Psychology of Language and Communication 2022, Vol. 26, No. 1

sciendo

DOI: 10.2478/plc-2022-13

Steven Coats
English, University of Oulu, Finland

A database of North American double modals and self-repairs from YouTube

Sequences of two modal verbs in spoken English can represent use of a nonstandard syntactic feature (double modal) or a corrected utterance in which a speaker begins with one modal auxiliary, but switches to another (self-repair). This article presents the Double Modals and Self-Repairs (DMSR) database, a table of naturalistic double modals and self-repairs in videos from local government entities in North America, created from the *Corpus of North American Spoken English* (*CoNASE*). The paper describes the procedures used for the database's creation, discusses potential uses, and presents an exploratory analysis in which a logistic regression classifier is trained with *CoNASE* data to distinguish authentic double modals from self-repair sequences on the basis of local discourse context. The analysis demonstrates how large corpora of speech can be used to investigate the links between syntactic and pragmatic phenomena and shows specifically that double modals are an interactive device, while two-modal sequences as self-repairs may be the result of high cognitive load. The paper concludes with a discussion of multimodal corpus creation from YouTube for the study of lexical, syntactic, and interactional phenomena in speech as well as for the analysis of complex, multilevel computer-mediated communication (CMC) phenomena.

Key words: multimodality, corpus linguistics, dialectology, modal verbs, self-repair.

Address for correspondence: Steven Coats
English, University of Oulu, Oulu, Finland
E-mail: steven.coats@oulu.fi

This is an open access article licensed under the CC BY NC ND 4.0 License.

Modal auxiliaries in English exhibit several formal characteristics that distinguish them from full verbs, one of which is that in standard language, they do not co-occur within a verbal phrase.[1] Some dialects of English, however, exhibit the use of constructions such as such as "*I might could help you with that.*" Double modals have been documented in Scotland, Northern Ireland and Northern England (Beal, 2004; Brown, 1991; Corrigan, 2010) and Caribbean Englishes and creoles (see Zullo et al., 2021 for a summary), but they are a syntactic feature primarily associated with the spoken English of the Southern United States (Bernstein, 2003; Montgomery, 1989, 1998; Schneider 2003, 2004). Due to the rareness of double modal constructions, however, most studies of their geographical distribution, possible combinatorial types, and syntactic behavior in the United States have been based on relatively limited data, such as elicited grammaticality judgments, typically administered in a specific place (e.g., Butters, 1973; Close, 2004; Hasty, 2012, 2014; Williamson, 2018), or responses to worksheet items asked of informants during the data collection phase of two major 20th-century linguistic atlases: *The Linguistic Atlas of the Middle and South Atlantic States* (*LAMSAS*, McDavid & O'Cain 1980), and *the Linguistic Atlas of the Gulf States* (*LAGS*, Pederson et al. 1986–1992). Evidence for multiple modal usage outside the American South, for example, in New England, the Midwest, the mountain states of the West, or the Pacific coast states, is limited (for a few examples, see Antieau, 2006; Di Paolo, 1989). As far as is known, the feature has not been considered in Canadian English (see, however, Coats, 2022). Much of the existing data has been summarized in tabular form in Reed and Montgomery (2016).

Although grammaticality judgments and linguistic atlas survey data, as well as a few sociolinguistic studies (Feagin, 1979; Labov, 1972) demonstrate a rich diversity of double modal types in North American speech, they may not adequately consider the range of possible combinations: Data collection for *LAMSAS* and *LAGS* included a worksheet item designed to elicit *might could*, but not other double modals, and grammaticality judgment surveys have typically also presented participants with a limited number of two-modal combinations rather than all two-tier possibilities. As for the geographical extent of double modals, studies have focused primarily on the American Southeast, where they are known to be a characteristic dialect feature, and have only rarely considered their use in other parts of the US. In addition, most of the data for double modals recorded in *LAMSAS* and *LAGS* do not represent naturalistic utterances, but rather suggested responses (Bailey & Tillery, 1999).

Sequences of two modal verbs in spoken English can be double modals, but they can also occur as corrections or self-repairs: utterances in which a speaker replaces an already-articulated modal verb with a different modal verb, often after a pause or a filler word such as *uh* or *um*. The phenomenon of correction or self-repair (Schegloff et al., 1977) has been analyzed from psycholinguistic perspectives, for example, in terms of what it can tell us about lexical access and production (Levelt, 1983;

1 For an overview of the formal criteria of modal auxiliaries, see, for example, Palmer (1990); Quirk et al. (1985).

Levelt & Cutler, 1983; Levelt et al., 1999), and also from conversation-analytical perspectives, where it is interpreted as a pragmatic cue that functions to maintain turn-taking regularity (Fox et al., 1996; Schegloff, 1987; Schegloff et al., 1977).

This article introduces a manually checked database of double modals and self-repairs found in naturalistic speech in publicly available videos from YouTube. The Table of North American Double Modals and Self-repairs (DMSR),[2] created from data from the *Corpus of North American Spoken English* (*CoNASE*, Coats, forthcoming), contains over 2,000 naturalistic utterances judged to be either authentic double modals or instances of self-repairs. Each entry table is linked to the URL of the video in which it occurs at a time three seconds prior to the utterance. In addition to its possible utility as a resource for the study of a rare syntactic feature, it may also be useful for the study of self-repair from psycholinguistic or pragmatic perspectives. By linking directly to video recordings of naturalistic usages, the resource exemplifies how publicly available data from video sharing sites can be utilized to create specialized multimodal corpora suitable for the study of complex computer-mediated communication (CMC) phenomena which, according to a theoretical classification proposed by Dynel (2014), can be manifest at multiple conceptual levels: the recorded spoken interaction between interlocutors in a given video, the "one-to-many" broadcast-style interaction between a streamer or channel owner and viewers of a stream or a video, and the interaction between viewers commenting on a particular video or on the content of a channel (see Section 5, Methodological considerations).

The article is organized as follows: In the next section, the *CoNASE* corpus from which the database was created is briefly introduced and the methods used to identify, verify, and filter double modals and to distinguish them from self-repairs are discussed. Next, a few perspectives for possible work with the database in dialectology, sociolinguistics, syntax, and pragmatics are provided. Then, the article presents an exploratory study in which a machine-learning model is trained to distinguish authentic double modals from self-repair instances on the basis of the local discourse context in which the utterance occurs. It can be demonstrated that the lexical types associated with authentic double modals are different from those that occur in the context of self-repairs, supporting previous interpretations of self-repairs as resulting from an increase in lexical planning difficulty brought about by high cognitive load (Bortfeld et al., 2001; Plevoets & DeFrancq 2018). Finally, the potential uses of the database and similar resources for the study of multimodal und multilevel CMC phenomena are noted and some caveats are discussed.

Data and Methods

The data for the study come from *CoNASE* (Coats forthcoming, 2019, 2020), a 1.25-billion-word corpus of automatic speech recognition (ASR) transcripts

2 Available at https://stcoats.github.io/dm_table1.html.

from 2,572 YouTube channels of local governments in the United States and Canada created from 2018–2020. Channels were identified primarily via the implementation of a web-scraping script that searched homepages of local government entities for links, as well as via targeted searches and calls to the YouTube API. Transcript files were downloaded using scripts from the open source software *YouTube-DL* (https://youtube-dl.org). Manual examination and custom scripts removed various types of channel false positives, such as channels not from the targeted municipality, channels of municipalities with the same name, but not from North America, transcripts not generated by YouTube's ASR algorithm, or automatically-translated transcripts of non-English-language videos. Geolocation of channels was achieved with a geocoder. A custom script created word timing information from the transcript files. Part-of-speech tagging with tags from the Penn Treebank tagset was done using *spaCy* in Python (Honnibal et al., 2020). The corpus contains transcripts from all 50 U.S. states, the District of Columbia, and all 13 Canadian provinces and territories.

Regular Expressions

Regular expressions and manual checking of video recordings were used to identify authentic double modals and self-repairs and distinguish them from various types of false positives, such as homonyms, ASR errors, or clause or speaker overlaps.

Combining the 13 modals or semi-modals *can*, *could*, *may*, *might*, *must*, *shall*, *should*, *will*, *would*, *ought to*, *oughta*, *'ll*, and *used to* and excluding repetitions (e.g., *can could*, *can may*, *can might*, etc., but not *can can*) results in 156 different two-tier permutations. These 156 types were used to create regular expressions that capture two modal verbs in sequence, but also strings with intervening negators, personal pronouns, and perfect auxiliaries (e.g., *might not could* or *can't should*; *might he should*; *will they might*; *shouldn't we've not ought to*, etc.). Each of the 156 regular expressions was used to search the entire *CoNASE* corpus. The resulting 24,530 hits were combined into an editable HTML table with columns indicating the state/province/territory, the name of the channel from which the video was downloaded, the two-modal combination, the exact string captured by the regular expression (including negators, auxiliary verbs, and personal pronouns, if applicable), the URL of the video in which the sequence can be found at a time three seconds prior to its utterance, and two columns for manual annotation (see Figure 1). Approximately 3,400 of these hits were then checked manually in order to verify their status as double modals, as self-repairs, or as false positives of various types.

Manual Filtering

Many of the regular expression hits in the corpus were not double modals or self-repairs: for example, homonyms such as "*food from a **can would** not taste*

Table 1. Excerpt From Annotation Table

6786	Florida	CityofDestin	might would	might would	https://youtu.be/ZmSqPDHe2OU?t=2250	fp a1 sr p	"and my... would be"
6787	Florida	CityofDestin	might would	might would	https://youtu.be/MSSwEeb71DM?t=3118	t	"mought would" (?)
6788	Florida	CityofVeroBeach	might would	might would	https://youtu.be/z-pkih_HnIk?t=915	t	
6789	Florida	Flagler County	might would	might would	https://youtu.be/MrzQvPb36TY?t=1004		
6790	Florida	Flagler County	might would	might would	https://youtu.be/1MMQNOH3ZS8?t=7286	fp i	Video unavailable

as good" or "***Will*** (personal name) ***might*** *help you*" are not double modals, and nor are some two-modal sequences that overlap phrase/clause boundaries (e.g., "*if I* ***might, could*** *I remark that…*"). In the annotation table, such instances were recorded with the codes *fp hn* and *fp o1* (for *false positive homonym* and *false positive overlap 1*). Because YouTube ASR transcripts are not diarized (i.e., contain no indication of speaker), some hits corresponded to a first speaker ending an utterance with a modal, then a second speaker beginning an utterance with a different modal; these were annotated with *fp o2*. Misidentifications of a word by the YouTube ASR algorithm (for example, *Mike* as *might*) were annotated with *fp a*. The code *fp u* was used for audio in which speech was unintelligible, and the code *fp i* for video errors (for example, if the video had been removed or made private after the transcript was downloaded).

Self-repairs were annotated with two different codes: those with perceptible prosodic disfluency, but without pauses, were annotated with *sr d* (for *self-repair disfluency*). Those with an intervening pause, including filled pauses such as *uh* or *um*, were annotated with *sr p* (*self-repair pause*).

Excerpts were assigned the code *t*, for "true" double modal, on the basis of two criteria: lack of perceptible disfluency and semantic/discourse coherence. Because disfluencies are almost always accompanied by phenomena such as a raised fundamental frequency (Shriberg et al., 1992), longer pronunciations (Clark & Fox Tree, 2002), or other prosodic marking (Levelt, 1983; Lickley, 2015; Postma, 2000), video segments in which the speech was fluent, semantically coherent in the context of local discourse, and without perceptible prosodic marking were deemed to be authentic double modals. The 2,072 instances containing the codes for either true double modals or for self-repairs were used to create the DMSR. Table 1 summarizes the frequency of double modal types by status as self-repair or true double modal, in order of decreasing total frequency.

Potential Applications

Dialectology and Sociolinguistics

Previous research into North American double modals mainly used data from Southern informants, not from inhabitants of Western, Northern, or New England states or of Canada. Thus, the data in the DMSR can be used to reassess the geographical

Table 1. Two-Modal Sequences by Status as Self-Repair or Double Modal

	type	sr	t		type	sr	t		type	sr	t
1	might could	22	242	31	might 'll	9	4	60	'll can	1	2
2	might can	23	233	32	will must	7	5	61	must should	2	1
3	will can	109	36	33	can would	10	2	62	'll shall	3	0
4	might would	34	107	34	will 'll	10	0	63	could used to	2	1
5	would should	94	26	35	will shall	8	0	64	may shall	3	0
6	would could	79	16	36	can might	3	5	65	shall can	3	0
7	would might	36	54	37	could would	7	1	66	'll may	1	2
8	'll will	64	12	38	will should	7	1	67	should must	1	1
9	will could	51	12	39	would will	8	0	68	would oughta	0	2
10	could might	28	32	40	would may	3	4	69	should will	2	0
11	will might	31	26	41	must may	3	4	70	can must	0	2
12	may could	19	28	42	could can	7	0	71	could must	2	0
13	may can	16	28	43	can may	4	3	72	must would	1	1
14	may should	8	28	44	should could	6	0	73	should may	1	1
15	may will	26	9	45	should can	5	1	74	might oughta	1	1
16	might should	8	23	46	can used to	4	2	75	'll should	1	0
17	can could	27	2	47	'll might	1	5	76	can shall	1	0
18	will would	28	0	48	may would	1	4	77	could should	0	1
19	should might	12	13	49	will may	2	3	78	shall would	0	1
20	can will	20	3	50	should would	5	0	79	'll could	0	1
21	might will	12	10	51	may ought to	0	5	80	shall will	1	0
22	must can	17	3	52	must shall	4	0	81	oughta can	0	1
23	may might	12	5	53	could ought to	2	2	82	oughta could	0	1
24	might must	12	5	54	must will	3	1	83	oughta should	0	1
25	would can	11	6	55	could may	2	2	84	oughta would	0	1
26	could will	15	2	56	should shall	3	1	85	shall may	1	0
27	might may	9	6	57	will ought to	1	2	86	should oughta	0	1
28	must might	10	5	58	would must	2	1	87	shall should	1	0
29	may must	12	2	59	would shall	3	0	88	must could	0	1
30	can should	13	1								

distribution of double modals in North American speech (see Coats 2022).

The DMSR may also shed light on the history of the feature. While it is generally accepted that double modals in North America are an inheritance of the speech of settlers from the British Isles, mainly of Scots-Irish from Ulster, in Southern American English, double modals in which the first modal is *might* or *may* are the most common types, but in Northern England, Scotland, Northern Ireland, or the Caribbean, the type *will can*, which is only infrequently attested in the US South, is more common (Montgomery & Nagle 1994). Transatlantic differences in double modal inventories, as well as the larger number of attested types in American varieties, have led some analysts to propose that at least some American double modal types may be 19th-centry American innovations (for a discussion, see Fennell & Butters, 1996; Montgomery & Nagle, 1994; Schneider, 2003, 2004). The new attestation data

in the DMSR could serve as a starting point for a reconsideration of the historical circumstances surrounding the transmission and maintenance of this feature.

The DMSR data is not annotated with speaker identity traits, but it would be a straightforward procedure to annotate the video excerpts with perceived speaker attributes such as age (e.g., under 25, 25–50, 50 or above), ethnicity (e.g., White, Black, Asian, Hispanic), or gender. These categories could then be employed as variables in a sociolinguistic analysis and, for example, regressed against double modal or repair frequency using standard sociolinguistic analytical procedures.

While the entries in the DMSR record only the double modals (including intervening auxiliary verbs, personal pronouns, and negators), simple scripts can be used to retrieve the immediate discourse contexts of double modal-containing utterances from the *CoNASE* corpus. An exploratory analysis is also presented below, which describes the extraction of a local discourse window for double modals and self-repairs.

Syntax and Semantics

The DMSR data provide evidence for patterns of negation and question formation that have previously been proposed to be unlikely according to some accounts (e.g., Close, 2004; Hasty, 2012, 2014). For example, in this data, double modals are found to exhibit a remarkably flexible syntax in terms of constituent ordering in question forms. For a basic non-question double modal exemplified by *might could*, question patterns equivalent to "*might you could help me?*" as well as *might could you…*, *could you might…*, and *could might you…* are attested. The data in the DMSR resource thus may be of interest for theoretical accounts of the syntax and constituency of double modals, which have been proposed to represent single lexical items (Di Paolo, 1989), adverbial modifiers (Labov, 1972), bi-clausal constructions (Nagle, 2003), the specification of a TP (tense phrase; Battistella 1995), or a merged MP (modal phrase) above T (Hasty, 2012, 2014). Additional evidence for the syntactic structure of double modals in the data comes from negative sentences: the two principal negational possibilities are *might not could* and *might could not*. These are both attested in the naturalistic data.

From a semantic perspective, the naturalistic North American double modals in the database may provide evidence for an analysis of the changing status of modal auxiliaries in English, a phenomenon which has been proposed to be particularly evident in North America (Leech, 2003; Leech et al., 2009; Myhill, 1995). Manual annotation of the modal meaning of the instances in the dataset could inform a study of the semantics of modality in *CoNASE* data or in other large corpora of contemporary spoken North American English.

Disfluencies and Corrections/Self-Repairs

In the course of normal interaction, speakers often modify their speech to correct

or repair their statements. Shriberg et al. (1992) found that approximately 10% of utterances contained self-repairs, and Bortfeld et al. (2001) found that approximately 6% of words in a corpus of transcribed English conversations from California consisted of repeats, repairs, or filler terms. Levelt (1983) introduced the terms *reparandum* (the item to be corrected) and *repair* (the material used to replace the reparandum) for the analysis of self-repairs, which may also contain an *editing term* such as a filled pause (*uh*, *uhm*, *er*), although it is also possible to repair an utterance without an editing term. In Example 1, a self-repair occurring during a meeting in Prattville, Alabama, in 2016 (https://www.youtube.com/watch?v=r2oK0ybEhn8&t=1092s), the speaker utilizes the editing term *uh,* repairing *can* to *might*.

	Reparandum	*Editing term*	*Repair*		
"… they're saying what they think and what we	can,	uh,	might	do about it."	(1)

Some repairs consist of a *retracing*, in which the speaker rearticulates an already uttered segment. Example 2, excerpted from the transcript of a meeting in Flagler Country, Florida (https://www.youtube.com/watch?v=QOAFBd7SItk&t=733s), shows a sequence in which a speaker has repaired the segment "*I can't ma(ke)*" by retracing back to the personal pronoun, without inserting an editing term.

	Reparandum	*Repair*		
"you can't make the building any smaller".	"I can't ma-	I could	probably make it shorter, but I couldn't…"	(2)

In Example 3, a speaker in a meeting in Santa Barbara, California (https://www.youtube.com/watch?v=4i_ii1mMOGo&t=7182s), makes a repair in which no retracing or editing terms are employed. In the sequence, the reparandum *would* is replaced with the repair *could*, but with a noticeable disfluency: a stutter or stammer consisting of a perceptible repetition of the initial consonant in *could*.

	Reparandum	*Repair*		
"a land trust acquired it so that it	would	could	continue…"	(3)

In the terminology of Levelt (1983), Example 3 would represent an "error repair", in which a speaker corrects an erroneous utterance (the modal *would*), as opposed to an "appropriateness repair", in which a speaker elaborates on the reparandum for the purposes of optimal communication, but does not propose a fundamentally different meaning.

Nakatani and Hirschberg (1993) showed that disfluencies can be identified solely on the basis of prosodic properties such as pitch, duration, or pause, without taking lexical or syntactic information into account. In self-repairs, disfluencies are characteristic for the reparandum, but the repair move itself exhibits prosodic marking mainly in the case of error repairs (Levelt & Cutler, 1983) rather than appropriateness

repairs: the reason is that in the former case, successful communication depends on the hearer understanding that the meaning of the utterance has been changed, whereas in the latter case, the speaker does not need to emphasize a change in meaning.

In general, self-repairs have been seen in the field of conversation analysis (CA) as a device for speaker turn maintenance or local meaning negotiation (Fox et al., 1996; Schegloff, 1987; Schegloff et al., 1977) or as a strategy used by speakers to make meanings "more specific, explicit, and clear" (Kaur 2011, p. 2712). Analyses of self-repair in the CA tradition have typically been undertaken using qualitative approaches and considering small numbers of conversations. In the DMSR data, self-repairs could be analyzed using corpus techniques in which geographical locations, combinatorial types, or categories that can be annotated such as speaker traits, modal meanings, or self-repair types (error repair or appropriateness repair) are considered as the independent variables. Self-repairs could also be considered in terms of the discourse contexts in which they occur: Are there particular topics or discourse contents that are more likely to prompt self-repair of speaker utterances?

Predicting a Double Modal or a Self-Repair Using Discourse Windows

In order to determine if the local discourse context of the utterance can be used to distinguish double modals from self-repairs, two simple "bag-of-words" machine learning models were created using Scikit-learn (Pedregosa et al. 2011).[3] In the first model, term frequency-inverse document frequency (tf-idf) values for discourse windows centered upon the utterance were used as model inputs. In the second model, state or province-level location was additionally taken into account.

The following steps were used to create the models: first, in order to capture the discourse context of the double modal/self-repair utterances, a text window of 100 words (50 tokens before the utterance and 50 tokens after) was retrieved for each of the 2,072 entries in the database by accessing the corresponding transcript from *CoNASE*. Then, the discourse windows for the 2,072 samples were converted into tf-idf matrices (Manning et al., 2008, pp. 117–120) containing values for the 9,658 unique word types in the set of 100-word windows.[4] These values were then used to create a binary logistic regression model in order to predict whether a given 100-word span containing a sequence of two modal verbs was more likely to have been annotated as a true double modal or as an instance of self-repair. The second model included 63 additional binary features: state or province-level locations, encoded as 1 if the utterance was from a video from that state/province, and as 0 otherwise.

The models were trained using 80% of the samples (1,656 discourse windows) and tested using 20% (415 discourse windows). The first model, based solely on tf-idf information, correctly classified 263 of the discourse windows (113 of the self-repairs and 150 of the double modals), for an accuracy of 63%. Inclusion of the location at

3 In a "bag-of-words" approach, the lexical types in a text are represented as model features without consideration of word order.

4 To reduce overfitting, the targeted modal forms themselves were excluded.

Table 2. Evaluation of Models

	Model 1 (only tf-idf)			Model 2 (tf-idf + state/province)			Ground-truth labels
	Precision	Recall	F1	Precision	Recall	F1	(test data)
Double modal	0.63	0.57	0.60	0.72	0.72	0.72	198
Self-repair	0.64	0.69	0.66	0.74	0.75	0.74	217
Accuracy		0.63			0.73		

state/province level increased accuracy of the model to 73% (see Table 2).[5]

The accuracy scores show that the models are not perfect, but suggest that there may indeed be differences in the relative frequencies of different word types in the immediate discourse contexts of double modals when compared to self-repairs. This becomes apparent when the weights of the most important features for the first model are examined (Table 3). These values, which correspond to the coefficients with the greatest magnitude in the formula for the logistic regression, distinguish the lexical items associated with true double modals from those associated with self-repairs in the testing data.

The grammatical functions and semantic values of the items in Table 3 suggest that double modals are associated with highly frequent word types such as conjunctions, prepositions, and adverbials, as well as some core verbs of cognition, perception, and interaction, such as *see*, *work*, *come*, *know*, and *tell*. Self-repairs, on the other hand, are associated with lexemes specific to the context of local government, such as *council*, *zoning*, *property*, *street*, *members*, *city*, and *motion*. The cardinal number words *one* and *two*, which can occur in a broad range of contexts as adverbial specifiers or as nouns, are associated with double modals. The ordinal number words *first* and *second* are associated with self-repairs: compared to *one* and *two*, *first* and *second* are more restricted in the range of grammatical contexts in which they can occur. Interestingly, the form *yeah* is also associated with self-repairs—perhaps as an editing term or as a polar response particle.

One possible interpretation for this pattern is that a double modal, as a non-standard feature, may be more likely to occur during relatively informal discourse occurring in official meetings, such as when council members are not attending to specific agenda items. Such discourse is often formulaic in nature and does not necessarily involve specific information. Conversely, when negotiating highly specific content, such as the text of local ordinances or decisions on budget items, speakers make errors due to the relatively high demands imposed upon their short-term memory and a resultant impingement of their speech planning facility.

5 Other classifiers, such as a Support Vector Machine, a Bernoulli Classifier, and a Naïve Bayes Classifier, were also tested. They achieved similar F1 scores. For baseline comparison, a "dummy" classifier in which all text windows were classified as the majority class (self-repairs) was created. This baseline classifier achieved an overall accuracy of 0.52 (= proportion of ground-truth "self-repairs" in the test data). Precision, recall, and F1 scores of the dummy classifier were 0 for the "true" label, and 0.52, 1, and 0.69 for the "self-repair" label.

Table 3. Model Weights for Vocabulary in tf-idf Model

feature	weight	feature	weight
do	1.528568	council	-1.35507
you	1.298725	the	-1.00066
but	1.276029	so	-0.92935
way	1.143866	zoning	-0.83738
one	1.095031	yeah	-0.82878
see	0.994165	by	-0.79852
work	0.887884	property	-0.77293
two	0.878656	street	-0.73308
not	0.873683	second	-0.73253
county	0.862826	very	-0.72794
some	0.862771	members	-0.71661
came	0.834398	city	-0.65825
things	0.82835	motion	-0.65402
help	0.797828	many	-0.65257
me	0.78735	six	-0.6517
know	0.785432	first	-0.65119
three	0.758408	bringing	-0.64172
budget	0.733508	strong	-0.6157
if	0.732846	others	-0.6118
tell	0.725491	great	-0.60491

A few examples of discourse windows which have been highlighted with colors to indicate the relative weights of individual lexical items in the model's prediction for that window can help understand the role of discourse context as a prompt for double modals and self-repairs. The visualizations for the examples below were generated with *eli5* (https://github.com/TeamHG-Memex/eli5), a library for the analysis of machine-learning "black box" algorithms based on the LIME framework (Ribeiro et al., 2016).

Classification as double modal

The first example (see Figure 2) shows the 100-word window around an occurrence of *might not can* which was manually annotated as a true double modal. The logistic regression classifier correctly assigned the label for double modal to the excerpt based on the model weights. The words marked with green have contributed towards the classification (in this case, as a double modal), whereas those in red are weighted towards the other binary outcome (as a self-repair); the color saturation indicates the relative weight of each lexical item.

In the transcript excerpt, a speaker at an event hosted by a local school board accepts an award for achievements as a schoolteacher. The speaker directly addresses the audience, and particularly students, expressing appreciation and positive emotion. Word types often associated with interactivity are weighted in

Figure 2. *Discourse Window for an Acceptance Speech*

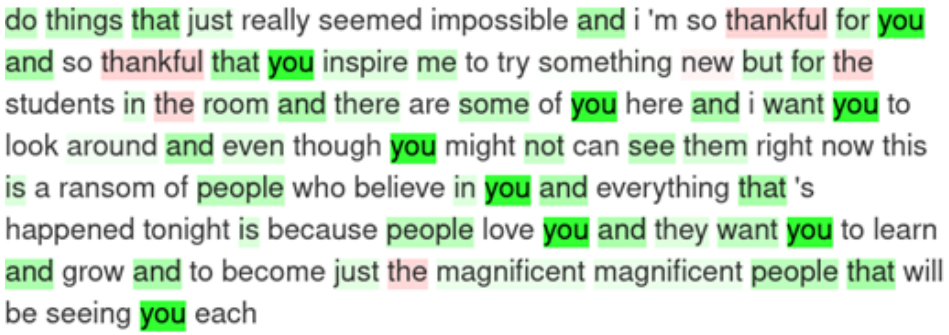
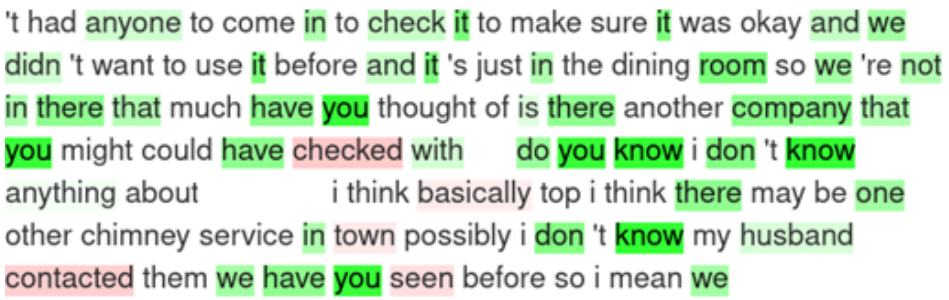


Figure 3. *Discourse Window for a Historical Preservation Discussion*



the model to favor use of a double modal instead of a self-repair. In this passage, the word *you* contributes the most to the classification of the window as a double modal context, along with function words such as the conjunction *and*, the relativizer *that*, and the verbs *want* and *see*, among others.[6]

The second double modal example (see Figure 3) is an excerpt from a meeting of a historic preservation committee in which a citizen and a committee member are discussing an application for historical status for the citizen's home. The passage represents a conversational exchange in which the applicant responds to a specific query about the home's chimney by utilizing a narrative mode in order to explain why a chimney cleaning service has not been contracted. The words in the window contributing to the classification as a double modal are the pronouns *you, it, we*, and *there*, and the verbs *know* and *have*, among others.[7]

6 The word *ransom* is an ASR error; the speaker says *roomful*. In this case, the reason for the ASR error is likely to be a perceptible clipping of the acoustic signal in the corresponding video at the moment of utterance of this item.

7 The gaps in the text are personal names, with no weight in the model, that have been manually obscured.

In the third example (see Figure 4), a councilperson has just noted that it is the Fourth of July weekend (the Independence Day holiday in the United States) and is telling a story about her recent encounter with military personnel preparing for a foreign deployment. The narration has an interactive quality, as the narrator playfully addresses another councilperson in the chamber ("*i didn't shake their hand you would have done that but i hugged them you might not could have gotten away with that*"). Again, interactive features such as pronouns (*me, you, it*), pronouns or relativizers (*those, that*), and function words (*and, but, if*) contribute the most towards the classification of the window as a double modal context.[8]

Classification as Self-Repair

Figures 5–7 show discourse windows that have been annotated as self-repairs and correctly identified as such by the logistic regression model. Again, the green-backgrounded words are those that contribute towards this classification, whereas red-backgrounded terms contribute towards a classification as the other category (as a double modal). In these examples it is evident that lexical items used in the context of council meetings to refer to council organization and procedure, words used to describe administrative tasks and legal concepts, and words used to impart specific information are likely to prompt a self-repair instead of a double modal.

In Figure 5, the two-modal sequence *will shall* was manually annotated as a self-repair during the initial creation of the DMSR dataset. The window is an excerpt from a transcript in which a councilperson reads a portion of a proposed development ordinance and then makes a comment on the text. The weights assigned to terms such as *development* and *party*, the discourse organizers *the*, *by*, and *of*, and verbs including *agree* and *terminate* result in the classification of the excerpt as a self-repair.

In Figure 6, the sequence *may must* occurs as a self-repair in the context of a discussion of a proposed change to the text of the local building code.[9] Lexical items specific to official meetings such as *motion*, *amended*, or *counselor*, and items used for imparting specific information such as articles, prepositions, and the verbs *change* and *maintain* contributed to the classification.

Finally, in Figure 7, the self-repair *will can* occurs in an excerpt from a meeting in which a participant introduces detailed information pertaining to city environmental policies. In the corresponding video segment, the speaker, who is reading from a prepared text, makes two additional self-repairs recorded in the transcript window: for *eighth floor one 104,* the item *one* is realized with disfluency; the speaker then repairs by repeating the item in *104*.[10] In the sequence *switching out their life their current lightbulbs*, *life* is the reparandum,

8 The word *hooked* in the passage is an ASR error (*hugged*).

9 The gaps represent personal names, with no weights in the model, that have been manually obscured.

10 The utterance "eight four one- one oh four" (84104), a local post code for Salt Lake City, Utah, has been transcribed by the ASR as a combination of numeral words and numerals. The word *floor* is an ASR error.

Figure 4. *Discourse Window for an Interactive Narration*

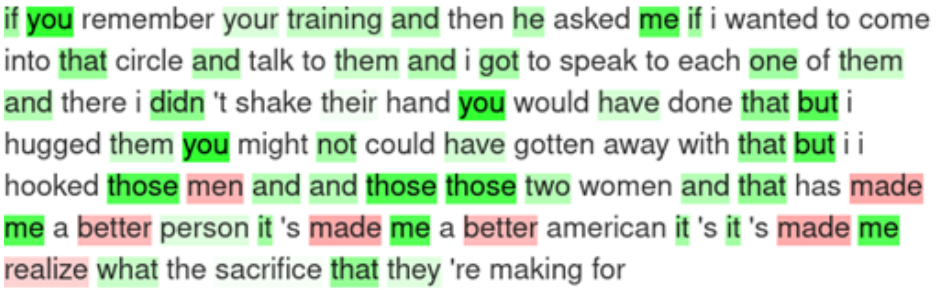


Figure 5. *Discourse Window for a Spoken Legal Text*

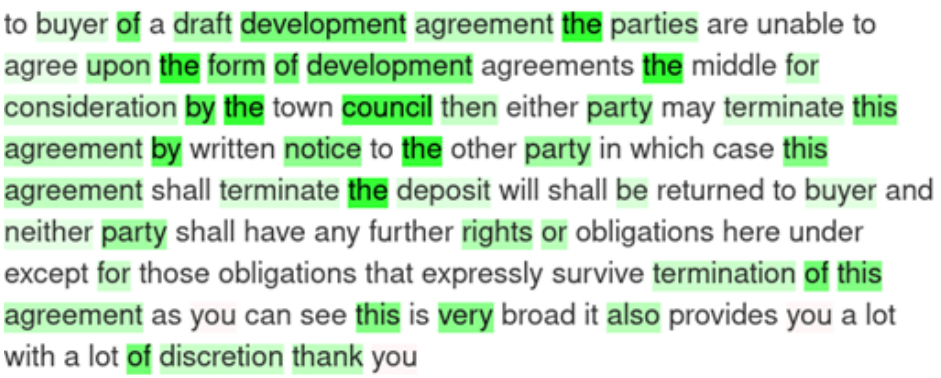


Figure 6. *Discourse Window for a Proposed Ordinance Change*

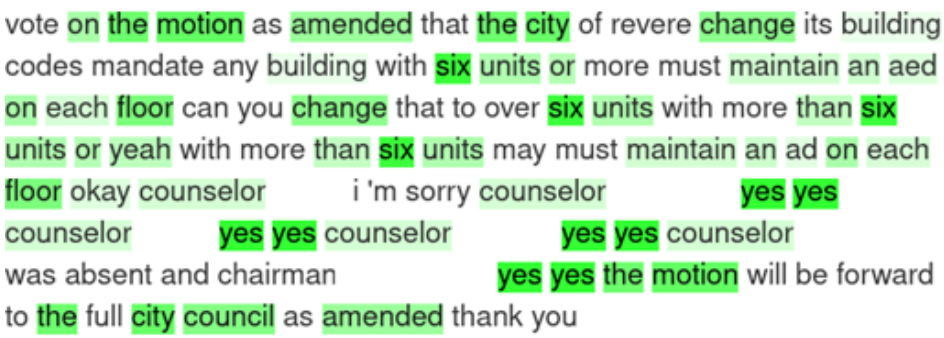
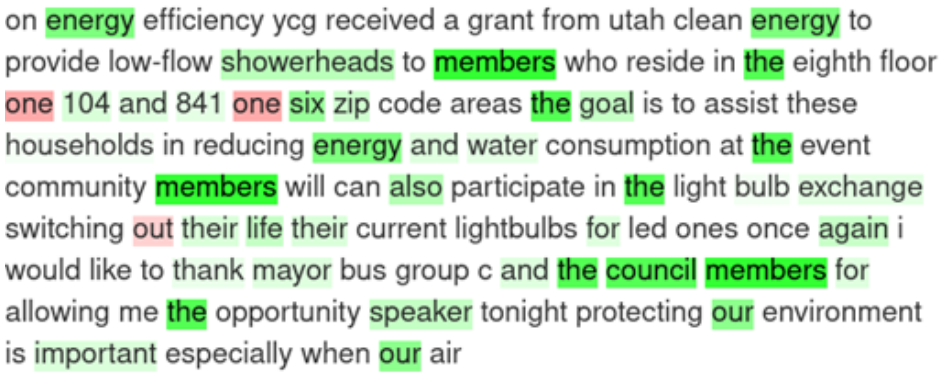


Figure 7. *Discourse Window for a Grant Discussion*



and the repair move retraces to repeat *their* and issue the repair item *lightbulbs*. Information-laden terms such as *energy, members, council, 104*, and *841*, among others, have contributed to the classification as a self-repair.

Discussion of Classification Based on Discourse Context

The distinction proposed between involved and informational discourse (Biber, 1994; Biber & Finegan 1988, 1989, 1992), proposed to correspond to differences in fundamental stance configuration that are manifest as frequencies of lexical and grammatical features in different kinds of written and spoken genres, helps to explain the classifications shown above. In general, the terms in the model associated with double modals are more likely to be used in interactive or involved discourse, whereas those associated with self-repairs are more likely to be used for the communication of specific information.

Public meetings of local councils can consist of different types of spoken interaction between participants. Council meeting transcripts contain relatively informal, spontaneous exchanges, in which interlocutors make reference to the immediate local discourse context and discuss personal feelings and attitudes, but they also contain passages in which more formal stances are used to impart specific information, for example by reading prepared texts or making reference to meeting procedures, agenda items, legal ordinances, and other types of content, some of which is informationally dense. In the examples provided above, we can observe that double modals tend to occur in spontaneous utterances that communicate emotional states and attitudes or involve narration, whereas self-repairs occur when prepared texts are read aloud, ordinances or other legal content is discussed, or highly specific information is imparted: in other words, in contexts in which "heavier demands are placed on the speech planning system" (Bortfeld et al., 2001, p. 141). Speech contexts representing informational stances

such as this are more likely to prompt disfluencies than are more involved stances in which interlocutors can plan their speech without simultaneously preparing and delivering informationally dense content.

For the purposes of this exploratory analysis, the finding that discourse context can predict whether a given two-modal sequence represents a double modal or a self-repair is in line with previous findings from studies of speech planning in which cognitive load was found to be a significant predictor of disfluency (Bortfeld et al., 2001; Levelt & Cutler, 1983; Levelt et al., 1999; Plevoets & DeFrancq, 2018).

With additional annotation, it may be possible to utilize the DMSR in a similar manner for discourse-context-based predictive modeling of (for example) the modal meanings of double modal expressions in terms of epistemic, deontic, or dynamic modality.

Methodological Considerations

In the previous two sections it has been shown that the DMSR may be useful for corpus-based or corpus-assisted studies of two specific phenomena: double modality and self-repair. In addition, some of the methods used to create the database (and to create *CoNASE*) may be suitable for the creation of specialized corpora for the analysis of specific types of multimodal and multilevel CMC interaction on YouTube or other online platforms. In this section some considerations pertaining to the creation of multimodal and multilevel corpora from YouTube and other platforms are briefly discussed.

YouTube videos can be rich sources of data for linguistic analyses, whether qualitative (e.g., Androutsopoulos, 2013) or quantitative (e.g., Coats, 2020), but communication on the platform is not limited to the utterances recorded in the site's video content. *CoNASE* and the DMSR provide direct access to video and audio content via metadata associated with each transcript file. This metadata can be used to download the corresponding video for transcripts of interest, which then allows the consideration of acoustic and visual aspects of communication in addition to the orthographic representations of speech. Thus, it would be a simple endeavor to search for a particular lexical item in *CoNASE*, assemble examples from across North America, download the corresponding video files, and convert them to audio segments, for example, in order to examine vowel quality or other acoustic properties of the speech signal. Video data, accessed in the same manner and analyzed using automated methods, may open new perspectives for the study of aspects of physical presence such as facial expression or gesture and how they may vary according to discourse content.

The participatory framework of YouTube has been described from a theoretical perspective as "mediated quasi-interaction" occurring on three communicative levels (Bou-Franch et al., 2012, Dynel, 2014), each of which comprises a different configuration of communicative parameters. The first level

can be conceived of as events recorded in the videos viewable on the platform. For videos that record spoken interaction, the communicative configuration can be characterized in terms of contextual and interlocutor factors such as location, occasion, participant status, topic, or other parameters known to affect face-to-face communication (see e.g., Goffman, 1981; Levinson, 1988). The communicative situation of the second level resembles in some ways that of traditional broadcast media, particularly in terms of the one-to-many configuration of sender and recipient (cf. Bell, 1991; Hutchby, 2006), but with the important difference that some recordings on YouTube are broadcast as live streams in which the sender may have access to additional interactive modalities during the broadcast, such as comments entered by viewers into a chat window. At the third level, according to the typology proposed by Dynel, interaction makes use of additional affordances of the YouTube platform, such as chatting or commenting on a particular video, subscribing to a channel, or liking a video.

The DMSR and *CoNASE* provide access to text transcripts and URLs of videos of interactions and can serve as a basis for the analysis of interactions occurring at the first level, and in some cases at the second level.[11] Data from comment threads, live chat transcripts, video like counts, subscriber counts, and other third-level phenomena are available through YouTube's API or by conducting targeted scrapes of the public-facing web server. Although *CoNASE* and the DMSR do not contain third-level materials, they do include video IDs and URLs as metadata, which means that simple scripting pipelines can be created to gain access to third-level phenomena.

In summary, the data collection and annotation methods used for the creation of *CoNASE* and the DMSR represent a possibility to expand upon first-generation corpora of CMC containing primarily textual content such as email, SMS or instant messenger texts, chat or bulletin board posts, Twitter texts, wiki discussion pages, blogs, or similar content (cf. Herring, 2007). By providing a basis for access to transcripts, video and audio content, as well as various types of platform interactions, the *CoNASE* and DMSR open up new possibilities for the corpus-based study of multimodal and multilevel CMC interaction.

Caveats and Summary

The *CoNASE* corpus from which the database was created, as it consists of ASR transcripts, contains many errors. However, because the database has been manually verified, it excludes ASR errors and instances in which phrase/clause or speaker overlap results in two modal auxiliaries in sequence. Nevertheless, several caveats should be noted. First, the speech genre represented by the corpus, which contains mainly transcripts of local government meetings, is not necessarily a

11 For example, some of the data in *CoNASE* are transcripts of recordings that have been streamed live and contain utterances in which meeting attendees (i.e. Level 1 participants) instruct stream viewers (Level 2 participants) "write in the chat".

good indicator of the range of conversational genres that can occur in spoken North American English. Government meetings, although they can occasion informal interaction, tend to be somewhat more formal and prepared than some other speech contexts. However, because such meetings typically involve careful negotiation, they may provide a suitable pragmatic context for the occurrence of double modals, which are proposed to occur primarily in "subtle give-and-take negotiations and sensitive facesaving situations in which highly conditional and indirect speech takes place" (Montgomery, 1998, p. 96).

The method used to evaluate the authenticity of the double modals—instances that occur in fluent, non-prosodically-marked-speech which is semantically coherent—may also capture cases of repairs in which a retracing of one word (the modal form) has occurred, but neither the reparandum (the first modal), nor the repair (the second modal) is prosodically marked. However, for some speakers, it may be possible that an error repair has taken place, but prosody is unaffected (Lickley 2015). This issue, while it cannot be addressed in the context of this short introduction to the database, may represent a subject for further study, one which could shed light on the semantic relationship between careful epistemicity and the phenomenon of repair in general.

To summarize, the DMSR provides links to 2,072 double modals and self-repairs in naturalistic speech. The data, drawn from word-timed tokens in *CoNASE*, demonstrates how this corpus may fruitfully be used to quickly assemble a large number of multimodal, verifiable recordings of speech utterances. The database provides evidence for a wider geographic extent of double modals in North American English than has previously been documented, and it can serve as a starting point for corpus-based analyses of a common conversational phenomenon, self-repair. The methods used to create the resource demonstrate how multimodal and multilevel CMC corpora can be created from publicly available data on YouTube or other platforms, opening up new possibilities for dialectological, sociolinguistic, syntactic, or pragmatic research.

Acknowledgments

Thank you to two anonymous reviewers for helpful comments on the manuscript and to Finland's Centre for Scientific Computing for computational resources and storage.

Conflict of Interest Disclosure

The author does not have any conflicts of interest to report.

Funding

The author has no funding sources to disclose.

Research Ethics Statement

The Author declares that no Research Ethics Statement was required.

Authorship Details

Steven Coats: research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

References

Androutsopoulos, J. (2013). Participatory culture and metalinguistic discourse: Performing and negotiating German dialects on YouTube. In D. Tannen & A. M. Trester (Eds.), *Discourse 2.0: Language and new media* (pp. 47–72). Georgetown University Press.

Antieau, L. (2006). *A distributional analysis of rural Colorado English* [Doctoral dissertation, University of Georgia] UGA Athaneum. https://esploro.libs.uga.edu/esploro/outputs/doctoral/A-distributional-analysis-of-rural-Colorado-English/9949332921302959

Bailey, G., & Tillery, J. (1999). The Rutledge effect: The impact of interviewers on survey results in linguistics. *American Speech, 74*(4), 389–402.

Battistella, E. L. (1995). The syntax of the double modal construction. *Linguistica Atlantica, 17*, 19–44.

Beal, J. (2004). English dialects in the north of England: Morphology and syntax. In B. Kortmann, K. Burridge, R. Mesthrie, E. W. Schneider, & C. Upton (Eds.), *A handbook of varieties of English vol. 2: Morphology and syntax* (pp. 114–141). Mouton de Gruyter.

Bell, A. (1991). *The language of news media.* Blackwell.

Bernstein, C. (2003). Grammatical features of southern speech: Yall, might could, and fixin to. In S. J. Nagle & S. L. Sanders (Eds.), *English in the Southern United States* (pp. 106–118). Cambridge University Press.

Biber, D. (1994) An analytical framework for register studies. In D. Biber, D. & E. Finegan (Eds.), *Sociolinguistic perspectives on register variation* (pp. 31–56). Oxford University Press.

Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes, 11*, 1–34. https://doi.org/10.1080/01638538809544689

Biber, D., & Finegan, E. (1989). Drift and the evolution of English style: A history of three genres. *Language, 65*, 487–517. https://doi.org/10.2307/415220

Biber, D., & Finegan, E. (1992). The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, T. Nevalainen, & I. Taavitsainen (Eds.), *History of Englishes: New methods and interpretations in historical linguistics* (pp. 688–704). Mouton de Gruyter.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123–147. https://doi.org/10.1177/002383090104400201 01

Bou-Franch, P., Lorenzo-Dus, N., & Garcés-Conejos Blitvich, P. (2012). Social interaction in YouTube text-based polylogues: A study of coherence. *Journal of Computer-mediated Communication, 17*, 501–521. https://doi.org/10.1111/j.1083-6101.2012.01579.x

Brown, K. (1991). Double modals in Hawick Scots. In P. Trudgill & J. Chambers

(Eds.), *Dialects of English: Studies in grammatical variation* (pp. 74–103). Longman.

Butters, R. (1973). Acceptability judgments for double modals in Southern dialects. In R. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in linguistics*. Georgetown University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111. https://doi.org/10.1016/s0010-0277(02)00017-3

Close, J. (2004). *English auxiliaries: A syntactic study of contraction and variation* [Doctoral Dissertation, University of York]. White Rose eTheses Online. https://etheses.whiterose.ac.uk/9870/1/424126.pdf

Coats, S. (2019). A corpus of regional American language from YouTube. In C. Navarretta, M. Agirrezabal, & B. Maegaard (Eds.), *Proceedings of the 4th Digital Humanities in the Nordic Countries Conference, Copenhagen, Denmark, March 6–8, 2019* (pp. 79–91). CEUR.

Coats, S. (2020). Articulation rate in American English in a corpus of YouTube videos. *Language and Speech, 63*(4), 799–831. https://doi.org/10.1177/0023830919894720

Coats, S. (forthcoming). Dialect corpora from YouTube.

Coats, S. (2022). Naturalistic double modals in North America. *American Speech*. https://doi.org/10.1215/00031283-9766889

Corrigan, K. (2010). *Irish English, volume 1: Northern Ireland.* Edinburgh University Press.

Di Paolo, M. (1989). Double modals as single lexical items. *American Speech, 64*(3), 195–224. https://doi.org/10.2307/455589

Dynel, M. (2014). Participation framework underlying YouTube interaction. *Journal of Pragmatics 73*, 37–52. https://doi.org/10.1016/j.pragma.2014.04.001

Feagin, C. (1979). *Variation and change in Alabama English: A sociolinguistic study of the white community.* Georgetown University Press.

Fennell, B. A., & Butters, R. R. (1996). Historical and contemporary distribution of double modals in English. In E. W. Schneider (Ed.), *Focus on the USA: Varieties of English around the world* (pp. 265–88). John Benjamins.

Fox, B. A., Hayashi, M. & Jasperson R. (1996). Resources and repair: A cross-linguistic study of syntax and repair. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 185–237). Cambridge University Press.

Goffman, E. (1981). *Forms of talk.* Blackwell.

Hasty, J. D. (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), 1716–1738. https://doi.org/10.1016/j.lingua.2012.09.005

Hasty, J. D. (2014). We might should be thinking this way: Theory and practice in the study of syntactic variation. In R. Zanuttini & L. R. Horn (Eds.), *Micro-syntactic variation in North American English* (pp. 269–293). Oxford University Press.

Hasty, J. D., Hesson, A., Wagner, S. E., & Lannon, R. (2012). Finding needles in the right haystack: Double modals in medical consultations. *University of Pennsylvania Working Papers in Linguistics, 18*(2), 41–47.

Herring, S. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet, 4*. https://www.languageatinternet.org/articles/2007/761/

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength natural language processing in Python.* https://doi.org/10.5281/zenodo.1212303

Hutchby, I. (2006). *Media talk: Conversation analysis and the study of broadcasting.* Open University Press.

Kaur, J. (2011). Raising explicitness through self-repair in English as a lingua franca. *Journal of Pragmatics, 43*, 2704–2715. https://doi.org/10.1016/j.pragma.2011.04.012

Labov, W. (1972). *Language in the inner city: Studies in the black English vernacular.* University of Pennsylvania Press.

Leech, G. (2003). Modality on the move: The English modal auxiliaries 1961–1992. In R. Facchinetti, F. Palmer, & M. Krug (Eds.), *Modality in contemporary English* (pp. 223–240). De Gruyter Mouton.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in contemporary English: A grammatical study*. Cambridge University Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*(4), 41–104. https://doi.org/10.1016/0010-0277(83)90026-4

Levelt, W. J. M., & Cutler, A. (1983). Prosodic marking in speech repair. *Journal of Semantics, 2*, 205–217. https://doi.org/10.1093/semant/2.2.205

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–75. https://doi.org/10.1017/s0140525x99001776

Levinson, S. (1988). Putting linguistics on a proper footing: Explorations in Goffman's participation framework. In P. Drew & A. Wootton (Eds.), *Erving Goffman: Exploring the interaction order* (pp. 161–227). Polity Press.

Lickley, R. J. (2015). Fluency and disfluency. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 445–469). Wiley-Blackwell.

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.

McDavid, R. I., & O'Cain, R. K. (1980). *Linguistic atlas of the Middle and South Atlantic states, fascicles 1–2*. University of Chicago Press.

Montgomery, M. (1989). Exploring the roots of Appalachian English. *English World-Wide, 10*(2), 227–278. https://doi.org/10.1075/eww.10.2.03mon

Montgomery, M. (1998). Multiple Modals in LAGS and LAMSAS. In M. Montgomery & T. E. Nunnally (Eds.), *From the Gulf States and beyond: The legacy of Lee Pederson and LAGS* (pp. 90–122). University of Alabama Press

Montgomery, M., & Nagle. S. J. (1994). Double modals in Scotland and the Southern United States: Trans-Atlantic inheritance or independent development? *Folia Linguistica Historica, 14*(1–2), 91–108. https://doi.org/10.1515/flih.1993.14.1-2.91

Myhill, J. (1995). Change and continuity in the function of the American English modals. *Linguistics, 33*, 157–211. https://doi.org/10.1515/ling.1995.33.2.157

Nagle, S. J. (2003). Double modals in the southern United States: Syntactic structure or syntactic structures? In R. Facchinetti, F. Palmer, & M. Krug (Eds.), *Modality in contemporary English* (pp. 349–371). De Gruyter. https://doi.org/10.1515/9783110895339.349

Nakatani, C., & Hirschberg, J. (1993). A speech-first model for repair detection and correction. *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics* (pp. 46–53). https://doi.org/10.3115/1075671.1075748

Palmer, F. R. (1990). *Modality and the English modals (2nd ed.)*. Routledge.

Pederson, L., McDaniel, S. L., & Adams, C. M. (1986–1992). *Linguistic Atlas of the Gulf States (7 vols.)*. University of Georgia Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Plevoets, K. & Defrancq, B. (2018). The cognitive load of interpreters in the European Parliament: A corpus-based study of predictors for the disfluency uh(m). *Interpreting, 20*(1), 1–28. https://doi.org/10.1075/intp.00001.ple

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition, 77*, 97–131. https://doi.org/10.1016/s0010-0277(00)00090-1

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language (2nd ed.)*. Longman.

Reed, P., & Montgomery, M. (Eds.). (2016). *MultiMo: The database of multiple modals.* http://artsandsciences.sc.edu/multimo/welcome

Ribeiro, M. T., S. Singh, & C. Guestrin. (2016). Why should I trust you?: Explaining the predictions of any classifier. *arXiv*:1602.04938 [cs.LG].

Schegloff, E. A. (1987). Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In G. Button & J. R. E. Lee (Eds.), *Talk and social organisation* (pp. 70–85). Multilingual Matters.

Schegloff, E. A., Jefferson, G., & Sacks H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361–382. https://doi.org/10.2307/413107

Schneider, E. W. (2003). Shakespeare in the coves and hollows? Toward a history of Southern English. In S. J. Nagle & S. L. Sanders (Eds.), *English in the Southern United States* (pp. 17–35). Cambridge University Press.

Schneider, E. W. (2004). The English dialect heritage of the southern United States. In R. Hickey (Ed.), *Legacies of colonial English: Studies in transported dialects* (pp. 262–309). Cambridge University Press.

Shriberg, E., Bear, J., & Dowding, J. (1992). Automatic detection and correction of repairs in human-computer dialog. In M. Marcus, (Ed.), *Proceedings of the Fifth DARPA Speech and Natural Language Workshop* (pp. 419–424). Harriman.

Williamson, S. L. (2018). *Might should we consider this? Patterns of double modal inversion in Southern United States English* [Master's thesis, Simon Fraser University]. SFU Summit. https://summit.sfu.ca/item/19080

Zullo, D., Pfenninger, S. E., & Schreier, D. (2021). A pan-Atlantic 'multiple modal belt'? *American Speech, 96*(1), 7–44. https://doi.org/10.1215/00031283-8620506