The necessity modals have to, must, need to and should: using n-grams to help identify common and distinct semantic and pragmatic aspects. 11.2: 220-243

Bert Cappelle, Ilse Depraetere, Mégane Lesuisse

The necessity modals *have to*, *must*, *need to* and *should*: using n-grams to help identify common and distinct semantic and pragmatic aspects[1]

Bert Cappelle, Ilse Depraetere and Mégane Lesuisse
Univ. Lille, CNRS, UMR 8163 - STL - Savoirs Textes Langage, F-59000 Lille, France

2019. Special issue of *Constructions and Frames* 11. Finkbeiner, Rita (ed.) Pragmatics and Constructions. 220-243. https://doi.org/10.1075/cf.00029.cap

Abstract
When an ambiguous lexical item appears within a familiar string of words, it can instantly receive an appropriate interpretation from this context, thus being saturated by it. Such a context may also short-circuit illocutionary and other pragmatic aspects of interpretation. We here extract from the British National Corpus over 500 internally highly collocating and high-frequency lexical n-grams up to 5 words containing *have to*, *must*, *need to* and/or *should*. These contexts-as-constructions go some way toward allowing us to group these four necessity modals into clusters with similar semantic and pragmatic properties and to determine which of them is semantico-pragmatically most unlike the others. It appears that *have to* and *need to* cluster most closely together thanks to their shared environments (e.g., *you may have/need to…*, expressing contingent, mitigated necessity), while *should* has the largest share of unique n-grams (e.g., rhetorical *Why shouldn't I…?*, used as a defiant self-exhortation).

Keywords: necessity modals, n-grams, mutual information, lexically restricted saturation, short-circuiting, illocutionary force, hierarchical clustering analysis, intersubjectivity

1. Introduction: modal constructions, lexically restricted saturation and short-circuited meaning

Modal verbs, or modals for short, are notoriously ambiguous (see, among many others, Ruppenhofer and Rehbein 2012; Marasovic et al. 2016) and can be used in utterances with often specific pragmatic effects (see, for instance, Leech 2014). This paper aims to understand the role and nature of 'context' in interpreting a modal verb, here specifically one of the necessity modals *have to*, *must*, *need to* and *should*. In Cappelle and Depraetere (2016a), it was argued that a modal and the immediate context in which it occurs – one or more words to the left and/or to the right of it – may often be considered a construction. 'Construction' was used there in the sense, as is customary in Construction Grammar, of a form-function unit that is likely to be stored in a speaker's mental repository of words, idioms, recurrent sequences and semi-fixed patterns (cf. Goldberg 2006). This is also the sense of 'construction' adopted here. Our focus is on constructions towards the lexical pole of what is typically referred to as the syntax-lexicon continuum. We look at lexically specified strings, that is, contiguous sequences of actual words (not syntactic categories), one of which is a modal. However, most of these sequences do not constitute complete utterances. For instance, *it should be noted that* (one of the lexical constructions whose mental storage we posit) still needs to be followed by a clause to be used legitimately. In that sense, the vast majority of the constructions we deal with here are just lexical 'chunks' that are part of larger units.

We believe that a close investigation of these chunks may shed some light on how each of the four necessity modals can be characterized semantically as well as pragmatically. We do not pretend, though, that this focus on collocational preferences is the one and only way to reach such characterizations of the modals concerned. We do not claim either that speakers and hearers have access only to these very specific lexical chunks and not also to more schematic constructions with the necessity modals. A more general construction with *must*, for instance, probably specifies that,

[1] This paper was funded through a grant from the *Agence Nationale de la Recherche* (ANR-16-CE93-0009 REM).

disregarding the concrete words this modal is used with, the necessity it expresses is speaker-driven, while a more general construction with *have to* specifies that the necessity is circumstantial, one with *should* that the necessity is of a relatively weak degree, and so on. What we do claim, however, is that, alongside these general constructions, many of the lexically more specific sequences definitely have their cognitive reality. It is even to be expected that in some cases, the function of a lexically specific string overrides the meaning inherited by the more general construction containing that modal. For instance, *it must be noted that…* is hardly any higher in strength or more obviously speaker-directed than *it should be noted that…*.

Although the scarce attention to pragmatics in constructionist accounts of modals (for exceptions, see e.g. see e.g. Stefanowitsch 2003; Boogaert 2009; De Haan 2014; Boogaert and Fortuin 2016) might suggest that the integration of pragmatics in modal constructions is problematic, it is certainly possible to include pragmatic information in a modal construction's representation. After all, constructions can come with pragmatic specifications (for surveys, see Kay 2004; Nikiforidou 2009; the introduction to the papers in Bergs and Diewald 2009; Lee-Goldman 2011; Cappelle 2017; introduction to this issue). What is typical of the pragmatic value of a modal construction is that it can be 'short-circuited', to use Morgan's (1977) imagery of how pragmatic aspects of our understanding of an utterance can be conventionally associated with specific expressions. For instance, *Can you please… ?* is *immediately* understood as a request, without the hearer having to compute this interpretation on the basis of a more basic ability meaning that *can* might have (see, e.g., Searle 1975; Aijmer 1996; Leech 2014, Ruytenbeek 2017). Short-circuiting is in some respects similar to the process of conventionalisation that affects frequent invited inferences (cf. Traugott and König's 1991). However, even if a short-circuited implicature is one that is conventionally associated with the form of a construction, short-circuited pragmatic aspects of meaning do not become part of a construction's semantics, as they can still (albeit with effort or facetiousness) be cancelled in context.

What is more, Cappelle and Depraetere (2016a) argue that modal verb constructions can also short-circuit 'non-pragmatic' aspects of interpretation. For instance, the modal in the common expression *if I could just say*... is standardly understood as expressing permission, which is why it sounds funny when this string of words ends up being used to express another kind of root possibility, such as ability (e.g. Homer Simpson: "If I could just say a few words… I'd be a better public speaker").[2] So, while there is a semantic core of 'possibility' meaning (referred to as the 'context-independent semantic layer' in Depraetere 2014), in context, this meaning is made more specific and the hearer chooses between one of several limited options (a process which constitutes 'the context-dependent semantic layer' in Depraetere 2014). That is, assigning to a modal a particular meaning such as 'have permission to' rather than 'have the ability to', 'have the opportunity to', etc. happens on the basis of contextual cues. Nevertheless, we adhere to the view that the output of disambiguation is a *semantic* level of interpretation (Recanati 2004; Depraetere 2010). Besides, there is little or no pragmatic *reasoning* involved when the disambiguating context is a well-entrenched sequence of words. When modals feature in a construction (a stored form-function unit), as in *If I could just say…*, the selection of the right meaning is 'short-circuited' by the familiar context around the modal.

Depraetere and Salkie (2017) consider this disambiguation to be a case of saturation (Carston 2009: 49, Recanati 2004: 7, Recanati 2010: 70), a pragmatic process which is obligatory: it is necessary in order to complete an incomplete proposition. As the search space for a contextually appropriate interpretation is constrained, Depraetere and Salkie (2017: 26) use the term *lexically restricted* saturation. In order to determine the propositional content of an utterance, one of several (unrelated or otherwise distinct) meanings has to be selected and the context gives clues as to what meaning is appropriate. For instance, *There's a bat in the cupboard* (cf. Clark 1991: 10) will be understood differently depending on whether it is reply to *Why is she screaming*? or to *Let's go out and play a*

[2] *If I could just say a few words* usually also exemplifies insubordination (cf. Verstraete and D'Hertefelt 2016; D'Hertefelt 2018). Treating this sequence as an ordinary subordinate clause in a main clause, as does Homer Simpson in this example, is therefore already odd in itself (although one does very occasionally find sentences such as *If I could say a few words, that'd be awesome*).

game. Specific contexts trigger interpretations but while there is a choice to be made, the choice for one rather than another meaning can be fairly immediate. To illustrate with a familiar textbook example, while selecting the meaning of 'financial institution' rather than 'shore of a river' for *bank* is a context-dependent process, arriving at the right semantic value for *bank* can hardly be called a pragmatic affair if we find this noun followed by the noun *crisis*. This is not to say that *bank crisis* can never be given another interpretation. For instance, a river's banks on the brink of collapsing due to erosion may be referred to by a journalist as a *bank crisis*. Yet, this will then have the flavour of a (poor) pun.

Just like *bank crisis* is stored as a lexical unit in which *bank* already comes with an appropriately disambiguated value, *If I could just say…* is a stored sequence that immediately helps the hearer to determine which semantic value of *could* is intended among a range of possible candidates.[3] An expression with a disambiguated modal meaning can still convey extra pragmatic information, that is, information at what Levinson (2000) and Recanati (2012) call the post-semantic level (but this term's suggestion that the pragmatic information is added at a later stage than the disambiguation is not something we would subscribe to). In the case of *If I could just say…*, what is meant is more than just 'If I had permission to say…'; this sequence of words is a standard way of *asking for* permission (i.e., the expression functions as a request) and of expressing that the speaker assumes that this permission is in fact granted (i.e., the expression also functions as a floor-taker) (Cheng 2007: 206; Cappelle and Depraetere 2016a: 27). In Depraetere's (2014) multi-layered approach to modal meaning, this layer is called the 'context-dependent, pragmatic layer'.

If I could just say… is just one of very many recurrent sequences containing a modal. Cappelle and Depraetere (2016b) present a methodology to identify such 'lexical' constructions with modals, based on the mutual information of the modal and words occurring to the left and right of it, as well as of the words to the left and right of these context words, and so on. This methodology is a simple way of extracting from a corpus the lexical n-grams (i.e., continuous strings of 2, 3, or any number *n* of words) around a word such that each word entertains a strong collocational connection with its neighbouring words. Recurrent multi-word sequences have been referred to as 'lexical bundles' in the corpus-linguistic literature (cf. Biber and Conrad 1999), a term that we find potentially useful, even though we would like to stress that the n-grams retrieved with the methodology referred to above are not just high-frequency ones but especially also ones that are high in internal collocational strength, in terms of mutual information.[4]

Must served as a test case for Cappelle and Depraetere (2016b). The present paper applies the methodology described there to a larger range of necessity modals: *have to, must, need to* and *should*. We here aim to use this methodology as a tool to capture semantic and any pragmatic aspects that these modal verbs share or fail to share, by comparing the highly collocating lexical environments in which they occur. This study complements existing research into the semantics and pragmatics of modal verbs in at least two ways. First, as far as we know, a contrastive analysis of the lexical sequences with these modal verbs is new, at least for sequences of a size larger than two words. Our general approach has points in common with Hilpert's (2016) and Flach and Hilpert's (2017), but these studies of modal meanings focus on 2-grams. In the approach that is adopted here, there is potentially no limit on the size of the lexical sequences retrieved from the corpus. Secondly, previous studies paying attention to specific syntactic, semantic and pragmatic features only considered necessity modals in pairs (*should* and *must*, e.g. Rivière 1981; Groefsema 1995; Keckses & Kirner-Ludwig (2017); *have to*

[3] We are aware that the use of the term 'disambiguated' is somewhat unfortunate here, as our view is that modals illustrate polysemy (related meanings) rather than ambiguity (unrelated meanings). For want of a less ambiguous verb, we will continue to use it, bearing in mind that in the case of modals, the choice is among one of several *polysemes* (Cruse 2011: 115).

[4] As pointed out by a reviewer, our perspective aligns in a general way with Hunston and Francis's (2000) pattern grammar approach. Given the emphasis in our paper on phraseological units, we acknowledge that this may turn out to be the case, although we have at this stage not made a systematic comparison between our methodology and theoretical stances and Hunston and Francis's.

and *must*, e.g. Larreya 1982; Depraetere and Verhulst 2008; *should* and *ought to*, e.g. Cappelle and De Sutter 2010). Our study fills this gap by taking into consideration four necessity modals at the same time.

In keeping with one of (Cognitive) Construction Grammar's basic tenets (Goldberg 2003), we allow constructions to be not just highly schematic templates but also specific lexical strings that consist of words whose co-occurrence can be observed significantly more often than could be expected, based on their individual frequencies. Of course, we would not go as far as to say that, for instance, *it should be noted that* is really a full construction. As this n-gram is not a complete syntactic constituent and as speakers know it has to be followed by a clause, what speakers have stored as a unit in the mind is really [*it should be noted that* + clause]. Our methodology therefore does not yield lexical *constructions* necessarily – it typically allows us to retrieve lexical sequences that are part of constructions containing both fixed lexical material and at least one open slot yet to be filled in (cf. Cappelle and Grabar 2016). That constructions can be hybrid, combining lexical items and syntactic categories, is a common assumption in Construction Grammar and related frameworks (cf., e.g., Jackendoff 1997). Accordingly, multiple levels of abstraction should be considered when we chart speakers' knowledge of constructions (cf., e.g., Trousdale 2008, Gyselinck 2018). In view of this, we do not claim either that the modal constructions we here posit, by syntactically completing, where needed, the retrieved lexical sequences around a modal verb, are the *only* constructions with these modal verbs. Some modal verb constructions stored in the mind of the speaker will be considerably more abstract and general than the 'syntactically extended' patterns found here. They may, for instance, consist of just a modal verb preceded by a syntactic position to be filled by a subject NP and a syntactic position to be filled by a VP. Some other stored modal verb constructions may be lexically even more specific than the ones our methodology allowed us to retrieve. An example may be *You should try this*, which is possibly a stored exemplar of *you should try {NP / to VP}*.

The remainder of our paper is organized as follows. In Section 2, we describe the methodology used to identify (partial) modal constructions in the BNC (British National Corpus). In Section 3, we present the corpus findings and analyse the lexical n-grams, viewed as (partial) constructions, in an attempt to shed light on shared and different semantic and pragmatic aspects of the necessity modals studied. We provide a conclusion in Section 4.

2. Methodology

2.1. Extracting highly collocating lexical n-grams

The methodology adopted in Cappelle and Depraetere (2016b) is as follows. In Mark Davies's BYU-BNC interface (Davies 2004-) two-word strings (bigrams) with *must* were extracted that occur at least 50 times in the corpus and that have a mutual information (MI) score of at least 3. This latter cut-off point has been proposed by some researchers working on collocations (Hunston 2002: 71; Xiao and McEnery 2006: 105). The resulting strings were then used as search items in the next round of n-gram retrieval. The procedure remained the same: we again set 50 as the minimal frequency and 3 as the minimal MI score. The procedure was repeated until no more strings were found that occur at least 50 times in the corpus with an MI of 3. The recursive nature of the procedure would in principle allow the retrieval of n-grams to be carried out more or less automatically. For the present study, however, we did not write a script but used Davies's interface. N-grams were extracted, in the way just described, for *must*, *should* and the inflected forms and base form of *have to* and *need to*.

2.2. Identifying shared and unique n-grams for the necessity modals

In order to explore to what extent the four modals are found in the same or distinct n-grams, we created four sets (i.e., lists) of retrieved n-grams, one for each modal (with multiple inflectional

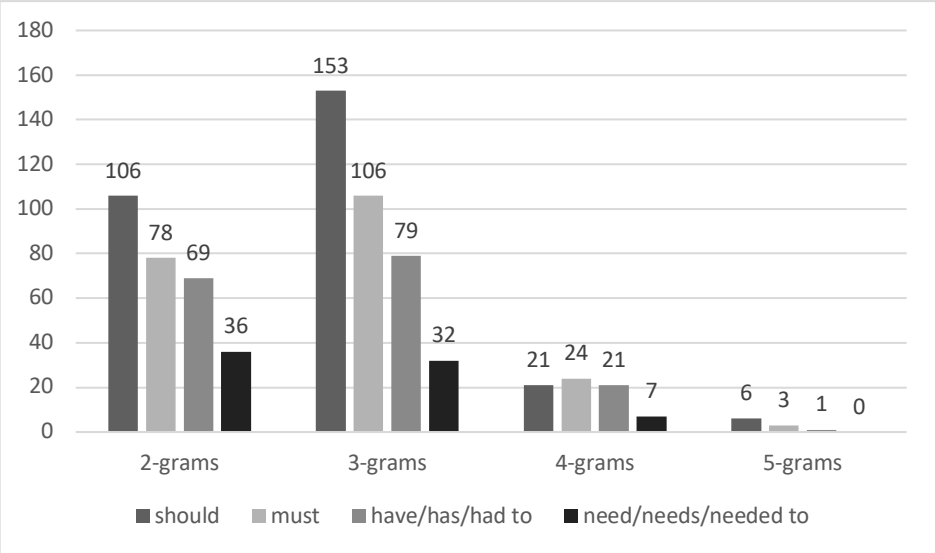
forms of *have to* and *need to* collapsed into single lemmas[5]). In each set of n-grams, we replaced the modal verb in each n-gram by the generic label 'MOD'. With the help of an online tool for creating Venn diagrams, we then determined the size and members (if any) of the intersection(s) of these four sets of n-grams, as well as the n-grams that are unique to a particular set.[6]

We also performed hierarchical cluster analysis to find out which modals are most alike with respect to their lexical environments, as captured by the n-grams they share. For this, we used the function `hclust` in the R package `pvclust`, by which a dendrogram can be plotted. It is assumed here that similarity in collocational behaviour indicates similarity in function (Harris 1954, Firth 1957, McDonald and Ramscar 2001, Sahlgren 2008).[7]

3. Results

3.1 Quantitative results

Out of the four necessity modals studied, *should* is the one to be found in the highest number of different n-grams (cf. Figure 1): 106 different bigrams (e.g. *applicants should*, *should specify*), 153 different 3-grams (e.g. *maybe we should*, *local authorities should*), 21 different 4-grams (e.g. *I think we should*, *consideration should be given*) and 6 different 5-grams (e.g. *it should be noted that*, *no reason why you should*).[8]



[5] This lemmatization was done for ease of analysis but is also justified in that the n-grams we retrieved with the forms *has* and *had* and for *needs* and *needed* are also to be found in the list of n-grams with the forms *have* and *need*, respectively. It appeared, in other words, that the present tense third person singular use and the past tense use of the modals *have to* and *need to* do not exhibit lexical preferences that are specific to these inflectional forms only.

[6] http://bioinformatics.psb.ugent.be/beg/tools/venn-diagrams, last accessed 6 March 2018.

[7] A reviewer rightly comments that distributional similarity can also point to a paradigmatic opposition. For example, antonyms (e.g. *increase* and *decrease*) are known to have similar collocational profiles. However, as argued by Janda (2009), "[a]ntonyms are synonyms that differ in just ONE value, they are the next closest thing to synonyms." Moreover, in the context of our study, there is little to no risk that any of the four necessity modals (*have to*, *must*, *need to* and *should*) is a paradigmatic opposite of any of the others.

[8] Among the unique n-grams with *should*, there are also a small number of sequences in which *should* is not, in fact, a necessity modal but an alternative of *would* (e.g. *I should imagine*, *I should think so*, *I should be grateful if*). We have not discarded these few occurrences.

Figure 1. Number of high-collocation n-grams containing each of the four necessity modals *should, must*, *have to* and *need to*, extracted from the British National Corpus

Need to is the necessity modal with the lowest number of n-grams retrieved: 36 bigrams, 32 3-grams, 7 4-grams and no 5-grams. Except for *need to*, the number of extracted 3-grams containing a necessity modal is higher than the number of extracted bigrams with a necessity modal. This may be surprising, as one would expect there to be fewer collocational sequences as the number of words in the string increases – an expectation in fact supported by the observation that there are no sequences longer than 5 words to be obtained with our extraction method. The main reason why the 3-grams nonetheless outnumber the 2-grams is that for a single bigram such as *should be*, a large number of 3-grams could be retrieved with a strongly collocating past participle or adjective (*should be added*, *should be adopted*, *should be capable*, *should be careful*, …).

Once we replace the actual modal by a generic label ('MOD'), we find that there are 517 different n-grams, some of which occur with multiple modals (see Appendix 1). Of these 517 n-grams, however, only 20 (3.9%) are shared by all four necessity modals (cf. Figure 2). 357 n-grams (69%) are unique to one particular modal. That more than half of the n-grams are specific to one modal or another suggests that the necessity modals have their own lexical profiles. This specificity is most pronounced for *should*, for which 171 out of the 286 n-grams (or 60%) are not shared by any other modal, and least so for *need to*, for which only 20 out of 75 n-grams (or 26.6%) are unique to this modal.

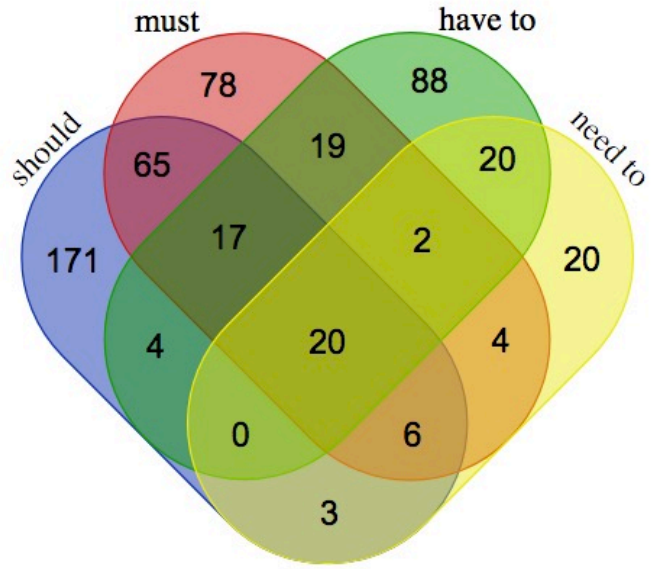


Figure 2. Venn diagram showing the number of n-grams that are common to two or more modals or that are unique to each modal.

We can make the following observations about *must*, *have to* and *need to*:

(i) *must* and *have to* share 9% (19 out of 211) and 11% (19 out of 170) of their n-grams, respectively, with one another;

(ii) *have to* and *need to* share 11.7% (20 out of 170) and 27% (20 out of 75) of their n-grams, respectively, with one another; together with the previous observation, this suggests that *have to* has more n-grams in common with *need to* than with *must*;

(iii) *must* shares only 2% (4 out of 211) of its n-grams with *need to*, and *need to* 5.3% (4 out of 75) of its n-grams with *must*.

If we bring *should* into the picture, we find the following facts:

(iv) *must* shares as many as 30.8% (65 out of 211) of its n-grams with *should* and *should* 22.7% (65 out of 286) of its n-grams with *must*;

(v) *have to* shares 2.4% (4 out of 170) of its n-grams with *should* and *should* only 1.4% (4 out of 286) of its n-grams with *have to*;

(vi) *need to* shares 4% (3 out of 75) of its n-grams with *should* and *should* only 1% (3 out of 286) of its n-grams with *need to*.

Table 1 summarizes these findings by showing for each necessity modal whether its unique and shared n-grams represent a surplus (in green) or shortage (in red), when observed values are compared to statistically expected values.

Table 1. Surplus and shortage (in percentage) of observed over expected n-grams

	should	must	need to	have to
should	37%			
must	10%	0%		
need to	-34%	-8%	29%	
have to	-50%	-10%	46%	64%

The mutual relatedness of the four modals is best brought out by the dendrogram shown in Figure 3, which is the output of our hierarchical cluster analysis. This statistical analysis measures the closeness between the modals in terms of their occurrence in n-grams. The degree of similarity is represented on the Y-axis. The dendrogram shows that *have to* is more closely related to *need to* than to *must*. *Should* is the least connected to the other modals, *must* being a somewhat closer 'relative' of *should* than of *have to* and *need to*.

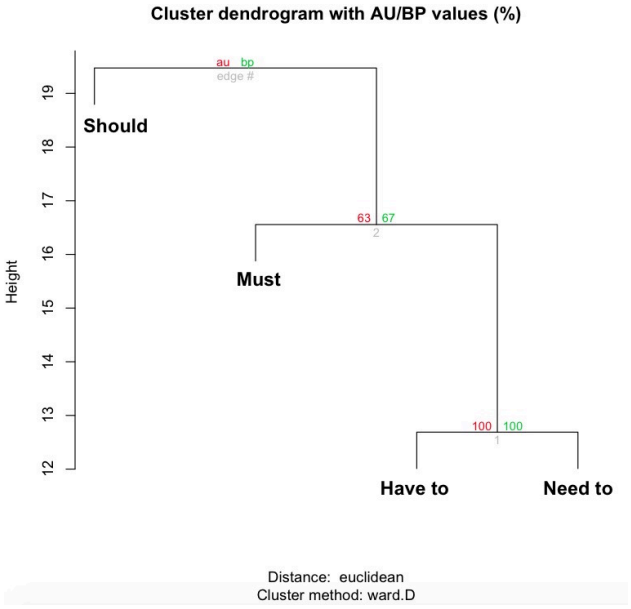


Figure 3. Cluster dendrogram with AU/BP values (%)

The figures in red and green correspond to *p*-values. The ones in red show the AU (approximately unbiased) *p*-value, while the ones in green show the BP ("bootstrap probability") *p*-value. The former of the two is more accurate. The p-values indicate how much support there is for distinct clusters at each split. Thus, that *must* forms a cluster of its own (leaving *should* behind on its own as well) is actually less strongly supported by the data than that *have to* and *need to* form a cluster together. Indeed, the cluster of *need to* and *have to* has maximum AU/BP values, *i.e.,* 100%, which means that this cluster reliably follows from the data.

3.2 Qualitative results

In this section we zoom in on a few qualitative generalizations that emerge from the sets of n-grams-as-constructions. We especially concentrate on pragmatic effects relating to illocutionary force.

3.2.1. *Why should(n't)…*: rhetorical questions

Should stands out among the other necessity modals in that it often occurs with *why*: there are 12 (partially overlapping) n-grams containing both *why* and *should(n't)*.[9] These function as rhetorical questions denying the necessity to do something (as in (1a)) or not to do something (as in (1b)); as the negation here is situated at a higher level (negating the justification of the proposition), it is compatible with both *should* and *shouldn't*. Pragmatically, sequences containing *Why shouldn't…* often function as exhortations to do the opposite of what is expressed in the underlying proposition:

(1) a. 'I've had enough of you!' she cried vehemently, not waiting for his response. '*Why should I* care what you feel when all the time you behave like a monster?' she asked rhetorically. (= 'I see no reason why I should care…'; hence: 'I won't (allow myself to) care…')

b. I felt so cross with myself, so stupid for wearing a short skirt. But on the other hand, I was furious; I thought: this is ridiculous, *why shouldn't I* go out wearing what I want without getting hassle? (= 'I see no reason why I shouldn't go out…'; hence: 'I will go out wearing…')

Why should(n't)… questions may also be used in cases where *should* is epistemic. Here, too, the pragmatic function is that of denying that the underlying proposition is justified:

(2) She thought, looking dubious. 'It doesn't sound likely. Why should there be a conspiracy? (…)'

The illocutionary force is then something like 'Let's not assume/conclude too easily that…'.

3.2.2. *Should* with a passive infinitive: low intersubjectivity

Another striking finding is that *should* is particularly prone to combining with passive infinitives, as in this example:

(3) Any visual aid *should be* carefully *selected* and *planned* to add clarity to the presentation.

In 47 out of the 171 n-grams that s*hould* does not share with any other modals (27%), we find a passive infinitive (compared to 14 out of 78 unique n-grams with *must* (14%) and no passive infinitives in the unique n-grams with *have to* and *need to*). This suggests that *should*, more than any other of the necessity modals, occurs in contexts in which intersubjectivity (Verhagen 2005; Davidse et al. 2010; Traugott 2010; Brems et al. 2012) is low, in the sense that the utterance does not explicitly represent the addressee as the agent of the action. (Introducing the addressee as the agent in a *by*-PP would be very odd: *??Any visual aid should be carefully selected… by you*).

3.2.3. *Must surely*; *should always/never*; *perhaps you/we should*; *I don't think we should*: strength of modality

[9] In COCA, punctuation marks are treated as separate items. Therefore, *Why should I?* could have been considered as a 4-gram. We have decided, though, not to count this sequence separately from *Why should I*, although in actual fact, one might argue that [*Why should I?*] and [*Why should I + VP?*] are different constructions, as they have somewhat different discourse-functional properties. (The former can only be used as a defiant retort to a suggestion just made; the latter is used to challenge a suggestion made in the clause itself.)

Given that *must* expresses strong necessity, we shouldn't be surprised to find that it collocates well with the adverb *surely*, which is harmonious with strong epistemic necessity:

(4) She must surely be the most beautiful in the world.

This adverb only occurs with *must* in the complete set of n-grams and suggests that *must* can be used for intersubjective modality in the sense of Nuyts (2012: 58), that is, the modal evaluation is "presented as being shared between the assessor and a wider group of people, possibly (but not necessarily) including the hearer".

Should, which is generally seen as expressing a mitigated, weakened form of necessity, is nevertheless also used with with strong adverbs, like *always* and *never*. *Never* only occurs in combination with *should* in the complete set of n-grams.

(5) a. You *should always* be very careful when dealing with persuasive equipment sales people.
b. Teaching strategies and classroom organization *should never* be viewed as ends in themselves.
c. The simple fact of the matter is that we *should never* have been there in the first place.

These combinations can be seen as cases where the relatively weak necessity expressed by *should* receives pragmatic (or at least contextual) strengthening.

On the other hand, *should* combines with *think* in hedging contexts involving negative raising, as in the following example, where it cannot be replaced by *must* (*I don't think we should…* has 62 occurrences in the BNC, as opposed to none for *I don't think we must…*):

(6) It was kind of Anne to think of it and not condemn us, but – well, we've managed so far being discreet about it, and *I don't think we should* change things now.

This ties in with *should* being much more suitable than *must* for hedged expressions of root necessity. Note also that the mitigating adverbs *maybe* and *perhaps* are only found with *should* in our set of n-grams (*maybe we should*, *perhaps you should*, *perhaps we should*).

4.2.4. The modals with *I*/*you*/*we*: hedged performatives, positive politeness and speaker inclusion

When it comes to the Subjects that each of the modals collocate with, the following patterns can be observed. First, *I* typically occurs in combination with performative verbs in so-called hedged performatives (Fraser 1975).

(7) I {have to / must} {say / tell (you)}
I must admit, I must ask, must admit I
I have to confess
I should {imagine / think / think so}[10]

The other n-grams with the first person singular pronoun similarly occur in hedging contexts in combination with the verb *think*. Here, we find the modal *should* (*think we should*, *I think we should*, *I don't think we should*, *I think they should*, *I think you should*), as we mentioned in Section 3.2.2, but also the verb *need* (*think we need to*, *I think we need to*).

Second, *you MOD* is at the intersection of the four sets of modal verb n-grams; in other words, the four modals are commonly used to issue a directive to the addressee. However, the communicative

[10] In fact, these cases do not, strictly speaking, involve *should* as a necessity modal but as an alternative to *would*.

effect achieved by these constructions is quite different. Two verbs that collocate with *must* in particular are typically or often used as part of positive politeness markers (cf. Brown and Levinson 1987) aimed at bonding:

(8) a. You must come and have supper sometime.
b. But you must tell me all about it, I can't wait to hear.

In these cases, *must*, which is generally considered as the strongest modal, giving rise to very forceful orders, is not actually used with its unmarked, 'psychologically prototypical' (Coates 1983) force: it serves as a sign of comradery, whereby the speaker shows interest for the hearer.

Third, *we MOD* is likewise an n-gram that is common to the four modals. Here, the self-exhorting character of the construction is likely to be part of a strategy that mitigates the effect of the directive by including the speaker into the set of addressees to whom the order is addressed.

3.2.5. *Must* and *should* with a perfect infinitive: epistemic necessity vs. reproach

Must and *should* share many n-grams with a perfect infinitive. What these shared environments by themselves do not reveal, however, is that the meaning involved is very different. *Must* in this case usually expresses epistemic necessity, while *should* in these environments is used to make a reproach:

(9) a. I'm sure Suzannah must have known him.
b. But I should have known better than to accept anyone recommended by you.

3.2.6. *May have to* and *may need to*: contingent necessity

A syntactic property that distinguishes the semi-auxiliaries *have to* and *need to* on the one hand from *must* and *should* on the other is their combinability with auxiliaries (*don't*, *doesn't*, *didn't*, *will*, *'ll*, *shall*, *won't*, *would* and *may*). Of particular interest is *may*, which in combination with *you* and *have*/*need to* helps to mitigate the addressee-directed necessity by presenting it as a contingency, not as a certainty:

(10) a. Your financial situation can improve, but *you may have to* wait until next year before you are entirely out of the woods.
b. If you have serious drain or waste blockage problems, *you may need to* use specialist equipment or to call in a specialist drain clearing firm.

There is a clear semantico-pragmatic difference between *you may have to wait* and *perhaps you should wait*: the former expresses a weak degree of epistemic possibility that it will be necessary for you to wait due to external circumstances, while the latter is a way of suggesting politely that, as far as the speaker is concerned, it is advisable for the addressee to wait.

4. Conclusion

In this paper, we have laid out a particular approach to the pragmatic meaning of modal constructions: we have argued that there is a pragmatic layer that complements the semantic layer in the functional component of the constructions. Lexically restricted saturation results in the selection of one from a restricted set of meanings potentially communicated by the modal and the modal construction 'short-circuits' the choice. Likewise, lexically specific modal constructions can give rise to short-circuited pragmatic effects.

We have also explained and applied a specific methodology to identify such lexically specific constructions (or at least lexically specific *parts* of constructions), in an attempt to pin down the semantico-pragmatic profiles of four necessity verbs. As argued in Cappelle and Depraetere (2016b) a

constructional analysis of the type presented here is very useful from the perspective of second language acquisition in that it tells us what collocational patterns should be insisted upon. Section 2 shows that it is possible to make some precious generalisations about typical usage contexts of each of the modals on the basis of the range of lexical constructions of each of the modals. While this methodology has revealed interesting insights about lexical bundles and the relative kinship among the four modals (made explicit in the Venn diagram and in the dendrogram presented in Section 3), it does not in itself provide a comprehensive picture of the factors that lead the speaker to use say, *have to*, rather than *need to*, or *should* rather than *need to*. It follows from this that an additional feature analysis is needed of syntactic, semantic and pragmatic features that are significant if we want to find out what unites and sets apart the modal verbs.

It would also be worthwhile to use the n-grams retrieved here as predictors in a multinomial regression model to see how well they classify (i.e., predict the actual use in context of) the four modals in a different (but similar) corpus. In addition, we should compare or combine such n-grams as predictors with the predictive power provided by a model that implements a so-called Bag-of-Words approach (where one simply checks which words occur how often in a text, independently of these words' positions in the text – word order plays no role). In the Natural Language Processing community, such a Bag-of-Words model seems to be treated as a standard baseline for classification performance. Obviously, if we want to validate the insights from the present paper, a classifier that (also) incorporates lexical n-grams should perform better than a classifier that is (only) based on a Bag-of-Words approach.

The present study has not presented a methodology to retrieve *all* the constructions that contain the modals concerned. There are, obviously also some more general modal verb constructions, which allow speakers to go beyond the set, large though it may be, of stored lexical sequences. For instance, *Do you need to speak French to work at Disneyland Paris?* is an utterance that is not a stored lexical sequence and thus can only have been formed on the basis of a more general *need to VP* construction, in combination with other smaller or larger constructions of a lexical or more syntactic nature (*you*, *speak French*, *Disneyland Paris*, *work at* NP, *in order to* VP, the inversion construction, etc.). If we want to adequately represent speakers' knowledge of how each of the four necessity modals discussed here are used, we should do more than just look at the lexical n-grams they are found in. One possible way of complementing our approach is to retrieve constructions towards the more syntactic pole of the syntax-lexicon continuum, for instance by finding out the preferred part-of-speech n-grams for each of the four necessity modals (cf. Cappelle and Grabar 2016). However, it will then still be necessary to perform a careful analysis of actual examples of these modals in context to get a complete picture of their semantic and pragmatic properties. Further research is therefore needed to meet this challenge.

References

Aijmer, Karen. (1996). *Conversational routines in English: convention and creativity*. New York, NY: Addison Wesley Longman.

Bergs, Alexander & Diewald, Gabriele. (2009). *Contexts and Constructions*. Amsterdam: John Benjamins.

Biber, Douglas & Conrad, Susan. (1999). Lexical bundles in conversation and academic prose. In Hilde Hasselard & Signe Oksefjell (Eds.), *Out of Corpora: Studies in Honor of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.

Boogaart, Ronny. (2009). Semantics and pragmatics in construction grammar: the case of modal verbs. In Alexander Bergs and Gabriele Diewald (Eds.), *Contexts and constructions* (pp. 213–41). Amsterdam: John Benjamins.

Boogaart, R. & Fortuin, E. (2016). Modality and mood in cognitive linguistics and construction grammars. In J. van der Auwera & J. Nuyts (Eds.), *The Oxford handbook of mood and modality* (pp. 514–533). Oxford: Oxford University Press.

Brems, Lieselotte, Lobke Ghesquière & Freek Van de Velde. (2012). *English Text Construction*, Special Issue. Intersections of intersubjectivity, 5(1).
Brown, Penelope & Stephen C. Levinson. (1987) [1978]. Politeness: Some Universals in Language Usage. Cambridge: Cambridge University Press.
Cappelle, Bert. (2017). What's pragmatics doing outside constructions? In Ilse Depraetere and Raphael Salkie (Eds.), *Semantics and pragmatics. Drawing a line* (pp. 115–151). Cham: Springer.
Cappelle, Bert & Depraetere, Ilse. (2016a). Short-circuited interpretations of modal verb constructions: Some evidence from *The Simpsons*, *Constructions and Frames*, Special Issue. Modal meaning in Construction Grammar, 8(1), 7–39.
Cappelle, Bert & Depraetere, Ilse. (2016b). 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 & Grabar, Natalia. (2016). Towards and n-grammar of English. In In: Sabine De Knop and Gaëtanelle Gilquin (Eds.), *Applied Construction Grammar* (Applications of Cognitive Linguistics) (pp. 271–302). Berlin / Boston : De Gruyter Mouton.
Carston, Robyn. 2009. The explicit/implicit distinction in pragmatics and the limits of explicit communication, *International Review of Pragmatics*, 1(1), 35–62.
Cheng, Winnie. (2007). "Sorry to interrupt, but…": pedagogical implications of a spoken corpus. In Mari Carmen Campoy & María José Luzón (Eds.), *Spoken corpora in applied linguistics* (pp. 199–215). Bern: Peter Lang.
Clark, William. 1991. *Relevance theory and the semantics of non-declarative sentences*. Ph.D. dissertation. University College London.
Coates, Jennifer. (1983). *The semantics of the modal auxiliaries*. London and Canberra: Croom Helm.
Cruse, Alan D. (2011). *Meaning in language*. (3rd edition). Oxford: Oxford University Press.
Davidse, Kristin, Lieven Vandelanotte & Hubert Cuyckens (Eds.) (2010). *Subjectification, Intersubjectification and Grammaticalization* (Topics in English Linguistics). Berlin: De Gruyter Mouton.
Davies, Mark. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Available online at https://corpus.byu.edu/bnc/.
De Haan, Ferdinand. (2012). The relevance of constructions for the interpretation of modal meaning: the case of *must*. *English Studies*, 93(6): 700–728.
Depraetere, Ilse. (2010). Some observations on the meaning of modals. In Bert Cappelle and Naoaki Wada (Eds.), *Distinctions in English Grammar, Offered to Renaat Declerck* (pp. 72–91). Tokyo: Kaitakusha.
Depraetere, Ilse. (2014). Modal meaning and lexically-regulated saturation, *Journal of Pragmatics* 71: 160–177.
Depraetere, Ilse and Raphael Salkie. (2017). Free pragmatic enrichment, expansion, saturation, completion: a view from linguistics. In Ilse Depraetere and Raphael Salkie (Eds.), *Semantics and Pragmatics. Drawing a Line* (pp. 11–38). Cham: Springer.
D'Hertefelt, Sarah. (2018). *Insubordination in Germanic. A Typology of Complement and Conditional Constructions*. Berlin: De Gruyter.
Firth, John Rupert. (1957). *A synopsis of linguistic theory 1930-1955* (Studies in Linguistic Analysis). Oxford: Philological Society, 1–32. Reprinted in Frank Robert Palmer (ed.) 1968. Selected Papers of J.R. Firth 1952-1959. London: Longman.
Flach, Susanne and Martin Hilpert (2017). From big data to small data and back again: Using token-based semantic vector spaces for corpus-linguistic analyses. Talk presented at 7th International Conference on the Linguistics of Contemporary English (BICLCE7), Vigo, Spain, 28-30 September 2017.
Fraser, Bruce. (1975). Hedged Performatives. In Peter Cole and Jerry L. Morgan (Eds.), *Syntax and Semantics* 3, (pp. 187–210). New York, NY: Academic Press.
Gyselinck, Emmeline. (2018). The role of expressivity and productivity in (re)shaping the constructional network. Ph.D. dissertation. Ghent University.

Goldberg, Adele E. (2003). Constructions: A new theoretical approach to language. *Cognitive Science*, 7(5): 219–224.
Goldberg, Adele E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Groefsema, Marjolein. (1995). *Can*, *May*, *Must* and *Should*: A Relevance Theoretic Account, *Journal of Linguistics*, 31(1), 53–79.
Harris, Zellig. (1954). Distributional structure. *Word* 10(23), 146–162.
Hilpert, Martin. (2014). *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.
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.
Hunston, Susan. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
Hunston, Susan and Gill Francis. (2000). *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
Jackendoff, Ray S. 1997. Twistin' the night away. *Language* 73, 534–559.
Janda, Laura. (2009). Linguistic profiles and construction grammar. Paper presented at the Conference "Russkij jazyk: konstrukcionnye i leksiko-semantičeskie podxody", St. Petersburg, Russia, 24-26 March 2009.
Kay, Paul. 2004. Pragmatic aspect of grammatical constructions. In: Laurence Horn and Gregory Ward (Eds.), *The Handbook of Pragmatics* (pp. 675–700). London: Blackwell.
Keckses, Istvan and Monika Kirner-Ludwig. 2018. It would never happen in my country I must say: A corpus-pragmatic study of Asian learners' preferred uses of *must* and *should*, *Empirical Pragmatics*, 1(2): 91–134.
Lee-Goldman, Russell. (2011). *Context in Constructions*. Ph.D. dissertation. University of California at Berkeley.
Leech, Geoffrey. (2014). *The Pragmatics of Politeness*. Oxford: Oxford University Press.
Levinson, Stephen (2000). *Presumptive Meanings*. Cambridge, Mass: MIT Press/Bradford Books.
Marasović, Ana, Mengfei Zou, Alexis Palmer and Anette Frank. (2016). Modal sense classification at large. paraphrase-driven sense projection, semantically enriched classification models and cross-genre evaluations. *Linguistic Issues in Language Technology* 14. http://csli-lilt.stanford.edu/ojs/index.php/LiLT/article/view/65.
McDonald, Scott and Michael Ramscar. (2001). Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 611–616.
Morgan, Jerry L. (1977). Two types of convention in indirect speech acts. Technical report No. 52. University of Illinois at Urbana-Champaign. https://www.ideals.illinois.edu/bitstream/handle/2142/17765/ctrstreadtechrepv01977i00052_opt.pdf?seque.
Nikiforidou, Kiki. (2009). Constructional analysis. In: Frank Brisard, Jan-Ola Östman and Jef Verschueren (Eds.), *Grammar, Meaning and Pragmatics* (pp. 16–32). Amsterdam: John Benjamins.
Nuyts, Jan. (2012). Notions of (inter)subjectivity. *English Text Construction*, Special Issue. Intersections of intersubjectivity, 5(1), 53–76.
Recanati, François. (2004). *Literal Meaning*. Cambridge: Cambridge University Press.
Recanati, François. (2010). *Truth-conditional pragmatics*. Oxford: Clarendon.
Recanati, François. (2012). Pragmatic enrichment. In Gillian Russell and Delia Graff Fara (Eds.), *The Routledge Companion to Philosophy of Language* (pp. 67–78). New York: Routledge.
Rivière, Claude. (1981). Is *should* a weaker *must*? *Journal of linguistics* 17: 179-195
Ruppenhofer, Josef and Ines Rehbein. (2012). Yes we can!? Annotating English modal verbs. In *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'12).
Ruytenbeek, Nicolas. (2017). The comprehension of indirect requests: Previous work and future directions. In Ilse Depraetere & Raphael Salkie (Eds.)*, Semantics and Pragmatics. Drawing a Line* (pp. 293–322) Cham: Springer.

Sahlgren, Magnus. 2008. The Distributional Hypothesis. *Rivista di Linguistica* 20(1): 33–53.
Searle, John R. (1975). Indirect Speech Acts. In Peter Cole and Jerry L. Morgan (Eds.), *Syntax and Semantics. Vol. 3, Speech Acts* (pp. 59–82). New York: Academic Press.
Stefanowitsch, A. (2003). A construction-based approach to indirect speech acts. In Klaus-Uwe Panther and Linda L. Thornburg (Eds.), Metonymy and pragmatic inferencing (pp. 105–126). Amsterdam: John Benjamins.
Traugott, Elizabeth Closs. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. In Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (Eds.), *Subjectification, Intersubjectification and Grammaticalization* (Topics in English Linguistics) (pp. 29–71). Berlin: De Gruyter Mouton.
Traugott, Elizabeth Closs and Ekkehard König. (1991). The semantics-pragmatics of grammaticalization revisited. In Elizabeth Closs Traugott and Bernd Heine (eds.), *Approaches to Grammaticalization*. Vol. 1, 189–218. Amsterdam: Benjamins.
Trousdale, Graeme. 2008. A constructional approach to lexicalization processes in the history of English: Evidence from Possessive Constructions. *Word Structure* 1(2): 156–177.
Verhagen, Arie. (2005). *Constructions of Intersubjectivity*. Oxford: Oxford University Press.
Verstraete, Jean-Christophe and Sarah D'Hertefelt. (2016). Running in the family. Patterns of Complement Insubordination in Germanic. In Nicholas Evans and Honoré Watanabe (eds.), *Dynamics of Insubordination* (pp. 65–88). Amsterdam: Benjamins.
Xiao, Richard and Tony McEnery. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics* 27(1): 103–129.

Appendix 1. Overview of n-grams extracted from the BNC

1. Intersection *have to* - *must* - *need to* - *should*: 20 n-grams
MOD be able, MOD be considered, MOD be done, MOD be made, MOD be taken, MOD ask, MOD be, MOD consider, MOD go, MOD make, MOD take, MOD write, MOD learn, you MOD go, you MOD make, you MOD, we MOD look, we MOD, we MOD make, we MOD get,

2. Intersection *have to* - *must* - *should*: 17 n-grams
MOD accept, MOD bear, MOD be careful, MOD be carried, MOD be seen, MOD be kept, MOD be paid, MOD carry, MOD choose, MOD decide, MOD give, MOD leave, MOD pay, MOD prove, MOD stay, MOD stop, we MOD go

3. Intersection *must* - *need to* - *should*: 6 n-grams
MOD be aware, MOD be given, we MOD consider, MOD develop, MOD ensure, MOD ensure that

4. Intersection *have to* - *must* - *need to*: 2 n-grams
MOD find, MOD keep

5. Intersection *have to* - *should*: 4 n-grams
MOD explain, MOD send, MOD stand, MOD wear

6. Intersection *have to* - *must*: 19 n-grams
MOD admit, MOD agree, MOD be able to, MOD confess, MOD make sure, MOD remember that, I MOD say, MOD tell, MOD go back, MOD pass, MOD rely, MOD say, MOD tell you, I MOD tell, I MOD tell you, you MOD remember, you MOD take, MOD wait, we MOD find

7. Intersection *have to* - *need to*: 20 n-grams

doesn't MOD, didn't MOD, don't MOD, 'll MOD, may MOD, -n't MOD, shall MOD, won't MOD, would MOD, will MOD, you don't MOD, you MOD do, you MOD do is, you may MOD, you MOD get, MOD buy, MOD get, MOD look at, MOD move, MOD worry

8. Intersection *must* - *should*: 65 n-grams
MOD add, MOD allow, MOD also, MOD also be, MOD always, MOD always be, MOD apply, MOD avoid, MOD be added, MOD be addressed, MOD be based, MOD be allowed, MOD be borne, MOD be borne in, MOD be borne in mind, MOD be capable, MOD be established, MOD be exercised, MOD be included, MOD be interpreted, MOD be maintained, MOD be noted, MOD be obtained, MOD be placed, MOD be provided, MOD be recognized, MOD be regarded, MOD be remembered, MOD be sent, MOD be stressed, MOD be prepared, MOD be treated, MOD begin, MOD contain, MOD continue, MOD continue to, MOD expect, MOD follow, MOD have, MOD have been, MOD have done, MOD have gone, MOD have known, MOD have seen, MOD have taken, MOD have thought, MOD include, MOD involve, MOD provide, MOD recognize, MOD remain, MOD seek, MOD take place, MOD therefore, MOD try, MOD try to, applicants MOD, candidates MOD, care MOD be, care MOD be taken, user MOD, they MOD, we MOD also, we MOD take, we MOD try

9. Intersection *must* - *need to*: 4 n-grams
you MOD know, MOD understand, MOD remember, MOD look

10. Intersection *need to* - *should*: 3 n-grams
think we MOD, I think we MOD, MOD check

11. *should*: 171 n-grams
attention MOD be, authorities MOD, client MOD, consideration MOD, consideration MOD be, consideration MOD be given, courts MOD, don't think we MOD, everyone MOD, gentleman MOD, hon gentleman MOD, how they MOD, I don't think we MOD, I MOD be grateful if, I MOD imagine, I MOD think, I MOD think so, I think they MOD, I think you MOD, it MOD be noted that, local authorities MOD, managers MOD, maybe we MOD, MOD adopt, MOD advise, MOD aim, MOD aim to, MOD attend, MOD be adopted, MOD be applied, MOD be assessed, MOD be available, MOD be avoided, MOD be banned, MOD be brought, MOD be carefully, MOD be checked, MOD be completed, MOD be consulted, MOD be dealt, MOD be directed, MOD be discussed, MOD be doing, MOD be drawn, MOD be emphasised, MOD be encouraged, MOD be entered, MOD be entitled, MOD be examined, MOD be excluded, MOD be extended, MOD be fitted, MOD be granted, MOD be grateful, MOD be grateful if, MOD be informed, MOD be introduced, MOD be judged, MOD be noted that, MOD be permitted, MOD be pointed, MOD be ready, MOD be reduced, MOD be referred, MOD be removed, MOD be replaced, MOD be restricted, MOD be retained, MOD be returned, MOD be sought, MOD be spent, MOD be sufficient, MOD be taught, MOD be undertaken, MOD be used, MOD be viewed, MOD bear in, MOD bear in mind, MOD concentrate, MOD concentrate on, MOD consult, MOD contact, MOD cover, MOD die, MOD discuss, MOD draw, MOD enable, MOD encourage, MOD enjoy, MOD enter, MOD focus, MOD happen, MOD have known better, MOD have told, MOD have won, MOD ideally, MOD imagine, MOD inform, MOD n't, MOD never, MOD never be, MOD never have, MOD never have been, MOD normally, MOD not, MOD not be, MOD not expect, MOD note, MOD note that, MOD occur, MOD pay for, MOD proceed, MOD receive, MOD reduce, MOD refer, MOD reflect, MOD seek to, MOD serve, MOD specify, MOD stick, MOD therefore be, MOD think, MOD treat, MOD write to, -n't think we MOD, no reason why you MOD, nor MOD, patient MOD, perhaps we MOD, perhaps you MOD, procedures MOD, pupils MOD, pupils MOD be, purchaser MOD, reader MOD, readers MOD, reason why they MOD, reason why you MOD, reference MOD, reference MOD be, so why MOD, teachers MOD, tenant MOD, that they MOD, that they MOD be, the drafter MOD, think they MOD, think we MOD be, think you MOD, we MOD be, we MOD be able, we MOD be able to, we MOD expect, we MOD have, we MOD not, whether they MOD, whether we MOD, why MOD, why MOD I, why MOD n't, why MOD n't I, why MOD we, why they MOD, why we

MOD, why you MOD, you MOD always, you MOD ask, you MOD n't, you MOD n't have, you MOD never, you MOD try

12. *Must*: 78 n-grams
care MOD, I MOD admit, I MOD ask, I MOD go, if you MOD know, it MOD be admitted, it MOD be remembered, it MOD be remembered that, it MOD be stressed, it MOD be stressed that, MOD address, MOD admit I, MOD also have, MOD assume, MOD be accepted, MOD be accompanied, MOD be accompanied by, MOD be admitted, MOD be based on, MOD be capable of, MOD be joking, MOD be mad, MOD be prepared to, MOD be regarded as, MOD be remembered that, MOD be satisfied, MOD be seen as, MOD be stressed that, MOD be understood, MOD come, MOD comply, MOD comply with, MOD decide whether, MOD depend, MOD exist, MOD have felt, MOD have heard, MOD have looked, MOD have seemed, MOD inevitably, MOD learn to, MOD lie, MOD make a, MOD meet, MOD necessarily, MOD realize, MOD satisfy, MOD speak, MOD surely, MOD surely be, MOD surely have, MOD take into, plaintiff MOD, somebody MOD, someone MOD have, there MOD, there MOD be, there MOD be some, there MOD be something, there MOD have, there MOD have been, they MOD also, they MOD be, they MOD have, they MOD have been, we MOD ask, we MOD ensure, we MOD ensure that, we MOD keep, we MOD remember, we MOD remember that, you MOD be, you MOD be joking, you MOD come, you MOD have, you MOD have been, you MOD provide, you MOD tell

13. *Have to*: 88 n-grams
MOD wait for, MOD admit that, you MOD keep, she didn't MOD, 're gonna MOD, MOD struggle, don't MOD worry about, MOD fill, MOD make do, MOD contend with, will MOD be made, MOD face, MOD get up, MOD give up, MOD cope, you won't MOD, MOD turn, MOD put up, MOD pull, MOD run, 'll MOD find, MOD live with, MOD compete, gonna/wanna MOD, you MOD pay, MOD travel, she MOD, we MOD do, MOD listen, MOD suffer, so we MOD, you MOD put, MOD put up with, she MOD get, MOD walk, MOD be found, MOD be taken into, MOD go into, 'm gonna MOD, what we MOD do, we MOD find, -n't MOD worry about, all you MOD do is, MOD resort, you MOD look, you MOD say, MOD deal, MOD sell, MOD accept that, she MOD admit, you don't MOD worry, I MOD confess, MOD spend, 'll MOD ask, MOD go out, MOD endure, gonna MOD, MOD sit, MOD offer, MOD be content, MOD fight, I shall MOD, MOD get out, MOD put, MOD cope with, MOD listen to, MOD resort to, MOD decide whether, MOD tell me, MOD call, MOD settle, MOD wait until, she MOD go, MOD do something , someone MOD, MOD contend, he MOD say, 'd MOD, it doesn't MOD, MOD watch, MOD go through, MOD worry about, you don't MOD go, it doesn't MOD be, MOD deal with, we shall MOD, MOD rely on, MOD live

14. *Need to*: 23 n-grams
all you MOD do, all you MOD know, may MOD be, MOD know, MOD know about, MOD know how, MOD know what, MOD talk, MOD use, really MOD, we don't MOD, we MOD know, we MOD look at, we'll MOD, what MOD be done, will MOD be, would MOD be, you MOD know about, you will MOD, you'll MOD