

## What makes us choose which modal? Putting semantic, syntactic and lexical factors in the balance

Cyril Grandin  
Bert Cappelle  
Ilse Depraetere  
Université de Lille  
{cyril.grandin,bert.cappelle,ilse.depraetere}@univ-lille.fr

Keywords: corpus linguistics, modal verbs, multifactorial analysis, usage-based linguistics, entrenchment

One of the key properties humans possess is their ability to communicate states of affairs that are not straightforward facts but that *should* or *could* be the case. The words and constructions that allow people to verbalize non-factual meanings belong to modality, a domain that is central to grammar and that, accordingly, has spawned a voluminous literature. Despite the wide attention given to this familiar area, we still do not know which linguistic features exactly allow speakers to choose one modal expression over another. Why, for instance, does one tend to say, “Sometimes you need to choose your battles” (rather than “... you should choose...”) but “It should be noted that...” (rather than “It needs to be...”) We here report interim findings from an ongoing extensive corpus-based study that aims at exploring this issue, by simultaneously measuring the impact of semantic, syntactic and lexical factors on speakers’ choices of modal verbs.

From the Corpus of Contemporary American English, we randomly extracted 5000 occurrences with English modal expressions in 150-word contexts, equally balanced across six core modal auxiliaries (*may*, *might*, *can*, *could*, *should*, and *must*), three semi-modals (*have to*, *need to*, *ought to*), and a periphrastic construction (*be able to*). These are being annotated for 35 variables that previous smaller-scale theoretical and empirical studies argued to be relevant to modal verb selection. Some of the variables are primarily semantic in nature, pertaining, for example, to whether or not the modal verb expresses epistemic modality (e.g., *I must have been confused* vs. *What needs to be done?*), narrow-scope or wide-scope modality (e.g., *You should try this* vs. *There must be blood*), and subject-internal-source or subject-external-source modality (e.g., *Well, if you must know, ...* vs. *You must work hard to succeed*). Other variables are tied to formally more easily detectable features, such as the presence/absence of a negator, or the use or not of subject-auxiliary inversion. To model variation between the modal expressions, we use logistic regression techniques. With about two thirds of the 175,000 coding decisions completed and an interrater reliability for the relatively hard-to-code variables reaching at least 85%, we are in a position to confirm that many, though not all, of the previously proposed variables play a significant role in a speaker’s selection of a particular modal expression.

However, while most of the literature on modality has stressed the usefulness (supported in our research) of binary categorical distinctions such as those mentioned above, we argue that a model of speakers’ choices should take into account insights from recent theoretical developments in linguistics. In particular, the frameworks of Cognitive Construction Grammar and usage-based linguistics predict that modal expressions and their surrounding lexical context can form entrenched chunks (Hilpert 2016, Cappelle and Depraetere 2016, Cappelle, Depraetere and Lesuisse forthcoming). For instance, language users frequently encounter sequences such as *all you need to know* (rather than *all you should know*) and *perhaps you should...* (rather than *perhaps you need to...*). We argue that this lexical knowledge should be incorporated into the model and demonstrate how this can be done.

### References

- Cappelle, Bert and Depraetere, Ilse. 2016. Response to Hilpert, Martin. Change in modal meanings. Another look at the shifting collocates of *may*, *Constructions and Frames*, Special Issue. Modal meaning in Construction Grammar, 8(1), 86–96.
- Cappelle, Bert, Ilse Depraetere and Mégane Lesuisse. Forthcoming. The necessity modals *have to*, *must*, *need to* and *should*: using n-grams to help identify common and distinct semantic and pragmatic aspects. *Constructions and Frames*.
- Hilpert, Martin. 2016. Change in modal meanings: Another look at the shifting collocates of *may*, *Constructions and Frames*. Special Issue. Modal meaning in Construction Grammar, 8(1): 66–85.