



HAL
open science

Annotation Guide REM Project (Re-thinking English Modal Constructions: From feature-based paradigms to usage-based probabilistic representations)

Mégane Lesuisse, Benoît Leclercq, Bert Cappelle, Ilse Depraetere, Cyril Grandin

► To cite this version:

Mégane Lesuisse, Benoît Leclercq, Bert Cappelle, Ilse Depraetere, Cyril Grandin. Annotation Guide REM Project (Re-thinking English Modal Constructions: From feature-based paradigms to usage-based probabilistic representations). 2023. hal-04698213

HAL Id: hal-04698213

<https://hal.science/hal-04698213v1>

Submitted on 15 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Annotation Guide

REM Project

Mégane Lesuisse, Benoît Leclercq, Bert Cappelle, Ilse Depraetere and Cyril Grandin

Re-thinking English Modal constructions: From feature-based paradigms to
usage-based probabilistic representations.

Vers une nouvelle conception des constructions modales en anglais : Des paradigmes
basés sur les traits distinctifs à la représentation probabiliste basée sur l'usage.

ANR 16-CE93-0009

Preface

The coding guidelines were developed in the context of the ANR-FNS funded project REM – ‘Re-thinking English Modal constructions: From feature-based paradigms to usage-based probabilistic representations’ (<https://anr.fr/Project-ANR-16-CE93-0009>), coordinated by Ilse Depraetere (ULille) and Martin Hilpert (Université de Neuchâtel).

One of the main aims of the project was to get a clearer view on the factors (lexical, formal, semantic, and pragmatic) that enable English speakers to choose one among the following modal expressions: *can*, *could*, *may*, *might*, *be able to* – on the possibility side – and *should*, *must*, *have to*, *need to* and *ought to* – on the necessity side. For this study, 36 predictors were identified and defined. The overview and description that follow are meant to help the annotator analyse English modal sentences in terms of these predictors. The REM dataset, which contains 5,000 instances extracted from COCA, each of which was coded in terms of these features, is available from the companion website of the book *Models of Modals*.

Table of contents

Preface.....	2
Table of contents.....	3
Level 1. The modal verb.....	8
1.1 Form.....	8
1.1.1 Predictor 1: Following constituent.....	8
1.1.2 Predictor 2: <i>do</i> -support for <i>need to</i> and <i>have to</i>	10
1.1.3 Predictor 3: Forms of <i>be able to</i> , <i>need to</i> , and <i>have to</i>	11
1.2 Semantics.....	12
1.2.1 Meaning of the modal.....	12
1.2.2 Modality and time.....	26
Level 2. The verbal complement.....	29
2.1 Predictor 12: Present or perfect infinitive.....	29
2.2 Predictor 13: The meaning of the perfect infinitive.....	29
2.3 Predictor 14: Progressive aspect.....	31
2.4 Predictor 15: Voice.....	31
2.5 Predictor 16: Situation type.....	32
2.6 Predictor 17: Lemma form of the verb.....	32
Level 3. The grammatical subject.....	33
3.1 Form.....	33
3.1.1 Predictor 18: Subject extraposition.....	33
3.1.2 Predictor 19: Person and number.....	33
3.2 Semantics.....	34
3.2.1 Predictor 20: Animacy.....	34
3.2.2 Predictor 21: Agentivity.....	35
3.2.3 Predictor 22: Genericity.....	36
Level 4. The clause.....	37
4.1 Form.....	37
4.1.1 Predictor 23: Clause type.....	37
4.1.2 Predictor 24: Subclause introduction.....	38
4.1.3 Predictor 25: Declarative, interrogative, exclamative or imperative.....	39
4.1.4 Predictor 26: Polarity.....	39

4.1.5 Predictor 27: Scope of negation.....	41
4.2 Semantics.....	41
4.2.1 Predictor 28: Habituality.....	41
4.2.2 Predictor 29: Counterfactuality.....	42
4.2.3 Predictor 30: Hypothetical meaning.....	42
4.3 Pragmatics.....	43
4.3.1 Predictor 31: Contextual effects (strengthening and weakening).....	43
4.3.2 Predictor 32: Speech act.....	44
4.3.3 Predictor 33: Directness and indirectness.....	46
Level 5. Adverbials.....	48
5.1 Predictor 34: Strength.....	48
5.2 Predictor 35: Position.....	48
5.3 Predictor 36: Lemma form of the adverb.....	49
References.....	50

Introduction

The guidelines can be used for the annotation of sentences with modal verbs in English. In the REM project, it was designed for the analysis of sentences that contain one of the following verb types: (i) a ‘core’ modal (namely *can*, *could*, *may*, *might*, *should* and *must*), (ii) a ‘semi-modal’ (namely *have to*, *need to* and *be able to*), or (iii) a ‘peripheral modal’ (namely *ought to*).

Extracting the data

For each of these ten modal expressions, a random selection of 500 sentences was extracted from COCA (Corpus Of Contemporary American English) via Mark Davies’s interface (<https://www.english-corpora.org/coca/>), resulting in a dataset of 5,000 examples. The random selection was facilitated by the randomization function in COCA (‘Find sample: 500’ function in the interface). For the six modal auxiliaries, Part of Speech (POS) tagging was used to reduce noise. The tag *_vm** was added at the end of each query to narrow down the search to modal verbs only and ignore homonyms (e.g. the metal container *can*, the month of *May*, ‘strength’ *might*, and nominal *must*). For the semi-modals *have to* and *need to*, the extracting procedure was as follows: first, a quick search enabled us to identify the different forms of the verbs (finite and non-finite) and their overall frequency in COCA; then, a random sample of each verb form (e.g. *have to*, *has to*, *had to*, etc.) was extracted in such a way that their distribution in our dataset corresponded to that in COCA. This procedure was also applied to include a representative share of examples with *have got to* in our sample. In the case of *need to*, the extracted sample also contains modal auxiliary uses of *need* (e.g. *Cheerleading purists need not worry*), which were identified with the POS tag *_vm**. Again, the size of the share of auxiliary uses of *need*, relative to the semi-modal uses of *need to*, was determined on the basis of a randomly selected search, enabling us to have an idea of how many examples of each form had to be included. The data for *ought to* and *be able to* were extracted from the list of contexts that matched the queries ‘ought’ and ‘able to’, which likewise allowed us to obtain frequency data of the different morpho-syntactic realizations of these modal verbs (e.g. *She ought always to remember that; You oughtn’t stay in the shade with those wounds; I am able to stand and walk; We were finally able to spend time together*).

The sets of 500 concordances were saved on the interface via the function ‘Create new list’. There, the left/right context of each verb was expanded from 60 to 150 words to ensure access to a maximum of contextual information during the annotation. Finally, all ten samples of 500 concordances were copy-pasted into an Excel spreadsheet.

Excel: formatting issues

Some formatting issues emerged when importing the data into Excel. The following is an example of the result obtained, with numbers corresponding to rows:

1	COCA:2014:SPOK	Concordance 1
2	SpeakingTeacherForInstance	
3	COCA:2002:SPOK	Concordance 2
4	CNNNews	

For each concordance, the associated metadata appeared in different rows (2nd column) rather than in different columns. To rearrange the data correctly, a number of steps were taken. First, we selected the ‘concordance’ column and split the cells. This resulted in the following layout:

1	COCA:2014:SPOK	Concordance 1
2	SpeakingTeacherForInstance	
3	COCA:2002:SPOK	Concordance 2
4	CNNNews	

Second, using the ‘filter’ function on the second column, we made sure the spreadsheet only showed the source rows which were not associated with the concordance anymore:

2	SpeakingTeacherForInstance	
4	CNNNews	

We copy-pasted this column in another tab, then deleted these rows from the mother tab. Once the filter removed, we obtained this configuration:

1	COCA:2014:SPOK	Concordance 1
2	COCA:2002:SPOK	Concordance 2

Third, we added a new column, to which we copied the ‘source’ information saved earlier on in another tab:

1	COCA:2014:SPOK	SpeakingTeacherForInstance	Concordance 1
2	COCA:2002:SPOK	CNNNews	Concordance 2

This is the layout needed for the annotation.

Preparing the annotation

The lay-out of the excel file is as follows: there are a number of columns that precede the actual predictors.

- The column labelled ‘ID’ contains the number which the annotator assigned to each of the examples.
- The column labelled ‘IDorigin’ contains a number corresponding to a prior sorting of examples.

- The column labelled 'MOD' provides the modal expression that occurs in the concordance.
- The column labelled 'Source 1' specifies the date and the subgenre of the concordance, such as "COCA:1995:SPOK".
- The column labelled 'Source 2' provides more details about the subgenre source, such as "CBS_Sixty".
- The column labelled 'Genre' mentions the COCA subgenre only, such as "SPOK", "NEWS" or "ACAD".

Additional columns were added for each of the predictors to be hand-annotated.

Organization of the guidelines

The predictors are described in the rest of this coding booklet, and it is made explicit how they have been operationalized for data analysis in the context of the REM project. The organization is as follows:

1. The predictors are organized in five categories which represent different levels of analysis: (i) the modal verb, (ii) the verbal complement, (iii) the subject, (iv) the clause, and (v) adverbials. These levels will be presented in turn.
2. For each level, predictors will be defined and coding values will be provided. The values relate to either formal (F), semantic (S), or pragmatic (P) properties of the level studied. These properties have been integrated into the coding system adopted: all predictors in the dataset begin with one of the following tags: F_, S_ or P_, even if these tags are not systematically used throughout these guidelines, mainly for reasons of readability.

Tip to the annotator

To facilitate annotation of the data, the cells of each predictor can be programmed to display a drop-down list menu with the different values from which the annotator can choose. We implemented this option in the REM project. The only exceptions were predictors 17, 23 and 36, all of which are lexical in nature, for which manual annotation was needed. The creation and modification of a drop-down list menu requires the use of a separate sheet in the Excel document. The annotator is referred to <https://support.office.com/en-us> for details, as the procedure may differ depending on the version used.

Level 1. The modal verb

In the first level of analysis, attention is given to the modal verb used by the speaker. In the REM project, 10 verbs were analysed: the modal auxiliaries *may*, *might*, *can*, *could*, *should* and *must*; the semi-modals *have to*, *need to* and *be able to*, and the peripheral modal *ought to*.

For reasons of clarity and presentation, some predictors are grouped artificially. For instance, predictor 2, which pertains to the presence of the periphrastic *do* in *do*-support contexts, has been included under Level 1: the modal verb, while we acknowledge that other variables, beyond the modal verb itself, predict the use of the periphrastic *do*. In a similar way, the meaning of a modal verb (e.g., root vs. epistemic, predictor 4) is determined in context and by the co-text. So again, while the label given pertains to the modal verb, it will be clear that more encompassing ‘levels’ in the sentence are at stake here as well.

1.1 Form

1.1.1 Predictor 1: Following constituent

The first predictor concerns the form of the constituents that directly follow each modal. There are five possible values. First, the modal can be followed by a verb (cf. e.g. (1)). In this case, use the label `vb`.

- (1) Conversation with children should include new, rare words, and teachers **should encourage** the students to use descriptive words and grammatically correct statements. (ACAD, 2015)

The same label is used when a modal auxiliary is followed by a modal verb:

- (2) The guy pleaded with me to be his friend, and he kept on calling. At first I thought, I should be gracious. I **should be able to** be this guy’s friend. But my stomach turned every time he called. (MAG, 1999)
- (3) And if it gets to a point where the damage is not reversible, you **might have to** have surgery. (SPOK, 2007)
- (4) Aid to Syrians and Iraqis to protect buildings, monuments, libraries and museums **may need to** be distributed covertly. (NEWS, 2015)

Second, the modal verb can be directly followed by the subject in inversion contexts (cf. e.g. (5)). In this case, use the label `subjinv`.

- (5) “Why **should I** go on living?” she asked. “All my hopes for a little happiness are gone.” (FIC, 2015)

Third, the modal verb can be directly followed by an adverb (cf. e.g. (6)). In this case, use the label `adv`.

- (6) Music education curriculum objectives for students with disabilities **might wisely** be focused on their emotional well-being. (ACAD 2015)

Fourth, the modal verb can be used on its own, with an elided verb phrase (cf. e.g. (7)). This property, known as ‘code’, refers to the use of a modal as an “operator in reduced clause” (Quirk *et al* 1985: 137), the content of which is recoverable from the context. In this case, use the label `co`.

- (7) As for you, press your tongue flat against the roof of your mouth, covering as much as you **can**. (MAG, 2005)

A final possibility is for the modal to be followed by *not*-negation (cf. e.g. (8)), contracted or otherwise.

- (8) A cone snail doesn’t sting or bite. It stabs. This tiny sea snail is deadly. It carries nearly 100 toxins inside its shell. This snail **can’t** swim or move fast. It buries itself in the sand. (MAG, 2012)

For this kind of example, the annotator also needs to specify what constituent follows *not* and complete the label accordingly:

- `neg.vb` when negation is followed by a verb (e.g. *It’s the boys who should be accused and asked why they do this. They **shouldn’t do** this.* SPOK, 2015)
- `neg.subjinv` when negation is followed by a subject (e.g. ***Should not you and a senator** do that, challenge the electoral votes?* SPOK, 2004)
- `neg.adv` when negation is followed by an adverb (e.g. *We fully recognize that most students **might not voluntarily** pick up the poems of Langston Hughes, for example.* ACAD, 2015)
- `neg.co` when negation is used in the context of ‘code’ (e.g. *His book might ignite another firestorm, or it **might not**.* NEWS, 2013)

Tip to the annotator

This first round of annotation is meant to facilitate the rest of the annotation procedure. Indeed, some predictors apply only in the presence of one of the five possible following constituents. In this way, predictor 1 can be used as a filter throughout the annotation. For instance, it can be used to display only those sentences in which the modal is followed by

not-negation for predictor 27. Likewise, it will be useful for the predictors in Level 2, in which attention is given to the verbal complement: examples of modals annotated as ‘code’ can easily be filtered out. Note that predictor 1 is not an exhaustive treatment of negation or verbal complements. The predictors that follow will be crucial to establishing the finer-grained distinctive features of the contexts of use of the modals. For instance, predictor 26 is meant to identify all types of negative operators (and not only *not*-negation).

1.1.2 Predictor 2: *do*-support for *need to* and *have to*

The second predictor focuses on the semi-modals *need to* and *have to*, both of which require ‘*do*-support’ in the context of *not*-negation (cf. e.g. (9)), inversion (cf. e.g. (10)), code (cf. e.g. (11)) and emphasis (cf. e.g. (12)).

- (9) You **don’t need to** believe that “men are from Mars, women are from Venus” in order to accept that men and women have different nutritional needs. (MAG, 2006)
- (10) How many times **do I have to** tell you? Come on, you just don’t listen. (SPOK, 2011)
- (11) You don’t have to cook the vegetables too much, I gather? — No, you **don’t have to**. (SPOK, 1999)
- (12) I **do need to** say this to all the mothers out there: I hope that our stories inspire them to work towards independence. (SPOK, 2002)

In cases where the *do*-support is used, use the label *y*. When there is no *do*-support, use *n*. The predictor cannot be applied to the other modal verbs, for which the label *NA* is used. This includes uses of modal *need*, which does not require *do*-support:

- (13) Instinctively, David ran towards him, trying to save him from the falling tree. He **needn’t** have bothered, however. Zack simply sidestepped the tree, as well as David, who had to quickly alter his course with a jump to avoid being crushed. (WEB, 2012)

Tip to the annotator

When the predictor is not applicable, i.e. for all the other auxiliaries, the feature *NA* can be generated automatically through a filter.

1.1.3 Predictor 3: Forms of *be able to*, *need to*, and *have to*

Unlike the other verbs studied in the REM project, and disregarding the fact that *could/might* are the preterite forms of *can/may*, the three semi-modals *be able to*, *need to* and *have to*, have different inflectional forms. Predictor 3 is designed to identify each of them:

Tense	Label	Modal
Present	pres	<i>am/are/is able to</i>
		<i>need/s to</i>
		<i>have/s to</i>
Present Perfect	prespf	<i>have/s been able to</i>
		<i>have/s needed to</i>
		<i>have/s had to</i>
Preterit	pret	<i>was/were able to</i>
		<i>needed to</i>
		<i>had to</i>
Past Perfect	pastpf	<i>had been able to</i>
		<i>had needed to</i>
		<i>had had to</i>
Future	fut	<i>will be able to</i>
		<i>will need to</i>
		<i>will have to</i>
Future Perfect	futpf	<i>will have been able to</i>
		<i>will have had to</i>
		<i>will have needed to</i>
Past Future	pastfut	<i>would be able to</i>
		<i>would have to</i>
		<i>would need to</i>
Past Future Perfect	pastfutpf	<i>would have been able to</i>
		<i>would have had to</i>
		<i>would have needed to</i>

Table 1. Finite forms of *be able to*, *need to*, *have to*

Form	Label	Modal
To-infinitive	toinf	<i>to be able to</i>
		<i>(it is silly) to have to...</i>
		<i>to need to</i>
Infinitive	inf	<i>(should) be able to</i>
		<i>have to</i>
		<i>need to</i>
-ing form	ing	<i>being able to</i>
		<i>(ended up) having to...</i>
		<i>needing to</i>
Past participle	pastpp	<i>(tasks) needed to (be carried out)</i>

Table 2. Non-finite forms of *be able to*, *need to*, *have to*

Note that past participles that form part of a perfect tense form should be annotated as such (i.e. as *prespf*, *pastpf*, *futpf* or *pastfutpf*).

When this predictor does not apply (i.e. with all remaining modals in the REM project), use the label NA. (This can be generated automatically through a filter.)

1.2 Semantics

1.2.1 Meaning of the modal

1.2.1.1 Predictor 4: Epistemic vs. root meaning

This predictor is the first of a series that focuses on the meaning conveyed by the modal expression used. It concerns the central distinction between the categories of epistemic and non-epistemic (or ‘root’) modality.

Depraetere & Reed (2011: 3) define epistemic modality as the speaker’s judgment “that a proposition is possibly or necessarily true.” In other words, epistemic modality indexes the speaker’s assessment on the likelihood of a situation being the case, and thus relates to the truth of a proposition, also called the *residue* (i.e. “what is left of the meaning expressed in an utterance of the clause when the modality is abstracted away”, Huddleston 1984: 168). The example in (14) receives such an epistemic interpretation: the speaker expresses her confidence in the fact that ‘the rock has been dropped there after the door had been opened’. In this case, use the label *ep*.

- (14) It didn’t fit with Horace’s story. The rock **must** have been dropped there after the door had been opened. Faced with this evidence, Horace admitted he did it. (MAG, 2003)

In comparison, root modality is not concerned with the factual status of a situation but with its actualization. Depraetere & Reed (2006: 274) define root modality as “the speaker’s judgments about factors influencing the actualization of the situation referred to in the utterance.” In other words, root modality indexes the speaker’s assessment on the possibility/necessity for a situation to materialize. The sentence in (15) is an example of root modality: the speaker indicates that it is possible for college graduates who meet certain criteria to transfer as juniors into one of the CSU schools (i.e. actualization is possible). In this case, use the label `root`.

- (15) If community college graduates meet criteria for these pathways, they **can** transfer as juniors into one of the California State University schools and won’t have to repeat courses. (NEWS, 2014)

While epistemic and root modality form distinct categories, there are examples in which the two meanings seem to co-exist (cf. e.g. (16)). Coates (1983) uses the term ‘merger’ to discuss such examples, for which disambiguation is not necessary since the two meanings are “mutually compatible.” In this case, use the label `merg`. Here are some examples from Coates (1983):

- (16) Speaker A: Newcastle Brown is a jolly good beer.
Speaker B: Is it?
Speaker A: Well it **ought to** be at that price.
(Coates, 1983: 17)
- (17) It is important to note that where high concentrations are theoretically possible in the plant evaporator, the time required to build them **may** be considerable.
(Coates 1983: 145)

As we explain in Chapter 1 of the book *Models of Modals*, “[i]n [16], speaker A either means that given its price, Newcastle Brown is necessarily good beer (epistemic), or, alternatively, “the speaker is referring to the maker’s obligation to provide good beer” (Coates 1993: 17) (root, general situation necessity). The sentence in [17] either expresses what we [call] general situation possibility (the situation of the time required being considerable is possible), or it is interpreted in terms of epistemic likelihood: it is possible that the time required will be considerable. What is typical of merger is that the two meanings can co-exist and that the ‘ambiguity’ does not need to be resolved. Merger is not to be understood as an actual blend of the two meanings. It occurs in contexts in which it just so happens that either of the two (still distinct) meanings make full sense. Still, the speaker, if pressed for an answer, would be able to say which meaning she had (most clearly) in mind.”

Finally, in some cases the annotator may simply not be able to decide which of the two categories is conveyed by the modal expression (cf. e.g. (18)). These are cases of indeterminacy. Such unclear examples will be labelled with a question mark (?).

- (18) Right, right. What was the issue? - What was the issue why she couldn't have a - why weren't they, why didn't they have sex and have a baby? Don't know. GEORGE-STEPHANOPOU# (Off-camera) I don't know the answer to that. LARA-SPENCER# (Off-camera) I don't know if they had broken up. And just wanted, thought they **could**... That would be great, you don't have to have intimate relations with someone. Can I just have some sperm? You meet somebody at a bar. Hi. (SPOK, 2013)

Tip to the annotator

Instead of a question mark, using a label such as `uncl` (for 'unclear') may be considered, as this could prove to cause fewer problems for later statistical processing of the annotated data.

1.2.1.2 Predictor 5: Possibility vs. Necessity

In addition to the root/epistemic distinction, modal meaning is also marked by an opposition between the conceptual categories of possibility and necessity. Possibility modals refer to situations that are likely to be true or to happen. Necessity modals refer to situations that are guaranteed to be true or required to happen. Table 3 lists the different possibility and necessity modals studied in the REM project. For possibility modals, use the label `pos`. For necessity modals, use the label `nec`.

	POSSIBILITY	NECESSITY
EPISTEMIC	<i>may</i> <i>might</i> <i>could</i> <i>can't</i> <i>(can)</i>	<i>must</i> <i>should</i> (Coates 1983) <i>have to</i> <i>ought to</i>
ROOT	<i>be able to</i> <i>can</i> <i>may</i> <i>might</i> <i>could</i>	<i>must</i> <i>should</i> <i>have to</i> <i>ought to</i> <i>need to</i>

Table 3. Possibility and Necessity: classification of auxiliaries

Note that *can*, which is often said to express root meaning only, has also been included (in brackets) in the category of epistemic possibility modals. Indeed, previous studies have shown that the modal expression can receive an epistemic interpretation in specific

contexts in American English (Coates 1995, Collins 2009). As the REM project uses COCA, such cases may be found too.

1.2.1.3 Predictor 6, 7 & 8: Root possibility/necessity

Predictors 4 and 5 enable us to capture, in fairly broad terms, the meaning expressed by the modal verb. While the category of epistemic modality is fairly homogeneous, this is, however, not the case of root modality, which is used as a cover term for a number of different meanings. Predictors 6, 7 and 8 are meant to help determine these meanings.

Depraetere & Reed (2011) and Depraetere (2014) use three defining criteria in their taxonomy of root modality: (i) scope, (ii) source, (iii) and potential barrier. Each criterion will be discussed in turn. The taxonomy of root meanings will then be presented in Table 4.

(a) Scope of modality (Predictor 6)

Modal expressions are used to express a judgment about the likelihood that a situation is true or the likelihood of actualization. Predictor 6 is concerned with the scope of this assessment and tries to answer the following question: does the modality apply to the VP only (narrow-scope modality) or to the entire residue (wide-scope modality)? The sentence in (19) is an example of narrow-scope modality. The speaker asks if it is possible (for the subject referent) to ‘recall the last day she felt completely well’. In this case, use the label *ns* (for narrow scope).

- (19) **Can** you recall for me the last day that you felt completely well?
(SPOK, 2014)

The sentence in (20) illustrates wide-scope modality. Here, the speaker says that it is necessarily the case that ‘people in the terminal thought I was crazy.’ In this case, use the label *ws* (for wide scope).

- (20) “I jumped up and down of happiness,” Rouzbeh Pirouz said by telephone. An Oxford-educated investment fund manager, Mr. Pirouz, 43, received the news just after landing on the Mediterranean island of Majorca. “People in the terminal **must** have thought I was crazy,” he added. (NEWS, 2015)

The sentences in (19) and (20) are examples of root and epistemic modality respectively. Note, however, that the difference narrow- vs. wide- scope does not align with the distinction between root and epistemic modality. Although epistemic modality always has wide scope, root modality can have either narrow or wide scope (cf. Nordlinger & Traugott 1997, Depraetere & Reed 2011). The sentences in (21a) and (21b) are examples of wide-scope root possibility (see predictor 8 for further details). In (21a), the speakers

says that ‘for you to be an award winning journalist covering wars and to be gay’ is possible. In (21b), the speaker indicates that the situation of ‘them being employed in the United States or obtain driver’s or professional licenses’ is not possible. These examples thus have to be coded *ws*.

- (21) a. That being said, it’s great that, you know, that folks come out and they show that you **can** be an award winning journalist covering wars and you can be gay. You can be a hip-hop artist singing about macho lyrics and women and whatever else and be gay, that it doesn’t define you. (SPOK, 2012)
- b. As students, when they graduate from high school, they often can not go on to college because they are not eligible for financial aid and must pay higher tuition rates. If they do succeed in graduating from college, regardless of their academic accomplishments, they **can** not be legally employed in the United States or obtain driver’s or professional licenses. (NEWS, 2012)

Tip to the annotator

Since epistemic modality always has wide scope, a filter can be used to code the value *ws* automatically.

(b) Source of modality (Predictor 7)

The next predictor focuses on the source of the modality, which has to do with ‘who’ or ‘what’ makes actualization possible or necessary (Depraetere & Reed 2011: 9).

In the case of epistemic modality, the source of the modality is always the speaker: it is always the speaker who contextually infers the likelihood of a particular state of affairs. For this type of subjective source (see below), the label *disint* is used (which stands for discourse-internal).

In the case of root modality, five different types of sources can be distinguished. (The overview below is based on Depraetere & Verhulst (2008), Depraetere & Reed (2011) and Leclercq *et al* (2022).) Note that this predictor cannot be applied to cases of merger and indeterminacy, in which case the label *NA* is used.

[1] The source of modality can be **subject-internal**, as in (22): the source of ability lies within the referent of the grammatical subject who, as a result, has the ability to use the word. Use the label *subjint*. Ability can also occur with inanimate subjects (cf. e.g. (23)). (See below, in the discussion of example (52), for further comments about subject-internal necessity.) The example in (24) illustrates a subject-internal source in a sentence that expresses necessity meaning.

- (22) When I understand the word so well that I **can** use it in my own vocabulary, that is my victory. (ACAD, 2015)
- (23) We use the term base word for free morphemes, one-morpheme words that **can** stand on their own. (ACAD, 2014)
- (24) We've finally reached the stage where he takes his time-out himself. If he sees he's getting upset, he says, "I **need to** be alone." (MAG, 1998)

[2] The source of modality can be **discourse-internal** (i.e. subjective, or speaker- or hearer-oriented) as in the following example in which the speaker gives the addressee the permission to smoke inside. In this case, use the label `disint`.

- (25) The Indians have a much more positive view of us, because we have great trade relations with them. We have a common foe in the Islamic extremists. CARLSON: Yes. We **ought to** let more of them in. I like them." (SPOK, 2006)

A discourse-internal source is not only found in examples that express permission, as in (25) and (27), but is also found in the contexts with necessity verbs as well. So in general terms, if there is a discourse-internal source, it is the speaker who is at the origin of the possibility or necessity in declarative sentences (cf. e.g. (26)) and it is the hearer in interrogative sentences (cf. e.g. (27)):

- (26) She wheeled on me. She said, "Do you honestly believe money will make up for what I went through? Visiting all our high-class neighbors, throwing myself on their mercy, pleading with them not to press charges?" "I never asked you to do that," I said. "Well, Mrs. Gaitlin, we'll **need to** think this over," she said, putting on a pinched and simpering tone of voice. "We'll need to give it some thought," they told me. (FIC, 1998)
- (27) Why are we ignoring the fact that in fact they haven't contained in northern Iraq, haven't contained in Syria? We have taken back land from them and that's been totally ignored. (CROSSTALK) GRENNELL# **Can** I say something here, Eric? BOLLING# Very quickly, Rick. (SPOK, 2015)

[3] In some cases, the possibility or necessity results from a **rule or a regulation**, as in the following examples. In this case, use the label `rul`.

- (28) In the District, under this new law, you **can** have two ounces of marijuana, which is more than you can have legally at any one time in Washington or Colorado or Alaska. (SPOK, 2015)

- (29) I gently explain to my client that if his wife has not deserted him or committed adultery-he responds with almost comic indignation that she has not-and will not agree to a divorce, as the law stands he will either **have to** wait five years to obtain his freedom without her consent, or else cobble together a charge of unreasonable behaviour. (FIC, 2008)
- (30) To the extent that the Eighth Amendment also entails a unitary proportionality principle, anti-harmfulness principle, or anti-arbitrariness principle, such a principle **might** likewise operate based on the Court’s “independent judgment,” even without any national consensus. (ACAD, 2010)
- (31) There was a trust fund built on safe, boring investments, and yes, the name on that fund looked rather like my name. But it wasn’t. Grandpa left a robust fortune to a twenty-eight-year-old man who hadn’t yet come into existence. I could curse him all I wanted, but I had to survive until that ripe old age before I **could** claim the millions. (FIC, 2013)

Note that moral and social norms (cf. e.g. (32) and (33)) are also included in this category:

- (32) The United States and the North Atlantic Alliance are probably stronger than ever, so helping the reformers in Russia and in the former Soviet Union is not letting our guard down, it’s not doing anything that we shouldn’t do. It’s doing exactly what we **ought to** be doing. (SPOK, 1993)
- (33) Oh, I could n’t agree more with the senator. He’s absolutely correct about foreign policy. He’s absolutely correct about the whole missile problem. He’s absolutely correct about what we **ought to** do about the southern half of the whole planet, you know, how we can help continents like Africa and those underdeveloped countries. (SPOK, 2000)

[4] **Circumstances** constitute another source of the modality (cf. e.g. (34) and (35)). In this case, use the label *circ*.

- (34) Exploring a variety of drum sounds may help students develop self-expression. For example, the taiko drum sound is very powerful. However, general music teachers **might** not have access to, storage for, or experience with taiko drums. (ACAD, 2015)

- (35) Melody wasn't surprised that Keith Taylor made Sarah's list of latest lovers. Melody picked up on Keith's disturbed nature the first time she laid eyes on him. She wouldn't dare leave her room until Sarah arrived. Luckily she didn't **have to** wait. She heard Sarah rushing upstairs. // "Hey, Mel, you in there?" Sarah knocked. (FIC, 2008)

It is important to add that circumstances may be of different types: they may refer to (a) (one-off) arrangements or particular situations (cf. e.g. (36)) that necessitate the actualization of a particular situation (or that make is necessary for someone to do something, in the case of narrow scope necessity) or that make a particular situation possible (or that make possible for someone to do something, in the case of narrow scope possibility), (b) habitual situations (cf. e.g. (37)), or (c) 'the nature of things'. The final case refers to the literal interpretation in the sense of the necessities resulting from the laws of nature (cf. e.g. (38)), as well as a broader interpretation in the sense of the only logically possible outcome given the state of the world (cf. e.g. (39)).

- (36) When placing a pre-service teacher (PST), certain factors **need to** be addressed. (ACAD, 2006)
- (37) In fact, looking young is expensive. Botox costs an average of \$400 per session. The procedure **must** be repeated every few months. (SPOK, 2002)
- (38) But just because all men **must** die, that doesn't mean we can't have some fun (MAG, 2017)
- (39) Every hermit crab, land or sea, **must** bear the burden of carrying its home on its back whenever it travels (MAG, 2014)

In addition, cases of *hedged performatives* (cf. e.g. (40) to (41)), i.e. the juxtaposition of a modal and a performative verb (Fraser 1975), and optatives (cf. e.g. (42)) were also coded as having a 'circumstantial' source.

- (40) Well, Jane, first of all, thank you so much for having me on the program. I **have to tell** you that I went one on one with the Orange County sheriff's office detectives and CSIs, who investigated the murder of Caylee for three years. (SPOK, 2011)
- (41) SIMON: I believe I've seen the mural down by ground zero, 240 feet tall. Mr-SMIRNOFF: Correct. SIMON: And I **must say**, I didn't think it was you. (SPOK, 2003)
- (42) "Prost," she concluded, raising her coffee cup. "**May** you get a lively book out of our dysfunctional family. It's about time the Crosses were good for something besides turning a dishonest dollar." (FIC, 2011)

[5] **Conditions** constitute the final source of the modality that we use when coding this predictor, as in the following examples. In this case, use the label `cond`.

- (43) If you want to continue living in poverty without clothes and food,” I told them, “then go and drink in the shebeens. But if you want better things, you **must** work hard. We can not do it all for you; you must do it yourselves.” (MAG, 1995)
- (44) If there is to be a continuation of the cease-fire, the borders will **have to** be opened, and one way or another, someone has to deal with Hamas. (NEWS, 2009)

In this type of example, a situation is said to be necessary in order to reach a specific aim or with a view to a specific purpose (‘if Y is to actualize, X is necessary/possible’). The conditioning situation is posterior to the modal meaning of necessity (X is necessary at t in order for Y to actualize at t+1). The following are possibility examples with a conditional source:

- (45) “Be that as it may,” said the AI, “if you are to restore teleportation to the entire galaxy, you **may** need to compromise your principles.” (FIC, 2010)
- (46) “I think you’re jumping at conclusions, officer.” “To. Jumping to conclusions. If you want to imply that we are looking to fix blame prematurely you **might** say ‘at,’ as in the phrase ‘he’s looking at likely suspects.’ But I think you meant to say jumping to conclusions, a common and therefore less troubling prepositional phrase.” (FIC, 2014)

Unlike what examples (43) to (46) may suggest, the source ‘conditional’ does not hinge exclusively on the presence of an *if*-conditional clause but can also be found in other contexts (the preceding clause in (47) or a subclause of purpose in (48)):

- (47) Do you want to use retirement benefits as a tool to recruit and retain the best workers? You’ll **need to** research and match the benefit packages provided by competitors. (NEWS, 2007)
- (48) Kohut (1959) was one of the first to articulate that both the conscious (for example, perspective taking) and implicit or unconscious (for example, emotion sharing) processes are vital to empathy and **must** be integrated to achieve a true empathic reaction or response. (ACAD, 2011)

In a similar way, the use of an *if*-conditional clause does not systematically entail that the modal source of necessity is a conditional source. In the following example, the source is a rule (coded as `rul`):

- (49) However at my hospital it is policy that if you are not feeling well in any way you **MUST** wear a mask. (BLOG, 2012)

While a ‘conditional’ source and ‘rules and regulations’ could be considered as a kind of ‘circumstantial’ source, the approach taken here can be summarized as follows:

- Subject-internal (SI) source
- Discourse-internal (DI) source (subjective)
- Discourse-external (DE) source
 - o Rules and regulations
 - o Conditional source
 - o Circumstantial source

In other words, a circumstantial source is a discourse external source that is not ‘rules and regulations’ and not a ‘conditional’ source.

It is to be noted that source is not always explicitly present in the sentence: it can be implicit but context should help the annotator to put a label on the source. If it is not possible to identify the source, a question mark (?) is used.

Likewise, it is important to disentangle modality from speech act when identifying source. While some speech acts (cf. predictor 32) add a subjective tone to the utterance, it does not entail that modality is itself subjective (i.e. ‘discourse-internal’). In examples (50) and (51), for instance, the speech act may be directive, but the source is ‘circumstantial’.

- (50) Soften the edges Civic Center has benches and trees, but rarely is there a bench under a tree. It has lots of grass but no native grasses. It has flowers in horticulturally correct squares, but they don’t look happy. **Couldn’t** we replace some of those hard concrete sidewalks with gravel paths? (NEWS, 015)
- (51) Everyone is concerned about the deficit, and that’s why we **have to** be careful how we frame this. (SPOK, 2010)

A final comment concerns the annotation of sentences with *need to*. In the literature, this verb has often been associated with a kind of irresistible urge or a need that is not controlled by the subject referent. Our corpus contains hardly any examples of this type. In most cases, the subject referent appears as the agent in control of the situation referred to (and therefore constitute the necessity counterpart to ability examples). In (24), repeated here in (52), the subject referent intentionally wants to ‘be alone’. The label *subj int* is used for cases like this.

- (52) We’ve finally reached the stage where he takes his time-out himself. If he sees he’s getting upset, he says, “I **need to** be alone.” (MAG, 1998)

Agentivity, in the sense of Cruse's (1973) 'volitive' (i.e. "is present when an act of will is stated or implied" p.18)¹, is taken as the defining criterion for 'subject-internal' source in sentences with an animate subject referent. Examples in which the necessity does not originate in the subject referent's conscious volition, but is rather the result of an irresistible character trait (which could be called *irresistible urge necessity*) illustrate a specific kind of 'circumstantial' source:

- (53) What do you do when you **need to** kind of chill, when you need like to breathe, what does Miley do? (SPOK, 2008)

Need to expresses circumstantial necessity (the subject referent is required to do something, either as a result of an irresistible character trait (*irresistible urge necessity*, a kind of circumstantial source), or because there is a specific situation, not specified in this example, which required him to do so) rather than necessity originating in the subject referent's conscious volition, which can be paraphrased by *want to* (*subject-internal necessity*).

Tip to the annotator

A filter can be applied to code semi-automatically the epistemic sentences as `disint`.

(c) Potential barrier and the taxonomy of root meanings (Predictor 8)

The third and final criterion that Depraetere & Reed (2011) use to establish their taxonomy of root meanings is 'potential barrier'. With this criterion, the goal is to know whether the source of the modality has source status because "it is capable of imposing a barrier to the subject referent's performing a particular action or to the actualization of the situation" (Depraetere & Reed 2011: 13). In other words, could the source of the modality prevent or block actualization of the residue? Answer to this question can be either positive or negative. In (54), the 'discourse-internal' source (the speaker) can potentially impose a barrier to actualization. In (55), the 'circumstantial' source cannot potentially impose a barrier.

- (54) You **can** take a small camera along, but SLRs aren't allowed. A photographer/guide is taking pictures, and he'll sell you a \$20 CD with photos and videos at the end. (NEWS, 2011)
- (55) "You're lucky," the nurse said. "Didn't hit any vitals. But we **had to** stitch you up and give you a shot for infection. That was a pretty rusty shank." (FIC, 2000)

¹ Cruse uses *agentive* as a broader category: the feature *agentive* "is present in any sentence referring to an action performed by an object which is regarded as using its own energy in carrying out the action" (1973: 21). Van Valin & Wilkins (1996) use the term 'effector' to capture agentive as defined by Cruse: "roughly the dynamic participant doing something in an event" (1996: 291).

While this criterion does not need to be coded separately, it is crucial to identify the different root (possibility) meanings in Depraetere & Reed’s taxonomy that we adopt here. Table 4 shows how the three criteria (scope, source and potential barrier) contribute to distinguishing root modal meanings. For each meaning (cf. e.g. (56) to (63), use the associated code for predictor 8.

Tip to the annotator

As the value for predictor 8 partly depends on the values attributed to the concordance for predictors 5, 6, and 7, filters can be used to code the attestations more efficiently.

	Root meanings	Scope	Source	Pot. barrier	Label
Possibility	Ability	Narrow	SI	-	ab
	Opportunity	Narrow	DE	-	op
	Permission	Narrow	DE <i>or</i> DI	+	perm
	General Situation Poss.	Wide	DE	-	gsp
	Situation permissibility	Wide	DE <i>or</i> DI	+	sitperm
Necessity	Subject-internal Nec.	Narrow	SI	-	subjintn
	Subject-external Nec.	Narrow	DE <i>or</i> DI	-	subjextn
	General Situation Nec.	Wide	DE <i>or</i> DI	-	gsn

Table 4. Predictors for root modality meanings

(56) ABILITY

- a. If the student encountered the term biodegradable in text or during a lecture and could not remember the term, they would be explicitly taught to remember the silly image presented by the KMS, and its embedded remembering system. (ACAD, 2013)
- b. When they return, they’re more mature and better able to accept the rigors of college. That goes for athletes, too. Older athletes generally are better able to handle the daunting challenge of mixing studies with sports, essentially two full-time jobs. (NEWS, 2006)

(57) OPPORTUNITY

- a. “I put this driver in the bag last October. It has a bigger head, a longer face and spins a little bit more than my previous one the original Covert Tour, which makes this more playable and more controllable, so I can hit the shots I want to hit. It’s been a really good addition.” (MAG, 2014)

- b. All that discipline led Bill to a tour of duty in Iraq, where he interned for the non-denominational minister and, like the prophets of scripture, hear his calling to the Presbyterian clergy in the desert wind. Meanwhile, I was busy turning down a career in technical writing so I could move with my girlfriend to Iowa and work in a copy store. (FIC, 2014)
- (58) PERMISSION
- a. “She is dead, Zezen,” he said. “She will not be less dead by taking her inside the shiro. Leave her. Tomorrow, or perhaps the next day, when the air is safe, you may do with her what you wish. (FIC, 2011)
 - b. In Texas, where prescription drug overdose deaths peaked at 577 in 2006 the Legislature acted earlier than many other states, putting more regulations on pain clinics in 2010, including restrictions on who can own and work in them. (NEWS, 2015)
- (59) GENERAL SITUATION POSSIBILITY
- a. Like an unwatered lawn and a driveway full of rolled-up newspapers, NBC’s one-hour dramedy “Camp” (also premiering Wednesday) is a clear sign that nobody’s home at the network for the rest of the summer. You could break in and take what you like. (NEWS, 2013)
 - b. Egg-sharing arrangements are also common, where the cost of one woman’s IVF treatment might be met by a group of other women, who would then use surplus eggs. (ACAD, 2007)
- (60) SITUATION PERMISSIBILITY
- a. Under canon law someone who tortures an accused party may be held guilty of a capital offence if the victim dies as a consequence. (MAG, 2014)
 - b. In Linder v. United States (1925), the Court stated addicts should be able to obtain medical care like anyone else. (ACAD, 2009)
- (61) SUBJECT-INTERNAL NECESSITY
- a. I have to know what Parenting readers think: I am sickened by the trend of “push” presents! A friend’s husband bought her a ring after she gave birth. Save the money for the kid’s college! (MAG, 2007)
 - b. Need to tame your inner foodie? Book a table at the River Garden Cafe (802/626-3514; rivergardencafe.com), where creatively prepared meat and seafood await. (MAG, 2008)

- (62) SUBJECT-EXTERNAL NECESSITY
- a. Because half of China’s population is still poverty-stricken -- a reality with explosive, revolutionary possibilities -- the Communist Party feels that it must produce economic growth to stay in power. (ACAD, 2010)
 - b. The muriqui, almost a third of its captor’s weight, was a heavy load for a boy of eight, but he was a hunter now. Right and duty dictated that he should carry it. (FIC, 2014)
- (63) GENERAL SITUATION NECESSITY
- a. And in a country that celebrates the 200th anniversary of the First Amendment this year, we ought to be reminded that the whole purpose of the thing is to have those kinds of decisions made by journalists, about whether and how the news is covered, and not by government officials who are themselves involved in the process. (SPOK, 1991)
 - b. The level of frustration is elevated to the point where the original slow pace is not adequate. I believe in gradual reform, but I think it must also be sustained with a clear time line. It’s not that the king doesn’t want it, but I believe he must lead a process that would accelerate the current pace. (MAG, 2013)

1.2.1.4 Predictor 9: Actualization

Next, closer attention is given to cases of narrow-scope root modality (i.e. ‘ability’, ‘opportunity’, ‘permission’, ‘subject-internal necessity’ and ‘subject-external necessity’). In this case, the annotator needs to identify whether actualization is expressed alongside modality, i.e. whether the situation referred to in the VP materialized or not. There are three possibilities: (i) it is clear that there is actualization (cf. e.g. (64)), in which case the label *y* is used; (ii) it is clear that there is no actualization (cf. e.g. (65)), in which case the label *n* is used; (iii) it is not clear whether there is actualization, in which case a question mark ? is used.

- (64) # What could be bigger than the Super Bowl for a kid correspondent? # “The presidential campaign,” she said, matter-of-factly. # “I **was able to** interview some of the people running for president such as Hillary Clinton. (News, 2009)
- (65) I clutched my temples, regretful that I had been so hard on him. He was all the family I had left, even if he was only a simulacrum, a phantom recreated through uber-symmetry. I probably **should** have told him that I still loved him, in spite of everything. (FIC, 2014)

In the case of examples of wide-scope root modality in the dataset, we do not consider this predictor, and the label NA will accordingly be used. (This can be generated automatically through a filter.)

1.2.2 Modality and time

1.2.2.1 Predictor 10: Temporal location of the modal situation

Having defined the different modal meanings, the next step consists in identifying the time sphere in which modality is located. There are four possibilities. First, modality can be located in the present time sphere (cf. e.g. (66), ‘it *is* not possible for Bob to drive’). In this case, use the label `pres`.

- (66) He stepped across the room and took his jacket from the closet. “I’m going,” he said. “Where are you going?” “Home.” “You **can**’t drive, Bob. You had too much to drink.” “I’m going,” he said again. (FIC, 2007)

Second, the modality can be located in the future time sphere (cf. e.g. (67), ‘it *will be* possible for you to begin your real work with us shortly’). Note that not all examples of future-time modality hinge on the reference to future time expressed by a verb like *will* or *be going to*. In example (68), *have to* is also used with future time reference (‘it *will be* necessary for Russel Wilson to fight for his job’). In this case, use the label `fut`.

- (67) You will **be able to** begin your real work with us shortly. (FIC, 2015)
- (68) The Wolfpack will be the surprise team in the ACC Atlantic after a strong finish. Russell Wilson was the All-ACC quarterback but may **have to** fight for his job against Mike Glennon, a highly recruited freshman who redshirted last season. (NEWS, 2009)

Third, the modality can be located in the past time sphere (cf. e.g. (69), ‘it *was* necessary for us to take care of things ourselves’). In this case, use the label `past`.

- (69) When Mahmoodi was 13, his father was in a car accident and died. It was a very difficult time from then on. It was a large family. The man was gone, and we **had to** take care of things ourselves. (MAG, 1996)

Note that in narratives, when the past tense is used in the main clause, the reference to past time is expected to carry over to the subordinate clause. The sentence in (70) illustrates this observation: the past reading of *should* in the subordinate clause is

triggered by the past form of *think* in the main clause. In such cases, a more specific label is used: `pastsub` (where ‘sub’ stands for *subordination*).

- (70) Venable sent them to me because you were involved, and he thought you **should** be advised. I take it you worked with Jantzen? (FIC, 2015)

Examples of this type can in particular be found in novels, in reported and free indirect speech (or thought).

Tip to the annotator

Note that the historical past form of modals (e.g. *could* for *can*, *might* for *may*, etc.) do not necessarily have past time reference in present day English. In the sentence *She could already be home*, modality is located in the present even though the past form *could* is used.

1.2.2.2 Predictor 11: Temporal relation between the modal situation and the residue

Another time-related dimension that requires attention is the temporal relation between the modal situation (M) and the residue (R). There are three such temporal relations: simultaneity, posteriority, and anteriority. First, the modal situation and the residue can be simultaneous. In example (71), for instance, they are both located in the present time sphere (M = ‘It *is* possible’; R = ‘They *are* already on the water again’). In this case, use the label `sim` (for simultaneous).

- (71) He’s fine, Father. He says they didn’t lose any ships to the freeze, so they’re expecting to be on the water again. They **might** already be. (FIC, 2012)

Second, the residue can be posterior to the modal situation. In example (72), modality is located in the present time sphere and the residue in the future (M = ‘It *is* possible’; R = ‘My book *will help* bring out some of the work we do’). In this case, use the label `post` (for posterior).

- (72) But mountain rescue is only in the mountains. It’s a low-profile kind of thing. We don’t really talk to the press, and we don’t advertise when we go out. We don’t do press releases telling everybody what we do. Maybe we should. My book **might** help bring out some of the work we do. (NEWS, 2007)

Third, the residue can be anterior to the modal situation. In example (73), the modal situation is located in the present and the residue in the past (M = ‘It *is* possible’; R = ‘Forbes *asked* for a plea deal’). In this case, use the label `ant` (for anterior).

(73) Shaklee said Forbes **may** have asked for a plea deal to avoid the death penalty or going to prison as a sex offender. (NEWS, 2011)

In this example, the use of the perfect infinitive *have asked* is a clear indicator of this anteriority. However, as will become clear in the discussion of predictor 13, perfect infinitives are not always used to express anteriority.

Level 2. The verbal complement

The second level of analysis is concerned with the verb that follows the modal used. Predictors 12, 14 and 15 are formal in nature, predictors 13 and 16 are semantic, predictor 17 is lexical.

Tip to the annotator

Sometimes, there is no verbal complement (e.g. in ‘code’ contexts). In such cases, the label NA will be used for all the predictors of this level.

2.1 Predictor 12: Present or perfect infinitive

The verbal complement of a modal is always used in its infinitival form. The question that predictor 12 aims to answer is whether a *present* infinitive or a *perfect* infinitive form of the verb is used. In the case of a present infinitive (see e.g. (74)), use the label `presvb`.

- (74) Do you worry when you go out and campaign for him that people **might bring** those things up? (SPOK, 2008)

In the case of a perfect infinitive, two options are possible. Either the auxiliary *have* is used in its full, non-contracted form (see e.g. (75)), in which case the label `pfvbnbc` applies (which stands for present perfect verb non-contracted). Or the contracted form of the auxiliary *have* is used (see e.g. (76)), in which case the label `pfvbc` applies (which stands for present perfect verb contracted).

- (75) The store **may have smelled** like dried fish and old cheese, but at least it was air-conditioned. (FIC, 2013)
- (76) “I **must’ve been** about ten when my father sat me down in front of a mirror, pulled a Ronald Reagan Halloween mask over his head, pinned a defunct pair of Trans World Airlines captain wings to his lab coat, and proclaimed himself a white authority figure.” (FIC, 2015)

2.2 Predictor 13: The meaning of the perfect infinitive

Predictor 13 applies to perfect infinitives and their functions/meanings only. Attestations with a present infinitive are disregarded. While the term ‘perfect’ may seem to have temporal implications, this form can actually also be used for a number of (non-temporal) meanings. Depraetere (2017) has identified four different contexts in which perfect infinitive *have* can be found. First, *have* can be used to locate the modal situation, i.e. the possibility/necessity, in the past time sphere (cf. e.g. (77)). In this case, use the label `past`.

- (77) She seldom stopped talking, and it was an astonishing, multihued and never-ending stream of esoteric information. Who else **could have combined** a passion for low-dimensional topology with a passion for squash? (FIC, 2014)

The fact that *have* serves to establish past time reference becomes clear when you change the perfect infinitive into a present infinitive (e.g. *could combine*). Such a change coincides with a change in the temporal location of the modal situation.

Second, *have* can be used to establish a temporal relation of anteriority between the modal situation and the residue. In example (78), modality is located in the present time sphere ('It is possible') and the residue is in the past ('Streams *dried up*'). In this case, use the label *ant*.

- (78) Others blame climate change. The lagoons, marshes, and streams **may have dried up**, since even small shifts in rainfall can have a dramatic effect on water sources in the area. (ACAD, 2011)

Third, the perfect infinitive can also be used to express counterfactuality. Counterfactual statements refer to situations that were possible or necessary but did not actualize. In example (79), the speaker indicates that although Gary Johnson was not at the debate, his presence was necessary. In this case, use the label *count*.

- (79) COLMES: Eight against one is like me being on the show here. But look, you are not wrong -- you are not wrong because the problem is --(CROSSTALK) O'REILLY: All right, go ahead. COLMES: -- that people like Gary Johnson **should have been** in the debate. (SPOK, 2011)

Fourth, somewhat paradoxically, perfect infinitive *have* can also be used to put emphasis on actualization (cf. e.g. (80)). In this case, use the label *act*.

- (80) Now, in the office, he said to me, "I have to be good now. I've decided I have to be good. I can't talk to you ever." "What do you mean, "ever"?" I said. "Ever," he said. I didn't believe him. What people always want to know, now, is, "How **could you have done** it? Didn't you think of the cost to the other people involved?" And of course I did. But when, over our time together, I'd asked P. to describe his marriage, the picture he painted was chilling and hellish-no communication, no love, no respect. (MAG: 1998)

Tip to the annotator

When the predictor cannot be applied (i.e. with present infinitives and when there is no following verb), the feature NA can be generated automatically through filters.

2.3 Predictor 14: Progressive aspect

The next predictor focuses more on the aspectual form of the verbal complement and specify if it is used with progressive or non-progressive aspect. In example (81), the infinitive form occurs in the progressive, in which case the label *y* applies. In example (82), a non-progressive infinitive is used, in which case the label *n* applies.

- (81) He **might be helping** some corporation that hopes to cover up an illegal toxic burial or spill. (FIC, 2008)
- (82) We should have no greater priority than the goal of obtaining high-quality education for every child. # To achieve this goal, we **must work** together in trust and mutual respect. Our children deserve no less. (NEWS, 2004)

2.4 Predictor 15: Voice

This predictor specifies whether the verbal complement occurs in the active or passive voice. The active voice is used in (83), and is coded *act*. The passive voice is used in (84), and is coded *pas*.

- (83) The world's people may **have to take** matters into their own hands – up to and including breaking the law. (MAG, 2014)
- (84) Surveys are available at the library at 1938 Lohmans Crossing and via its website at laketravislibrary.org. Surveys **may** also **be mailed** to interested individuals who call the library and request one. (NEWS, 2014)

Note that the 'get V-ed' construction can also be used to build a passive. Examples such as those in (85) are thus also coded *pas*.

- (85) Sometimes you need a showman in the show. However, even in opera bouffe, showmen **can get thrown out of** the show if they cross certain boorish lines. Mr. Trump just can't stop himself. (NEWS, 2015)

2.5 Predictor 16: Situation type

This predictor zooms in on the verbal complement. The goal here is to determine what situation type (cf. Vendler 1957) is encoded by the VP: dynamic or stative.

The category ‘dynamic’ refers to situations that require some input of energy and usually involve a change of state: activities (e.g. *to swim in a pool*), achievements (e.g. *to break a glass*) and accomplishments (e.g. *to run a marathon*). In this case (cf. e.g. (86)), use the label *dyn*.

- (86) There’s a few people who shouldn’t be taking that - people who are allergic to aspirin; people who have ulcers, for example. People who are onprescription blood-thinners **ought to ask their doctor about it**. (SPOK, 1991)

By contrast, the category ‘stative’ refers to situations in which no input of energy is required (cf. e.g. (87)). In this case, use the label *st*.

- (87) Monitor the water temperature to warm the tissue slowly for approximately 25 to 40 minutes. Do not pour warm water directly on the affected area because it will warm the tissue too quickly causing tissue damage. After the area is warmed, it may have a feeling of burning or numbness and **may be puffy and blister**. (MAG, 2011)

2.6 Predictor 17: Lemma form of the verb

The last predictor of Level 2 is lexical in nature: it lists the lemma of the verbal complement. For each example, the lemma form of the lexical verb has been listed in capital letters in the dataset, e.g. FOCUS in (88) and COPY in (89).

- (88) You **can** only really **focus** on one thing at once. (SPOK, 2015)
(89) Not even Poli was able to diagnose this abnormal behavior of the software; it **might** have been badly **copied** onto that dubious hard disk. (FIC, 2011)

Level 3. The grammatical subject

The third level of analysis zooms in on the grammatical subject of the modal sentence. For each of the predictors, use NA for subjectless sentences.

3.1 Form

3.1.1 Predictor 18: Subject extraposition

In most declarative sentences in English, the subject precedes the verb at the beginning of the sentence. When the subject is a clause however, extraposition can happen: the clause can be moved to the end of the sentence and replaced in initial position by ‘dummy’ *it* (cf. e.g. (90)) or existential *there* (cf. e.g. (91)). In this case, use the label *y*.

- (90) As noted above, ***it may*** be that schools emphasized interventions that focused on behavioral issues rather than emotional issues, leading to the finding that the impact was on these types of problems only. (ACAD, 2015)
- (91) And what I was surprised, when I first walked in, is that ***there must*** be people coming to see this two or three times because people were already applauding before the play started. (SPOK, 2009)

Examples with regular, non-extraposed subject (cf. e.g. (92)) will be coded *n*.

- (92) **Government officials were able to** see that the count was off, particularly in the count of black men of a certain age group in the South, because they were using census data to plan for how many would be registering to fight in World War II, Sparks said. (NEWS, 2012)

3.1.2 Predictor 19: Person and number

The next goal is to identify the grammatical person and number of the subject used. Use the following table to code the data:

Person	Number	Label	Examples
1 st	Singular	1s	- Pronoun <i>I</i>
2 nd	Singular	2s	- Pronoun <i>You</i>
3 rd	Singular	3s	- Pronouns <i>It, She</i> and <i>He</i> - Martinez died on the scene and Alvarez , who was also stung, was able to run away and refused medical treatment, according to the lawsuit. (NEWS, 2017)

1 st	Plural	1 _P	- Pronoun <i>We</i> - I tried to ask again -- I had made a reservation, my daughter and I could share a bed -- but now he got angry, repeating an English word he knew very well: No. (FIC, 2013)
2 nd	Plural	2 _P	- Pronoun <i>You</i>
3 rd	Plural	3 _P	- Pronoun <i>They</i> - But with enough resources and enough patience, there is little doubt that American troops and their Afghan allies will be able to secure key areas of southern Afghanistan that have slipped out of the government's grasp. (NEWS, 2009)

Table 5. Subject number and person

Examples with subject extraposition are coded as follows. Cases of *it*-extraposition (cf. e.g. (90)) are labelled 3_S (3rd person singular). Cases of *there*-extraposition are labelled 3_S (3rd person singular, e.g. (93)) or 3_P (3rd person plural, e.g. (94)), depending on the number of the complement.

- (93) Finally, I hope that all of us will support the notion that **there ought to** be some period after which we end welfare as we know it. (SPOK, 1996)
- (94) In a single teaspoon of garden soil, **there can** be many yards of hypha. (MAG, 2014)

3.2 Semantics

3.2.1 Predictor 20: Animacy

Predictor 20 is concerned with animacy. The subject can be either animate (cf. e.g. (95)), in which case the label *y* is used, or inanimate (cf. e.g. (96)), in which case the label *n* is used.

- (95) These organelles are stuffed with their own DNA, which **examiners can** analyze for patterns and use to link hair samples from crime scenes to suspects. (MAG, 2015)
- (96) A healthier economy is what's boosting prices. **Rates would have to** rise sharply to make a mark. (MAG, 2013)

The following classification is used (cf. Garretson *et al* 2004, Zaenen *et al* 2004):

- Animate subjects: human beings (including imaginary entities that look/act like humans), organizations, animals (including viruses and bacteria), intelligent machines (computers, robots etc.), and vehicles treated as living beings in some linguistic contexts.

• Inanimate subjects: concrete entities (e.g. body parts, a table), places (potential locations for humans), times, and other non-concrete entities such as air, voice, wind and other intangibles.

Tip to the annotator

Remember that the focus here is on the grammatical subject. Thus, in the case of subject extraposition (cf. e.g. (90)), the grammatical subject (*it/there*) is inanimate. As first- and second-person subjects can safely be set as ‘animate’, a filter can be applied to fill for this value semi-automatically.

3.2.2 Predictor 21: Agentivity

A distinction also needs to be made between agentive and non-agentive subjects. An agentive subject has control over the action denoted by the VP (cf. e.g. (97)). In this case, use the label *y*.

- (97) Just a minute, whoever you are. **You can** say what you want about me, but leave my dog out of it. (FIC, 2014)

Non-agentive subjects lack control over the action denoted by the VP (cf. e.g. (98)). In this case, use the label *n*.

- (98) To be successful in the general curriculum, **students with disabilities should** be provided additional supports and services. (ACAD, 2015)

Cruse (1973) shows that agentivity can be assessed with a simple *do*-test, which reveals relations of entailment. When the subject is agentive, the VP can be replaced by ‘do something’. In (97), for example, “**You can** say what you want about me” entails that “**You can do something**”. However, in (98), “**students with disabilities should** be provided additional supports and services” does not entail that “**students with disabilities should do something**”.

A final comment concerns examples such as in (99) and (100). Depending on how one takes the metonymy into account, the subjects could be viewed as either inanimate/non-agentive (train, linguistics) or animate/agentive (train drivers, linguists). In the REM project, examples of this type were coded as inanimate (predictor 20) and non-agentive (predictor 21).

- (99) Under CSX’s own rules, a train **must** sound its horn until it enters a crossing. (NEWS, 2004)

- (100) The only things I can assert with confidence are these: linguistics **must** deal with all such cases, because they are part of our ability to use and understand language. (ACAD, 2006)

When this predictor cannot be applied, use the label NA.

3.2.3 Predictor 22: Genericity

With the last predictor of level 3, the aim is to distinguish between generic subjects and non-generic subjects. A generic subject does not refer to an item or a group of items from a class of like items but instead refers to a kind or the entire class (cf. e.g. (101) and (105)). Such NPs are called ‘kind-referring’ or ‘generic’ NPs (Carlson 1995), and the predications in sentences involving such NPs are called ‘kind predications’. In this case, use the label γ .

- (101) In the northernmost states, potatoes **can** be planted anytime in summer and harvested in fall. (MAG, 2014)
- (102) To capture a grown hare, an animal with a maximum speed almost half again as fast as its own, the fox **must** depend either on stealth or ambush. (MAG, 1994)
- (103) Well, this is the cheapest way to go but because you never know what you're going to get, you **might** end up with a box full of turnips and parsnips. (SPOK, 2009)
- (104) Such maxims had such an effect because they revealed that these events were beyond the control of human beings. Hadot captures this when he writes: We **must** confront life's difficulties face to face, remembering that they are not evils, since they do not depend on us. (ACAD, 2006)
- (105) Thus, one **could** conclude that delaying age of entry into formal education could have a positive impact on the child due to increases in self-regulating behaviors. (ACAD, 2012)

By contrast, non-generic subjects refer to an item or a group of items (individual or object) (cf. e.g. (106)). In this case, use the label n .

- (106) As the outsider, Bush **has been able to** strike appealing chords -- a promise to be “different” from other leaders, a promise to be “a compassionate conservative”, a promise to be “a uniter, not a divider”. (NEWS, 2004)

Level 4. The clause

We now turn to the clausal level, that is, the clause in which the modal verb appears.

4.1 Form

4.1.1 Predictor 23: Clause type

The goal of this predictor is to determine the kind of clause in which the modal occurs: main clause or subclause. When the modal is used in the main clause (cf. e.g. (107)), use the label `maincl`.

- (107) To be a good reliever, you **should** have the nerves of a surgeon and the control of a robot. With these qualities, you can do what you want. (NEWS, 2015)

When the modal occurs in a subclause, the annotator has to use Quirk *et al*'s (1985: 1047) classification and specify which type of subclause is used. There are four types. First, the modal can be used in a 'nominal clause'. These are subclauses that serve functions similar to those of regular NPs (i.e. subject, object, complement and appositive). Nominal clauses can be of different types (see Quirk *et al* 1985: 1048-1068), among which *that*-clauses and subordinate interrogative clauses, subordinate *to*-infinitive and *-ing* clauses. In all these cases, use the label `sub.nomcl`:

- (108) **That you can read this post** means you're alive, right? (WEB, 2012) (*that*-clause)
- (109) I don't know **why you have to create these artificial distinctions**. (SPOK, 2010) (subordinate interrogative clause)
- (110) **Being able to vote in our free society** carries with it a certain minimum responsibility. (SPOK, 2000) (subordinate *-ing* clause)

Second, the subclause may be an adverbial clause (cf. e.g. (111)). In this case, use the label `sub.advcl`.

- (111) If you didn't make 90 percent, practice some more. **As soon as you can copy the first two characters with 90 percent accuracy**, add a third character to your practice. (WEB, 2012)

Third, the subclause may consist of a relative clause (cf. e.g. (112)). In this case, use the label `sub.relcl`.

(112) That's also good for smaller companies which might not necessarily have the resources to bring on a chief-of-staff or assistant role. (MAG, 2016)

Fourth, the subclause may consist of a comparative clause (cf. e.g. (113)). In this case, use the label `sub.comp1`.

(113) He was more collaborative **than you could ever imagine**, and he probably was more collaborative **than he needed to be**. (NEWS, 2013)

(114) It's not as good **as it should be**. (SPOK, 2018)

Note that some 'subordinate' clauses are not attached to a main clause (cf. e.g. (115)), in which case we talk of 'insubordination'. In such cases, use the label `insub`.

(115) We understand. **If you could just answer a couple more questions**. We heard Rameen's headscarf is missing. (TV, 2018)

4.1.2 Predictor 24: Subclause introduction

When the modal is used in an 'adverbial' clause or a 'nominal' (cf. supra), the annotator then needs to identify, in the former case, the subordinating conjunction, and in the latter, the main lexical element that the nominal clause relates to. For instance, the annotator has to identify the subordinating conjunction **ALTHOUGH** in (116).

(116) **Although** the specific terminology can be adjusted to reflect current lingo, surface management techniques are still used universally in classrooms. (ACAD, 2014)

Note that *so* and *so that* were both coded as **SO THAT** in the dataset. Examples like the following were coded as **SO ADJ THAT**:

(117) He said his problems with Post Traumatic Stress Disorder, the product of combat in Iraq, were **so** bad he **could not** work during that time

In the case of a nominal clause, the annotator has to identify the main lexical element(s) to which the clause relates. In examples (108) to (110) above, this means **MEAN**, **KNOW** and **CARRY**. In examples (118) to (120), the items identified are **RESULT**, **FEAR** and **BE AWARE**.

- (118) But the exciting **result**, he says, was that even with the technical differences, researchers **were** still **able to** reliably differentiate samples from sick people from those of well people. (MAG, 2015)
- (119) [...] worse than the **fears** I have when I lie awake at 3:00 in the morning (that I won't **be able to** make enough money [...]) (FIC, 2005)
- (120) Even if your home has grounded wiring, you should **be aware** that the quality of your electrical circuitry **can** significantly impact the performance of power-line networking devices. (MAG, 2015)

4.1.3 Predictor 25: Declarative, interrogative, exclamative or imperative

When the modal appears in the main clause, specify whether it is a declarative, an interrogative, an exclamative clause, or an imperative. For declaratives (cf. e.g. (122)), use the label `decl`. For interrogatives (cf. e.g. (123)), use the label `int`. For exclamatives (cf. e.g. (124)), use the label `excl`. For imperatives (cf. e.g. (121)), use the label `imp`.

- (122) You **should** gather as much detail as possible so that you are able to provide specific answers to questions about the current status of your program. (ACAD, 2011)
- (123) “What **can** be done?” He heard Carver breathing. Frank knew he was thinking. (FIC, 2015).
- (124) I came to find the flock! And I came here to get you before you got hurt! We **have to** go! (TV, 2012)
- (125) “Open up” means no matter where the ball is, **be able to** face it, but keep your man in sight. (MAG, 1999)

Use the label `NA` when the modal is used in a subordinate or insubordinate clause (see predictor 23).

4.1.4 Predictor 26: Polarity

The next predictor focuses on the polarity of the clause, positive or negative. Positive polarity clauses do not contain a negator (cf. e.g. (126)); the label `pos` is used in this case.

- (126) A contract **should** address expectations, job standards, what is to be paid and when. (NEWS, 2014)

By contrast, negative polarity clauses contain a negator. There are four ways in which negation can be realized. First, *not*-negation can be used. When the operator occurs in its

full form (cf. e.g. (127)), the label `not` is used. When the contracted form *n't* is used (cf. e.g. (128)), the label `notc` applies.

- (127) You are twenty-three, Leila! You must get married sometime. It will be shame for our family if a beautiful girl like you **can not** find a husband. (FIC, 2009)
- (128) Eventually we lapsed back into argument: “We **shouldn't** have to scold him when he does something out of line around you, Mom. You can tell him, you're his grandmother.” (FIC, 2013)

Second, the modal clause can have negative polarity if it occurs in a negation raising context, also called ‘neg-raising’ (cf. e.g. (129)). In this case, although negation conceptually bears on the modal complement clause (e.g. ‘I don't need to convince anyone...’), the negative *not* operator is located in the main clause. In this case, use the label `n.rais`.

- (129) I **don't** think I **need to** convince anyone that America's current troubles are a result of not only governmental breakdown but also failing families and a decline of public virtue. (NEWS, 2011)

Third, the modal clause can also be negative when the speaker uses a negative polarity subject (cf. e.g. (130)). In this case, use the label `n.subj`.

- (130) It belongs, you know, to the kind of rumors that **nobody is able to** control, that never threatens the United States to do something, the United States doesn't do it. (SPOK, 1990)

Finally, the modal clause can have negative polarity when an adverb of negation is used. This includes the use of adverbs like *never*, *hardly*, *barely* and *only* (cf. e.g. (131) and (132)). In this case, use the label `n.adv`.²

- (131) But he was not out of prison. And, Mike soon warned him, he **may never** get out. (NEWS, 2010)
- (132) She **could barely** make out the structure behind overgrown hedges and untrimmed trees. (FIC, 2014)

² One might point out that *not* is also an adverb of negation and should perhaps have been included in this category. In the REM project, *not*-negation was singled out for a number of reasons, one of which is formal: unlike other adverbs of negation, *not* is syntactically more constrained and usually appears after an auxiliary.

4.1.5 Predictor 27: Scope of negation

Once negative polarity clauses have been identified, the annotator checks the scope of negation. This predictor applies to the attestations with the negative particle *not* only. That is, the aim is to determine whether *not*-negation applies to the modality (*‘It is {not possible/necessary} {for something to happen/that something is true}.’*) or to the proposition expressed (*‘It is {possible/necessary} {for something not to happen/that something is not true}’*). When *not*-negation has scope over the modality (cf. e.g. (133); *‘It is not necessary for Democrats to campaign against Bush.’*), use the label `negmod`.

- (133) The point is that the Democrats don’t **need to** campaign against Bush or engineer a landslide to keep the Senate or take back the House. (MAG, 2002)

When *not*-negation has scope over the proposition (cf. e.g (134); *‘It is possible for such initiatives not to produce success.’*), use the label `negprop`.

- (134) However, while Fullan (2007) acknowledges the impact of bottom-up reform on a local level, he cautions that such initiatives **might not** produce success on any scale. (ACAD, 2012)

The label `NA` is used for positive polarity clauses as well as for sentences with a negator other than ‘not’.

Tip to the annotator

The coder can apply a filter on Predictor 26 to add the value `NA` for Predictor 27 for all the sentences with positive polarity or a negator other than ‘not’.

4.2 Semantics

4.2.1 Predictor 28: Habituality

The next set of predictors focuses on more functional aspects of the clause in which the modals are found. The first goal is to determine whether the residue refers to a habitual situation. Habituality is a feature that describes “the situation as characteristic of the referent of the subject NP over a certain period of time” (Declerck 2006:34). When that is the case (cf. e.g. (135)), use the label `y`. Otherwise, the label `n` will be used for non-habitual situations (cf. e.g. (136)).

- (135) When Jeff was not living at the hospital, we still had to go there for day-long clinic visits two to three times a week, and often we **had to** monitor IV infusions of antibiotics or chemo throughout the night. (FIC, 2001)

- (136) This morning at 5 AM, we **were able to** get Adelina a passport.
(SPOKE, 2002)

4.2.2 Predictor 29: Counterfactuality

A modal clause can also be used to introduce a counterfactual statement. There is counterfactuality when the speaker refers to a situation that was possible or necessary but did not actualize. There are two ways in which counterfactuality can be expressed in a modal clause. The modal verb can be directly followed by a perfect infinitive (cf. e.g. (137)), or it can be triggered by the verb *wish* (e.g. (138)). In this case, use the label *y*.

- (137) But immigration reform was important, too, and it **should have** been given a greater priority within the Obama administration.
(MAG, 2012)
- (138) I **wish I could** read what it says on the side, but the letters are upside down and backwards. (FIC, 2012)

When counterfactuality is not expressed, the label *n* is used.

4.2.3 Predictor 30: Hypothetical meaning

The next predictor concerns a relatively similar phenomenon. The aim is to identify whether the residue is construed as hypothetical. A situation is hypothetical when its actualization potential is not located in the real world but in some imaginary world. When the residue is non-hypothetical (cf. e.g. (139)), the label *n* is used.

- (139) She **could** see the downtown business owners had done their best to decorate their charming little clapboard-and-brick storefronts.
(FIC, 2014)

When the situation is hypothetical, the value *y* is used. The hypothetical situation may be introduced by an *if*-clause, such as in:

- (140) **If you need to** use the bathroom, you should go now. Cause I have a lot to say. (TV, 2000)

A situation may be construed as hypothetical due to the use of the (morphologically past) modal auxiliaries *could* or *might* (cf. e.g. (141) and (142)).

- (141) There are also concerns that his arrest **could** lead to another outbreak of violence over the next year as rival traffickers try to take over his routes –David. (SPOK, 2014)

- (142) And I think that yogurt cultures are a really good place to start. You **might** wind up with something that's insanely lemony and sour and you can tweak from there. But I think it's a great jumping off point. (SPOK, 2009)

The hypothetical situation may also be introduced by an *unless*-clause (cf. e.g. (139)).

- (143) Martok is still sleeping, his great soft body sprawled facedown across the mattress. The Grays are nearly deaf. Martok won't wake **unless** John **has to** use the rifle. (FIC, 2019)

Finally, some hypothetical situations involve subject-auxiliary inversion (cf. e.g. (140)).

- (144) A basic, inexpensive policy will protect your money **should you** need to cancel a trip before it begins (or your trip is interrupted) for a medical reason (you or immediate family member) or for other covered reasons. (NEWS, 2014)

4.3 Pragmatics

4.3.1 Predictor 31: Contextual effects (strengthening and weakening)

This is the first of a series of predictors that focus on pragmatic features of the modal verbs. Here, the goal is to identify whether the context in which a modal is used has any effect on the strength of the modal value: strengthening, weakening, or no effect. First, the context may have a strengthening effect. In example (145), the possibility modal *may* is used to convey a kind of obligation. Yet such meaning is usually expressed by necessity modals, which are stronger than possibility modals. The use of *may* results from a pragmatic (politeness) strategy: semantically, possibility meaning is communicated, but in this context, its meaning is strengthened and communicates shades of necessity meaning (*we advise you to, we suggest that you...*). In this case, use the label `str` (which stands for strengthening).

- (145) When you get the decision notice, you should read it and all reviewer comments carefully to determine what you need to do. If the decision is not to accept your manuscript, you **may** want to reconceptualize the content and style and prepare a new article for submission at a future date; or, you may be able to use the feedback to improve the manuscript and submit it to a different journal. If the decision is accept or revise and resubmit, you need to know what changes you have been asked to make by the reviewers and the editor and by what due date you must submit the revised manuscript. (ACAD, 2013)

Second, the context may have a weakening effect. In (146), the necessity modal *must* is used to make an invitation. The meaning of obligation encoded by *must* gets backgrounded, thus resulting in a weaker value. In this case, use the label *weak*.

- (146) Coma stood up to leave.’ You **must** come and see us, doctor.
Kaldren has so much he wants to show you. (FIC, 2014)

Finally, when there is no pragmatic effect (cf. e.g. (147)), use the label *noeff*.

- (147) Her career, her art, always came first, and their relationship **could**
not survive that. (FIC, 2015)

4.3.2 Predictor 32: Speech act

Attention is then given to speech acts in predictor 32. In particular, the aim is to pin down which illocutionary act the speaker intends to perform. In the REM project, we follow Searle (1975) and distinguish between (i) assertives (*representatives* in Searle’s terms), (ii) directives, (iii) commissives, (iv) expressives, and (v) hedged performatives. Table 6 provides a brief description of each category.

Speech Act	Tests / description
Assertive	“[commits] <i>the speaker (...)</i> to the truth of the expressed proposition” (Searle, 1979: 12)
Directive	“ <i>attempts (...)</i> by the speaker to get the hearer to do something” (Searle, 1975: 355)
Commissive	“ <i>acts whose point is to commit the speaker to some future course of action</i> ” (Searle, 1975: 356)
Expressive	Used “ <i>to express the [speaker’s] psychological state (...)</i> about a state of affairs specified in the propositional content.” (Searle, 1975: 356)
Performative	“ <i>Explicit performative utterances are characterised by the fact that the main verb describes the act being performed by the utterance. (...) [They] are characterised by the following features: the first-person pronoun, the simple-present tense and the possibility of inserting the adverb ‘hereby’.</i> ” (Jary, 2007: 209) (Exceptions: pronoun <i>we</i> / passive voice)

Table 6. Speech act categorization

Finer-grained distinctions can be drawn within each category. Table 7 identifies those categories. In each case, an example is provided and appropriate coding labels are given.

Speech Act	Label	Example
Assertive	ass	Where do we go from here? Who knows, but each of them has assets and flaws alike that could send them in any number of directions in the near-future. (MAG, 2017)
Directive		
Order	dir.ord	Papa reached inside and filled a small lamp with kerosene. “You must never, never touch the lamp,” he warned. (MAG, 1997)
Suggestion	dir.sug	He really should just garage the car, wait out his probation, do the right thing. (FIC, 2012)
Request for action	dir.areq	“This is an emergency. Could you please connect me to the front desk?” (FIC, 2005)
Request for information	dir.ireq	“My car’s right outside. David, should I get you an umbrella?” (FIC, 2005)
Granting/ Denying Permission	dir.perm	He finally lifted his head toward Val but then looked away once more. # “You can go,” he said. (FIC, 2017)
Commissive	com	CARDENAS # You wouldn’t know her address, would you? # JOANNE # No – but I can get it for you. (FIC, 1993)
Expressive	exp	“Bunnies! I love them!” # “How can you love them? You don’t even know what they are,” Rachel teased. (FIC, 2017)
Performative		
HedPerf. Ceremonials ³	hedperf.cer	As a literary prototype of this movement, we must nominate the hero Bazarov of Turgenev’s Fathers and Sons, published in 1862, just three years after Darwin’s Origin of Species. (MAG, 1996)
HedPerf. Assertives	hedperf.ass	I have to say cutting my hair was liberating, but that’s an individual choice. (SPOK, 2015)
HedPerf. Directives	hedperf.dir.ord	However I must request that you refrain from referring to Marines as “soldiers.” Just as a soldier probably wouldn’t like to be called a Marine, even if they make an amphibious assault, Marines take pride in being Marines.

3 Cf. Fraser (1975: 189, fn. 3).

		(MAG, 2008)
	hedperf.dir.sug	Sir.. the Mondoshawan do not belong to the federation. We do not know their intention.. I must recommend a full trinuclear assault.. (FIC, 1997)
	hedperf.dir.areq	Look, I know I did you a big favor, letting you stay here. But I need -- um, I'm going to -- I need to do some remodeling now. So I have to ask you to leave. Please. (FIC, 2014)
	hedperf.dir.ireq	I don't have time to debate the law with you. I need to ask you about Nicholas LeBeck." // "What about him? You got the report." (FIC, 2008)
	hedperf.dir.perm	He ate with surprising daintiness for someone so obviously undernourished. # "I can allow you to stay overnight. You would need to leave quite early, " Wynne offered. (WEB, 2012)
HedPerf. Commissive	hedperf.com	"I can offer you lemonade or water, " she said. (FIC, 2014)
HedPerf. Expressive	hedperf.exp	David scanned it. "Dr. Porter, I must apologize, but a situation has arisen requiring my attention. We can continue at a later date, or you can address the remainder of your questions to one of my colleagues. (FIC, 2006)

Table 7. Speech act categories and code

4.3.3 Predictor 33: Directness and indirectness

The last pragmatic predictor concerns the directness of the speaker's illocutionary act. The aim is to identify whether the speech act identified previously is directly or indirectly conveyed by the speaker (cf. Searle 1979: 30-37). A speech act is direct when it is conventionally associated with the constructions used by the speakers. This feature is typically found in discussion on basic sentence types. In example (148), for instance, the declarative sentence is used to make a direct assertion. In example (149), the directive speech act (request for information) is made directly available from the interrogative morpho-syntax of the utterance. In example (150), another directive speech act (an order) can be directly recovered from the imperative form of the clause. Finally, example (151) shows that exclamative sentences are typically used to perform an expressive speech act. In all these cases, use the label *dir*.

- (148) We use the term base word for free morphemes, one-morpheme words that **can** stand on their own. (ACAD, 2014)
- (149) I know you've been really, really supportive of her. **Have** you guys **been able to** talk? (SPOK, 2015)
- (150) And finally! **BE ABLE TO** RELAX!! (WEB, 2012)
- (151) We use empty wine bottles to keep chilled water in the refrigerator. You **should** have seen the looks we got from the kids the first time we took a "pull" straight from the bottle! (MAG, 2012)

A speech act is indirect when it is not conventional and has to be inferred from contextual clues. In example (152), the directive speech act (request for action, '*come here*'), is pragmatically derived. In this case, use the label *indir*.

- (152) NARRATOR 1: A few days later, the team is in the gym. COACH: OK, everyone. That's it. See you tomorrow. Julie, I **need to** talk to you. NARRATOR 2: Julie walks over to the coach. He looks mad. (FIC, 2000)

Level 5. Adverbials

The final level of analysis focuses on the use of adverbials. For each of the predictors discussed, use the label NA for examples that do not contain adverbials.

5.1 Predictor 34: Strength

In some cases, the adverbial used by the speaker affects the strength of the modality. Adverbials can be used to strengthen the modal value (cf. e.g. (153), ‘*it was really necessary for me to be in graduate school*’), in which case the label `advstr` is used, or they can be used to weaken the modal value (cf. e.g. (155), ‘*it is probably necessary for you to hold off a new lawn*’), in which case the label `advweak` is used.

- (153) Whenever I doubted whether I **should really** be in graduate school, whether the world needed yet another dissertation on feminist Victorian writers, she was there for me, reassuring me, that yes, I did have a unique perspective and I was talented enough to succeed in graduate school. (FIC, 2019)
- (154) We’re entering the hot, dry summer months, so if you’re shooting for the more common cool season grasses, you **should probably** hold off a new lawn until fall, when chillier weather cuts down on watering costs and allows sod or seeds to take root without constant irrigation. (NEWS, 2015)

It is important to note that in both these examples, the adverb somewhat qualifies the modal meaning of necessity or possibility. Not all adverbs in a modal sentence affect the modality though. In some cases, the adverb qualifies the main verb (cf. e.g. (155), ‘*it is possible that hormone therapy slightly increases the risk of breast cancer*’). This kind of adverbial is not our focus of attention here, however, and attention should only be given to modal adverbs, i.e. adverbs that affect the modal value. Use the label NA for sentences like (155).

- (155) It warned that, in some cases, hormone therapy **could slightly** increase the risk of breast cancer, stroke, blood clots and cognitive decline. (MAG, 2006)

5.2 Predictor 35: Position

Having identified modal adverbs, the next goal is to define the position of the adverb with regard to the modal verb. There are 5 possibilities: (i) at the beginning of the sentence, (ii) between the subject and the modal verb, (iii) between the modal verb and the lexical verb, (iv) after the lexical verb, and (v) at the end of the sentence (cf. Quirk *et al* 1985:

490). The position of the adverb is to be coded as indicated in Table 8. It will be clear that not all positions are made use of equally frequently. The medial positions cannot always be distinguished, for instance, when there is a less complex form in the VP following the auxiliary. This is why the difference between the three medial positions has been neutralized in the REM project (cf. Quirk *et al* 1985: 495, fn. *a*).

Position (Hoye 1997)	Label	Example
Initial	i	Possibly they may have been sent to London.
iM	im	They possibly may have been sent to London.
M (medial)	m	They may possibly have been sent to London.
mM		They may have possibly been sent to London.
eM		They may have been possibly sent to London.
iE (initial End)	ie	They may have been sent possibly to London.
E	e	They may have been sent to London possibly .

Table 8. Position of the adverb

Note that with *be able to*, adverbs can be used *within* the modal expression itself, just after *be* (cf. e.g. (156)) or after *able* (cf. e.g. (157)). In this case, use the label *m* (medial).

- (156) When an object was moved in front of his face, he **was barely able to** track it with his eyes. (NEWS, 2017)
- (157) Last year, four American League hitters slugged their teams to victories with three home runs, but National Leaguers **were able only to** win two and lose two. (NEWS, 1995)

5.3 Predictor 36: Lemma form of the adverb

The final predictor is lexical in nature: we list the lemma of the modal adverb in capital letters in the dataset, e.g. PROBABLY in (158).

- (158) Cyber-weapons development **should probably** be removed from military control and handed over to civilian agencies more responsive to privacy and cyber-defense concerns, the official argues. (NEWS, 2014)

References

- Carlson, G. N. (1995). Truth conditions of generic sentences: Two contrasting views. In G. N. Carlson & F. J. Pelletier (Eds.), *The Generic Book*, 176-223. Chicago: University of Chicago Press.
- Coates, J. (1983). *The semantics of modal auxiliaries*. London/Cambridge: Croom Helm.
- Coates, J. (1995). The expression of root and epistemic modality in English. In J. Bybee & S. Fleischmann (Eds.), *Modality in grammar and discourse*, 55-66. Amsterdam: John Benjamins.
- Collins P. (2009). *Modals and quasi-modals in English*. Amsterdam: Rodopi.
- Cruse, D. A. (1973). Some thoughts on agentivity. In *Journal of Linguistics* 9: 11-23.
- Davies, Mark. (2008-). *The Corpus of Contemporary American English (COCA): 1 billion words, 1990-2019*. Available online at <https://www.english-corpora.org/coca/>.
- Declerck, R. (2006). *The grammar of the English verb phrase. Vol. 1: The grammar of the English tense system*. Berlin/New York, Mouton de Gruyter.
- Depraetere, I. (2014). Modals and lexically-regulated saturation. In *Journal of Pragmatics* 7: 160-177.
- Depraetere, I. (2017). The meanings of *have* and the semantics/pragmatics interface. In I. Depraetere & R. Salkie (Eds.), *Semantics and pragmatics: Drawing a Line*, 265-291. Cham: Springer.
- Depraetere, I. & S. Reed. (2006). Mood and modality in English. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics*, 269-290. Oxford: Blackwell.
- Depraetere, I. & S. Reed. (2011). Towards a more explicit taxonomy of root possibility in English. In *English Language and Linguistics* 15(1): 1-29.
- Depraetere, I. & A. Verhulst. (2007). Source of the modality: A reassessment. In *English Language and Linguistics* 12(1): 1-25.
- Fraser, B. (1975). Hedged performatives. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, 3: Speech acts*, 187-210. New York: Academic Press.
- Garretson, G., M. C. O'Connor, B. Skarabela & M. Hogan. (2004). *Coding practices used in the project optimal typology of determiner phrases*. Unpublished ms., On-line, Boston University: <http://npcorpus.bu.edu/documentation/index.html>.
- Hoye, L. (1997). *Adverbs and modality in English*. London and New-York: Longman.
- Huddleston, R. (1984). *Introduction to the grammar of English*. Cambridge: Cambridge University Press.
- Jary, M. (2007). Are explicit performatives assertions? In *Linguistics and Philosophy* 30(2): 207-234.

- Nordlinger, R. & E. Traugott. (1997). Scope and the development of epistemic modality: Evidence from *ought to*. In *English Language and Linguistics* 1(2): 295-317.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Searle, J. (1975). A taxonomy of illocutionary acts. In *Language, Mind and Knowledge*: 344–369.
- Searle, J. (1979). *Expression and meaning*. Cambridge: Cambridge University Press.
- Van Valin, R. D. & D. P. Wilkins. (1996). The case for ‘effector’: Case roles, agents, and agency revisited. In M. Shibatani & S A. Thompson (Eds.), *Grammatical constructions: Their form and meaning*, 289-322. Oxford: Clarendon.
- Vendler, Z. (1957). Verbs and times. In *Philosophical review* 56: 143-60.
- Zaenen, A., J. Carletta, G. Garretson, J. Bresnan, A. Koontz Garboden, T. Nikitina, C. O’Connor & T. Wasow. (2004). Animacy encoding in English: Why and how. In D. Byron & B. Webber (Eds.), *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, 118-125. Barcelona, July 2004.