and the context of Scots in Scotland, see e.g. Aitken (1984), Johnston (2007), Smith (2012).

In Aitken (1984), linguistic features of Scots are divided into COVERT, OVERT and VULGARISMS, generally mapping to Labov's indicators, markers and stereotypes. Many morphosyntactic features of Scots are covert – used by speakers across class boundaries 'unself-consciously'. Speakers are often unaware of alternative (standard English) ways to express the same meaning (Aitken 1984: 18). For example, Scots covertly employs modal verbs *can* and *will* in place of *may* or *shall* respectively. Overt Scots features are used freely at the broad Scots end of the continuum, but can also be employed stylistically in more Standard Scottish English 'on occasions when it seems desirable to claim membership of the in-group of Scots' (Aitken 1984: 22); for example, in proceedings of the Scottish Parliament, using the Scots-specific form *kent* in place of the (Standard Scottish) English *knew* in 'Burns kent better' (quoted in Corbett & Stuart-Smith 2012: 86). Finally, vulgarisms are features of Scots that are often considered 'Bad Scots' and condemned across social contexts (Aitken 1984: 24). These tend to be variables that are most associated with urban Scots varieties and/or younger speakers, despite the fact that they may be used across social and regional groups – such as the pronoun *youse* (see also Corbett & Stuart-Smith 2012 for how these different categories are realised in both spoken and written data).

It is important to remember when considering Aitken's distinctions that there is considerable variation within Scots itself, with distinct regional varieties found across the country. In different regions, there are different features that will be overt and covert, and that may have different levels of social stigma attached to them (see e.g. Smith 2001).

Recalling Labov's (1996) proposed reasons as to why 'intuitions may fail', detailed in section 2.3 above, the social context of Scots and English in Scotland therefore provides us with an ideal testing ground to tease apart the most important factors. Features of Scots' morphosyntax of course vary in *frequency*, but there are also distinctions between highly local features and region-wide non-standard features, as well as standard (English) ways of 'saying the same thing', allowing us to explore the role of *regionality*. There is also detailed existing research into the salience and perception of many grammatical variables across Scots varieties (Aitken 1984), allowing us to take account of how social *stigma* may affect judgments. Attempting to combine these factors leads us to the choice of features investigated in this study, detailed in section 4. Before this, we present SCOSYA and the data that will be used in this article.

3.2 *Overview of SCOSYA*

The *Scots Syntax Atlas* (SCOSYA) builds on the growing enterprise of dialect syntax atlases, e.g. SAND (Barbiers & Bennis 2007), ASIS (Benincà & Poletto 2007), the *Nordic Syntax Database* (Lindstad *et al.* 2009) and YGDP (Zanuttini *et al.* 2018). While these atlases were built primarily from judgment data, SCOSYA also

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

incorporates the practice of a number of dialect syntax corpora, e.g. AAPCAppE (Tortora *et al.* 2017), gathering both judgment data and spoken data from the same speakers.[7] In total, SCOSYA combines 275 hours of sociolinguistic interview data with over 100,000 acceptability judgments across 200 morphosyntactic phenomena found in varieties of Scots. The data were collected between December 2015 and July 2018 from 530 speakers in 146 locations across Scotland (figure 1).

We divide the data into 15 broad geographic regions, based on traditional dialectological areas of Scotland (Grant 1931; Aitken 1984; Johnston 1997; Miller 2007). See scotssyntaxatlas.ac.uk/linguists-atlas for more information.

The data were collected by community insider fieldworkers, as insider status is key to gaining access to the relevant non-standard forms (Labov 1972). The fieldworkers were responsible for recruiting participants using a standard set of sociolinguistic criteria (Labov 1984): participants were born and brought up in the area; had not spent any significant time away; had parents who were from the area; and had not gone on to higher education.[8]

Dialect syntax atlases have tended to focus on older speakers to capture the 'traditional' dialects of a particular language or region. When younger speakers have been included, such as in the *Nordic Dialect Corpus*, there is no direct comparison with older speakers in the same community. In SCOSYA, pairs of participants were recruited in two age groups: 65+ and 18–25 (thus four participants in each location), in order to systematically investigate change in apparent time. Each pair generally consisted of friends or family members, and were usually also friends or family members of the fieldworker. Data collection was conducted in participants' homes to encourage relaxed, naturalistic data in a comfortable setting (Labov 1984). By following these criteria in the data collection process, the SCOSYA data hones in on participants' most vernacular speech, rather than a more formal style that might elicit the Standard Scottish English end of their speech continuum. We now detail the two data collection methods used.

3.3 *The judgment task*

The first part of the data collection process was an acceptability judgment task using a questionnaire developed by the SCOSYA team. The questionnaire targeted key non-standard morphosyntactic forms. All were features of Scots dialects; we use the term 'non-standard' to indicate the range of variation within the forms, from hyperlocal dialect features to more general English 'vernacular universals', that were tested. For example, some features were noted in the literature to occur throughout Scotland, such as possessives or determiners where standard English would have a bare noun (6). Other examples were attested for particular dialect areas, such as *gonnae* imperatives

[7] The *Nordic Dialect Corpus* (Johannessen *et al.* 2009) contains recordings which complement the *Nordic Syntax Database*, with some speakers contributing to both.

[8] Some flexibility was permitted on the latter two criteria, e.g. a younger participant who continued to live locally but commuted to college.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

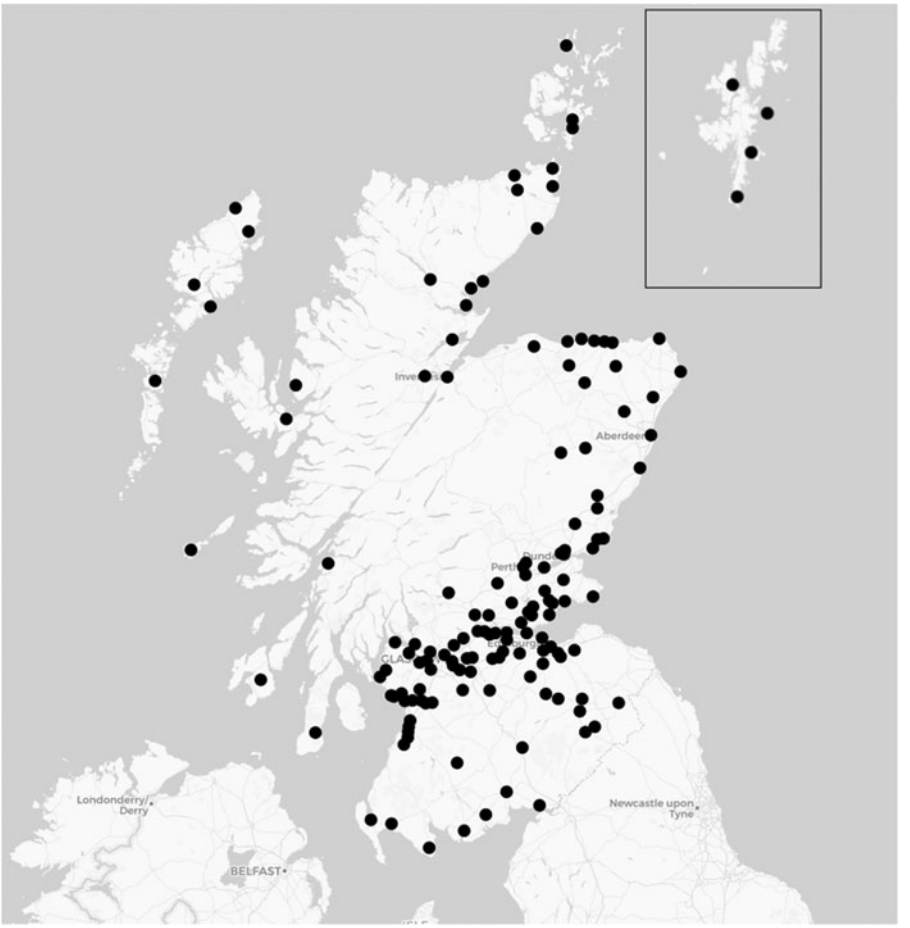

Figure 1. All locations in the SCOSYA data

(7), attested in Glasgow. Some pandialectal forms found in varieties of English both across the UK and beyond were also included, such as non-standard preterit verb forms (8).

(6) I'm going to **my** bed.
(7) **Gonnae** you leave us alone!
(8) I **seen** that last week.

A total of 182 features were tested with all participants. Additional items of specific interest were judged in certain locations only, so each questionnaire contained approximately 200 examples. In total, over 100,000 judgments were given.

The questionnaire was delivered in an adapted version of the interview method (Barbiers & Bennis 2007). Each target example was presented following a short context which included relevant referents and contextual information (9):

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

1	2	3	4	5
I would never say that	I probably wouldn't say that	I might say that	I would probably say that	I would definitely say that
You would never hear that here	You probably wouldn't hear that here	You wouldn't hear that a lot here	You would hear that occasionally here	You would hear that a lot here
It's very unnatural	It's unnatural	It's not very natural	It's quite natural	It's very natural
It's bad	It's weird	It's a bit weird	It's OKish	It's fine

Figure 2. SCOSYA five-point Likert scale

(9) You're telling me you saw me and a friend earlier. You say:
I saw **youse** earlier on.

The fieldworker read out the contexts and examples to the first participant in each pair. The participant verbally rated each example on a five-point Likert scale (figure 2), indicating their intuitions about whether/how the sentence would be used in their community. Each point on the scale was labelled, and participants were given a copy of the scale to refer to throughout the task. Scores were recorded by the fieldworker. The first participant subsequently adopted the role of 'fieldworker' and delivered the questionnaire to the second participant.

3.4 The interview

The second part of the data collection process was a standard one-hour sociolinguistic interview, conducted between the pairs of speakers in each age group, with the fieldworker present. Due to the insider status of the fieldworker, the Observer's Paradox was reduced as far as possible, and conversation was generally relaxed and open (see further in Labov 1984).

Example (10) demonstrates the kind of talk that arose in these interviews, with bold indicating some of the non-standard forms:

(10) Lawrence: I drove for, what, another two **month**?
Delia: **Wasnae** even that.
Lawrence: And then I **gied** up the car. I said 'that's me'.
Delia: 'Cause you **werenae** confident.
Lawrence: I got the thing in when I was 70, eh, **for to** put in for the—my license again and I just tore it up. **I never done it** so that was me, I never drove since then.
(Ayrshire, O)

The recordings were text-to-speech transcribed in Transcriber (Barras *et al.* 2001). The full corpus of over 3 million words is available at scotssyntaxatlas.ac.uk/about/accessing-the-spoken-corpus

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

4 The data

4.1 *Selecting the variables*

We focus on three variables (11a–c) which differ across geographic and social dimensions in Scots. Unless indicated, all examples are taken from the SCOSYA corpus.

(11) (a) My piece box actually doesnae **need washed**. (*need* + PAST)
(b) So what **div** you do in Hong Kong like? (*div* for *do*)
(c) We **didnae** know **nothing** about her background. (negative concord)

Choosing to investigate these features allows us to differentiate the roles of factors suggested by Labov (1996) as leading to the 'failure' of intuitions. Regarding frequency, while we do not have prior information on the relative frequency of variants, we do know about their likely absolute frequency. Variable contexts where *need* + PAST can possibly occur are known to be 'low frequency' (Strelluf 2022) but possible contexts for periphrastic *div* may be much higher. Negative concord (NC) may be situated somewhere in between.

The features also vary with regard to regionality. *Div* is highly local within communities, while *need* + PAST is supralocal across Scots varieties and NC is supralocal across Englishes. Crucially, all also have a 'standard' variant, which speakers may also have access to.

Finally, the features vary with regard to their salience and social stigma. *Need* + PAST is recognised as a covert feature of Scots (e.g. Aitken 1984:21). It is therefore unlikely to be salient, nor stigmatised. We know *div* is a socially salient feature of Tyneside English (Rowe 2007: 362; Pichler 2009: 290); however, there is little evidence regarding the salience of *div* in Scots. The features defining *div* as salient for Tyneside varieties of English – grammaticalisation to set phrases, strong usage by young working-class men (Rowe 2007; Pichler 2009) – may not be relevant for *div* in Scots. Our third feature, NC, is one of the most socially salient variants in any variety of English (Cheshire 1982; Labov 2001; Smith 2001; Anderwald 2005; Blanchette 2017). Aitken (1984:25) categorises NC as a Scots vulgarism, subject to 'explicit condemnation', while Smith & Holmes-Elliott (2022) find speakers in Buckie change their rates of NC if speaking to a community insider as opposed to an outsider. This controlled usage highlights its salience, and the associated stigma.

We now provide a more detailed description of what is already known about each of these features, before presenting the results of the SCOSYA corpus search for each and contrasting it with the judgments.[9]

[9] Note that the judgments here each come from a single example, which we use to establish broad acceptability. To understand the underlying grammar(s) of these forms would require testing more examples across syntactic contexts.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

4.2 need + *PAST*

4.2.1 Background

In English, a passive construction can be embedded under a matrix verb. Usually, the infinitival form *to be* combines with the past participle (12a). It is, however, also possible to have a present participle follow the matrix verb (12b).

(12) (a) He looks like he **needs to be washed** all the time. (Borders, Y)
(b) Her granny doesn't **need looking after.** (Lothian, Y)

There is a third variant available in some varieties in which certain matrix verbs (*need*, *want* and *like* (Murray & Simon 2002)) combine directly with a past participle (12c–e).[10]

(c) Aye the floor **needs renewed.** (Lothian, O)
(d) She'll **want subbed**. (Highlands, Y)
(e) Babies **like cuddled**. (Murray & Simon 2002)

We adopt the term *need* + PAST (Strelluf 2020) for this construction, as we focus purely on cases with *need*.

need + PAST constructions have been attested in some US Englishes (Stabley 1959; Murray, Frazer & Simon 1996; Maher & Wood 2011; Duncan 2021), and Irish English (Montgomery 1997). In Scots, *need* + PAST is understood as a feature of Scots generally (Brown & Millar 1980; Montgomery 1997); we might thus expect it to be attested and judged highly across Scots varieties.

4.2.2 need + *PAST: spoken data*

All three variants are used in the SCOSYA spoken corpus:

(13) (a) We've got a piano that **needs tuned**. (Dumfries, Y)
(b) I'll no say dated, but it **needs to be upgraded**. (Borders, O)
(c) Mute the bits that **need muting**. (Tayside & Angus, Y)

Table 1 shows how these are distributed across the data. Table 1 reveals that *need* + PAST (13a), the non-standard form, makes up over 57 per cent of examples. In this case, the non-standard variant is produced at the highest rate in the corpus.

Figure 3 plots where *need* + PAST occurs. The faded circular dots indicate communities where conversation was recorded, but no attestations of *need* + PAST were found. The black dots indicate communities where *need* + PAST was attested. Figure 3 shows that attestations of *need* + PAST are spread across Scotland. We note that absence of a form does not necessarily mean no use, but instead may arise from the general infrequency of any of the three constructions with *need*.

[10] For many speakers, the construction is only available with *need* (Murray & Simon 2002; Maher & Wood 2011), though Duncan (2021) argues there may be a greater range of verbs for some US speakers.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Table 1. *Proportion of* need *constructions in the SCOSYA corpus*

	N	%
need + PAST	27	57.4
need + *to be*	18	38.3
need + present participle	2	4.3
	47	100

4.2.3 need + *PAST: judgment data*

Participants judged one example of a *need* + PAST construction.

(14) The postman pulls up in his van and it's filthy. You say:
His van **needs washed**.

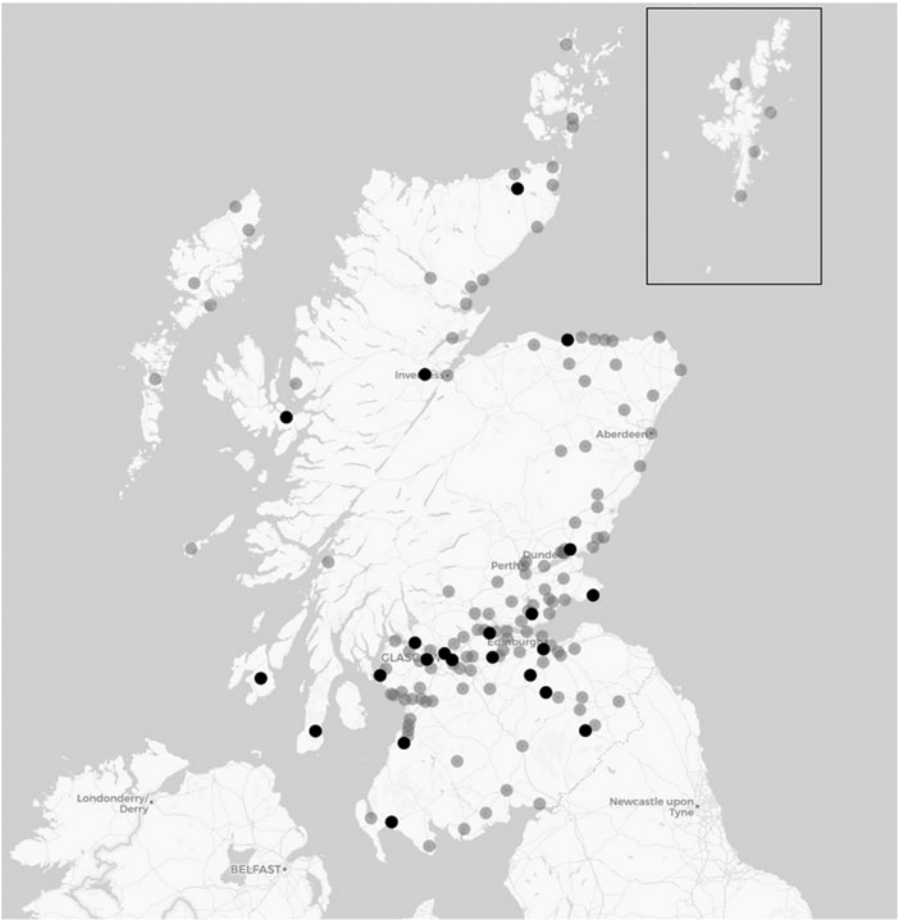

Figure 3. Locations with attestations of *need* + PAST in SCOSYA

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Recall that participants judged items on a 1–5 scale. '5' indicated the participant would 'definitely say' it and it was 'very natural', while '1' indicated they 'wouldn't say' it, and it was 'very unnatural'. We treat locations where two or more participants gave an example 4 or 5 as 'acceptable', following e.g. Zanuttini *et al.* (2018) and Thoms *et al.* (2019). We do not treat the data as continuous, and report median and mode statistics throughout.

Participants generally gave the example in (14) high ratings. The median and mode scores were 5, while the individual-level acceptance rate (percentage of speakers who rated the example 4 or 5) was 80 per cent. Figure 4 shows the widespread acceptance, with dark dots indicating locations where at least two participants rated the example 4 or 5.

Comparing figures 3 and 4 demonstrates that speakers' intuitions align with respect to *need* + PAST. It is both accepted and used throughout Scotland, despite the low absolute frequency of the construction in the corpus.

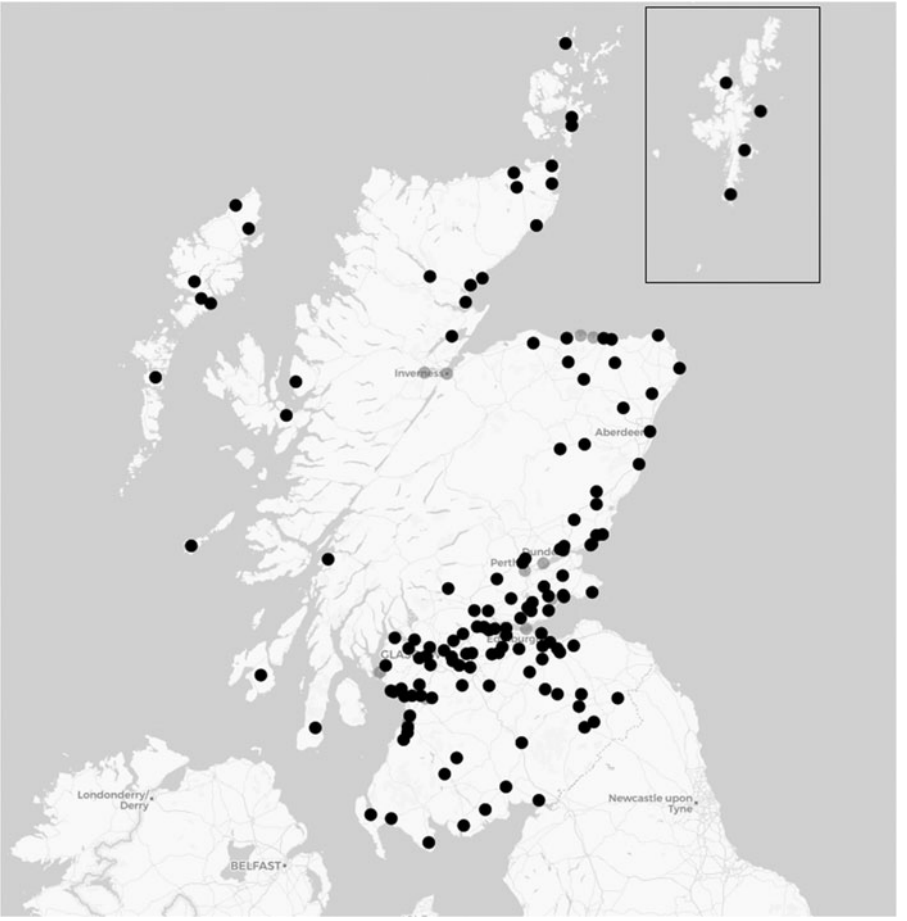

Figure 4. *His van needs washed*, accepted by at least two participants

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

4.3 *Div* for *do*

4.3.1 *Background*

In some varieties of Scots, *div* varies with periphrastic *do* in the present tense in all subject types except third-person singular (15a–e).

(15) (a) **I div** miss him when he goes home. (North East, O)
(b) Sometimes **you div** need it. (North East, O)
(c) **We div**, but as long as we're not in the same house. (Borders, Y)
(d) **They div** go occasionally. (North East, O)
(e) *He div eat sweeties.

Div can be used in negatives (16a), questions (16b) and tags (16c) but not imperatives (16d).

(16) (a) My crowd **divna** like barley. (Buckie, Smith 2000: 241)
(b) What **div** you do in Hong Kong? (North East, O)
(c) You seem to get four winds, **divn't** you? (North East, O)
(d) *Divna eat that scone!

In terms of geographic spread, *div* is attested in the North East (Smith 2000; *Dictionary of the Scots Language* 2004) and the Borders.

4.3.2 Div: *spoken data*

In the corpus, *div* is found in emphatic positives (17a), questions (17b) and tags (17c).

(17) (a) I **div** miss him when he goes home. (North East, O)
(b) **Div** you always go to the one below? (North East, O)
(c) You hear them fighting and swearing […], **div** you? (North East, O)

There are over 15,000 possible contexts in which *div* could alternate with *do*. Table 2 shows the relative frequency of the two forms.

Despite thousands of contexts where *div* could be used, it is very rarely used. Figure 5 plots the attestations. The faded circular dots indicate communities where conversation was recorded, but *div* was not attested. The black dots indicate communities where *div* was attested. Figure 5 shows the majority of attestations are in the North East. In the Borders, we also find attestations. The distribution of *div* in the SCOSYA corpus aligns with what we would expect based on previous reports of the form.

Table 2. *Proportion of periphrastic* do/div *in the SCOSYA corpus*

	N	%
Div	69	0.5
periphrastic *do*	15,097	99.5
	15,166	100

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

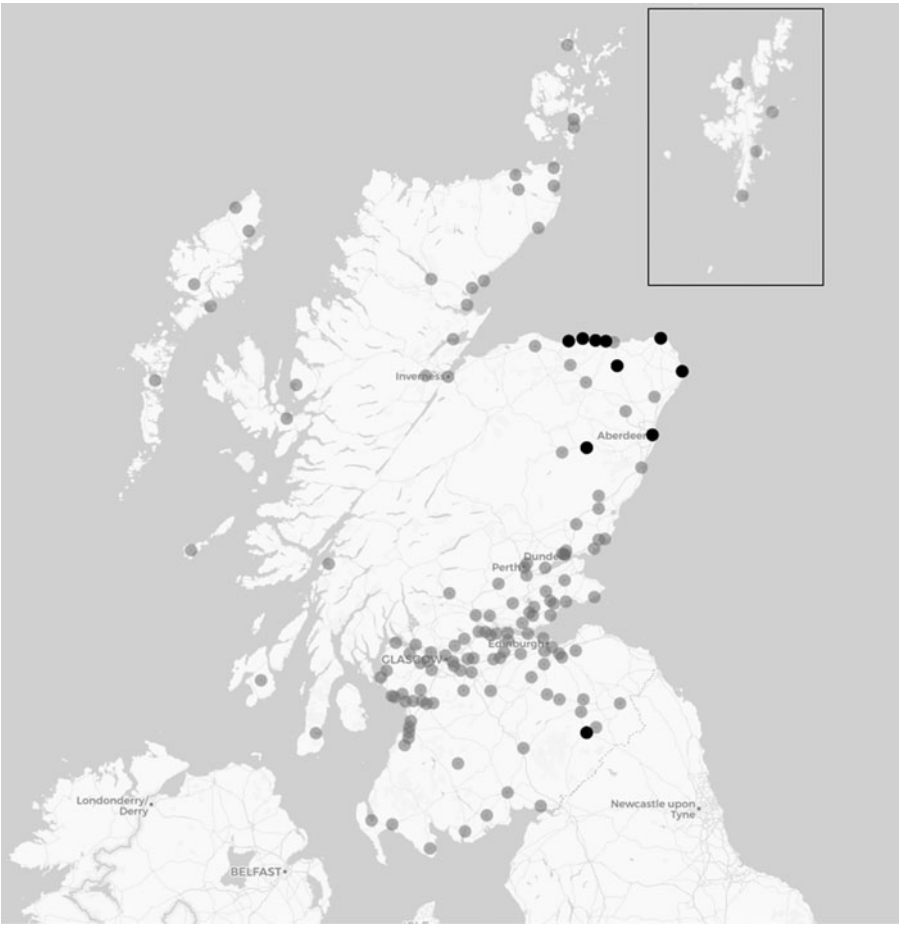

Figure 5. Locations with attestations of *div* in SCOSYA

4.3.3 Div: *judgment data*

Participants judged one example of *div* in an emphatic polarity context.

(18) You're sitting down with a cup of tea and a scone. You say:
I **div** like a scone!

Figure 6 shows the distribution of judgments. In figure 6, the dark dots indicating acceptance are geographically clustered. While *div* was not rated highly across the whole data set (median and mode both 1), in the North East, there is a median and mode of 5, while in the Borders the mode is 5 and the median is 3.5.

Just as with *need* + PAST, the spoken data in figure 5 and the judgment data in figure 6 align with respect to *div*, suggesting reliable intuitions regarding this feature.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

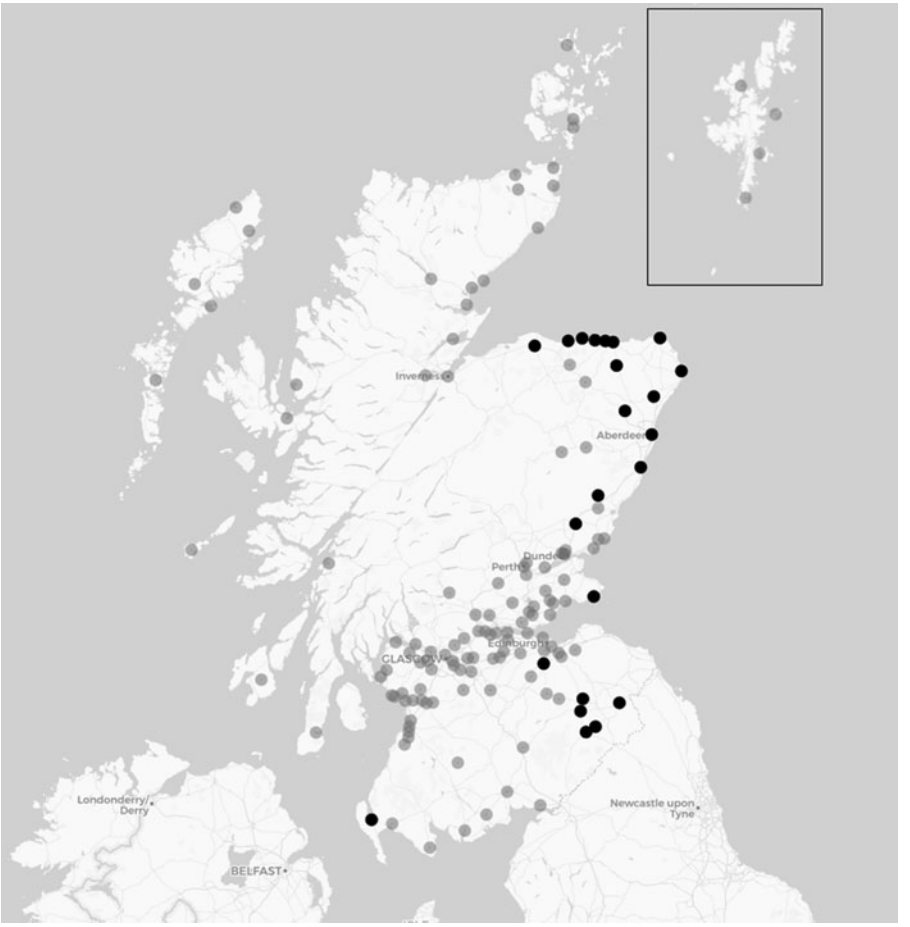

Figure 6. *I div like a scone*, accepted by at least two participants

4.4 Negative concord

4.4.1 Background

Standard English includes two ways of marking negation with indefinites. With sentential negation (19a), the negative marker *not* or *-n't* appears after the verb, and scopes over an indefinite *any-* form. Alternatively, the negative is incorporated into the indefinite, realised as a *no-* form (19b). A further alternative exists in non-standard dialects of English: negative concord (NC).[11] In (19c), negation is marked on the negative marker *-n't* and the indefinite, *nothing*.

[11] There are different types of NC. We focus on forms such as (19c), found across most varieties of English, in which sentential negation combines with a lower negative indefinite.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

(19) (a) They did**n't** do anything. (Lothian, Y)
(b) They did **nothing** about it. (Tayside & Angus, O)
(c) They could**n't** do **nothing** about it. (Borders, O)

Despite denigration of forms such as (19c) over centuries, NC is widespread and 'recurs ubiquitously all over the world' (Chambers 1995: 242), with numerous contemporary reports (e.g. in Britain, Hughes & Trudgill 1979; Cheshire 1982; Coupland 1988; Beal 1993; Edwards 1993; Anderwald 2005). Examples like (19c) have been reported across Scots varieties (Macaulay 1991; Cheshire *et al.* 1993; Smith 2001; Anderwald 2005; Macafee 2011), although at varying rates (e.g. 49 per cent in the North East (Smith & Holmes-Elliott 2022), but 8 per cent in Glasgow (Childs 2017)). Based on previous research, we may expect NC would be produced and judged acceptable in Scots varieties, but perhaps at variable rates across communities.

Note that there are two forms of contracted negative marker in Scots, the broad Scots form *-nae* and the (Standard Scottish) English form *-n't*. Speakers who have *-nae* also have *-n't* (Smith, Durham & Richards 2013):

(20) I **shouldn't** have drank as much, I **shouldnae** have that kebab. (Ayrshire, Y)

For the purposes of NC, *-n't* and *-nae* have the same syntactic properties (see Thoms *et al.* 2023), and we include examples with both contracted negative markers.

4.4.2 Negative concord: spoken data

In the judgment task (section 4.4.3 below), we discuss judgments for NC with two postverbal indefinites: *nothing* (21a) and *nowhere* (21b). We also focus on these two forms in the spoken data, though note the corpus also contains examples of NC with other indefinites, such as *nobody*, *none* and *no* NP.

(21) (a) They **cannae** see **nothing** now. (Fife, O)
(b) Havena seen him out in his car, **havenae** seen him out **nowhere**. (North East, Y)

Table 3 shows NC by indefinite type.[12]

Turning first to NC with *nothing* (21a), figure 7 shows the distribution of the 148 tokens. The faded circular dots indicate communities where conversation was recorded, with no attestations of NC with *nothing*. The black dots indicate communities where NC with *nothing* was attested. Figure 7 shows widespread use throughout Scotland. As with *need* + PAST, we cannot rule out NC with *nothing* in the communities with no attestations due to low absolute frequency.

There was only one attestation of NC with *nowhere*, from the North East (21b), and the total percentage of NC in the *nowhere/anywhere* case is only 1 per cent. From the corpus data, then, we see variation in NC depending on the indefinite.

[12] We excluded 364 existential constructions from the *nothing/anything* count (10 NC) and 23 existential constructions from the *nowhere/anywhere* count (0 NC).

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Table 3. *Proportion of negative concord by type in the SCOSYA corpus*

	nothing/anything		*nowhere/anywhere*	
	N	%	N	%
Any- indefinite	1,024	66	97	81
Negative indefinite	370	24	22	18
Negative concord	148	10	1	1
	1,542	100	120	100

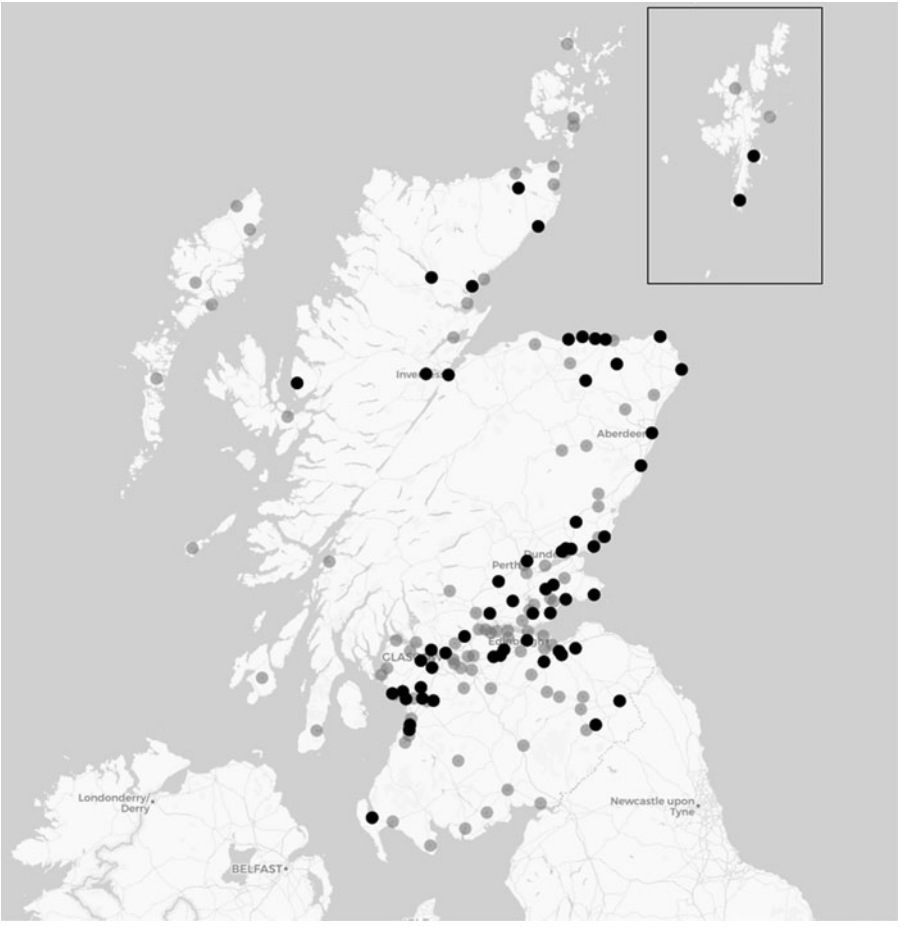

Figure 7. Locations with attestations of NC with *nothing* in SCOSYA

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

4.4.3 Negative concord in the SCOSYA judgment data

All participants judged examples of NC where sentential negation was combined with *nothing* (22a) and *nowhere* (22b).[13]

(22) (a) I ask if you can see my coat in the cupboard, but the light's off. You say:
I **cannae** see **nothing**.[14]

(b) You've been looking all over the house for an umbrella but had no luck. You say:
I **didnae** see it **nowhere**.

We will firstly discuss the example with *nothing* in (22a).

Cannae see nothing

There is a great deal of variability in the ratings for (22a). It was not rated highly across the regions (mode=1, median=3). Nevertheless, it had a 43 per cent acceptance rate (ratings of 4 or 5). Figure 8 presents an acceptability map for this example, with black dots indicating communities in which 2 or more participants gave the example 4 or 5.

It is clear from figure 8 that the moderate acceptability rate cannot be explained by variation across place; although there are some areas with attestations in almost every community (e.g. Tayside & Angus), there is not the strong geographic clustering of judgments we saw with *div* (see section 4.3.3).

Didnae see it nowhere

The example in (22b) was not rated highly across the regions (mode=1, median=2), with a 21 per cent acceptance rate. Figure 9 presents an acceptability map for this example, with black dots indicating communities in which 2 or more participants gave the example 4 or 5.

The map in figure 9 shows some geographic clustering. Although there are communities with acceptance dotted around e.g. Fife and Stirling, there is a particular concentration of black dots in the North East. There, the median is 4 and mode is 5, with an acceptance rate of 56 per cent. This geographic concentration of acceptance is also in the region where we find the singular attestation of NC with *nowhere* in the corpus.

4.4.4 Negative concord: combining spoken data and judgments

In contrast to *need* + PAST and *div*, figures 7 and 8 show little alignment between spoken attestations and judgments summed at community level with respect to NC with *nothing*. While *need* + PAST was infrequent in the corpus, the wide geographical distribution mapped to widespread acceptability, suggesting it was down to (in)frequency that it was not attested in our data from more communities. On the other hand, while attestations of NC with *nothing* in the spoken corpus are spread across the communities sampled, judgment data is patchier, with no clear distribution of (un)acceptability.

[13] Other types of NC were also tested; see scotssyntaxatlas.ac.uk/linguists-atlas

[14] In the Highlands and Western Isles, *-nae* negation is not used (see Thoms *et al.* 2023: 729). Participants here judged the examples in (22) with *-n't* negation, e.g. 'I can't see nothing'.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

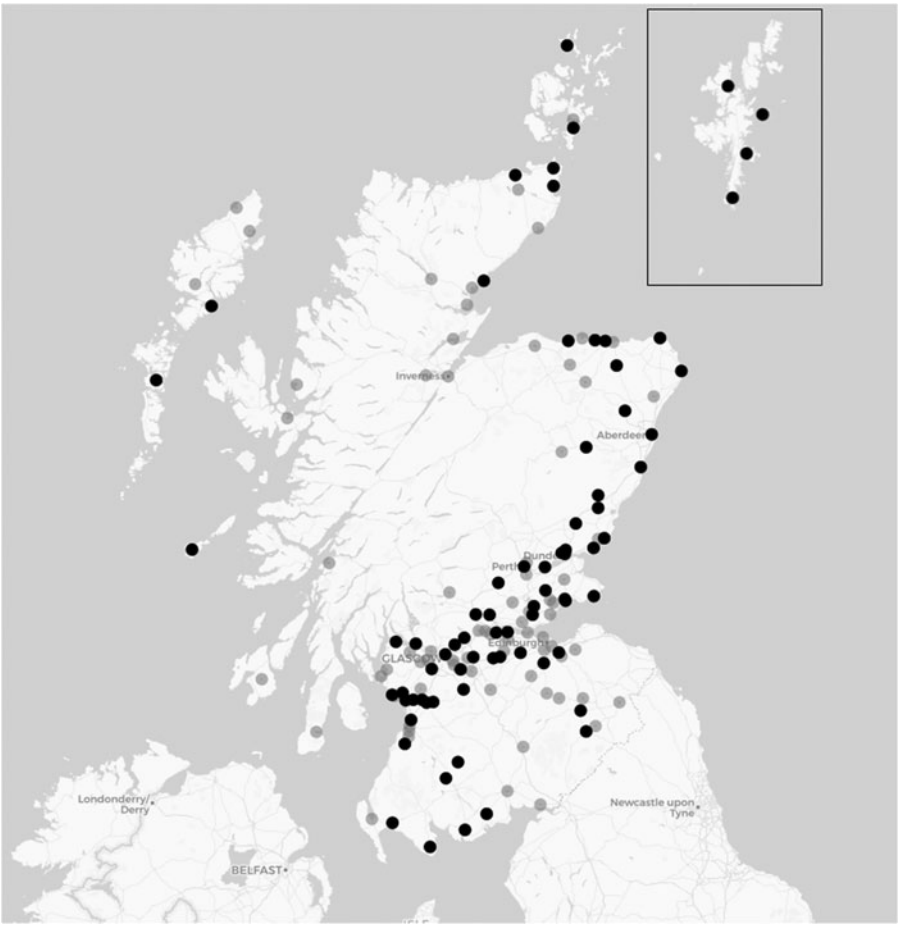

Figure 8. *I cannae see nothing*, accepted by at least two participants

It is worth exploring whether these are individual community-based distinctions. Comparing the maps, we can see that some communities attest NC with *nothing* and accept it. Some communities do not attest NC with *nothing*, but accept it in the judgment data. This is what we see for *need* + PAST and *div*, where we assume the phenomenon was not attested in those communities for frequency reasons. However, other communities *do* attest NC with *nothing*, but reject it in the judgment task – for example, this is a clear pattern for communities in Caithness and the Highlands. With NC with *nothing*, then, judgments do not clearly map to the spoken data.

As there was only one attestation of NC with *nowhere* in the corpus, it is difficult to draw conclusions about its distribution; however, the single attestation came from a location in the North East, where there is also a cluster of acceptability in the judgments. The judgment data for NC with *nowhere* maps more clearly to the spoken data than NC with *nothing*, though the picture is not as sharp as for *need* + PAST and *div*.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

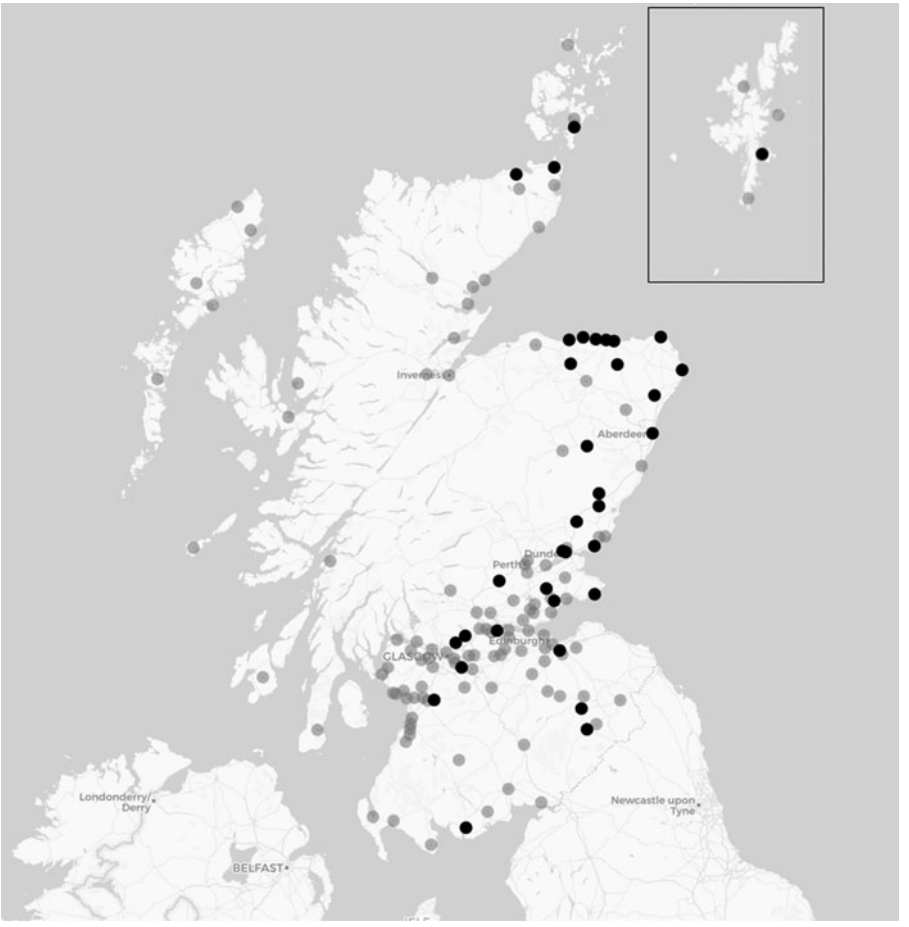

Figure 9. *I didnae see it nowhere*, accepted by at least two participants

4.5 Summary

We have examined three variables from SCOSYA: *need* + PAST, *div* for periphrastic *do* and NC. Each of the variables has a different pattern when comparing spoken data to judgments. *Need* + PAST is produced and accepted across regions. *Div* for *do* is geographically circumscribed to the North East and Borders in the corpus, and this is matched in the judgments. For NC with *nothing*, results are mixed, with attestations across regions but inconsistent acceptability in the judgment task. Here, it seems the results of judgment tasks diverge from the usage picture. However, NC with *nowhere* exhibits a clustering of acceptability around the single attestation in the North East. In section 5 we will revisit the social context of Scots and Scottish English in Scotland presented in section 3.1 to propose reasons for the varying levels of match/mismatch in the data.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

5 Discussion

In section 2.3, we outlined Labov's (1996) reasoning as to why speakers' intuitions might 'fail' to match speakers' usage in everyday speech, and section 4.1 we laid out how the particular features investigated in this study allow us to try to tease apart the importance of these factors. Here, we consider each of the factors individually – frequency, regionality and social stigma – and summarise how they do, or do not, explain the data. We also draw on evidence from other features investigated in the SCOSYA data collection where relevant patterns seem to emerge.

5.1 Frequency

The first possible reason for the disparity between the outcome of introspective judgment tasks and speakers' behaviour in conversational settings is (relative) frequency. Looking firstly at the data for *need* + PAST, despite low absolute frequency, *need* + PAST accounts for 58 per cent of possible contexts (that is, its relative frequency is high). *Need* + PAST is the most frequent variant in production, and – in alignment with this – judgments are accurate. However, when we look at the difference in accuracy between judgments on *div* and those on NC, we find variation that cannot be explained in terms of relative frequency alone. With *div*, across Scots varieties, relative frequency in our data is low – less than 1 per cent. Nevertheless, speakers are accurate in their judgments. However: at what level should we be measuring frequency? Frequency in the corpus overall does not necessarily indicate frequency at community level, given the wide regional distribution of participants. Indeed, the frequency of *div* varies by area: within the North East and Borders, the relative frequency of *div* increases to 4 per cent (73/1,676 possible contexts), while it falls to 0 per cent in the rest of the regions. *Div*'s rejection in the rest of the regions, combined with its lack of attestations there, is a fairly reliable indication it is unacceptable in those communities.

However even in the North East and Borders, *div* remains a relatively low-frequency variant. As discussed in section 2.3, what is considered 'low frequency' varies across studies, from 3 to 50 per cent. The frequency of *div* is at the lower end of this range even in the North East and Borders, and may provide evidence that the cut-off point for being 'rare' enough to affect ratings is really quite low.

While relative frequency is aligned with the accurate judgment patterns for *need* + PAST and *div*, at least at a local level, the same cannot be said of the judgments given to NC. NC with *nothing* occurs at a rate of 10 per cent across the corpus. While this is not a high rate, it is certainly above the 4 per cent that saw *div* rated as acceptable in the communities in which it was produced. And yet, as discussed above, judgments on it are inconsistent.

Given the variable production rates for NC across Scots varieties in the literature (see section 4.3.1), we should consider whether that 10 per cent production frequency is a flat rate across the regions. Perhaps unsurprisingly, it is not. In the North East, NC with *nothing* occurs at a rate of 34 per cent. In a further three regions, NC with *nothing* occurs at a rate over 10 per cent (Ayrshire, Fife and Tayside & Angus). At the other end of the scale, three regions exhibit NC with *nothing* at a rate between 0 and 3 per

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

cent (Orkney, Stirling & Falkirk and the Western Isles). In the remaining eight regions, the rate is 4–8 per cent.

However, the differences in rates do not necessarily parallel differences in judgments. In the North East and Tayside & Angus, which have high relative production rates, there are also high acceptance rates (53 per cent and 61 percent 4–5 ratings respectively). The highest acceptance rate (63 per cent) is, however, found in Shetland, where NC with *nothing* was attested in only 6 per cent of possible contexts. Participants in the Western Isles had a relative frequency of 1 per cent in production, arising from a single attestation, but an acceptance rate of 36 per cent (median=2, mode=1). In Fife, where the relative frequency is 13 per cent, the acceptance rate is 31 per cent (median=2, mode=1).

We cannot rule out the possibility of low relative frequency of a variant in a community leading to the failure of speakers' intuitions, as evidenced by the judgment tasks. However, frequency is not able to explain the judgments on NC with *nothing*. We therefore move on to the second factor, regionality.

5.2 Regionality

Labov (1996: 22) proposes that 'any grammatical pattern that is perceived as regional may be suppressed in introspection'. The features we have investigated are different in terms of their regional distribution: *div* for *do* is highly localised; *need* + PAST is supralocal across Scots varieties, while NC is supralocal across Englishes. Does this variation affect judgments?

As we saw in section 4.2, judgments for *div* mapped accurately to attestations. This success is replicated with other highly localised features. For example, double modals are attested as a feature of Southern/Borders Scots (Brown 1991), and there are two examples of double modals in the corpus, from southern varieties (23a–b).

(23) (a) You **used to could** go to the Gaiety for about nine pound. (Ayrshire, O)
(b) Some people **might can.** (Borders, Y)

In the judgment data, double modals are rated highly across these areas, and low elsewhere. It seems when a morphosyntactic feature has a local distribution, speakers accurately judge its presence, or lack thereof, in their community.

The acceptability of NC with different indefinites also supports this. While NC with *nothing* is found across Scots varieties, NC with *nowhere* appears limited to the North East. This reflects other work showing that production of NC holds at different rates for different lexical items in different varieties (Cheshire 1982; Smith 2001; Anderwald 2005; Robinson & Thoms 2021). For localised NC with *nowhere*, participants' judgments accurately reflect local spoken use; on the other hand, for supralocal NC with *nothing*, judgments diverge from local usage patterns. This might lead us to conclude that it is with supralocal features that judgments diverge from spoken usage patterns, but of course, this does not hold true for *need* + PAST.

Unlike *need* + PAST, NC with *nothing* is a vernacular universal (Chambers 1995) across Englishes. It is possible, then, that being a non-standard variant that is not

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Table 4. *Intuition 'success' and 'failure' in the SCOSYA judgment tasks*

	Intuitions succeed	Intuitions fail
Highly local features	*div* NC with *nowhere*	—
Supralocal features	*need* + PAST non-temporal *never*	NC with *nothing*

specifically a Scots feature could affect judgments. However, another vernacular universal tested in the corpus is non-temporal *never*, where *never* acts as regular sentential negation (24a–b).

(24) (a) I **never** found out about that for ages. (Orkney, Y)
(b) We **never** got a car til the mid-fifties. (Stirling & Falkirk, O)

Although not a feature of standard English, this form is widespread across varieties of English (Cheshire *et al.* 1993; Kortmann & Szmrecsanyi 2004) – including Scots (Macafee 2011). Judgments for non-temporal *never* in SCOSYA are very clear. It is the most highly rated example in the data set, with a median and a mode of 5 across all participants, and an acceptance rate of 88 per cent.

In summary, judgments for highly localised features are very reliable. When intuitions do seem to diverge from spoken usage, it is when judging more widespread features (table 4), but supralocality in itself does not predict 'failure'. We use Labov's shorthand of 'failure' and 'success' of intuitions to mean that they fail to track or succeed in tracking the statistical patterns in corpus data. We will explore this further in section 5.4.

5.3 *Social stigma*

The salience of and social stigma surrounding the variants in this study was detailed in section 4.1. Firstly, *need* + PAST is known to be a covert feature of Scots (Aitken 1984), and it is therefore unsurprising that social stigma does not interfere in speakers' intuitions here.

There was little previous evidence about the salience of *div* in Scots. From the SCOSYA data, we can see that it appears to be overt to speakers outside the North East and Borders. For example, there were four examples of *div* in our corpus that were instances of metalinguistic discussion (these examples were removed from our count of tokens of this variable) (25).

(25) Allan: I was just saying to Innes, you know, 'div you ken something', that's more Wick isn't it? 'I div.'
Innes: Mhm. 'Div you mind, my boy?'
Claire: Aye, I would say Wick is more 'div you ken'. (Caithness, O)

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

The speakers in (25) are from Caithness. Evidently the feature is salient to them, although they ascribe it to speakers in a different part of Caithness, who do not use the feature.[15] Within the North East and Borders, we don't see this same kind of discussion – indicating *div* may not be salient for speakers who actually use it. So, although we cannot rule out *div* as salient for speakers in the North East or Borders, there is no evidence to suggest it is. It falls out from this that *div* is unlikely to be stigmatised, and thus stigma is not relevant to judgments of this feature.

Our third feature, NC, is highly salient and highly stigmatised, and it seems that even with a judgment task designed specifically to reduce the Observer's Paradox as far as possible, an overtly stigmatised variant like NC is still at risk of diverging from vernacular patterns. However, as we saw in section 4.4, participants are relatively successful at judging NC with *nowhere*: this variant is only attested in the North East, where it was rated more highly in the judgment task as well. There is no evidence that NC with *nowhere* is subject to different social evaluation as NC with *nothing*. In Smith & Holmes-Elliott's (2022) study, for example, speakers in Buckie 'control' their usage of NC with outsiders regardless of the indefinite used (e.g. *nothing*, *nowhere*). It seems, therefore, that social stigma alone cannot account for speaker intuitions diverging from the speech patterns in corpora. Instead, in the final section, we will consider the interplay between social stigma, regionality and the construction of a local identity in judgment tasks.

5.4 Stigma x local identity

As with the features in our study, Aitken's (1984: 25) list of 'vulgarisms' contains features which vary in their regionality. Some are highly regional, such as *so it is* tags (26), attested in the south west of Scotland, and sentence-final *but* (27), which means something like 'though' and is attested in Glasgow/Ayrshire.

(26) It's brilliant, **so it is.** (Dumfries, Y)
(27) Does it play good music **but**? (Ayrshire, Y)

However, some are vernacular universals – namely, NC, and 'the well-known syncretism of past tense and past participle forms' (Aitken 1984: 25) (28a–b).

(28) (a) I **done** a really good job of it. (Lothian, Y)
(b) It was Monday I **seen** him on. (Western Isles, O)

As we saw, speakers across Scots varieties often did not match the attested speech patterns when judging NC with *nothing*. The same issue arises at least for older speakers when it comes to irregular past tense forms. Older speakers rate examples like (28a–b) low (*seen* – 35 per cent acceptance rate; median=2, mode=1; *done* – 47 per cent acceptance rate; median=3, mode=1), despite attestations across Scotland. Younger speakers, on the other hand, accept the construction at a higher rate (*seen* – 56

[15] Speakers' attempts to reason about dialects other than their own are often unsuccessful (Henry 1995: 56–7).

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

per cent acceptance rate; median=4, mode=5; *done* – 70 per cent acceptance rate; median=5, mode=5). We suggest this is due to a change in levels of stigma rather than a usage change, as irregular past tense forms have been part of vernacular Scots for at least one hundred years (Grant & Dixon 1921).

The results for these two features contrast with what we saw for NC with *nowhere*, despite it also being a vernacular universal. We thus propose that, given a well-controlled judgment task, a salient variant which is stigmatised can lead intuitions to fail *if* the variant is not able to contribute to the construction of a local (cultural) identity. We define local identity following Hazen (2002: 241): 'how speakers conceive themselves in relation to their local and larger regional communities'.

Judgment tasks are ultimately metalinguistic, allowing speakers to project their identity by expressing their intuitions – and identity formation is locally oriented both in production (e.g. Labov 1963; Eckert 2000; Hazen 2002; Stuart-Smith, Timmins & Tweedie 2007) and judgments (e.g. Jamieson 2020). Reflecting on the relationship between Scots and (Scottish) English (see section 3.1), for the participants in the SCOSYA corpus, we would expect a stronger association with the broad Scots end of the continuum. The participants came from families who had been in the community for generations, and had themselves remained in the communities and networks they grew up in. Their cultural identity is likely to be oriented to that community, and so when given the opportunity to construct a self through a metalinguistic task – particularly in a task which is administered by someone else from the community, and specifically designed to encourage them to access the broad Scots end of their linguistic continuum – it is this alignment we would expect to emerge.

Both covert and overt highly regional forms allow successful judgments that map to their attested usage patterns – even if they are stigmatised, like NC with *nowhere*. In relation to identity construction, this makes sense – if the form is covert, it is acceptable with little hesitation; if it is overt, participants may be aware of stigma, but can reframe this as pride in a local variant. However, it appears that at least in the Scots context, English vernacular universals which are used across varieties do not contribute to local identity, and so any desire to push past prescriptive stigma and index regional associations is lessened. If the variant is covert, this may not affect judgments – as we see with non-temporal *never*. This is simply perceived as a feature of the grammar. However, if the variant is overt, intuitions may fail – as with NC with *nothing*. Here, the feature is stigmatised in such a way that it is known as an 'incorrect' feature of *English*. Participants who wish to align themselves with the broad Scots end of the Scots language continuum also wish to distance themselves from the non-local, 'English' end of the continuum (e.g. Le Page & Tabouret-Keller 1985).

It may be the case that (overt) English vernacular universals cannot regularly contribute to identity construction more broadly across dialects; this would require further study. However, certainly in the case of the Scots language continuum, the pull to construct a local identity appears to lead judgments to fail when a feature is both stigmatised and used by speakers with whom the participants do not wish to align themselves.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

6 Conclusions

In this article we investigated how speakers' intuitions map to production data using SCOSYA, a large-scale data set which allows us to compare acceptability judgments from 530 speakers with over 3 million words of spoken data from those same speakers. Speakers' intuitions, as expressed through the judgment task, broadly matched the corpus patterns for *need* + PAST and *div*, but failed to do so for NC with *nothing*. However, introspective judgment tasks are more successful in tracking the patterning of NC with *nowhere*, a more localised variant, in the corpus. We discussed the results in terms of frequency, regionality and salience/stigma, arguing that speakers are generally successful in judging covert features regardless of whether they are stigmatised, while stigma attached to salient features can affect speakers' judgments. However, if a stigmatised variant can be ascribed as a marker of local identity, speakers are more likely to accept it, while overt vernacular universals are more likely to be rejected.

As noted in the introduction, Rickford (2019) calls for a:

> concentrated effort to determine what kind of intuitive judgments are more robust than others, what factors influence their variability, and what methods we might use for calibrating then against observational and other evidence. (Rickford 2019: 102)

The analysis presented here provides the foundations for future research across a much wider range of morphosyntactic variables in uncovering the complexities of when intuitions (don't) fail.

Authors' addresses:

Department of Language & Linguistic Science
University of York
Vanbrugh College
York YO10 5DD
UK
e.jamieson@york.ac.uk

Department of English Language & Linguistics
University of Glasgow
13 University Gardens
Glasgow G12 8QQ
UK
jennifer.smith@glasgow.ac.uk

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Department of Linguistics
Queen Mary University of London
Arts One
Mile End Rd
London E1 4NS
UK
d.j.adger@qmul.ac.uk

Department of Linguistics and English Language
University of Edinburgh
Dugald Stewart Building
3 Charles Street
Edinburgh EH8 9AD
UK
caroline.heycock@ed.ac.uk

Department of Linguistics
New York University
10 Washington Place
New York 10003
USA
gst1@nyu.edu

References

Adger, David & Jennifer Smith. 2010. Variation in agreement: A lexical feature-based approach. *Lingua* 120(5), 1109–34.
Aitken, Jack. 1984. Scottish accents and dialects. In Peter Trudgill (ed.), *Language in the British Isles*, 94–114. Cambridge: Cambridge University Press.
Alim, Samy, John Rickford & Arnetha Ball. 2016. *Raciolinguistics*. Oxford: Oxford University Press.
Anderwald, Lieselotte. 2005. Negative concord in British dialects. In Yoko Iyeiri (ed.), *Aspects of English negation*, 113–37. Amsterdam: John Benjamins.
Arppe, Antte & Juhani Järvikivi. 2007. Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3(1), 99–109.
Bader, Markus & Jana Häussler. 2010. Toward a model of grammaticality judgments. *Journal of Linguistics* 46(2), 273–330.
Barbiers, Sjef & Hans Bennis. 2007. The Syntactic Atlas of the Dutch Dialects: A discussion of choices in the SAND project. *Nordlyd* 34, 53–72.
Barras, Claude, Edouard Geoffrois, Zhibiao Wu & Mark Liberman. 2001. Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication* 33(1–2), 5–22.
Beal, Joan. 1993. The grammar of Tyneside and Northumbrian English. In Milroy & Milroy (eds.), 187–212.
Bender, Emily. 2001. Syntactic variation and linguistic competence. PhD dissertation, Stanford University.

https://doi.org/10.1017/S1360674323000679 Published online by Cambridge University Press

Benincà, Paola & Cecilia Poletto. 2007. The ASIS enterprise: A view on the construction of a syntactic atlas for the Northern Italian dialects. *Nordlyd* 34, 35–52.
Bermel, Neil & Luděk Knittl. 2012. Corpus frequency and acceptability judgments: A study of morphosyntactic variation in Czech. *Corpus Linguistics and Linguistic Theory* 8(2), 241–76.
Blanchette, Frances. 2017. Micro-syntactic variation in American English negative concord. *Glossa* 2(65), 1–32.
Bresnan, Joan. 2007. Is knowledge of syntax probabilistic? Experiments with the English dative alternation. In Sam Featherston & Wolfgang Sternefeld (eds.), *Roots: Linguistics in search of its evidential base*, 75–96. Berlin: Mouton de Gruyter.
Brown, Keith. 1991. Double modals in Hawick Scots. In Peter Trudgill & J. K. Chambers (eds.), *Dialects of English*: *Studies in grammatical variation*, 74–103. London and New York: Longman.
Brown, Keith & Martin Millar. 1980. Auxiliary verbs in Edinburgh speech. *Transactions of the Philological Society* 78(1), 81–133.
Buchstaller, Isabelle. 2009. The quantitative analysis of morphosyntactic variation: Constructing and quantifying the denominator. *Linguistics and Language Compass* 3(4), 1010–33.
Buchstaller, Isabelle & Karen Corrigan. 2011. How to make intuitions succeed: Testing methods for analysing syntactic microvariation. In Warren Maguire & April McMahon (eds.), *Analysing variation in English*, 30–48. Cambridge: Cambridge University Press.
Carden, Guy & Thomas Dieterich. 1981. Introspection, observation, and experiment: An example where experiment pays off. In Peter Asquith & Ronald Giere (eds.), *PSA 1980*, 583–97. East Lansing, MI: Philosophy of Science Association.
Chambers, J. K. 1995. *Sociolinguistic theory: Linguistic variation and its social significance*. Oxford: Blackwell.
Cheshire, Jenny. 1982. *Variation in an English dialect: A sociolinguistic study*. Cambridge: Cambridge University Press.
Cheshire, Jenny. 1999. Taming the vernacular: Some repercussions for the study of syntactic variation and spoken grammar. *Cuadernos de Filología Inglesa* 8, 59–80.
Cheshire, Jenny. 2005. Age and generation-specific use of language. In Ulrich Ammon, Norbert Dittmar, Klaus Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An introductory handbook of the science of language and society*, 1552–63. Berlin: Mouton de Gruyter.
Cheshire, Jenny & Sue Fox. 2009. *Was/were* variation: A perspective from London. *Language Variation and Change* 21(1), 1–38.
Cheshire, Jenny, Viv Edwards & Pamela Whittle. 1993. Urban British dialect grammar: The question of dialect levelling. In Milroy & Milroy (eds.), 53–94.
Childs, Claire. 2017. Variation and change in English negation: A cross-dialectal perspective. PhD dissertation, Newcastle University.
Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, Noam. 1966. *Topics in the theory of generative grammar*. The Hague: Mouton.
Coats, Steven. 2023. Double modals in contemporary British and Irish speech. *English Language & Linguistics* 27(4), 693–718.
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.
Cornips, Leonie & Karen Corrigan. 2005. *Syntax and variation: Reconciling the biological with the social*. Amsterdam: John Benjamins.
Cornips, Leonie & Cecilia Poletto. 2005. On standardizing syntactic elicitation techniques. *Lingua* 115(7), 939–57.
Coupland, Nikolas. 1988. *Dialect in use*. Cardiff: University of Wales Press.
Coupland, Nikolas & Hywel Bishop. 2007. Ideologised values for British accents. *Journal of Sociolinguistics* 11(1), 74–93.

Destruel, Emilie, David Beaver & Elizabeth Coppock. 2019. It's not what you expected! The surprising nature of cleft alternatives in French and English. *Frontiers in Psychology* 10(1400).
Dictionary of the Scots Language. 2004. 'Dae'. Scottish Language Dictionaries Ltd. www.dsl.ac.uk/entry/snd/dae_v (accessed 12 December 2023).
Divjak, Dagmar. 2017. The role of lexical frequency in the acceptability of syntactic variants: Evidence from that-clauses in Polish. *Cognitive Science* 41(2), 354–82.
Drager, Katie & Thomas Kettig. 2021. Sociophonetics. In Rachael-Anne Knight & Jane Setter (eds.), *The Cambridge handbook of phonetics*, 551–77. Cambridge: Cambridge University Press.
Duncan, Daniel. 2021. A note on the productivity of the alternative embedded passive. *American Speech* 96(4), 481–90.
Eckert, Penelope. 2000. *Linguistic variation as social practice*. Oxford: Blackwell.
Edelstein, Elspeth. 2014. This syntax needs studied. In Raffaella Zanuttini & Laurence Horn (eds.), *Micro-syntactic variation in North American English*, 242–68. Oxford: Oxford University Press.
Edwards, Viv. 1993. The grammar of southern British English. In Milroy & Milroy (eds.), 214–42.
Eide, Kristin & Tor Åfarli. 2020. Dialects, registers and intraindividual variation: Outside the scope of generative frameworks? *Nordic Journal of Linguistics* 43(3), 233–48.
Feagin, Crawford. 1979. *Variation and change in Alabama English: A sociolinguistic study of the white community.* Washington, DC: Georgetown University Press.
Featherston, Sam. 2005. The Decathlon Model of empirical syntax. In Reis & Kepser (eds.), 187–208.
Fisher, Sabriya. 2018. Change over time in the grammar of African American English. *University of Pennsylvania Working Papers in Linguistics* 24(2), 29–38.
Goodall, Grant. 2021. *The Cambridge handbook of experimental syntax*. Cambridge: Cambridge University Press.
Grant, William. 1931. *The Scottish national dictionary.* Edinburgh: Scottish National Dictionary Association.
Grant, William & James Dixon. 1921. *Manual of modern Scots*. Cambridge: Cambridge University Press.
Green, Lisa. 2010. *Language and the African American child*. Cambridge: Cambridge University Press.
Guy, Gregory & Robert Bayley. 1995. On the choice of relative pronouns in English. *American Speech* 70(2), 148–62.
Hazen, Kirk. 1996. Linguistic preference and prescriptive dictum: On the phonological and morphological justification of *ain't*. In Jennifer Arnold, Renée Blake, Brad Davidson, Scott Schwenter & Julie Solomon (eds.), *Sociolinguistic variation: Data, theory and analysis*, 101–12. Stanford, CA: CSLI.
Hazen, Kirk. 2002. Identity and language variation in a rural community. *Language* 78(2), 240–57.
Henry, Alison. 1995. *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford: Oxford University Press.
Henry, Alison. 2005. Non-standard dialects and linguistic data. *Lingua* 115(11), 1599–1617.
Hughes, Arthur & Peter Trudgill. 1979. *English accents and dialects: An introduction to social and regional varieties of British English*. London: Arnold.
Jamieson, E. 2020. Viewing dialect change through acceptability judgments: A case study in Shetland dialect. *Glossa* 5(19), 1–28.
Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor Anders Åfarli & Øystein Vangsnes. 2009. The Nordic Dialect Corpus – an advanced research tool. In Jokinen & Bick (eds.), 73–80.
Johnston, Paul. 1997. Regional variation. In Charles Jones (ed.), *The Edinburgh history of the Scots language*, 433–513. Edinburgh: Edinburgh University Press.

Johnston, Paul. 2007. Scots and Scottish English. In David Britain (ed.), *Language in the British Isles*, 105–21. Cambridge: Cambridge University Press.
Jokinen, Kristiina & Eckhard Bick (eds.). 2009. *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009*. Odense: Northern European Association for Language Technology.
Kempen, Gerard & Karin Harbusch. 2005. The relationship between grammaticality ratings and corpus frequencies: A case study into word order variability in the midfield of German clauses. In Reis & Kepser (eds.), 329–49.
Kortmann, Bernd & Benedikt Szmrecsanyi. 2004. Global synopsis: Morphological and syntactic variation in English. In Bernd Kortmann, Edgar Schneider, Clive Upton, Kate Burridge & Rajand Mesthrie (eds.), *A handbook of varieties of English*, vol. 2, 1142–1202. Berlin: Mouton de Gruyter.
Labov, William. 1963. The social motivation of a sound change. *Word* 19, 273–307.
Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1(1), 97–120.
Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.), *Language in use: Readings in sociolinguistics*, 28–66. Englewood Cliffs, NJ: Prentice Hall.
Labov, William. 1994. *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
Labov, William. 1996. When intuitions fail. In Lisa McNair, Kora Singer, Lise Dobrin & Michelle Aucon (eds.), *CLS 32: Papers from the parasession on theory and data in linguistics*, 77–106. Chicago: Chicago Linguistic Society.
Labov, William. 2001. *Principles of linguistic change: Social factors*. Oxford: Blackwell.
Labov, William, Paul Cohen, Clarence Robins & John Lewis. 1968. *A study of the non-standard English of Negro and Puerto Rican speakers in New York City*. Cooperative Research Report 3288. New York: Columbia University.
Labov, William, Sharon Ash, Maya Ravindranath, Tracey Weldon, Maciej Baranowski & Naomi Nagy. 2011. Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15, 431–63.
Langsford, Steven, Andrew Perfors, Andrew Hendrickson, Lauren Kennedy & Danielle Navarro. 2018. Quantifying sentence acceptability measures: Reliability, bias and variability. *Glossa* 3(1), 1–34.
Lavandera, Beatriz. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7, 171–83.
Le Page, Robert & Andrée Tabouret-Keller. 1985. *Acts of identity*. Cambridge: Cambridge University Press.
Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen & Øystein Vangsnes. 2009. The Nordic Dialect Database: Mapping microsyntactic variation in the Scandinavian languages. In Jokinen & Bick (eds.), 283–6.
Macafee, Caroline. 2011. Characteristics of non-standard grammar in Scotland. MS, University of Aberdeen.
Macaulay, Ronald. 1991. *Locating dialect in discourse: The language of honest men and bonnie lasses in Ayr*. Oxford: Oxford University Press.
Maher, Zach & Jim Wood. 2011. Needs washed. *Yale Grammatical Diversity Project*. http://ygdp.yale.edu/phenomena/needs-washed (accessed 12 December 2023).
Millar, Robert McColl. 2007. *Northern and Insular Scots*. Edinburgh: Edinburgh University Press.
Milroy, Jim & Lesley Milroy. 1993. *Real English: The grammar of English dialects in the British Isles*. London: Longman.

Montgomery, Michael. 1997. The Scotch-Irish element in Appalachian English: How broad? How deep? In H. Tyler Blethen & Curtis Wood (eds.), *Ulster and North America: Transatlantic perspectives on the Scotch-Irish*, 189–212. Tuscaloosa: University of Alabama Press.
Munro, Pamela. 2017. Field linguistics: Gathering language data from native speakers. In Mark Aronoff & Jamie Rees-Miller (eds.), *The handbook of linguistics*, 57–73. Oxford: Blackwell.
Murray, Timothy & Beth Simon. 2002. At the intersection of regional and social dialects: The case of *like* + past participle in American English. *American Speech* 77(1), 32–69.
Murray, Thomas, Timothy Frazer & Beth Simon. 1996. *Need* + past participle in American English. *American Speech* 71(3), 255–71.
Niedzielski, Nancy & Dennis Preston. 2003. *Folk linguistics*. Berlin: Mouton de Gruyter.
Parrott, Jeffrey. 2007. Distributed morphological mechanisms of Smith Island *weren't* levelling. *University of Pennsylvania Working Papers in Linguistics* 13(1), 23.
Pichler, Heike. 2009. The functional and social reality of discourse variants in a northern English dialect: I DON'T KNOW and I DON'T THINK compared. *Intercultural Pragmatics* 6(4), 561–96.
Poletto, Cecilia. 1993. Subject clitic–verb inversion in North Eastern Italian dialects. In Adriana Belletti (ed.), *Syntactic theory and the dialects of Italy*, 95–135. Turin: Rosenberg e Sellier.
Preston, Dennis. 1989. *Perceptual dialectology*. Dordrecht: Foris.
Reis, Marga & Stephan Kepser. 2005. *Linguistic evidence: Empirical, theoretical and computational perspectives*. Berlin: Mouton de Gruyter.
Rickford, John. 2019. *Variation, versatility and change in sociolinguistics and creole studies*. Cambridge: Cambridge University Press.
Rickford, John, Tom Wasow, Norma Mendoza-Denton & Juli Espinoza. 1995. Syntactic variation and change in progress: Loss of the verbal coda in topic-restricting 'as far as' constructions. *Language* 71(1), 102–31.
Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris.
Robinson, Mary & Gary Thoms. 2021. On the syntax of English variable negative concord. *University of Pennsylvania Working Papers in Linguistics* 27(1), 24.
Roland, Douglas, Frederic Dick & Jeffrey Elman. 2007. Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language* 57(3), 348–79.
Rowe, Charlie. 2007. 'He divnt gan tiv a college ti di that, man!': A study of *do* (and *to*) in Tyneside English. *Language Sciences* 29(2–3), 360–71.
Rupp, Laura & David Britain. 2019. *Linguistic perspectives on a variable morpheme: Let's talk about -s*. London: Palgrave.
Smith, Jennifer. 2000. 'You Ø na hear o' that kind o' things': Negative *do* in Buckie Scots. *English World-Wide* 21(2), 231–59.
Smith, Jennifer. 2001. Negative concord in the Old and New World: Evidence from Scotland. *Language Variation and Change* 13, 109–34.
Smith, Jennifer. 2012. Scottish English and varieties of Scots. In Bernd Kortmann & Kerstin Lunkenheimer (eds.), *The Mouton world atlas of variation in English*, 21–9. Berlin: Mouton de Gruyter.
Smith, Jennifer & Sophie Holmes-Elliott. 2022. They never did nothing: Negation and the sociolinguistic monitor. In Tanya K. Christensen & Juel Jensen (eds.), *Explanations in sociosyntactic variation*, 58–89. Cambridge: Cambridge University Press.
Smith, Jennifer, Mercedes Durham & Hazel Richards, 2013. The social and linguistic in the acquisition of sociolinguistic norms: Caregivers, children and variation. *Linguistics* 51(2), 285–324.
Smith, Jennifer, David Adger, Brian Aitken, Caroline Heycock, E Jamieson & Gary Thoms. 2019. *The Scots Syntax Atlas*. scotssyntaxatlas.ac.uk (accessed 12 December 2023).

Sprouse, Jon & Diogo Almeida. 2017. Setting the empirical record straight: Acceptability judgments appear to be reliable, robust and replicable. *Behavioural and Brain Sciences* 40 (E311).
Stabley, Rhodes. 1959. 'Needs painted' etc in western Pennsylvania. *American Speech* 34(1), 69–70.
Strelluf, Christopher. 2020. *needs*+PAST PARTICIPLE in regional Englishes on Twitter. *World Englishes* 39(1), 119–34.
Strelluf, Christopher. 2022. Regional variation and syntactic derivation of low-frequency *need*-passives on Twitter. *Journal of English Linguistics* 50(1), 39–71.
Stuart-Smith, Jane, Claire Timmins & Fiona Tweedie. 2007. 'Talkin' Jockey?' Variation and change in Glaswegian accent. *Journal of Sociolinguistics* 11(2), 221–60.
Tagliamonte, Sali A., Jennifer Smith & Helen Lawrence. 2005. No taming the vernacular! Insights from the relatives in northern Britain. *Language Variation and Change* 17(1), 75–112.
Thoms, Gary & Craig Sailor. 2018. When silence gets in the way: Extraction from do-ellipsis in British dialects. In Sherry Hucklebridge & Max Nelson (eds.), *Proceedings of NELS 48*, 145–54. Amherst, MA: GLSA.
Thoms, Gary, David Adger, Caroline Heycock & Jennifer Smith. 2019. Syntactic variation and auxiliary contraction: The surprising case of Scots. *Language* 95(3), 421–55.
Thoms, Gary, David Adger, Caroline Heycock, E Jamieson & Jennifer Smith. 2023. English contracted negation revisited: Evidence from varieties of Scots. *Language* 99(4), 726–59.
Tortora, Christina. 2006. The case of Appalachian expletive *they*. *American Speech* 81(3), 266–96.
Tortora, Christina, Beatrice Santorini, Frances Blanchette & C. E. A. Diertani. 2017. *The Audio-Aligned and Parsed Corpus of Appalachian English* (AAPCAppE). www.aapcappe.org (accessed 12 December 2023).
Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winifred Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics*, 95–195. Austin: University of Texas Press.
Wood, Jim & Raffaella Zanuttini. 2018. Datives, data and dialect syntax in American English. *Glossa* 3(1), 87.
Wood, Jim, Raffaella Zanuttini, Laurence Horn & Jason Zentz. 2020. Dative country: Markedness and geographical variation in Southern dative constructions. *American Speech* 95(1), 3–45.
Zanuttini, Raffaella, Jim Wood, Jason Zentz & Laurence Horn. 2018. The Yale Grammatical Diversity Project: Morphosyntactic variation in North American English. *Linguistics Vanguard* 4(1), 1–15.