

*English Language and Linguistics*, page 1 of 26. © The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

doi:[10.1017/S1360674323000126](https://doi.org/10.1017/S1360674323000126)

## Double modals in contemporary British and Irish speech<sup>1</sup>

STEVEN COATS 

*University of Oulu*

(Received 21 March 2022; revised 27 February 2023)

This article reports on the use of double modals, a non-standard syntactic feature, in the contemporary speech of the UK and Ireland. Most data on the geographic extent of the feature and its combinatorial types come from surveys or acceptability ratings or from older attestations focused on northern England, Scotland or Northern Ireland, with relatively few attestations in naturalistic data and from England and Wales. Manual verification of double modals in a large corpus of geolocated Automatic Speech Recognition transcripts from YouTube videos of local government channels from the UK and Ireland shows that the feature exhibits a larger inventory of combinatorial types than has previously been found and is attested in speech from throughout the UK and Ireland. The development may be related to ongoing changes in the semantic space occupied by modal auxiliaries in English.

**Keywords:** double modals, corpus linguistics, dialects, British Isles, spoken language

### 1 Introduction

The use of more than one modal auxiliary within a single verbal phrase (e.g. *I might could do that* or *you will can help me*) is a syntactic feature of non-standard English varieties in the British Isles, North America and elsewhere. Although double modals have been widely studied in the past seventy years, questions remain as to their syntactic behavior, semantics and history, as well as to the pragmatic contexts that favor their use, the inventory of possible double modal types and their geographic distribution in different English varieties. Double modals are of theoretical interest as they represent micro-syntactic variation that may shed light on the status of the verbal phrase: although several proposals have been made as to the constituency relations and phrase structure of double modal constructions (Boertien 1986; Battistella 1995; Hasty 2012; see also Morin *et al.* 2020 for a discussion of double modals from the perspective of

<sup>1</sup> The author wishes to extend thanks to Karen Corrigan for providing access to the NECTE/DECTE corpus data, and to the Hosting and Distribution Team for the *International Corpus of English* at the University of Zurich (Marianne Hundt, Hans Martin Lehmann and Gerold Schneider) for access to the ICE-Ireland corpus. Thanks also to two anonymous reviewers and to Laurel Brinton at *English Language and Linguistics* for helpful comments on a draft version of this article. Finland's Centre for Scientific Computing ([www.csc.fi](http://www.csc.fi)) provided access to computing and storage resources.

Construction Grammar), their ultimate status has not yet been conclusively resolved, and the semantic status of some double modals remains unclear. Pragmatically, most double modal types have been proposed to be restricted to specific contexts: they are used for ‘the negotiation of a speaker’s wants or needs’ in polite, cautious conversation in order to mitigate face-threatening situations (Mishoe & Montgomery 1994: 12; see also Montgomery 1998; Schneider 2005; Hasty 2015).

The feature can also be adduced as evidence for different postulates about the evolution of English syntax (Nagle 1993; Fennell & Butters 1996; Coupé & van Kemenade 2009): either as a fossilized relic of two-verb constructions, predating the grammaticalization of modals from full lexical verbs to auxiliaries, or as a feature that emerges as a result of grammaticalization in the Middle English period. Despite their status as a nonpareil object of theoretical interest, the inventory of possible double modal combinations and the extent of their use in regional varieties of English have not been well established by empirical studies. The feature is known to occur in speech in Scotland and northern England, and, to some extent, in Northern Ireland, but inventories of double modals differ between Southern US English and British Isles English varieties, a fact which poses challenges for a theory of simple transmission of the feature to Southern American English by Scots-Irish settlers (Schneider 2005; Zullo *et al.* 2021). Information about the feature’s use in the Irish Republic is limited, and it has been little studied in the English Midlands, southern England or in Wales. Double modals have not been a focus of most previous studies of the dialect syntax of British Isles varieties of English based on corpus materials (e.g. Hernández, Kolbe & Schulz 2011; Szmrecsanyi 2013).

A principal difficulty in studying double modals in spoken English is their infrequency: like many syntactic features, they occur only rarely in spontaneous speech, even in varieties in which they are known to be used, and may also be subject to specific pragmatic constraints such as those noted above. The rareness of the feature means that much of the data in studies of double modals from the United States and from the British Isles consists of responses to acceptability judgment survey items or forms elicited in the context of interviews. In the UK and Ireland, double modal usage has been surveyed in northern England (McDonald & Beal 1987), Scotland (Brown & Millar 1980; Miller & Brown 1982; Brown 1991; Bour 2014, 2015; Smith *et al.* 2019; Morin 2021b) and Ireland (Montgomery & Nagle 1994; Montgomery 2006; Hickey 2007), but only a few double modals are attested in corpora of transcribed speech (Beal 2004: 127; Szmrecsanyi 2013: 37), and recorded examples of naturalistic use are rare: for example, Brown & Millar report one recorded instance from interviews conducted in Scotland in the 1970s in the context of a project investigating modality (1980: 122), and the *Scots Syntax Atlas* (Smith *et al.* 2019) provides a single naturalistic recorded example.

Commenting on the data for northern England, Beal remarks that there are ‘huge gaps in the geographical coverage, which need to be filled by the collection of new data from cities such as Carlisle, Lancaster, Liverpool and Manchester, and the processing of data already collected elsewhere’ (2004: 139). The same could be said for other areas of England, for Ireland, for much of Scotland and for Wales.

This study attempts to provide a starting point for further research into double modal usage in the British Isles by documenting their naturalistic use in the *Corpus of British Isles Spoken English* (CoBISE; Coats 2022b, 2023), a collection of geolocated Automatic Speech Recognition (ASR) transcripts of recordings uploaded to the channels of local government bodies in the UK and Ireland. The study assesses which double modal forms are in use in contemporary speech in council meetings in the UK and Ireland and assigns them to locations based on corpus metadata, using methods similar to those presented in Coats (2022a) for North America. The rest of the article is organized as follows: in the next section, a brief overview of some previous work on double modals in Britain and Ireland is provided, focusing primarily on the geographical distribution of the feature and the inventory of attested types. Section 3 describes the data for the study, from CoBISE, and the procedure used to identify two-modal sequences in the corpus and distinguish authentic double modals from various kinds of false positives. Section 4 presents the results for England, Scotland, Wales, Northern Ireland and the Irish Republic in terms of frequencies of double modal types; exploratory maps showing frequency count, relative frequencies and a local autocorrelation statistic are also presented. In section 5, the findings are discussed and a preliminary interpretation undertaken; several caveats are noted. Section 6 summarizes the study and suggests some directions for future research.

## 2 Previous work

### 2.1 Scotland and northern England

A few older dialectological works document use of double modals in Scotland and England. Murray (1873) notes the forms *will can* and *will ought*. Wright's *Dialect Dictionary* (1898–1905) remarks on regional use of *will can/ll can* in southern Scotland and northern England and *wouldn't could* in northern England (I: 502). *Used to could* (and phonetic variants) is attested by Wright in northern England and the Midlands, as well as in Oxfordshire, but not in Scotland (I: 502; VI: 333). *Should ought* and *shouldn't ought* are attested in Nottinghamshire and Lincolnshire (IV: 365, V: 350). Double modal types with *may* or *might*, common in America, are not attested by Wright.

Brown & Millar (1980) report use of the double modal forms *will can* and *will no can* by an informant and provide the results of a survey of university students from Edinburgh: 86 percent of students were familiar with *he'll no can*, 35 percent with *will can* and 9 percent with the question form *will he no can?* (1980: 121). At least one informant had heard or used *would could*, *used to could*, *might could* or *should ought to*, but only the type *won't can* was used spontaneously in a recorded interview (1980: 122). Miller & Brown report that in Scotland, 'only *will* and *might can* be the first member [of a double modal]'; in terms of frequency, *will can* is reported to be 'very common' in speech and *might could* 'increasingly common' (1982: 13). The forms *might can*, *might should* and *might would* are also attested.

Brown (1991) reports on double modal usage in Hawick, Scotland: in addition to the forms listed by Brown & Millar (1980), the types *should can*, *must can*, *should could*, *would could* and *must could* are attested; *would can* and *will could* are reported to be possible. According to Brown, double modals exhibit tense congruency (i.e. *will can* or *would could* are possible, but not *will could* or *would can*) (1991: 97). It is not clear if the large number of examples provided by Brown represent naturalistic usages or are constructed sentences based on expert knowledge.

Miller reports the forms *'ll can*, *might could*, *might can*, *would could*, *might should* and *might would* to be used in Scotland (2004: 53). Melchers finds *'ll no can* in Orkney (2004: 40). Bour (2014, 2015) reports on a survey-based investigation of double modals in the Scottish Borders region. In addition to previously reported forms such as *will can*, *should can*, *would can* and *should could*, some informants were familiar with non-canonical types such as *may not could*, *may can*, *should ought to* and *mustn't could have* (2015: 55–6). A survey conducted in Hawick by Morin (2021b) found *might can* to have the highest rate of self-reported usage; *may could* and *may should*, not attested in most previous literature, were also familiar to some respondents.

Informants for the *Scots Syntax Atlas* (Smith *et al.* 2019)<sup>2</sup> rated their own use of seven short sentences containing double modals from 1 ('I would never say that') to 5 ('I would definitely say that').<sup>3</sup> Responses show that double modals are not commonly used, but are more frequent in the Scottish Borders region and in Dumfries and Galloway, compared to other parts of Scotland. In contrast to previous literature in which *will can* is reported to be the most common Scottish double modal type, the sentence with *must can* shows the highest average value in the survey. As far as corpus data are concerned, according to Corbett, no double modals are recorded in the spoken portion of the *Scottish Corpus of Texts and Speech* (2014: 270).<sup>4</sup>

In northern England, double modals are used in Newcastle and Tyneside. McDonald & Beal report the forms *might can*, *must can*, *will can*, *might could*, *must could* and *would could* to be possible (1987: 47). Studies based on acceptability judgments and self-reports of use suggest that use of the feature is in decline in Northumberland (Beal 2004: 128). Fifteen percent of informants surveyed from the north of County Durham rated sentences containing double modals to be acceptable (McDonald 1981, as cited in Beal 2004: 128). A survey of young people conducted in Northumberland found less than 10 percent of teenagers considered double modals to be acceptable (Beal 2010: 38).

The feature is not well attested in corpora of northern English speech. A single double modal (*wouldn't could*) was recorded in a corpus of 300,000 words of transcribed speech collected in the 1970s from Tyneside and rural environs (McDonald & Beal 1987: 47). In the *Newcastle Corpus of Tyneside English* (NECTE; Allen *et al.* 2007), a 440,000-word

<sup>2</sup> <https://scotssyntaxatlas.ac.uk>

<sup>3</sup> (1) *You used to could get the papers at the shop on the corner.* (2) *Will you can meet me at six?* (3) *He might can help you next week.* (4) *Why would I might know him?* (5) *They must could see us at some point.* (6) *He must can swim, I see him at the pool all the time.* (7) *He'll can do it next week.*

<sup>4</sup> [www.scottishcorpus.ac.uk](http://www.scottishcorpus.ac.uk)

corpus of interview transcripts from the 1960s–70s and the 1990s, one double modal is recorded (*you'll probably not can remember*; Beal 2004: 127). No additional double modals can be found in the data for the updated version of the corpus, the *Diachronic Electronic Corpus of Tyneside English* (DECTE; Corrigan *et al.* 2012).

## 2.2 Ireland

Compared to Scotland and northern England, double modals are less extensively attested in Ireland. They are found mainly in speech from Northern Ireland, where they were brought by northern English and Scottish settlers at the time of the Ulster Plantation beginning in the seventeenth century (Montgomery & Nagle 1994; Hickey 2007; Corrigan 2011). Gregg attests use of *I'll no can stay* in Ulster Scots (1972: 131), and Montgomery & Nagle report the types *might could*, *might should*, *should ought to*, *used to could*, *used to would* and *will can* to be in use in Ulster English in the 1990s (1994: 102). An informal survey of twenty informants from Belfast and County Antrim conducted in 1990 and 1991 showed that a few respondents were familiar with *might could*, *might can* and *will can*, but not with other double modal types (*ibid.*). Corrigan found no double modals in a 52,000-word database of transcribed personal narratives collected in the mid twentieth century in South Armagh, Northern Ireland (2000: 27). Hickey (2007) found no naturalistic uses of double modals in recordings made from the 1980s to the 2000s in the Republic of Ireland, mainly in Dublin and Waterford. The *Survey of Irish English Usage* found that 7 percent of informants rated the sentence *He might could come after all* to be 'no problem' (Hickey 2004: 191).<sup>5</sup> Hickey remarks that double modals seem to be a feature that has been retained in North America, but which has 'receded in Britain and Ireland' (2007: 391). A search for double modals in the ICE-Ireland corpus (Kallen & Kirk 2007) returned one result, in a transcript from the category 'broadcast discussions' from Northern Ireland: *I mean we'll may get into it in a moment or two*. The codes in the corpus identify the speaker as an older male from County Down in Ulster.

## 2.3 Midlands, southern England and Wales

Except for the attestations of *used to could* in Wright (1898–1905) noted above, few double modal forms have been attested in speech from the Midlands, southern England or Wales. The sampler of the *Freiburg English Dialects Corpus* (Szmrecsanyi & Hernández 2007), a subset of the FRED corpus (Hernández 2006; Anderwald & Wagner 2007), contains a few double modals, including *would could*, *would used to*, *used to might* and *used to could*, from the Midlands and the southwest of England, regions whose dialects have not previously been reported to be associated with double

<sup>5</sup> See also [www.uni-due.de/IERC/IERC\\_SIEU\\_Heatmaps.htm](http://www.uni-due.de/IERC/IERC_SIEU_Heatmaps.htm)

modals.<sup>6</sup> As far as is known, the feature has not been attested in Welsh English (e.g. in Filppula 2004; Penhallurick 2004; Paulasto 2006; Paulasto *et al.* 2020).

Many of the double modal usages reported in the studies above are included in the *Multimo* resource (Reed & Montgomery 2016),<sup>7</sup> an online table that includes contextual information such as location, medium, sentence context and other details, if available. *Multimo* lists 168 double modal instances from England or Scotland. A number of these are from literary texts dating from the thirteenth to the twentieth century; the spoken attestations are mostly from the sources noted above as well as from personal communications to the resource's creators. *Multimo* does not include any entries that have been tagged with the locations Wales, Northern Ireland or the Republic of Ireland.

Morin *et al.* (2020) and Morin (2021a) point out that double modals can be found in corpora collected from UK web texts and suggest looking for the feature in social media texts from Britain.

Overall, the data on double modal use in Britain and Ireland are heterogeneous: some types have been more systematically documented, especially in Scotland, but for the most part, the range of possible combinatorial types has not been extensively investigated, either by written surveys, face-to-face elicitation in interviews, or documentation in corpora, and there are few examples in the literature of naturalistic usages in speech. In general, double modals have not been studied in speech from the English Midlands, southern England or Wales, either by means of surveys or in corpora.

The present study addresses two main questions: first, where are double modals used in naturalistic speech in the British Isles? Are they encountered only in northern England, Scotland and Northern Ireland, as has been suggested by previous research? Second, what combinatorial types are used, and how prevalent are they? Naturalistic spoken language data from Britain and Ireland can shed light on these questions, demonstrating that a larger number of combinatorial types are used in a wider geographical context than has previously been reported.

### 3 Data and methods

The main data for analysis come from the *Corpus of British Isles Spoken English* (CoBISE), a 112-million-token corpus of 38,680 word-timed and part-of-speech-tagged ASR transcripts from 495 YouTube channels of local councils or other local government organizations in 453 locations in England, Scotland, Wales, Northern Ireland and the Republic of Ireland (Coats 2022b, 2023).<sup>8</sup> Table 1 shows the breakdown by country for the corpus in terms of number of sampled locations, number of sampled videos, and aggregate transcript length as number of words and video duration.

<sup>6</sup> Anderwald & Szmrecsanyi report an additional *might should* from one of the transcribed Scottish interviews in the complete FRED corpus (2009: 1136).

<sup>7</sup> <https://artsandsciences.sc.edu/multimo>

<sup>8</sup> <https://cc.oulu.fi/~scoats/CoBISE.html>



Table 1. *CoBISE size by subcorpus*

Country	Locations	Videos	Words	Length (h)
England	324	23,657	72,879,173	8518.39
Northern Ireland	10	1,898	6,508,505	774.17
Republic of Ireland	26	2,525	6,264,276	680.81
Scotland	75	8,135	17,111,396	1845.35
Wales	18	2,465	8,800,264	982.66

Much of the content of the corpus consists of recordings of public meetings of local councils. In a typical council meeting, councilors will discuss various agenda items pertaining to (for example) local rules and ordinances, building proposals, budget considerations and spending, and a range of other local concerns. The speech in council meetings is somewhat more formal than in other interactive contexts of daily life, but from a pragmatic perspective, council meetings abound in contexts that favor the use of double modals: councilors are engaged in the negotiation of wants and needs and must cooperate in order to achieve the work of the council. The political necessity of cooperation tends to favor careful, polite and face-preserving speech.

Although the scope of this study does not allow a description of the contexts for all occurrences of double modals in the CoBISE transcripts, an example demonstrates the sort of careful, polite interaction in which the feature occurs. In a video entitled ‘Mortlake, Barnes Common and East Sheen Community Conversation’, uploaded to the YouTube channel of the London borough of Richmond upon Thames, a local resident expresses objections to proposed changes to local building codes designed to encourage more environmentally sensitive development such as a requirement for the construction of bicycle lanes, remarking that in Brighton, this created problems for car parking. A councilor in favor of the proposed changes responds very carefully and politely,<sup>9</sup> stating

I think you’re absolutely right, sort of taking into account all of that stuff, but do you think... where I may slightly disagree... you **would might**... is that I do think, you know, particularly with what we’ve seen over the last few weeks, with, erm, campaigners such as Greta and, erm, you know, protests happening all over the world, including Afghanistan, you know, there’s climate change protests happening in Afghanistan, we do have to take this issue seriously.

This type of careful, face-preserving disagreement is relatively frequent in council meetings, and represents the kind of pragmatic context in which double modals are likely to occur. In this example, the use of two epistemic modals could represent a more careful or tentative expression of epistemic possibility compared to use of a single modal auxiliary.

<sup>9</sup> <https://youtu.be/ZRid2Pi5nkU?t=3459>

Although many of the transcripts in the corpus are of meetings, the content of individual channels varies according to council practices. Local government meetings in the UK are, in general, open to the public,<sup>10</sup> but there is no legislation mandating that they be recorded, or that recordings be made available online. Some UK councils use their YouTube channel to make video recordings of public meetings available, but others use their channels to upload videos of local news and events, public service announcements and reminders, or information about changes in local ordinances. Many YouTube channels of local councils in the UK contain both recordings of public meetings as well as additional video content of various types.

In the Republic of Ireland, the default legal situation is different – in contrast to the UK, there is no law stipulating that meetings of local government institutions must be open to the public – only that the minutes of meetings must be accessible.<sup>11</sup> This difference may, in part, explain why the YouTube channels of local councils in the Republic of Ireland mostly do not include recordings of governmental meetings. As of 2019, only 4 of the 31 local councils in Ireland either broadcast meetings through a streaming service or made recordings available.<sup>12</sup> The shift to online meetings brought about by the Covid-19 pandemic beginning in 2020 induced a few additional Irish local councils to make live streams or recordings of council meetings available to the public via YouTube or other platforms, but most of the Irish local council channels sampled in CoBISE contain primarily other types of content. The speech recorded in CoBISE is recent: the corpus includes transcripts of videos made during the time period spanning approximately 2000–21, but most of the content consists of recordings from the years 2017–21.

One drawback of YouTube ASR transcripts is that they contain no speaker metadata: demographic categories commonly used for sociolinguistic analyses, such as gender, age, educational level or ethnic background, are thus not available to the analyst in the raw transcripts without additional annotation work. Nevertheless, the transcripts in CoBISE, which are mostly of meetings, represent a relatively large body of careful conversational discourse and negotiation from a large number of locales in the British Isles. They are therefore good material for looking into the prevalence and distribution of double modals.

### 3.1 Corpus search

The procedure used to search for and verify double modals in CoBISE was similar to that described in Coats (2022a): double modals were sought in the part-of-speech-tagged and word-timed corpus using regular expressions. Starting from thirteen modal or semi-modal forms (*may, might, can, could, shall, should, will, would, must, ought to, oughta, used to* and the abbreviated form *'ll*), a regular expression was generated for all possible two-slot

<sup>10</sup> In the UK, S.I. 2012/2089 established the legal principle that council meetings must be open to the public.

<sup>11</sup> [Local Government Act, 2001, Schedule 10 \(irishstatutebook.ie\)](#).

<sup>12</sup> [www.thejournal.ie/local-authority-webcasting-ireland-4734752-Jul2019](https://www.thejournal.ie/local-authority-webcasting-ireland-4734752-Jul2019)



combinations, with the exception of repetitions (i.e. *may might, may can, may could*, etc., but not *may may*). In order to account for potential negations, past tense constructions and question forms, the expressions also captured sequences with intervening pronouns or negators and the auxiliary forms *have* and *'ve*, as well as adverbs ending in *-ly*.

A script then used the 156 regular expressions generated in this way to search for double modals in the corpus and created a table with a row for each hit and columns for the country, channel name, two-modal sequence and text captured by the regular expression from the transcript, as well as a URL linked to the corresponding video at a time three seconds before the utterance. Then, a subset of the search hits was manually inspected in the corresponding videos to check if they were double modals or false positives.

### 3.2 Manual verification

Manual annotation consisted of clicking on the URLs in the table described above, listening to the speech and categorizing the sequence by adding an annotation code. The search hits fall into three basic categories: false positives, self-repairs and authentic double modals. False positives were classified as ASR errors (the algorithm had incorrectly transcribed a word), overlap errors (e.g. due to the overlap of two clauses in utterances like *if I might, could I...* or the overlap of two speakers), homonyms or homophones (e.g. *Will can help you*, where *Will* is a name, or *wood can be taken*, where *wood* has been incorrectly transcribed as *would*), errors due to unintelligible audio, or instances in which the video was no longer available on YouTube.

The annotation code *sr* (for 'self-repair') was applied to instances where a speaker noticeably corrected an utterance while articulating the two-modal sequence, for example, by cutting off a word prior to completion and beginning with a different modal form or by pausing noticeably between two modal forms and articulating the second form with emphasis. True double modals were identified on the basis of semantic and articulatory considerations: because corrections or self-repairs in speech are typically associated with disfluencies in terms of articulatory timing and prosody (Levelt 1983; Postma 2000; Lickley 2015), two-modal sequences that did not fall into one of the categories described above, had no disfluencies, and were coherent in terms of semantic content and discourse context were judged to be authentic double modals. An asterisk was used to indicate authentic but questionable usages: utterances with two modals in sequence that were semantically coherent and exhibited only very slight disfluencies, such as minimal pauses. A few example annotations are provided in [table 2](#).

For additional details about the annotation scheme, see Coats (2022a), in which a similar procedure was employed.

The selection of videos for manual inspection was pseudo-random, but focused on videos more likely to represent authentic usages: transcripts in which two base modal forms occurred in sequence, with no intervening words. In total, 1,154 (22.5%) of the 5,119 search hits were checked and manually annotated with one or more of the codes.

Table 2. *Example annotations*

URL	Transcript	Annotation (comment)
<a href="https://youtu.be/q4dKA16obj4?t=4661">https://youtu.be/q4dKA16obj4?t=4661</a>	‘... the app <b>can could</b> appear the condition...’	fp (false positive due to ASR error: the speaker says ‘the applicant could appeal a condition’)
<a href="https://youtu.be/x4ir-es-MCA?t=1884">https://youtu.be/x4ir-es-MCA?t=1884</a>	‘... those of us who <b>can must</b> make...’	o (overlap, clause: not a true double modal)
<a href="https://youtu.be/2kFsQA2BwQo?t=4084">https://youtu.be/2kFsQA2BwQo?t=4084</a>	‘... I think it should be supported so I <b>shall could</b> I...’	o (overlap, speaker: first speaker ‘... so I shall.’ Second speaker: ‘erm, could I ask for...’)
<a href="https://youtu.be/8FR7C9Wc-i8?t=14">https://youtu.be/8FR7C9Wc-i8?t=14</a>	‘... and I <b>should will</b> therefore...’	sr d (self-repair, disfluency: judged to not be a true double modal)
<a href="https://youtu.be/n2gxOuULeQg?t=1402">https://youtu.be/n2gxOuULeQg?t=1402</a>	‘... Bristol poetry <b>can should</b> lose its...’	hn (homonym: <i>Bristol Poetry Can</i> is/was an organization)
<a href="https://youtu.be/-V4mGJOjfSc?t=2569">https://youtu.be/-V4mGJOjfSc?t=2569</a>	‘... and whether they <b>will can</b> come into this system...’	t (no disfluencies, judged to be a true double modal)
<a href="https://youtu.be/2kRSDJp7ors?t=541">https://youtu.be/2kRSDJp7ors?t=541</a>	‘... we can’t expand it to Manchester much as I <b>would might</b> want to do so...’	t (no disfluencies, judged to be a true double modal)
<a href="https://youtu.be/igL5ABLKqeA?t=2380">https://youtu.be/igL5ABLKqeA?t=2380</a>	‘... then they <b>would might</b> and I trust would be more likely...’	t (no disfluencies, judged to be a true double modal)
<a href="https://youtu.be/G71IikU1ico?t=3035">https://youtu.be/G71IikU1ico?t=3035</a>	‘... we <b>would</b> certainly <b>can</b> have a look...’	t (no disfluencies, judged to be a true double modal)

## 4 Results

This section considers the authentic double modal combinations and their geographical distribution in the CoBISE data. In total, 187 double modal utterances in CoBISE were judged to be authentic or possibly authentic, based on the criteria noted above. This represents 16.2 percent of the utterances that were manually examined in the corresponding videos – most sequences of two modal verbs in the manually checked transcriptions were self-repairs (42.5%), overlaps (19.2%), homonyms or homophones (3.7%), ASR errors (27.6%) or video errors (1.5%).

### 4.1 Types by frequency

Table 3 shows the frequencies of ‘true’ double modal types in order of decreasing frequency, as well as for each type the number of search hits manually checked, the percentage of search hits manually checked, the number of hits judged to be true double modals and the percentage of checked hits judged to be true double modals.

Table 3. *Double modal types*

Type	No. of tokens	Type	No. true	Type	No. of tokens
<b>would might</b>	34	<i>might would</i>	3	<b>can shall</b>	1
<i>will/'ll can</i>	26	<i>must can</i>	3	<b>might ought to</b>	1
<i>will/'ll might</i>	12	<b>could used to</b>	2	<i>should ought to</i>	1
<b>might may</b>	11	<b>will/'ll would</b>	2	<i>should could</i>	1
<b>would may</b>	9	<i>could can</i>	2	<b>should may</b>	1
<b>would should</b>	9	<b>'ll will</b>	2	<b>could should</b>	1
<i>will/'ll should</i>	8	<i>might can</i>	2	<b>will/'ll must</b>	1
<i>would could</i>	7	<b>can will</b>	2	<b>would used to</b>	1
<b>will/'ll may</b>	6	<b>can might</b>	2	<b>could will</b>	1
<b>would can</b>	5	<b>may would</b>	2	<b>should must</b>	1
<b>can should</b>	4	<b>could must</b>	1	<b>can could</b>	1
<b>will/'ll shall</b>	4	<b>would must</b>	1	<b>may can</b>	1
<b>would will</b>	4	<b>could would</b>	1	<b>would ought to</b>	1
<b>could might</b>	4	<b>can used to</b>	1		
<i>will/'ll could</i>	3	<b>could shall</b>	1		

Types in boldface text are ‘non-canonical’: those not discussed (or only minimally) in the previous literature summarized above. *Will* and *'ll* occurring as the first element in a double modal have been collapsed to a single type in tables 3–8.<sup>13</sup>

The most frequent type in the data, somewhat surprisingly, is *would might*, a combination not previously commented upon in detail in reports of double modals in the British Isles. *Will can* (and *'ll can*), known to be a traditional type in descriptions of double modals in Scotland, is the next most frequent overall, followed by *will might*; the next most frequent types, *might may* and *would may*, have not been discussed extensively in the existing literature. Overall, the data in table 3 suggest a tendency towards favoring double modals in which the first element is *will* or *would*, rather than types with the first-element forms *might* or *could*, which are more common in American speech.

In total, 44 different double modal types were judged to be authentic (48 if *'ll* is considered distinct). The inventory of possible double modal forms in Britain and Ireland is thus considerably larger than has previously been documented: for example, Corrigan (2006) lists 10 possible double modals for the British Isles; Montgomery & Nagle list 10 distinct combinations for Scotland and 6 for Northern Ireland (1994: 94, 104); and Brown lists 12 for Scotland (1991: 75).<sup>14</sup> Even discounting hapaxes (types judged to be authentic only once), 25 combinations remain as naturalistic authentic usages, and more complete manual annotation of the 5,119 search hits in CoBISE

<sup>13</sup> Except for *'ll will* instances. There are no instances in which *'ll* was judged to be an authentic second element in a double modal.

<sup>14</sup> This includes only forms with two modal elements, not constructions with one modal and *be to* or *have to*.

would possibly find additional combinatorial types such as *can must*, attested by Brown (1991) for Scotland but not attested among the examined search hits. The large inventory suggests that although double modals are infrequent and proscribed in standard speech, they exhibit a combinatorial propensity in the British Isles that is greater than has been previously estimated. The number of search hits is not sufficient to make robust inferences about regional patterns of use, but some tendencies can be noted.

#### 4.2 Types by location

Of the types judged to be authentic double modal tokens, 103 were from England, 41 from Scotland, 21 from Wales, 13 from the Republic of Ireland and 9 from Northern Ireland. The manually verified data from Scotland include 25 different double modal types judged to be authentic usages (table 4). *Will can* is the most frequent type, followed by *will might*, *would will* and *might may*. Four types are attested twice, and 17 once.

Table 4. *Scotland double modal types*

Type	No. of tokens
<i>will/'ll can</i>	6
<i>will/'ll might</i>	4
<i>would will, might may</i>	3 times each
<i>would might, would could, would can, will/'ll may</i>	2 times each
<i>should must, would may, 'll will, will/'ll should, will/'ll could, can might, can shall, should could, must can, may would, may can, could used to, could should, could must, can used to, can should, should may</i>	1 time each

The thirty distinct double modal types and their frequencies in transcripts from English channels are shown in table 5. Types with *would* and *will* or *'ll* as the first element are the most common; *could*, *can*, *may*, *must* and *should* are also attested as the first element. Of the six forms reported to be part of the repertoire of Tyneside English (McDonald & Beal 1987: 47), four are attested in these data: *might can*, *must can*, *will can* and *would could*. Of the types reported in older sources, *used to could* and *should ought*, the latter is attested.

As far as is known, double modals have not previously been attested in Welsh English. Table 6 shows that the most frequent types judged to be authentic in transcripts from Wales, *will/'ll can* and *would might*, are also the most common types in Scotland and England.

Relatively few naturalistic double modals are found in transcripts from Northern Ireland (table 7): 9 were judged to be authentic usages. In addition, 8 hits in the Northern Irish subcorpus could not be examined manually because the source videos had been made private in the time between the download of the transcript material for the corpus (December 2019) and the manual verification (September 2021). The

Table 5. *England double modal types*

Type	No. of tokens
<i>would might</i>	25
<i>will/'ll can</i>	12
<i>will/'ll might</i>	8 times each
<i>will/'ll should, would should</i>	6 times each
<i>would may</i>	5
<i>would could, will/'ll shall, might may</i>	4 times each
<i>might would, could might, will/'ll may</i>	3 times each
<i>must can, can should</i>	2 times each
<i>would must, would can, 'll will, would ought to, can might, will/'ll ought to, will/'ll must, should ought to, might ought to, might can, may would, could would, could shall, could can, can will, would will</i>	1 time each

Table 6. *Wales double modal types*

Type	No. of tokens
<i>will/'ll can, would might</i>	5 times each
<i>might may, would can</i>	2 times each
<i>could can, could used to, could will, will/'ll may, would could, would may, would should</i>	1 time each

Table 7. *Northern Ireland double modal types*

Type	No. of tokens
<i>will/'ll can, will/'ll could, will/'ll would, would may</i>	2 times each
<i>can will</i>	1

attested types have *will/'ll* or *would* as the first element; in addition, one instance of *can will* is attested. The Northern Ireland subcorpus is smaller than the other country-level subcorpora. More ASR transcript data from Northern Ireland may result in additional authentic double modals being found.

For the Republic of Ireland (table 8), 13 double modals were judged to be authentic or possibly authentic. The attested types are somewhat different from those found in the CoBISE data from Northern Ireland or Scotland: first-element *might* or *would* types are more common than types with *will/'ll* in the first position.

The text frequencies are not high enough to draw robust conclusions about local usage patterns for particular double modal types. It can, however, be noted that the more frequent

Table 8. *Republic of Ireland double modal types*

Type	No. of tokens
<i>might may, would might, would should</i>	2 times each
<i>can could, can should, could might, might can, will/'ll can, will should, would used to</i>	1 time each

types (*would might, will/'ll can*) seem not to be restricted to any particular region, but can be found in speech from all over the British Isles.

### 4.3 Corpus frequencies and maps

At country level, Scotland shows the highest overall relative frequency of double modals and England the lowest (table 9), a result in line with previous reports of the geographical distribution of the feature. The values, between 1.38 and 2.40 double modals per million words, are roughly comparable to those found for North American data, where at state/province level, relative frequencies of double modals have been found to range from zero, for several states and provinces, to a maximum of 5.86 per million words, in the US state of Alabama (Coats 2022a).

Table 9. *Overall frequency of double modals (per million words)*

Country	Double modals pmw
England	1.40
Northern Ireland	1.38
Republic of Ireland	2.08
Scotland	2.40
Wales	2.39

However, the rareness of the feature and the relatively small number of authentic double modals attested in the CoBISE data mean that the differences in relative frequencies at country level are not significant, according to standard statistical tests, with the exception of the difference between England and Scotland.<sup>15</sup> Methodological considerations may also affect the country-level comparison of relative frequencies: while the number of false positives in the annotated data (non-double modals

<sup>15</sup> Pairwise chi-square and log-likelihood tests for the five countries, based on number of double modals and number of word tokens in the corresponding subcorpus and applying the Bonferroni correction, show that only the England–Scotland difference is significant at  $\alpha=.05$ , and only for the chi-square test (for the log-likelihood test,  $p=.062$ ).



incorrectly transcribed by ASR as double modals) is zero, due to the manual filtering of the search hits, there may be false negatives in the corpus (actual double modals transcribed incorrectly by ASR), which would not be captured by searches using regular expressions. Considering the fact that the YouTube ASR algorithm is likely to be more inaccurate for Scottish speakers (Tatman 2017), there could be proportionately more false negatives in the Scottish data than in the English data; hence, the relative frequency of double modals in Scottish transcripts may be underestimated. Quantifying the false negative rate, however, would require comparison of ASR transcripts with verified manual transcripts, and manually transcribing a large sample of corpus transcripts is an undertaking beyond the scope of this study (see, however, Coats 2023, for a rough method to check for false negatives).

A cautious interpretation of the relative frequency data from the annotated transcripts would be that speakers in Scotland use more double modals, a finding in line with previous studies. More data, both in terms of overall number of transcripts and in terms of the range of speech situations reflected in the videos, from a greater variety of locations, as well as a set of manually transcribed videos allowing estimation of false negative rates, would be needed before the relative frequencies of double modals in Scotland, England, Wales and Ireland could be compared with some degree of certainty, and before the observed differences in relative frequencies could be considered robust.

The map in figure 1 depicts the number of manually annotated true double modals falling within the boundaries of administrative subregions in the UK and Ireland as relative frequencies per million words. Although the map suggests that double modals are more common in more northerly locations, the unequal sizes of the administrative areas make it difficult to perceive a spatial pattern, and some high relative frequencies, for example in the counties Meath, Westmeath, Roscommon, Sligo and Donegal in the Republic of Ireland, correspond to single occurrences of a double modal. It should also be noted that the geographical distribution may in part reflect the different composition of the channel subcorpora in terms of video types – as noted above, while council meetings may represent a suitable pragmatic context for the use of double modals, other types of transcripts, such as recorded speeches or informational news videos, do not.

One way to address outliers caused by small or unequal subsample sizes (i.e. the CoBISE channel sizes mapped to administrative regions) is to calculate for each administrative subunit a local spatial autocorrelation statistic based on an aggregate measure of nearby values. Spatial autocorrelation, which quantifies the tendency of values of a spatially distributed variable to exhibit similar values at nearby locations, can be used to identify local and regional patterns (Anselin *et al.* 2010). The local spatial autocorrelation statistic Getis-Ord  $G_i^*$  is a ratio: the weighted average of the values in neighboring locations divided by the sum of all values, then scaled as a Z-score. For a given location, determination of which locations are considered nearby can be made on the basis of polygon continuity (i.e. bordering administrative regions), distance within a certain cutoff threshold, the  $k$  nearest neighbors, or based on other functions, and there are also various methods for the calculation of the weighting

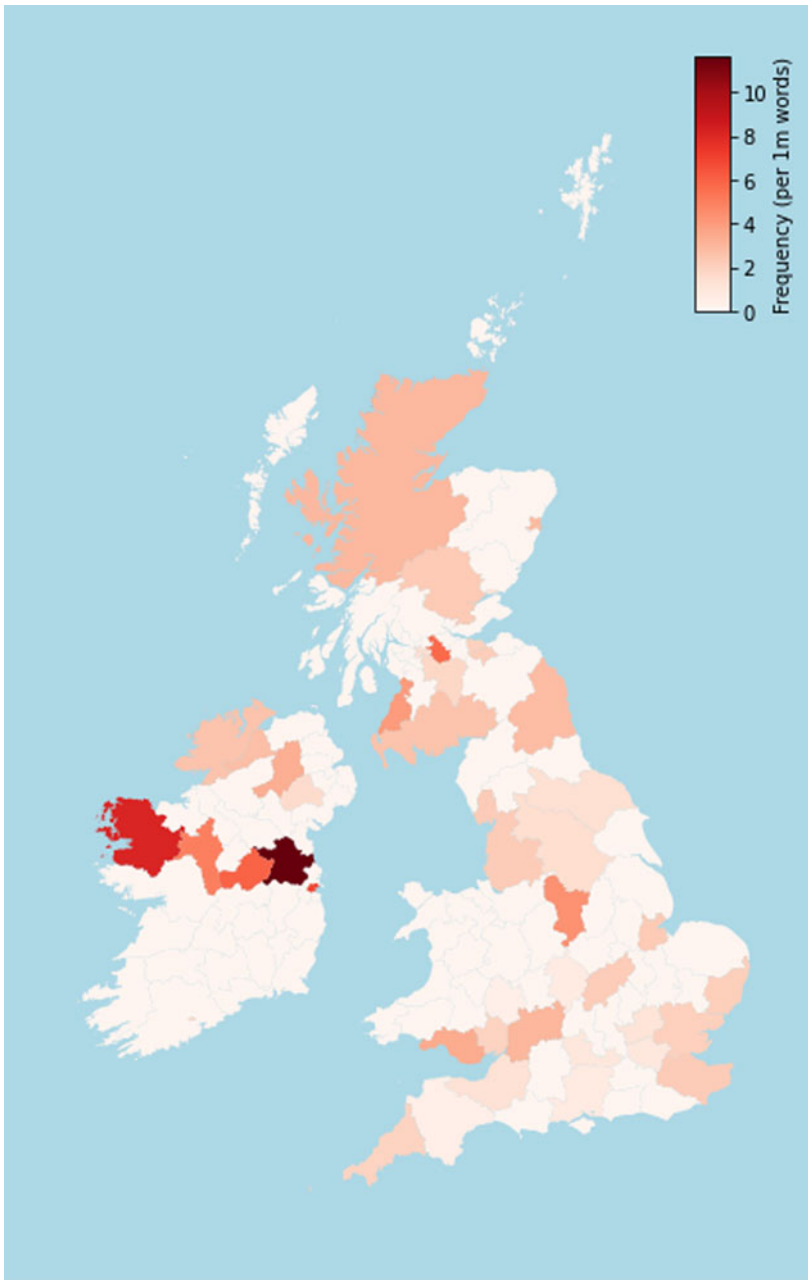


Figure 1. Double modals by county, relative frequencies

function (see Getis 2010). In the exploratory analysis of spatial data, the calculation of a local autocorrelation statistic is used to identify ‘hot spots’ and ‘cold spots’, whereby the weighting function serves to smooth values of nearby locations.

Figure 2 shows a map of Getis-Ord  $G_i^*$  values, calculated on the basis of relative frequencies of double modals, after removal of questionable tokens (i.e. those judged as ‘true’ but annotated with an asterisk during the manual inspection step). Now, a different picture emerges: the autocorrelation values largely correspond to previous reports on the frequency of double modals in the British Isles: relatively common in northernmost England and in Scotland, with the highest value found for the Scottish Borders, but with some attestations in the English Midlands and south. Wales, Northern Ireland and the Republic of Ireland exhibit lower values.

Although these exploratory maps seem to support the patchwork of results from previous studies, due to the small number of confirmed double modals in the CoBISE data, they cannot be considered robust. The values of the Getis-Ord  $G_i^*$  test statistic for the individual counties and shires do not generally meet the standard level of significance at  $\alpha=0.05$ , except for Northumberland, a few Scottish shires and two Irish counties.<sup>16</sup> In addition, the  $k$ -nearest-neighbors weights model, if it is to be conceived of as a model for communication patterns that underlies the spread of the feature, does not account for geographical features that impede communication, such as the Irish Sea. Nevertheless, mapping a spatial correlation statistic in an exploratory manner provides a result that can be reconciled with what is known from previous studies about the distribution of the feature.

## 5 Discussion

### 5.1 Double modal inventory and frequency

The data presented in section 4.1 show that the double modal inventory of the British Isles is larger than has previously been documented. Types known from earlier descriptions of dialectal speech and dialect dictionaries, as well as those more comprehensively documented in surveys, research interviews and corpora of transcribed speech, are attested in CoBISE, but the data also show use of double modal types that have hitherto either not been the focus of research attention or not been documented at all.

One of the most interesting findings is *would might*, a form mostly absent from previous accounts of double modals in the British Isles, as the most common type. It is attested in the current data in England, Scotland, Wales and Ireland, but also shows a broad geographic distribution in North America, where it is not concentrated in the American Southeast, the region traditionally most strongly associated with double modal use – California has the highest number of attestations (Coats 2022a). In terms of meanings, *would might* seems to be used to express epistemic possibility in the context of intrinsic volition or a hypothetical situation. Examination of the video segments from

<sup>16</sup> This significance level corresponds to a  $G_i^*$  value of  $\pm 1.645$ . Varying the number of nearest neighbors changes the administrative units that are statistically significant somewhat, but generally confirms higher usage in Scotland and N. England. See [https://cc.oulu.fi/~scoats/British\\_DM\\_KNN.html](https://cc.oulu.fi/~scoats/British_DM_KNN.html)

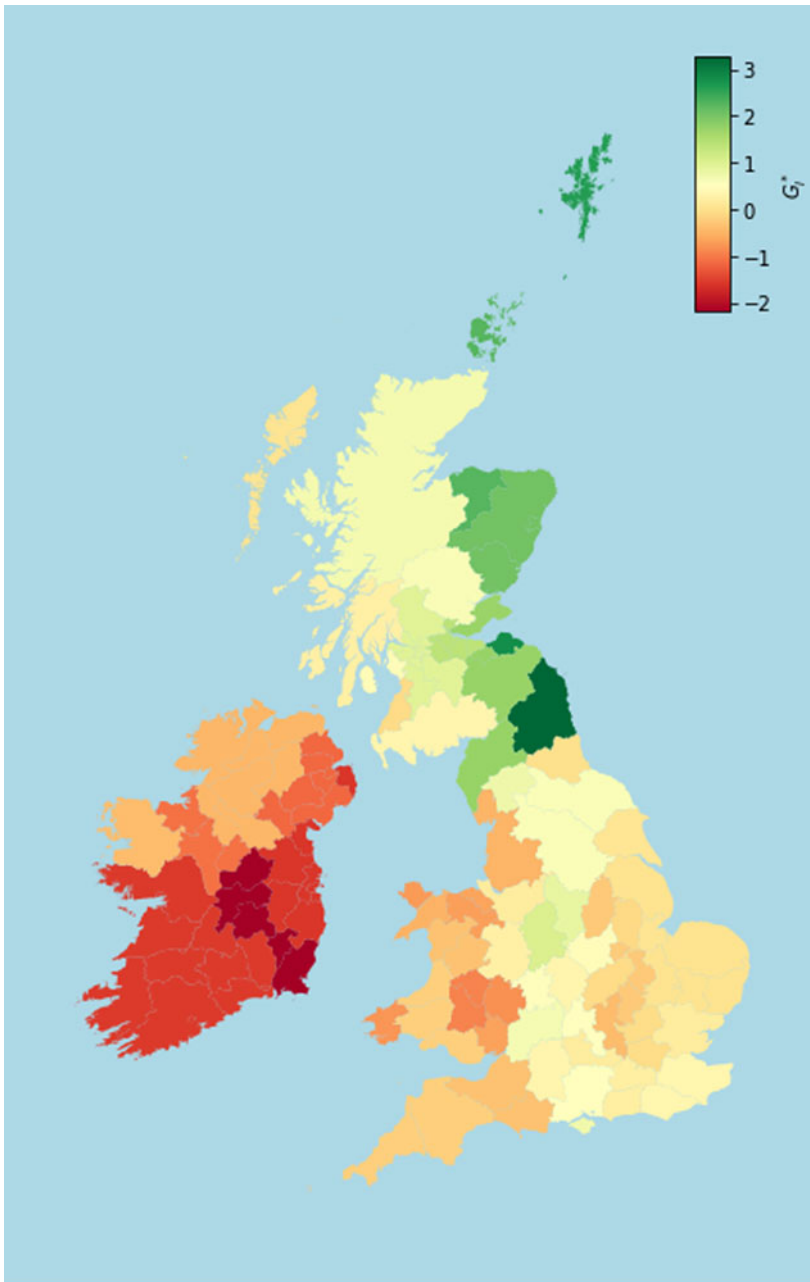


Figure 2.  $G_i^*$  score (20-nearest-neighbors binary weights matrix, questionable assessments removed) by county

CoBISE in which *would might* occurs suggests meanings approximately equivalent to conditional or hypothetical statements further marked for epistemicity through the addition of an epistemic stance adverbial, along the lines of *it would possibly* or *it would maybe*. *Would might* in CoBISE occurs in utterances such as *I would might like to make it clear; it would might need additional, other meetings; I'm not sure what the implications would might be, but they may not be too bad; that would might be my suggestion; we can't expand into Manchester, as much as I would might want to do so*, and others. The many occurrences of this type in the data and lack of geographic restriction, when considered alongside the naturalistic usages documented from North America and the dearth of previous attestations, suggest that *would might* may be an emergent global double modal for the expression of careful epistemicity. As for the frequencies of other types, those with *will* or *would* as the first element are more common in CoBISE, in contrast to North America, where types with *might* as the first element dominate, a finding in line with previous reports on double modal inventories in the British Isles and America.

## 5.2 Geographical distribution

The prevailing view of previous research is that double modals in British and Irish speech are restricted to Scotland, the north of England (Northumberland and Tyneside) and Northern Ireland. Indeed, some studies find that double modals only occur in these regions (e.g. Beal 2004: 127: 'Whilst these double modal constructions are found in Scots and in some dialects of the southern USA, the only area of England in which they occur is Northumberland and Tyneside'). According to Hickey, double modals are 'not found in the south of Ireland and only very rarely in the north nowadays' (2010: 86; see also 2005a: 102).

This study, however, demonstrates that double modals are used in naturalistic speech in the British Isles from a wide variety of locales, including in regions where the feature has not previously been systematically investigated. The data from CoBISE show that double modals are, at least in the context of semi-formal government meetings or public events, in use in conversational contexts not only in Scotland, Northern England and Ulster, but also in the English Midlands and south, Wales, and the Republic of Ireland. A preliminary analysis of double modals in data from the *Spoken British National Corpus 2014* (Spoken BNC2014; Love *et al.* 2017; Brezina *et al.* 2018) confirms the presence of double modals in southern and Midlands speech, as well as in Scotland and the English north, in informal conversation (cf. Morin 2021a). A total of 57 authentic double modals were found in the Spoken BNC2014. This corresponds to a higher relative frequency than in the CoBISE data (5 pmw vs. 1.67 pmw). In contrast to CoBISE, the Spoken BNC2014 consists of recordings of informal conversations held in private contexts, mostly between friends and family members; it is reasonable to assume that the lack of formality of context favors more use of non-standard features such as double modals (cf. Miller & Brown 1982: 13).

---

An additional piece of evidence comes from social media data, where preliminary findings show that in geolocated tweets, *will can* is used mainly in Scotland but also in the English Midlands and south and in Wales.<sup>17</sup>

In light of this, a reconsideration of the conceptualization of double modal inventories and their association with specific regional varieties or local dialects seems appropriate: given enough data, one is likely to find at least one use for most double modal types in most locations. The double modal data from this study support Kortmann's observation that variation in dialect syntax typically has a wide areal reach and is 'in many cases a matter of statistical frequency rather than the presence or absence of a feature' (2010: 842). Rather than a model in which a given double modal combination is held to be either present or absent in a given variety, the data from this study support an account in which individual double modal types, although rare, are theoretically possible in most varieties of English used in the British Isles or elsewhere, a conceptualization also in line with recent dialect-corpus-based approaches to the study of regional variation (e.g. Nerbonne 2009; Szmrecsanyi 2011, 2013; Grieve 2016). Thus, although *will can* can be found in naturalistic speech from Kent and Essex, it is more likely to be encountered in conversational speech in the Scottish Lowlands; likewise, *might could* could be spontaneously produced in a conversation in Ireland or in London, but is more likely to occur in a conversation in Georgia or Alabama in the US. As noted above in section 4, however, the number of double modals in the CoBISE data and the size of the channel subcorpora are not sufficient to provide reliable information about the geographical distribution of particular types, and can only provide a general overview of the tendency for double modals to be geographically distributed. In line with previous accounts, they are more common in Scotland, the north of England and Northern Ireland.

### 5.3 Caveats

Several considerations pertaining to the nature of the corpus data and the procedures used to identify and verify double modals should be noted. First, as discussed in section 4.3, although manual verification can remove false positives, the method cannot account for false negatives. Considering the fact that some Scottish modal verb constructions include enclitic forms that do not directly correspond to standard English (e.g. *wullna* 'will not', *canna* or *canny* 'cannot'), the relative frequencies of double modals calculated for Scotland in section 4 may be underestimates.

Another consideration concerns the speakers in the videos that comprise the corpus transcripts. Although the recordings are mainly of meetings or other events organized by councils and other local governmental organizations, there is no way to determine the residency of all the speakers in the transcripts. It is likely that some speakers in the corpus are not locals of the areas represented by the councils whose channels have been sampled in the corpus, but rather were raised elsewhere and moved to those

<sup>17</sup> <https://twitter.com/camhmorin/status/1350080811983319040>



localities. Indeed, during the manual verification process, a few instances of Scottish- or Irish-accented speech in English council sessions were encountered. Nevertheless, given the nature of local government, it is likely that most speakers are local residents, and given the size of most of the channel subcorpora, their linguistic signal would predominate in the corpus. Ideally, however, a corpus in which demographic information was known about individual speakers would be more suitable for making inferences about regional varieties.

As noted in section 2, relatively few local councils in the Republic of Ireland upload recordings of council meetings to their YouTube channels, reflecting in part the different legal framework for public access to local government meetings in Ireland compared to the UK, but possibly also different attitudes towards public speech. One Irish council rejected a proposal to broadcast or record council meetings because it could encourage ‘grandstanding’ by council members, who, possibly enticed by the prospect of online fame, might change their behavior in order to appeal to a potentially large online viewership, instead of speaking in a presumably more down-to-earth and businesslike manner to fellow council members in a non-broadcast council session.<sup>18</sup> For the analysis of double modals presented in this study, the fact that Irish council YouTube channels may be more likely to contain scripted dialogue such as speeches or news videos, rather than recordings of interactions in the pragmatic contexts in which double modals have been proposed to be more likely (*viz.* ‘situations of caution and sensitivity’ where politeness is required, Montgomery 1998: 96), means that the observed frequencies of double modals for Ireland may be underestimates. The examination of a larger number of transcripts, from a wider range of interactional types, would be necessary in order to further investigate this possibility.

The manual verification procedure used in this study relies on the judgment of a single annotator. Although the annotation criteria for the determination of true double modals (lack of prosodic marking, lack of semantic/discourse incoherence) were formal and not subject to arbitrary interpretation, a better approach, unfortunately not possible within the scope of this study, would be to use an annotator agreement measure for multiple annotators, ideally including persons familiar with local speech patterns in the subregions or regions of the sampled channels.

Despite these caveats, the evidence for diversity in the use of double modals in the British Isles seems strong, and it is hoped that the results presented here will be of interest as evidence in discussions of microsyntactic variation, modality in English and English language history, but also provide a new impetus for work along similar methodological lines that can help to document the use of infrequent features in English varieties.

## 6 Summary and conclusion

A large corpus of ASR transcripts of naturalistic speech from the UK and Ireland was used to identify instances of double modals using regular expressions and a manual verification

<sup>18</sup> See [www.thejournal.ie/local-authority-webcasting-ireland-4734752-Jul2019](http://www.thejournal.ie/local-authority-webcasting-ireland-4734752-Jul2019). The concern seems unfounded: view counts, at least for UK council meetings uploaded to YouTube, are typically fewer than 50.

procedure. The results show that double modals are more widely used in the British Isles than has previously been documented, both in terms of the size of the double modal inventory and the geographical distribution of the feature. Although raw relative frequency data for double modals do not show a clear regional patterning, a local spatial autocorrelation analysis based on a 20-nearest-neighbors spatial weights matrix, taking only unquestionable attestations into consideration, largely recapitulates the prevailing view that double modals are more commonly used in Scotland, the north of England and Northern Ireland. More data would be needed, however to confirm this tendency, as well as to ascertain the regional use of particular double modal types.

The type *would might*, the most frequent double modal in this study, may represent an emergent global double modal, based on its broad geographic distribution in these data and in data from North America, its absence from most earlier accounts of double modals and its likely meaning as an exponent of hypothetical epistemic possibility. The modal meaning of *would might* (and potentially of other combinatorial double modal types), in turn, may reflect complex ongoing changes in the semantics of the English modal auxiliaries, linked to processes of grammaticalization, that exemplify a shift towards increased epistemicity (Coates 1983; Traugott 1989; Leech 2003; Leech *et al.* 2009; Kranich & Gast 2015). In the context of further research into double modals, it may be useful to compare the frequencies of different markers of epistemic possibility, volition, ability or obligation as single modal auxiliaries, double modals, adverbials, adverbials or other expressions (cf. Kärkkäinen 2003; Kranich & Gast 2015) in a corpus where these elements have been disambiguated for modal meaning. The undertaking could represent a potential future use case for data from CoBISE.

Despite findings in recent decades that suggest double modals would be falling into disuse in parts of the British Isles, this study can show that the feature is, although rare, a relatively robust syntactic pattern in spoken language. The outlook for future investigations of double modals and modality in general in CoBISE and similar data sets is promising, given the size of the corpus and the likelihood that it contains a rich repository of expressions of modality, and in the future, continued improvements to ASR algorithms and growing access to streamed interactional data may enable more extensive investigations of naturalistic multiple modal use in the British Isles and beyond.

*Author's address:*

*English, Faculty of Humanities  
University of Oulu  
90014 University of Oulu  
Finland  
[steven.coats@oulu.fi](mailto:steven.coats@oulu.fi)*

## References

- Allen, Will, Joan C. Beal, Karen P. Corrigan, Warren Maguire & Hermann L. Moisl. 2007. A linguistic time-capsule: The Newcastle Electronic Corpus of Tyneside English. In Joan C. Beal,

- Karen P. Corrigan & Hermann L. Moisl (eds.), *Creating and digitising language corpora*, vol. 2: *Diachronic databases*, 16–48. Basingstoke: Palgrave Macmillan.
- Anderwald, Lieselotte & Benedikt Szmrecsanyi. 2009. Corpus linguistics and dialectology. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, vol. 2, 1126–39. Berlin and New York: Walter de Gruyter.
- Anderwald, Lieselotte & Susanne Wagner. 2007. FRED – The Freiburg English Dialect Corpus: Applying corpus-linguistic research tools to the analysis of dialect data. In Beal *et al.* (eds.), 35–53.
- Anselin, Luc, Yong Wook Kim & Ibnu Syabri. 2010. Web-based analytical tools for the exploration of spatial data. In Manfred M. Fischer & Arthur Getis (eds.), *Handbook of applied spatial analysis: Software tools, methods and applications*, 151–74. Heidelberg: Springer.
- Battistella, Edwin. 1995. The syntax of the double modal construction. *Linguistica Atlantica* 17, 19–44.
- Beal, Joan. 2004. English dialects in the North of England: Morphology and syntax. In Kortmann *et al.* (eds.), 114–41.
- Beal, Joan. 2010. *An introduction to regional Englishes: Dialect variation in England*. Edinburgh: Edinburgh University Press.
- Beal, Joan C., Karen P. Corrigan & Hermann Moisl (eds.). 2007. *Creating and digitizing language corpora*, vol. 1: *Synchronic databases*. Basingstoke: Palgrave Macmillan.
- Boertien, Harmon S. 1986. Constituent structure of double modals. In Michael Montgomery & Guy Bailey (eds.), *Language variety in the South: Perspectives in black and white*, 294–318. Tuscaloosa: University of Alabama Press.
- Bour, Anthony. 2014. Description of multiple modality in contemporary Scotland: Double and triple modals in the Scottish Borders. PhD dissertation, University of Freiburg/Breisgau.
- Bour, Anthony. 2015. Multiple modal constructions in the Western English-speaking world. *Linguistica Atlantica* 34(1), 45–59.
- Brezina, Vaclav, Robbie Love & Karin Aijmer. 2018. Corpus linguistics and sociolinguistics: Introducing the Spoken BNC2014. In Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus approaches to contemporary British speech: Sociolinguistic studies of the Spoken BNC2014*, 3–9. New York: Routledge.
- Brown, Keith. 1991. Double modals in Hawick Scots. In Peter Trudgill & Jack Chambers (eds.), *Dialects of English: Studies in grammatical variation*, 74–103. London: Longman.
- Brown, Keith & Martin Millar. 1980. Auxiliary verbs in Edinburgh speech. *Transactions of the Philological Society* 78, 81–133.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London and Canberra: Croom Helm.
- Coats, Steven. 2022a. Naturalistic double modals in North America. *American Speech*. <https://doi.org/10.1215/00031283-9766889> (accessed 27 February 2023).
- Coats, Steven. 2022b. The Corpus of British Isles Spoken English (CoBISE): A new resource of contemporary British and Irish speech. In Karl Berglund, Matti La Mela & Inge Zwart (eds.), *Proceedings of the 6th Digital Humanities in the Nordic and Baltic Countries Conference, Uppsala, Sweden, March 15–18, 2022*, 187–94. Aachen, Germany: CEUR. <http://ceur-ws.org/Vol-3232/paper15.pdf> (accessed 27 February 2023).
- Coats, Steven. 2023. Dialect corpora from YouTube. In Beatrix Busse, Nina Dumrukic & Ingo Kleiber (eds.), *Language and linguistics in a complex world*, 79–102. Berlin and Boston: Walter de Gruyter.
- Corbett, John. 2014. Syntactic variation: Evidence from the Scottish Corpus of Text and Speech. In Robert Lawson (ed.), *Sociolinguistics in Scotland*, 258–76. Basingstoke: Palgrave Macmillan.
- Corrigan, Karen P. 2000. ‘What bees to be maun be’: Aspects of deontic and epistemic modality in a northern dialect of Irish English. *English World-Wide* 21(1), 25–62.

- Corrigan, Karen. 2006. Might could English be ‘double Dutch’ too? Doubling and the Syntactic Atlas of British Isles’ Dialects. Paper presented at Workshop on Syntactic Doubling, Amsterdam, March 2006. [www.meertens.knaw.nl/projecten/edisyn/Online\\_proceedings/Paper\\_Corrigan.pdf](http://www.meertens.knaw.nl/projecten/edisyn/Online_proceedings/Paper_Corrigan.pdf) (accessed 27 February 2023)
- Corrigan, Karen P. 2011. Grammatical variation in Irish English. *English Today* 27(2), 39–46.
- Corrigan, Karen P., Isabelle Buchstaller, Adam Mearns & Hermann Moisl. 2012. *The Diachronic Electronic Corpus of Tyneside English*. Newcastle University. <https://research.ncl.ac.uk/dectec/> (accessed 27 February 2023).
- Coupé, Griet & Ans van Kemenade. 2009. Grammaticalization of modals in Dutch and English: Uncontingent change. In Paola Crisma & Giuseppe Longobardi (eds.), *Historical syntax and linguistic theory*, 250–70. Oxford: Oxford University Press.
- Fennell, Barbara A. & Ronald R. Butters. 1996. Historical and contemporary distribution of double modals in English. In Edgar W. Schneider (ed.), *Focus on the USA: Varieties of English around the world*, 265–88. Amsterdam: John Benjamins.
- Filppula, Markku. 2004. Irish English: Morphology and syntax. In Kortmann *et al.* (eds.), 73–101.
- Getis, Arthur. 2010. Spatial autocorrelation. In Manfred M. Fischer & Arthur Getis (eds.), *Handbook of applied spatial analysis: Software tools, methods and applications*, 255–78. Heidelberg: Springer.
- Gregg, Robert J. 1972. The Scotch-Irish dialect boundaries in Ulster. In Martyn F. Wakelin (ed.), *Patterns in the folk speech of the British Isles*, 109–39. London: Athlone.
- Grieve, Jack. 2016. *Regional variation in written American English*. Cambridge: Cambridge University Press.
- Hasty, J. Daniel. 2012. We might should oughta take a second look at this: A syntactic re-analysis of double modals in Southern United States English. *Lingua* 122(14), 1716–38.
- Hasty, J. Daniel. 2015. Well, he may could have sounded nicer: Perceptions of the double modal in doctor-patient interactions. *American Speech* 90 (3), 347–68.
- Hernández, Nuria. 2006. *User’s Guide to FRED: Freiburg Corpus of English Dialects*. Freiburg/Breisgau: University of Freiburg. <https://freidok.uni-freiburg.de/data/2489> (accessed 17 March 2022)
- Hernández, Nuria, Daniela Kolbe & Monika Schulz. 2011. *A comparative grammar of British English dialects*, vol. 2: *Modals, pronouns and complement clauses*. Berlin and Boston: Walter de Gruyter.
- Hickey, Raymond. 2004. *A sound atlas of Irish English*. Berlin and New York: Mouton de Gruyter.
- Hickey, Raymond. 2005a. Development and diffusion of Irish English. In Hickey (ed.), 82–120.
- Hickey, Raymond (ed.). 2005b. *Legacies of colonial English: Studies in transported dialects*. Cambridge: Cambridge University Press.
- Hickey, Raymond. 2007. *Irish English: History and present-day forms*. Cambridge: Cambridge University Press.
- Hickey, Raymond. 2010. The Englishes of Ireland: Emergence and transportation. In Andy Kirkpatrick (ed.), *The Routledge handbook of world Englishes*, 76–95. Abingdon: Routledge.
- Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation*. Amsterdam and Philadelphia: John Benjamins.
- Kallen, Jeffrey & John Kirk. 2007. ICE-Ireland: Local variations on global standards. In Beal *et al.* (eds.), 121–62.
- Kortmann, Bernd. 2010. Areal variation in syntax. In Peter Auer & Jürgen E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation*, vol.1: *Theories and methods*, 837–64. Berlin and New York: De Gruyter Mouton.
- Kortmann, Bernd, Edgar W. Schneider, Kate Burridge, Raj Mesthrie & Clive Upton (eds.). 2004. *A handbook of varieties of English*, vol. 2: *Morphology and syntax*. Berlin: Mouton de Gruyter.

- Kranich, Svenja & Volker Gast. 2015. Explicitness of epistemic modal marking: Recent changes in British and American English. In Juan R. Zamorano-Mansilla, Carmen Maíz, Elena Domínguez & M. Victoria Martín de la Rosa (eds.), *Thinking modally: English and contrastive studies on modality*, 3–22. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Leech, Geoffrey. 2003. Modality on the move: The English modal auxiliaries 1961–1992. In Roberta Facchinetti, Frank Palmer & Manfred Krug (eds.), *Modality in contemporary English*, 223–40. Berlin and Boston: Mouton de Gruyter.
- Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study*. Cambridge: Cambridge University Press.
- Levelt, Willem J. M. 1983. Monitoring and self-repair in speech. *Cognition* 14(4), 41–104.
- Lickley, Robin J. 2015. Fluency and disfluency. In Melissa A. Redford (ed.), *The handbook of speech production*, 445–69. Malden, MA: Wiley-Blackwell.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. In Tony McEnery, Robbie Love & Vaclav Brezina (eds.), *Compiling and analysing the Spoken British National Corpus 2014* (special issue of *International Journal of Corpus Linguistics* 22(3)), 319–44.
- McDonald, Christine. 1981. Variation in the use of modal verbs with special reference to Tyneside English. PhD dissertation, Newcastle University.
- McDonald, Christine & Joan Beal. 1987. Modal verbs in Tyneside English. *Journal of the Atlantic Provinces Linguistic Association* 9, 43–55.
- Melchers, Gunnel. 2004. English spoken in Orkney and Shetland: Morphology, syntax and lexicon. In Kortmann *et al.* (eds.), 34–46.
- Miller, Jim. 2004. Scottish English: Morphology and syntax. In Kortmann *et al.* (eds.), 47–72.
- Miller, Jim & Keith Brown. 1982. Aspects of Scottish English syntax. *English World-Wide* 3(1), 3–17.
- Mishoe, Margaret & Michael Montgomery. 1994. The pragmatics of multiple modal variation in North and South Carolina. *American Speech* 69(1), 3–29.
- Montgomery, Michael B. 1998. Multiple modals in LAGS and LAMSAS. In Michael B. Montgomery & Thomas E. Nunnally (eds.), *From the Gulf States and beyond: The legacy of Lee Pederson and LAGS*, 90–122. Tuscaloosa: University of Alabama Press.
- Montgomery, Michael B. 2006. The morphology and syntax of Ulster Scots. *English World-Wide* 27(3), 295–329.
- Montgomery, Michael B. & Stephen J. Nagle. 1994. Double modals in Scotland and the Southern United States: Trans-Atlantic inheritance or independent development? *Folia Linguistica Historica* 14(1–2), 91–108.
- Morin, Cameron. 2021a. Construire une triangulation méthodologique en variation syntaxique restreinte. *Echanges linguistiques en Sorbonne* 7, 48–71.
- Morin, Cameron. 2021b. Double modals in Scots: A speaker’s choice hypothesis. In Laura Baranzini & Louis de Saussure (eds.), *Aspects of tenses, modality, and evidentiality*, 165–82. Leiden: Brill.
- Morin, Cameron, Guillaume Desagulier & Jack Grieve. 2020. Dialect syntax in Construction Grammar: Theoretical benefits of a constructionist approach to double modals in English. *Belgian Journal of Linguistics* 34, 252–62.
- Murray, James. 1873. *The dialect of the southern counties of Scotland: Its pronunciation, grammar, and historical relations*. London: Asher & Co.
- Nagle, Stephen J. 1993. Double modals in early English. In Henk Aertsen & Robert J. Jeffers (eds.), *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics*, 363–70. Amsterdam and Philadelphia: John Benjamins.
- Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistics Compass* 3, 175–98.
- Paulasto, Heli. 2006. *Welsh English syntax: Contact and variation*. Joensuu: University of Joensuu Press.

- 
- Paulasto, Heli, Rob Penhallurick & Benjamin Jones. 2020. *Welsh English*. Berlin: De Gruyter Mouton.
- Penhallurick, Robert. 2004. Welsh English: Morphology and syntax. In Kortmann *et al.* (eds.), 102–13.
- Postma, Albert. 2000. Detection of errors during speech production: A review of speech monitoring models. *Cognition* 77, 97–131.
- Reed, Paul & Michael Montgomery (eds.). 2016. *MultiMo: The database of multiple modals*. <https://artsandsciences.sc.edu/multimo/welcome> (accessed 27 February 2023).
- Schneider, Edgar W. 2005. The English dialect heritage of the southern United States. In Hickey (ed.), 262–309.
- Smith, Jennifer, David Adger, Brian Aitken, Caroline Heycock, E. Jamieson & Gary Thoms. 2019. *The Scots Syntax Atlas*. University of Glasgow. <https://scotssyntaxatlas.ac.uk> (accessed 17 March 2022).
- Szmrecsanyi, Benedikt. 2011. Corpus-based dialectometry: A methodological sketch. *Corpora* 6, 45–76.
- Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press.
- Szmrecsanyi, Benedikt & Nuria Hernández. 2007. *Manual of Information to accompany the Freiburg Corpus of English Dialects Sampler ('FRED-S')*. Freiburg/Breisgau: University of Freiburg. <https://freidok.uni-freiburg.de/fedora/objects/freidok:2859/datastreams/FILE1/content> (accessed 27 February 2023).
- Tatman, Rachel. 2017. Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the First Workshop on Ethics in Natural Language Processing, April 4th, 2017, Valencia, Spain*, 53–9. Stroudsburg, PA: Association for Computational Linguistics.
- Traugott, Elizabeth Closs. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1), 31–55.
- Wright, Joseph. 1898–1905. *The English dialect dictionary*, 6 vols. London: Henry Frowde.
- Zullo, Davide, Simone E. Pfenninger & Daniel Schreier. 2021. A Pan-Atlantic 'Multiple Modal Belt'? *American Speech* 96(1), 7–44. <https://doi.org/10.1215/00031283-8620506> (accessed 27 February 2023).