English Today

cambridge.org/eng

Double modals beyond the Atlantic

New evidence from computational sociolinguistics

Steven Coats[1] and Cameron Morin[2]

[1]English, Faculty of Humanities, University of Oulu, Oulu, Finland and [2]Department of English, Ecole Normale Supérieure de Lyon, Lyon, France

Shorter Article

Cite this article: Coats S, Morin C (2024). Double modals beyond the Atlantic. *English Today* **40**, 294–299. https://doi.org/10.1017/S0266078424000191

First published online: 17 October 2024

Keywords:
Double modals; Twitter/𝕏; Australian English; New Zealand English

Corresponding author:
Steven Coats;
Email: steven.coats@oulu.fi

Abstract

This study investigates the use of double modals in Australian and New Zealand English using Twitter/𝕏 data. Double modals are rare grammatical constructions long believed to be limited to regional dialects in the Northern UK and the Southern US. Utilizing a geo-located corpus of over 80 million tweets, the study identifies 314 authentic double modal instances across 51 types, primarily occurring in informal tweets. Findings reveal widespread, albeit low-frequency, usage across both countries without clear geographical patterns. The results align with recent studies suggesting double modals are not confined to specific regions but are possible for most English speakers. The study also questions the traditional Scots-Irish origin theory, proposing an alternative view where the feature is a broader syntactic possibility. Future research should explore larger datasets and extend investigations to outer-circle English varieties to understand better the historical spread and syntactic nature of double modals.

1. Introduction

The double modal is a rare grammatical construction in the English language. As its name suggests, it denotes the use of two modals in a single tensed clause, such as *I **might could** pick you up at the station on Sunday* instead of *I **might be able to** pick you up at the station on Sunday.* In principle, double modals are disfavored by the 'rules' of Standard English Grammar (Huddleston and Pullum 2002: 107), which constrain clauses to contain a maximum of one tensed verb. Given that modals in English only feature tensed forms, their combination in a single clause is prohibited. By and large, double modals have thus long been considered unacceptable and avoided across varieties of English (Morin 2023). Indeed, many native speakers would consider a double modal 'ungrammatical' and would seek to replace it with alternative constructions such as a modal and an infinitive, as above, or a modal and an adverb (such as *maybe I could* instead of *I might could*) (Brown 1991; Morin 2021a).

This has not prevented double modals from being notable and interesting grammatical constructions for linguists and specialists of English grammar, because they have been recurrently attested in a small, distinct group of varieties and dialects of English around the world. Specifically, these constructions have often been cited as typical features of varieties of the Upper and Lower South of the United States (e.g. Fennell and Butters 1996; Bernstein 2003), Northern England and Scotland (Brown 1991, Smith et al. 2019; Beal 2004), and a small number of English-based creoles, such as Gullah and Jamaican Creole (Kortmann, Lukenheimer, and Ehret 2020). Studying double modals in these varieties has always been difficult because of the rarity of these constructions, as well as the limitations of available empirical methods to collect data on them, especially corpus-based and fieldwork-based approaches (Morin, Desagulier, and Grieve 2020; Morin 2021b). Survey-based research and recent corpus studies have found that the most common double modal types have epistemic meanings, and although dynamic and deontic meanings are also possible (Morin and Grieve 2024; Morin, Desagulier, and Grieve 2024), corpus findings have been interpreted as concomitant with a hypothetical 'epistemic expansion' in the recent history of English (Coats 2024). Nevertheless, much remains unclear about the meanings encoded by particular combinatorial types, especially rare combinations.

For the past half-century, double modals have been described as regionally restricted sets of infrequent forms: the most typical combinations considered in the literature involve *may/might* and *will/would* as a first modal and *can/could* as a second

© The Author(s), 2024. 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.

modal (Montgomery and Nagle 1993). Furthermore, because of the geographical circumscription of these constructions, experts have often assumed that they are historically related. More precisely, Scots would have been the original source of double modals, which would have then been imported into American and Caribbean varieties (Zullo, Pfenninger, and Schreier 2021).

However, a very recent wave of linguistic research on these constructions has found that this general picture is not the end of the story, and that we still have much to learn about double modals in the English language. Specifically, five new empirical studies have uncovered that double modals are more widespread in terms of modal combinations and regional distributions than has been documented in the past. These studies take advantage of a new and powerful range of methods in linguistic research known as corpus-based 'computational sociolinguistics' (Grieve et al. 2023). Specifically, they have relied on the investigation of newly created large corpora of geolocated social media data in order to observe double modals at scale for the first time in a number of varieties of English.

For example, Coats (2023, 2024) explored double modal use in two very large corpora of geolocated automatic speech recognition (ASR) transcripts from YouTube in North America, Britain and Ireland. The studies find that double modals are not only observed in the traditional dialect regions attested in the past, but can be found almost everywhere in these two broad areas, in a large number of low-frequency types with no clear constraints on their form. Similarly, Morin and Grieve (2024) and Morin et al. (2024) used two multi-billion-word corpora of geolocated American and British Twitter data to analyze double modals in these two varieties. Their results broadly align with those of Coats, as the authors find both regional clusters and an unexpected general, diverse distribution of low-frequency modal combinations across the two countries. Most recently, Morin and Coats (2023) have even found a general distribution of a rich and complex inventory of double modals in two varieties that had never been considered in the past: Australian and New Zealand English, using another corpus of geolocated YouTube data.

Collectively, these recent studies raise important new research questions and implications for our knowledge of double modals within the modal domain of English worldwide. Firstly, by contrast with previous assumptions in linguistic research, double modals appear to be far more widespread in English than has previously been appreciated. So far, we find that they are used without categorical constraints, albeit very rarely, in four of the most prominent inner-circle varieties of English: American, British, Australian, and New Zealand. There remain many other inner- and outer-circle varieties of English to investigate using computational sociolinguistic methods, in order to see whether these results will replicate in a more comprehensive sample. In addition, the unexpected finding of a seemingly general, low-frequency productivity of the feature suggests an alternative to the traditional Scots-Irish import theory of the historical origin and subsequent spread of double modals (Fennell and Butters 1996).

According to some theoretical accounts, syntactic features, including non-standard features such as double modals, have a 'wider areal reach' and are 'less restricted to very confined areas or individual dialects' (Kortmann 2010: 842). This conception moves away from the idea of static, location/variety-based inventories of features and towards cognitive frameworks in which most syntactic structures are possible for most speakers (cf. Adger and Trousdale 2007), and supports recent theories which emphasize the importance of speaker creativity and the potentially overlooked role of 'mistakes' in language diffusion and evolution (De Smet 2020). Thus, even rare syntactic constructions, including those which contravene prescriptive norms, are possible for most speakers, but speakers have different sensitivities to the acceptability of non-standard constructions such as double modals. From this perspective, it is possible that the acceptability of double modals may be an inheritance of Scots-Irish, but the ability to spontaneously produce the feature is more widespread. In line with this account is the recent analysis of Morin and Grieve (2024), who argue that some double modal types in American English may have emerged as innovations in African-American varieties of English in the Deep South of the United States. Likewise, Morin and Coats (2023) do not find conclusive evidence for the potential Scots-Irish origins of double modals in Australia and New Zealand. Taken together with the metalinguistic lack of awareness of the feature in these varieties, these factors suggest that double modals may be a rare but widespread syntactic construction for speakers of English worldwide.

In this article, we report the results of a new study of double modals on Australian and New Zealand Twitter. Our main aim is to explore whether i) results from a YouTube corpus of transcribed speech replicate in a new corpus compiled from a different medium, social media written texts, and ii) whether these results point towards register variation of double modal use online in these regions. Comparative discussions of results of this type have already been done for American and British English, confirming the underestimated productivity of the double modal system in these varieties, but also revealing subtle variation in double modal types used across these two platforms (Coats 2023, 2024; Morin and Grieve 2024; Morin et al. 2024). In addition, our aim is to illustrate the computational sociolinguistic approach to double modals in English worldwide for a broad audience, encouraging future research in this vein for studying rare and previously elusive morpho-syntactic constructions in varieties and dialects of English.

2. Data and methods

We implemented a location filtering procedure for the identification of potential Australian or New Zealand Twitter/𝕏 accounts. Starting with a global seed corpus of 653,457,659 tweets with 'place' metadata, collected from November 2016 to June 2017 from the Twitter Streaming API using Tweepy (Roesslein 2015), we identified 184,451 unique accounts which had authored a tweet with a 'place' field from Australia or New Zealand. 109,882 of these accounts

https://doi.org/10.1017/S0266078424000191 Published online by Cambridge University Press

were still in existence in April 2023; using Twitter/𝕏's API, all available tweets were downloaded from these accounts in April and May 2023, shortly before the free version of the API was closed in June 2023. Naturally, not every Twitter/𝕏 user who publishes a tweet with Australian/New Zealand 'place' metadata is a resident of those countries – they may be short-term visitors or simply commenting on news or other content from or related to Australia or New Zealand. We included users in our corpus if more than half of the 'place' tweets they had posted were from Australia or New Zealand; we assigned them to the latitude-longitude location represented by the centroid of the most common 'place' in their tweets. In total, the procedure resulted in a dataset of 80,157,335 tweets and 1,017,218,326 word tokens.[1]

This method of identification of Australia and New Zealand-based Twitter/𝕏 users entails several assumptions (see the discussion in Section 3, below), but manual inspection of a random selection of messages confirmed that for the most part, the procedure retrieved content from users who are active in Australia and New Zealand. In order to validate the method, we compared different types of location information for individual tweets, using a subset of the corpus: 5.5m tweets that contained not only a 'place' metadata field, but also exact GPS coordinates.[2] For these tweets, the median distance between the inferred location (the centroid of the account's most common 'place' entry) and the exact GPS location was found to be 25.5 km. We interpret this as evidence that for the most part, the accounts sampled in the dataset under consideration are associated with the inferred locations.

Regular expressions were used to identify all sequences of two modals or semi-modals in the tweet text entity, starting from the modal or semi-modal forms forms *may*, *might*, *can*, *could, shall*, *should*, *will*, *would*, *must*, *ought to*, *oughta*, *used to*, and the abbreviated form *'ll*. Manual filtering removed hits with repetitions (e.g. *might might*), cases of clause overlap (e.g. *. . . using it as much as you* ***can would*** *help . . .*), and cases of non-modals with the same word form (e.g. *a* ***will should*** *reflect your current domestic and financial situations #estateplanning*). Instances which were not filtered according to these criteria and which were coherent in terms of discourse were deemed authentic double modals. Additional tweet content such as images or videos were not considered.[3]

3. Results and discussion

In total, 314 of the 1,026 sequences of two modals were annotated as authentic double modals; these comprise 51 different combinatorial types. Examples (1) through (5) show authentic usages (usernames and URLs have been anonymized).

(1) @user What did you end up going with for your camera??? Given I'm a photographer I **may can** help if needed 😀
(2) I really dislike how my skin has broken out again. I'm hoping it **will can** clear with this new skincare product I'm currently trying out 😊.
(3) A Sweet Tooth I **will shall** remain.🤩☕. Also a photo comparison between 2 mid rangers, the Oppo R15 Pro (2018) and th . . . URL
(4) I know a few engineers **would might** sign up for that. URL
(5) Had he mentioned that as an option we **might could** have saved the day. He was in a hurry to get to the next customer. @user

In our data, double modals mostly occur in informal tweets such as personal status updates, responses to other users, or comments on other tweets, images, or links. We find that double modals are mainly restricted to non-institutional accounts and informal registers, contexts similar to those found for double modals in British and American Twitter (Morin et al. 2024; Morin and Grieve 2024). Figure 1 shows the frequencies for each type in the data.

Notable is the preponderance of types with first tier modals that denote future temporality or dynamic ability in terms of volition or prediction (*will*, *would*) among the most frequent types: In total, *will* or *'ll* is the first modal tier in 151 of the 314 attested double modals; *can* in 37 instances, and *would* in 35. The most frequent modals in the second tier mostly denote possibility or ability: *can* (53), *could* (43), *might* (43), and *would* (34). The most frequent double modal overall in this data, *will can / 'll can*, has traditionally been considered to be the most widely used double modal in Scotland (Miller and Brown 1982; Brown 1991); a finding replicated by recent studies of naturalistic speech and Twitter/𝕏 data from Scotland (Coats 2023; Morin et al. 2024). Although it is theoretically possible that the authors of DMs in our data are transplanted Scots, none of the authors of authentic double modals used the term 'Scotland' in their user profiles. Given that double modals with *will* or *would* as the first element are also the most frequent types in naturalistic speech from Australia and New Zealand (Morin and Coats 2023), we tentatively propose that double modals with the schematic form *dynamic modal* (*will*, *would*) + *epistemic modal* (*can*, *could*, *might*) are the default forms for Australia and New Zealand, a pattern that corresponds to British usage and contrasts with double modals in the US, where the first tier epistemic *might* is more common.[4] Given the low frequency of the feature overall, however, additional data would be necessary to corroborate this interpretation.

Figure 2 depicts the geographical distribution of the 314 verified double modals in the dataset. As can be seen, the feature is reasonably well distributed across Australia, occurring with higher absolute frequencies in the larger state capital regions of Sydney, Melbourne, Brisbane, Adelaide, and Perth. In New Zealand, double modals occur mostly in Auckland, Wellington and other North Island locations; one double modal is attested in the South Island in Christchurch. Overall, the feature does not appear to show a distinctive geographical pattern, a finding in line with those of a recent corpus study based on geolocated YouTube data from Australia and New Zealand (Morin and Coats 2023) and corresponding to the general research

https://doi.org/10.1017/S0266078424000191 Published online by Cambridge University Press

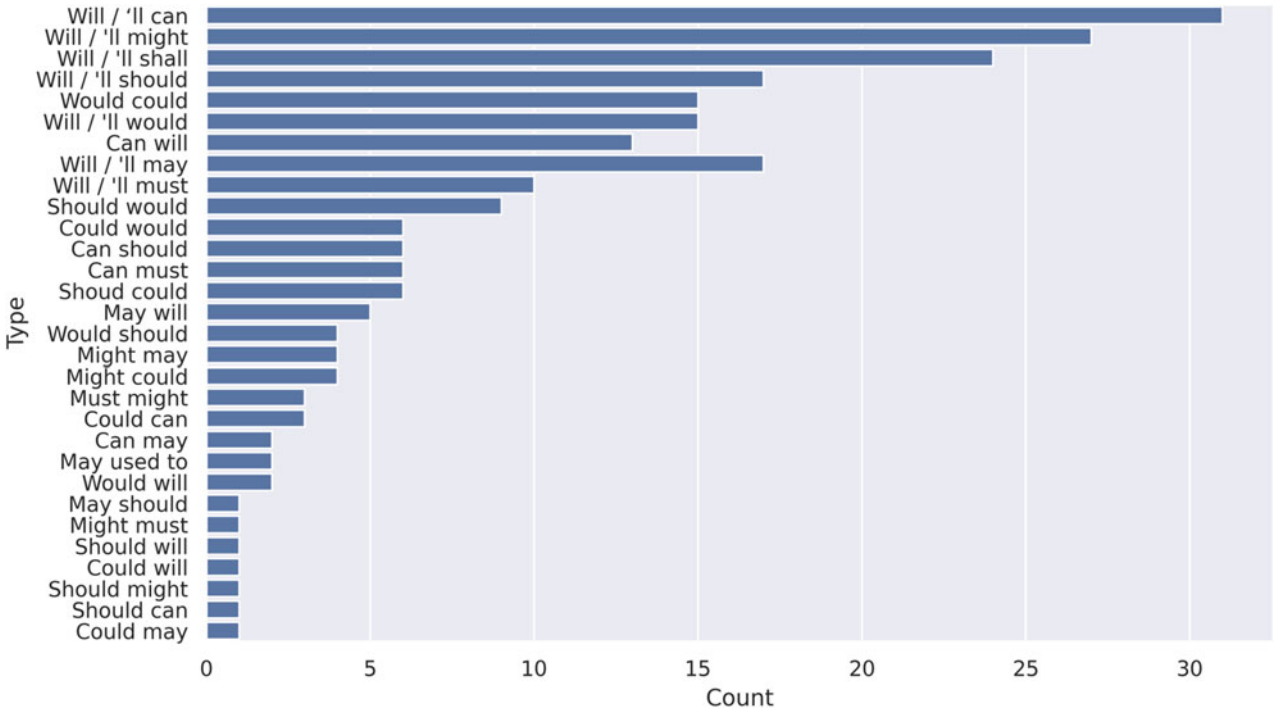


Figure 1. Double modal type frequencies.

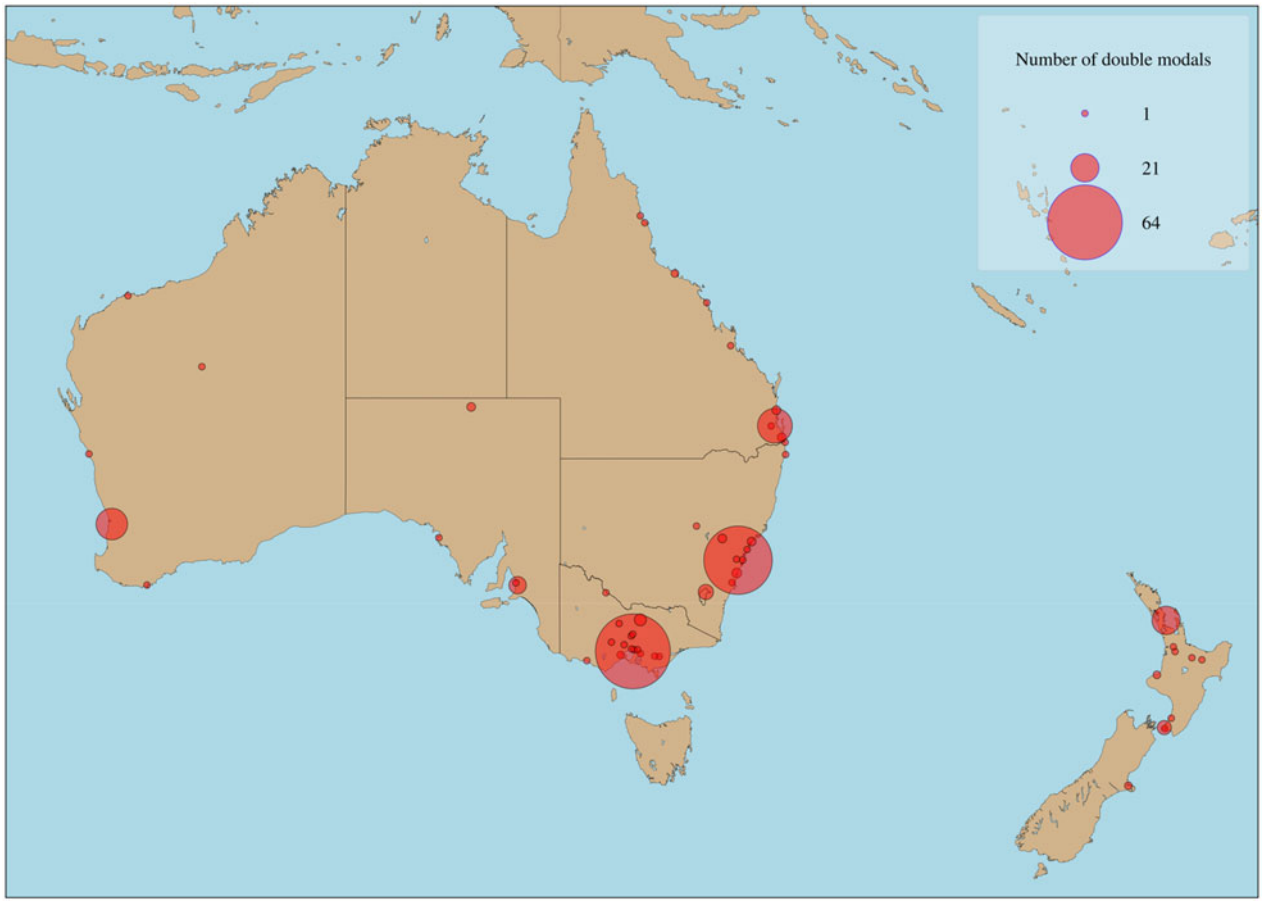


Figure 2. Locations of double modals.

https://doi.org/10.1017/S0266078424000191 Published online by Cambridge University Press

consensus that Australian and New Zealand varieties of English exhibit relatively little regional variation at the level of syntax (e.g. Hundt, Hay, and Gordon 2004; Murray and Manns 2020). The overall low frequency of occurrence of the feature in this data, both in Australia and New Zealand, however, is insufficient to conclusively demonstrate regional variation or a lack thereof.

4. Conclusion

Non-standard grammatical and syntactic features of English have attracted considerable research interest in recent years, and large corpora of naturalistic data, prepared from online sources such as video subtitles or social media messages, have opened up new perspectives for the analysis of rare syntactic phenomena. Double modals, a feature long thought to occur almost exclusively in English varieties of the Northern UK and the Southern US, have recently been analyzed in a new wave of research employing computational methods and large corpora of online content. The feature has recently been attested in English in broader geographic contexts and in a larger number of combinatorial types than has been previously attested on the basis of surveys and elicited data. Building on the results of Morin and Coats (2023) for streamed video content, we find double modals to be a rare but consistently used feature of social media writing on Twitter in English from Australia and New Zealand, a fact which supports our interpretation of the feature as not restricted to particular geographically defined varieties, but possible for most speakers of English, regardless of location. Our Twitter/𝕏 data suggest that the feature occurs mostly in informal online writing in Australia and New Zealand, with a type inventory dominated by forms in which the first modal is *will* or *would*, largely corresponding to the British pattern, rather than the American. No clear geographical patterning of the feature within Australia or New Zealand is evident in our dataset.

An important pathway for future work on double modals in Australia and New Zealand will be to examine larger data sets, such as that collected by Bruns et al. (2017); this may allow an analysis of potential differences between Australian and New Zealand usages, as well as consideration of social and demographic parameters which may correlate with double modals, topics which our data is not sufficiently large to explore. In addition, larger datasets may permit an interpretation of the modal meanings encoded by particular combinatorial types. For American and British English varieties, survey and elicitation data on double modal usage predated corpus analyses, but for Australia and New Zealand, no surveys have been conducted which investigate the grammatical acceptability of the feature or the possible modal meanings of particular combinatorial types. Collecting such data would provide further insight into the use of the feature in these varieties.[5]

From the perspective of the typology of English varieties, future work should also examine similar corpora of naturalistic speech and social media texts from 'outer circle' varieties of English, for example in India, Southeast Asia, Africa, and the Caribbean. If double modals can be attested in these varieties, the comparison of inventories with those of the UK, North America, and Australia and New Zealand may shed light on the historical provenance of the feature and its spread throughout English-speaking communities, as well as provide a basis for theoretical accounts of its syntax, for example in the context of Construction Grammar (Morin et al. 2020; Morin 2023).

Much remains to be explored regarding double modals, in Australia, New Zealand, and elsewhere. We expect that in the coming years, ongoing advances in the preparation and processing of naturalistic online data will continue to provide a rich basis for the analysis of this feature, and, in a broader sense, of rare features of English syntax in general.

Notes

1 A more comprehensive filtering-based approach to the identification of Australian Twitter/𝕏 accounts, utilizing the same basic approach but starting from a larger seed dataset and filtering on the basis of time zones and words in the user profile, is described in Bruns et al. (2017).
2 Until 2019, tweets broadcast from GPS-enabled mobile devices could, if the account owner had selected the appropriate setting, include exact latitude-longitude coordinates at the time the tweet was published.
3 See Coats (2023) and Morin et al. (2024) for details on the data processing steps.
4 Various proposals have been made as to the semantic classification of modals (and of double modals). In this report, we use a simplified version of the categories proposed by Biber et al. (1999), distinguishing between modals of possibility (epistemic), of ability or prediction (dynamic), and of obligation (deontic).
5 Given that double modals have not previously been described extensively in these varieties and are unlikely to be acknowledged as acceptable, such a survey could focus instead on meanings in the context of 'errors', providing material for an elaboration of theoretical accounts of errors as a factor in language change.

References

Adger, D., and Trousdale, G. 2007. "Variation in English syntax: Theoretical Implications." *English Language and Linguistics* 11 (2): 261–78.

Beal, J. 2004. "English Dialects in the North of England: Morphology and Syntax." In *A Handbook of Varieties of English, Volume 2: Morphology and Syntax*, edited by B. Kortmann, and E. Schneider, 114–41. Berlin: Mouton de Gruyter.

Bernstein, C. G. 2003. "Grammatical Features of Southern Speech: Yall, Might Could, and Fixin to." In *English in the Southern United States*, edited by S. Nagle, and S. Sanders, 106–18. Cambridge, UK: Cambridge University Press.

Brown, K. 1991. "Double Modals in Hawick Scots." In *Dialects of English: Studies in Grammatical Variation*, edited by P. Trudgill, and J. Chambers, 74–103. Essex: Longman.

Bruns, A., Moon, B., Münch, F., and Sadkowsky, T. 2017. "The Australian Twittersphere in 2016: Mapping the Follower/Followee Network." *Social Media + Society* 3 (4).

Coats, S. 2023. "Double Modals in Contemporary British and Irish Speech." *English Language and Linguistics* 27 (4): 693–718.

Coats, S. 2024. "Naturalistic Double Modals in North America." *American Speech* 99 (1): 47–77.

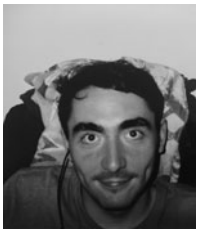
De Smet, H. 2020. "Are Changes Transmitted Mistakes?" *Cognitive Semiotics* 13 (1): 59–64.

https://doi.org/10.1017/S0266078424000191 Published online by Cambridge University Press

Fennell, B., and Butters, R. 1996. "Historical and Contemporary Distribution of Double Modals in English." In *Focus on the USA*, edited by E. Schneider, 265–88. Amsterdam: John Benjamins.
Grieve, J., Hovy, D., Jurgens, D., Kendall, T. S., Nguyen, D., Stanford, J. N., and Sumner, M. (eds.) 2023. *Computational Sociolinguistics*. Lausanne: Frontiers Media SA.
Huddleston, R., and Pullum, G. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
Hundt, M., Hay, J., and Gordon, E. 2004. "New Zealand English: Morphosyntax." In *A Handbook of Varieties of English, Volume 2: Morphology and Syntax*, edited by B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie, and C. Upton, 560–92. Berlin: Mouton de Gruyter.
Kortmann, B. 2010. "Areal Variation in Syntax." In *Language and Space: An International Handbook of Linguistic Variation, Volume 1: Theories and Methods*, edited by P. Auer, and J. E. Schmidt, 837–64. Berlin: De Gruyter Mouton.
Kortmann, B., Lukenheimer, K., and Ehret, K. 2020. *The Electronic World Atlas of Varieties of English*.
Miller, J., and Brown, K. 1982. "Aspects of Scottish English syntax." *English World-Wide* 3 (1): 3–17.
Montgomery, M., and Nagle, S. 1993. "Double Modals in Scotland and the United States: Trans-Atlantic Inheritance or Independent Development?" *Folia Linguistica Historica* 14 (1): 91–107.
Morin, C. 2021a. "Double Modals in Scots: A Speaker's Choice Hypothesis." In *Aspects of Tenses, Modality, and Evidentiality*, edited by L. Baranziniand and L. de Saussure, 165–82. Leiden: Brill.
Morin, C. 2021b. "Construire une Triangulation Méthodologique en Variation Syntaxique Restreinte." *Echanges Linguistiques en Sorbonne* 7: 48–71.
Morin, C. 2023. "Social Meaning in Construction Grammar: Double Modals in Dialects of English." PhD diss., Paris: Université Paris-Cité.
Morin, C., and Coats, S. 2023. "Double Modals in Australian and New Zealand English." *World Englishes* (Early View): 1–24.
Morin C., and Grieve, J. 2024. "The Semantics, Sociolinguistics, and Origins of Double Modals in American English: New Insights from Social Media." *PLoS ONE* 19 (1): e0295799. https://doi.org/10.1371/journal.pone.0295799.
Morin, C., Desagulier, G., and Grieve, J. 2020. "Dialect Syntax in Construction Grammar: Theoretical Benefits of a Constructionist Approach to Double Modals in English." *Belgian Journal of Linguistics* 34: 252–62.
Morin, C., Desagulier, G., and Grieve, J. 2024. "A Social Turn for Construction Grammar: Double Modals on British Twitter." *English Language and Linguistics* (Early View): 1–29.
Murray, L., and Manns, H. 2020. "Lexical and Morphosyntactic Variation in Australian English." In *Australian English Reimagined: Structure, Features and Developments*, edited by L. Willoughby, and H. Manns, 120–33. London: Routledge.
Roesslein, J. 2015. *Tweepy* (Python library). https://github.com/tweepy/tweepy.
Smith, J., Adger, D., Aitken, B., Heycock, C., Jamieson, E., and Thoms, G. 2019. *The Scots Syntax Atlas*. University of Glasgow. Accessed December 31, 2023. https://scotssyntaxatlas.ac.uk.
Zullo, D., Pfenninger, S., and Schreier, D. 2021. "A Pan-Atlantic 'Multiple Modal Belt?'" *American Speech* 96 (1): 1–46.



STEVEN COATS is a lecturer at the University of Oulu, Finland, with interests in corpus linguistics and the language of computer-mediated communication. He has created the Corpus of North American Spoken English (CoNASE), the Corpus of British Isles Spoken English (CoBISE), the Corpus of Australian and New Zealand Spoken English (CoANZSE), and the Corpus of German Speech (CoGS), as well as CoANZSE Audio (https://coanzse.org), a searchable online version of CoANZSE, which contains audio and alignments, in addition to speech transcripts. Email: steven.coats@oulu.fi



CAMERON MORIN is a temporary lecturer at the Ecole Normale Supérieure de Lyon, specialising in usage-based cognitive linguistics, sociolinguistics, dialectology, and language variation and change. His research includes notable publications on double modals in various English dialects, based on large corpora of geolocated social media data. His work also focuses on issues of social meaning in the theory of Construction Grammar. He recently completed his PhD in linguistics at the Université Paris-Cité in 2023.

https://doi.org/10.1017/S0266078424000191 Published online by Cambridge University Press