

EPISTEMIC *MUST*

Ferdinand de Haan
University of Arizona

1. introduction

It has long been assumed that the English core modal verbs form a more or less homogeneous group of verbs. In formal frameworks they are usually grouped together under an AUX-node (or a more updated version thereof). While it is clear that verbs such as *must*, *may*, *can*, *should* or *will* are closer together both syntactically and morphologically than later modal verbs like *have to*, *supposed to* or *want to* (due to the fact that the core modals were grammaticalized earlier), these similarities must not obscure their differences. This paper is devoted to an analysis of one modal of the core modal group, namely *must*. It will be shown that the uses of *must* differ per register (in the sense of Biber 19xx). While in written registers the deontic use of *must* prevails, in spoken conversations it is the epistemic sense of *must* which is by far the most frequent.

With the availability of large computer-based corpora, it has become possible to investigate individual modal verbs in greater detail (and faster) than was previously possible. The first such study is Ehrman (1966) who gave an analysis of the (American) modal system, based on a subset of the Brown corpus (Kucera ...). Other influential studies are Palmer (1979) and Coates (1983), both working on British English corpora. The most recent study is the section on modality in the Longman Grammar of Spoken and Written English (Biber et al. 1999), which is based on samples of written and spoken English, both British and American.

2. definitions

Most of the core modals have two separate meanings, which will be referred to here by their classical names *epistemic* and *deontic* modality. Deontic modality refers to the obligation or permission the subject of the sentence has to perform the action described, while epistemic modality deals with the degree of confidence a speaker has in his/her statement. Thus, a sentence such as *John must work* is ambiguous between a deontic or an epistemic interpretation. The deontic interpretation is something like *X wants for John to work*, where the subject, John, has an obligation, imposed by someone else or by an external force, such as a law or a custom. The epistemic interpretation involves the speaker, because the statement is made on his/her authority and goes something like *I firmly believe that John has a job*. Similar analyses can be given for most other core modals.

In recent years, other terms have been adopted by some scholars, e.g., *root* modality (Coates 1983) or *agent-oriented* modality (Bybee et al. 1994) for deontic modality, but these terms suffer from various defects (such as vagueness or impreciseness) and for these reasons the term deontic will be used here. This is not to say that the term *deontic* is problem-free (see De Haan, forthcoming), but it does have the advantage that it is most widely known. The term *epistemic* is often confused or used interchangeably with the term *evidentiality*. This is especially the case with a modal like

must, which appears at first glance to have evidential interpretations. The view adopted here (and defended in section ... below) is that *must* is not evidential, but rather purely epistemic. This sets *must* apart from its cognates in German and Dutch (*müssen* and *moeten*, respectively), which can have both an evidential and an epistemic interpretation. In other words, epistemic modality and evidentiality are essentially separate domains and should not be conflated (see also De Haan 1999, 2001 for further discussion).

In certain cases, it is not possible to determine, even within context, whether a given modal is epistemic or deontic. The classic example is the exchange of (7), from Coates (1983:17):

- (7) A: Newcastle Brown is a jolly good beer.
B: Is it?
A: Well it ought to be at that price.

In this exchange, the modal *ought* in the third sentence can be interpreted as either a deontic modal (the brewers of Newcastle Brown have an obligation to put out a good beer given the high price) or an epistemic modal (“It costs a lot, therefore it is good”). Such cases are called *indeterminate* cases by Coates (1983) and this term will be maintained here.

Even though a modal verb such as *must* can be analyzed as being inherently ambiguous, this is not necessarily so. A sentence such as *John must have been working* essentially forces an epistemic interpretation and any attempt to force a deontic interpretation would require a rather strained context. Such a context never seems to appear in natural language corpora and hence the structure of such sentences yields important clues regarding how modal interpretations are assigned to sentences.

This paper looks at the sentences containing *must* in two corpora, the Switchboard Corpus, which is a corpus of telephone conversations from the American South and Southwest, i.e. a spoken language corpus, and the Brown Corpus, a corpus of written American English. The corpora are used in the form provided on the Treebank 2 CD-ROM as distributed by the Linguistic Data Consortium (LDC 1999). These corpora were used because of their availability in annotated form, both syntactically in the Treebank format and tagged with part-of-speech information (see ...). The use of the Brown corpus evokes memories of the first corpus study of English modals, Ehrman 1966, who also used the Brown corpus, which at that time was not yet fully available electronically, and Ehrman used a subset of the entire corpus, a sample of 300,000 words from the one million total.

This paper is a synchronic study, but the application of it to diachronic research is

Corpus data

We start by looking at the raw data from both corpora. As far as raw numbers are concerned, it is clear from Table A that the distribution of modality in the corpora is almost a mirror-image of each other. In the Switchboard corpus, almost 80% of the occurrences of *must* is epistemic while deontic modality has a similar percentage in the Brown corpus. The number of indeterminate cases is identical percentage-wise, while

ungrammatical/irrelevant cases account for a negligible number. In the case of the Switchboard data, it mainly concerns cases of the type “a must-V N”, as in (...):

(...) ...those are two definite must see movies, I think.

The few cases in the Brown corpus are similar, plus a number of cases where another author is cited. Such a case is (...), where the 17th century author Jonathan Swift is quoted.

(...) In *The Publick Spirit of the Whigs*, it may be noted, Swift himself contemptuously dismissed Steele's reference to his friend at court: “I suppose by the Style of old Friend, and the like, it must be some Body there of his own Level ...” [B445]

Even though the sentence could pass for contemporary English, it was thought best to disregard this and similar cases!

Table A
Distribution of modality in both corpora

	Switchboard		Brown	
	Tokens	Percentage	Tokens	Percentage
<i>Deontic</i>	66	12.7	762	77.1
<i>Epistemic</i>	412	79.2	159	16.1
<i>Indeterminate</i>	31	6.0	60	6.1
<i>Ungrammatical or irrelevant</i>	11	2.1	7	0.7
Total	520	100.0	988	100.0

A comparison with other corpus linguistic studies on *must* reveals that the raw results of Table A do not significantly deviate from the results of these other studies.

In Coates' 200 token sample of the Survey of English Usage (Coates 1983:31-2), epistemic use was slightly less common than deontic use (92 versus 106 occurrences, respectively, with 2 indeterminate). In her other corpus, the 1 million Lancaster corpus, deontic usage is the common interpretation of *must* by about a 2:1 ratio (153 deontic, 74 epistemic and 9 indeterminate).

The LGSWE does not give exact figures, but from the histograms provided (p. 494) it can be seen that epistemic *must* is more common than deontic *must* by a factor of 3 in the conversation register, while deontic *must* is more common in the register of academic prose by about a 4:3 ratio (no specific data are provided for the other registers). Overall, *must* occurs about 800 times per million words in the conversation register, and about 1,200 times in the academic register (p. 489).

In Australian English, the same phenomenon can be observed. Collins (1991:154) reports that epistemic *must* is the prevalent usage in conversation. Of 104 occurrences of *must* in spoken Australian English, 88 were epistemic, and only 16 deontic. As expected, this situation is reversed in the written register: of the 72 occurrences of *must* in writing, 47 were deontic, 18 epistemic, and 7 indeterminate. The fact that *must* is relatively rare in

writing in Australian English means that *must* is the only modal in Australian English which is used more epistemically than deontically. It must be said, though, that the overall corpus was relatively small, especially by current standards, namely 225,000 words in total.

While there are some studies done on the occurrence of the modals in the British National Corpus, none deal with the specific topic of this study. Kennedy (2002) is a detailed analysis of the modals in the BNC, including data on the occurrence of modals with specific verbal constructions, but no differentiation is made between epistemic and deontic uses of the modals. Overall, *must* occurs about 72,000 times in the BNC, which is 4.9% of all modals. In the spoken part of the BNC, *must* has a frequency of 2.9% of all modals, while the percentage goes up to 5.3% in the written part (Kennedy 2002:77).

Correlation of person with modality

The next couple of sections of this study are devoted to a number of features that are thought to play a large role in the disambiguation of modals. This section deals with the correlation of person/subject and modality. It will be argued that this feature, alone and in combination with the verbal construction of *must* (discussed in section ...), is a significant factor in determining the modal interpretation.

Table B summarizes the data for this feature. The table gives the data for each possible person, including those sentences with impersonal subjects and sentences without subjects. Data for first and third persons were split according to number as well. Since the second person pronoun is ambiguous between singular and plural (and forms like *y'all* or *you guys* did not occur with *must*), the forms for the second person are not differentiated according to number.

Table B
Correlation of person and modality in the Switchboard corpus

	Deontic	Epistemic	Indeterminate
<i>1 SG</i>	39	14	1
<i>2 SG/PL</i>	10	67	5
<i>3 SG</i>	6	166	13
<i>1 PL</i>	2	13	1
<i>3 PL</i>	7	59	10
<i>No overt subject</i>	1	29	0
<i>Impersonal subject</i>	1	64	1
Total	66	412	31

While 62.1% (41/66) of all occurrences of deontic *must* involve a first person, only 6.6% (27/411) of all cases of epistemic *must* do. Conversely, 19.7% of all deontic *must* instances (13/66) have a third person, and 54.3% of epistemic *must* sentences do (223/411).

Also striking is the correlation of epistemic modality with sentences without overt subject or with an impersonal subject. This would seem to have a ready explanation in the fact that deontic modality by its nature needs a subject to serve as the entity acted upon,

while epistemic modality acts on the proposition as a whole and does not require a subject, overtly expressed or not. The lone deontic example in the data is:

(...) I think there ought to be a unanimous, uh, vote because by law and by constitution there must be proven beyond a shadow of a reasonable doubt that the person in question did this.

Table C shows the Brown correlation of modality and person.

Table C
Correlation of person and modality in the Brown corpus

	Deontic	Epistemic	Indeterminate
<i>1 SG</i>	23	2	2
<i>2 SG/PL</i>	26	9	5
<i>3 SG</i>	393	102	40
<i>1 PL</i>	101	0	0
<i>3 PL</i>	151	22	8
<i>No overt subject</i>	27	7	1
<i>Impersonal subject</i>	41	17	4
Total	761	159	60

In most cases the distribution of modality according to person in the Brown corpus is the reversal of the situation in the Switchboard corpus. The only case in which the preferred interpretation stays the same is first person singular, which is the only person in the Switchboard corpus with a preferred deontic interpretation. In all other cases, including that of sentences with an impersonal subject, the preferred interpretation is the opposite from that of the other corpus.

Coates (1983:37-8) found a strong correlation between deontic modality and first person subject, with a 92% correlation in the SEU corpus (and 83% in the Lancaster corpus). The present study cannot confirm that; while the correlation is indeed strong in the Brown data (even 100% for 1PL, 92% for 1SG), in the Switchboard data the correlation is weaker for 1SG (72%), and the correlation is reversed for 1PL subjects (87% epistemic). Coates also notes a correlation between second person and deontic modality in the Lancaster corpus (87%), but only a weak correlation in the SEU corpus (62%). In the Brown corpus, the correlation is very weak (71%), and in the Switchboard data the correlation runs the other way (87% epistemic).

In the remainder of this section, I will discuss two cases in more detail, because they provide interesting results and, in the case of the first person singular, because the data have relevance for current research into other areas of grammar, in this case evidentiality.

Impersonal subjects

This section deals with those constructions that have an impersonal subject. It may be supposed that impersonal subject have a tendency to select an epistemic interpretation because of the propositional nature of epistemic modality, and because of the fact that

deontic modality usually acts on subjects. Nevertheless, sentences with an impersonal subject have a preferred deontic interpretation in the Brown data, in line with the general preference of deontic modality in this corpus.

Deontic sentences

Impersonal sentences with a deontic modality fall into three groups:¹

- *it must be + passive* (20 cases)
- *there must be + N* (13 cases)
- *one must + V* (9 cases)
- *there must be + passive* (1 case, example (...) above)

Examples are (all from the Brown corpus):

a. *one must + V*

(...) Statistically this has represented a loss to the nation, although one must admit that in an individual case the decision in retrospect may have been a wise one.

(...) On the whole, however, one must wonder at just what it is that forces a beloved artist to besmirch her own reputation as time marches inexorably on.

(...) But to operate in American politics one must go a step further -- one must build a bridge to such names, establish a warmth, a personal connection.

b. *it must be + passive*

(...) But, with all due respects and allowances, it must truthfully be said that what they heard was more syrupy than sweet, more mannered than musical.

(...) If the content of faith is to be presented today in a form that can be “understood of the people” -- and this, it must not be forgotten, is one of the goals of the perennial theological task -- there is no other choice but to abandon completely a mythological manner of representation .

c. *there must be + N*

(...) In a pessimistic assessment of the cold war, Eden declared: “There must be much closer unity within the West before there can be effective negotiation with the East”.

(...) For an experiment to qualify as a binomial experiment, it must have four properties: (1) there must be a fixed number of trials, (2) each trial must result in a “success” or a “failure” (a binomial trial), (3) all trials must have identical probabilities of success, (4) the trials must be independent of each other.

The deontic nature of the (c) group can be understood by viewing the impersonal subject *one* as a placeholder for the real subject, which is, based on the sentences (...) – (...), usually the first person or, in case of (...) a generic person. The use of *one* in examples (...) and (...) instead of the pronoun *I* is due to the desire of the speaker to distance him/herself from the statement he/she is making and at the same time trying to make the statement more inclusive, more relevant to other people as well. Thus, in (...) the speaker wonders about something and at the same time believes that other people may be wondering about the same thing. The use of *one* in sentence (...) is thus partly a distancing device, as in (...), and partly a device of genericness, as in (...).

The use of impersonal *it* with the passive can be explained in similar ways. Biber *et al.* (1999:500) view the use of the passive “... to avoid explicit identification of the person who is obliged to act.” When this person can be identified as the speaker (i.e., first person), as is likely the case in (...), the result is that the passive and the impersonal *it* act as a distancing device to add a degree of softening in the statement.

Finally, the construction with the existential *there*, *there must be + N*, differs from the previous two in that this construction is used when there is no direct agent for the obligation to act on, but rather when there is an obligation for some conditioning factor to exist before some other condition is met. Thus, in example (...) the requirement is that the West is united before negotiations with the East are effective; in (...), before something can be called a binomial experiment, the requirement that there be a fixed number of trials has to be true.

Epistemic sentences

Very interestingly, the constructions used in impersonal sentences with *must* in an epistemic reading differ from those sentences that have a deontic interpretation. To begin with, impersonal sentences with the impersonal pronoun *one* and the modal *must* do not occur at all. This is probably due to the low frequency of impersonal *one* in conversation (for data see Biber *et al.* 1999:354-5). Sentences with *it* or *there* have different complements as well. The raw numbers are shown in Table D:

Table D
Impersonal constructions with an epistemic meaning

Construction	Switchboard	Brown
<i>There must be + N/Adj</i>	24	4
<i>It must be + N/Adj</i>	24	3
<i>It must have been</i>	7	3
<i>There must have been</i>	5	2
<i>It must + V</i>	1	0
<i>It must have V-ed</i>	1	2
<i>It must be V-ing</i>	1	1
<i>That must be</i>	1	0
<i>Somewhere must be</i>	0	1
Total	64	16

Thus, the most common epistemic impersonal constructions involve copula constructions and these are the most common in either corpus. Examples are shown below:

a. *there must be* + *N/Adj*

(...) Quite often, honeybees form a majority on the willow catkins. As we have already seen in the first chapter, bumblebees are bigger, hairier, and much more colorful than honeybees, exhibiting various combinations of black, yellow, white and orange. Let us not try to key them out at this stage of the game, and let us just call them *Bombus*. There must be several dozen species in the United States alone. If you really insist on knowing their names, an excellent book on the North American species is *Bumblebees And Their Ways* by O.E. Plath. [B646]

(...) ... Uh-huh. It's in, in an area that's within, you know, four or five blocks. There must be forty restaurants. I'm not kidding. [S266]

b. *it must be* + *N/Adj*²

(...) It must be that son who sat before her now, shriveled to half his size and half his senses. [B957]

(...) ... or, uh, inferior quality when the people involved really aren't inferior, so it must be the way we're managing the process ... [S396]

(...) It must have hurt her even to walk, for the sole was completely off her left foot and Morgan saw that it was bruised and bleeding . [B924]

In Coates' study, epistemic *must* and the existential subject *there* are strongly correlated (1983:43), in the SEU corpus, 86% of the sentences with *there must* are epistemic, and in the Lancaster corpus all such sentences are epistemic. As can be seen from the data in this section and the previous one, all but one example of *must* with the existential *there* are epistemic (24/25), but the majority of such cases in the Brown corpus are deontic. There is a difference between epistemic and deontic existential *there* sentences, however. One, the construction *there must have been* can only have an epistemic interpretation. This is due to the fact that the construction *must have been* is only interpreted epistemically (see section ... below) and not to the presence of existential *there*.

First person singular

Within modality in general, first person singular occupies a peculiar role. As shown in Tables ... and ... above, it is the only person that is more commonly associated with deontic interpretations than with epistemic interpretations, irrespective of register. The reason for this is due to the fact that in general, people are not usually uncertain about actions or states they have witnessed themselves, as sentences with a first person subject generally are. On the other hand, it is quite normal to describe obligations about oneself, so that first person singular deontic sentences are not at all unusual. This asymmetry is

found in other areas as well. For instance in evidential systems it is quite often the case that first person singular is treated differently from other persons, especially in indirect evidential constructions. Because indirect evidential constructions tend to signify that the speaker was not present when the action occurred, there is a mismatch when the subject itself is also the speaker. Since you typically cannot talk about actions you yourself performed with an evidential that signifies that you did not in fact witness it, languages typically leave a gap here in the evidential system. This is analogous to the situation in epistemic modality, where first person singular is less often associated with this interpretation.

That is not to say that first person singular does not occur with *must* in its epistemic sense at all.³ As can be seen from the data, it occurs at least 16 times in the combined corpora, twice in the Brown corpus, and 14 times in the Switchboard corpus, still accounting for about 21% of all cases.

Of the fourteen cases of epistemic *must* in the Switchboard corpus, eight have the construction *must have V-ed*, as in:

(...) And I mean, I did stuff but, to me it feels like I must have done nothing. [S157]

(...) Uh-huh, I did notice, uh, this Wednesday when I put them out, I must have gotten them out late because they didn't pick them up. [S206]

This construction is used to describe the speaker's evaluation of past events. The use of this construction in (...) signifies that the speaker draws a tentative conclusion (the taking out of the garbage occurred after the regularly scheduled garbage collection) from evidence (the garbage was not collected). Since there are other possibilities for the
A similar reasoning can be given for (...).

This is not the only use of *must* in these sentences, however. Consider example (...):

(...) Sure, but, uh, I had the car, it's an eighty-one Toyota Corona. I must have worked on it, like every third day ... [S391]

The speaker is recounting an event from his personal experience. Unlike sentence (...) this example was witnessed/experienced by the speaker directly and the presence of *must* cannot be explained by drawing a conclusion based on evidence. It is also hard to see any epistemic value in the usage, because the action itself is not in doubt. Rather, the modality concerns the measurement *every third day* (which is accentuated by the use of *like*). The speaker is unsure about exactly at what times he worked on the car, but that he worked on the car is certain to him. What probably also plays a role is the desire on the part of the speaker to exaggerate a little bit, and use the time measurement as an expression of *a lot*. Consequently, *must* in this sentence has more of a pragmatic role than a purely semantic one. It is used to liven up the conversation and possibly to engage the hearer more in the conversation. The semantic value of *must* is quite low, as evidenced by the fact that it can be safely omitted without any loss of content (partly also due to the

presence of *like*). This use of *must* may be typical when measurement phrases are involved. The other example from the Switchboard corpus is shown in (...):

- (...) - And I fished until three o'clock in the morning, and I must have caught
twenty salmon.
- Wow.
- And it was just great. [S427]

Example (...) is part of a conversation and the hearer's engagement in the action is evident. Again, the semantic value of *must* is low, and concerns not the action, but rather the result of said action, the caught fish.

The other six cases of epistemic *must* with a first person singular subject are rather equally divided among *must be V-ing* (2 examples, see (...) below), *must have* (2 examples), and one each of *must be* and *must get*.

- (...) That's the only place I was sore, and I thought, well, I must not be doing them right ... [S272]

There are only two examples of epistemic *must* with 1SG subject in the Brown corpus. They are:

- (...) "But I have no permission to re-enter France, and I have just left", I told him. "I must then be standing on the line between France and Germany." [B462]

- (...) I must have written to say how much I had enjoyed his fine book "The Building of Eternal Rome", and I found he had not regretted giving me the highest mark in his old course on the later Latin poets, ... [B502]

Sentence (...) shows a deduction from evidence. The speaker has just been told that he had left France, but also that he had not yet entered Germany. He then draws from this the conclusion shown in (...). Sentence (...) can likely be explained by the fact that the speaker had forgotten the letter he is referring to in the statement, and thus he makes a conjecture based on evidence and personal knowledge. These two examples are therefore genuine examples of epistemic modality without any additional pragmatics.

The sentences with a first person singular subject and a deontic interpretation are overwhelmingly of the shape *I must V*, where V is any main verb. 36 of 39 examples in the Switchboard corpus, and all 23 examples in the Brown corpus are of this nature. There is however a marked difference between the main verbs used in both corpora. The Switchboard corpus' use of this construction is basically limited to tag questions with three verbs, *admit* (19 occurrences), *say* (14 occurrences), and *confess* (one occurrence). This pattern is also observed by Coates (1983:38).

Examples are:

- (...) That's great. But, um, those things are rare, I must say. [S250]

(...) Yeah, I do some of that and I, I must admit that I do a certain amount of it. [S311]

(...) But I must, I must confess the only control I have over our budget is ... [S485]

By contrast, only one epistemic example with the *I must V* construction occurs in the Switchboard corpus, namely (...):

(...) But, you know, people, it's like your name must be on a list or something, because I know at my office, I get calls, I mean, I must get at least, I probably average at least a call a day, ... [S407]

One example of deontic *must* but not followed by a main verb, is (...) below, where the main verb is simply omitted:

(...) ... and I'm on, uh, I, everybody says I don't need to lose weight, but I feel I must. [S177]

The other examples are shown in (...) and (...). In (...), where the construction *must have* has a deontic interpretation, helped by the context of the verb *regulate*. Here, *have* functions more like a main verb of possession than an auxiliary. Sentence (...) shows a case of an external requirement (an exam) which forces the speaker to make a request of the hearer.

(...) What, they should also regulate how much sleep I must have. [S441]

(...) Right, I have an exam. I must have you baby-sit. [S480]

In the Brown corpus, a wider variety of verbs is used, as is to be expected. The verb *admit* is used only once, *say* twice, while the construction *I must tell you* occurs three times (although twice in the same sentence). Other verbs used are *state*, *speak*, *plead guilty* (...), *talk*, *make clear* (...).

(...) I must plead guilty to a special sympathy for nomias. [B649]

(...) To you, for instance, the word innocence, in this connotation, probably retained its Biblical, or should I say technical sense, and therefore I suppose I must make myself quite clear by saying that I lost – or rather handed over – what you would have considered to be my innocence two weeks before I was legally entitled, and in fact by oath required, to hand it over along with what other goods and bads I had. [B429]

There are a number of constructions which do not involve this subjective, self-imposed obligation. An example is (...), where the obligation is objective, the inevitable end result of a chain of events.

(...) Miriam Noel Wright said, "Here I am at my own home, locked out, so I must

stand in the road!!”

[B426]

The majority of cases are like (...)-(...) above, and so the majority of deontic sentences with 1SG subject can be considered the equivalents of the impersonal sentences discussed in section ... above. The speaker uses deontic *must* to express an internal obligation to perform a certain action. This use of deontic *must* is in some sense superfluous because its typical use is as tag clause (...) or matrix clause of a more important embedded sentence (...). Because it can be omitted without losing any semantic information, its function in these cases is a pragmatic one.

There is thus a large dichotomy between sentences with epistemic and deontic *I must*. Whereas constructions with a main verb point almost exclusively to a deontic interpretation, constructions with auxiliary verbs point to epistemic usage. This intersects with the preferred modal interpretation of such constructions in general, to be discussed in the next section.

The overall conclusion that can be drawn from the examination of the subject of *must*-sentences is that in general, the choice of modality is largely dependent on the register in which these sentences occur. The spoken register has predominantly epistemic utterances, while the written register has more deontic utterances. This is a tendency that can be observed with virtually every subject type. The only exception to this generalization is 1SG, for which the preferred modality is deontic in both registers. When the subject is identical to the speaker, the choice of modality can be judged from the verbal construction with which *must* cooccurs.

Occurrences of modality and construction

This section examines the relationship between the choice of modality and the verbal construction with which the verb *must* cooccurs. It will be seen that the choice of modality is partly dependent on the verbal construction, but also on the register involved. Table E shows the distribution of verbal constructions in the Switchboard corpus. Ten different constructions are involved.

Table E
Occurrences of *must* with verbal complements in the Switchboard corpus

	Deontic	Epistemic	Indeterminate
<i>Be + V- part.</i>	4		1
<i>V</i>	54	55	7
<i>Have</i>	3	34	4
<i>Have been</i>		70	
<i>Be</i>	2	195	5
<i>Be + V-ing</i>		9	1
<i>Have got + V-ed</i>		2	
<i>No V</i>	3	5	13
<i>Have + V-ed</i>		41	
<i>Have been + V-ing</i>		1	

A striking thing about the data shown in Table D is the fact that *must* is used in far more constructions in its epistemic sense than in its deontic sense (9 epistemic constructions as opposed to 5 deontic ones). In addition, there are four constructions that have the possibility of being analyzed both epistemically and deontically, one which is only deontic, and 5 which are only epistemic.

Of the four ambiguous constructions, two are heavily skewed toward epistemic uses, namely *must have* and *must be*, with three and two deontic examples, respectively, which means that there are only two constructions which are genuinely ambiguous. These are the constructions *must* + main verb, and *must* without any other verb present. These are also unsurprisingly the constructions with the largest number of indeterminate cases.

The cases with *must have* and *must be* involve objective deontic modality, such as (...), which involves an objective requirement, as it states a lawful rule or regulation.

(...) ... and you must be twelve years of age to go in and not more than two people at a time. [S172]

Two cases with *must have* were already discussed above (... - ...). The third case is shown in (...), where the speaker states a general rule.

(...) Yeah, and, but I believe the, a child must have a pet to grow up with ... [S307]

Another frequently observed phenomenon is the cooccurrence of epistemic *must* with the Progressive. Coates (1983:44-5) reports a 100% score for epistemic modality and the progressive in the SEU corpus and the Lancaster corpus. This holds for the Switchboard corpus as well (though not for the Brown corpus, see below), although it is not exactly a frequently found construction.⁴ It occurs a total of 11 times out of 519 possible cases. Ten of these are regular progressives (present progressives, none are past progressives), and one is a perfective progressive (example ...). Some examples are:

(...) I guess my husband got a letter at the office, I, I presume, and they must have been asking. [S77]

(...) That's the only place I was sore, and I thought, well, I must not be doing them right ... [S272]

(...) People are reading newspapers less than they used to and I guess that means they must be reading magazines and, you know, other written stuff less too. [S378]

There is also a correlation between the perfect and epistemic modality in the Switchboard corpus. There are a total of 113 cases in the corpus, distributed over four constructions (including the perfective progressive example (...) above), and none of the examples is deontic (there are no indeterminate cases either). This absolute figure is also reported by Coates for the SEU corpus (1983:44)

The most common perfect construction is *must have been*, while *must have V-ed* is about half as common. In addition, there are two examples with *must have got V-ed*.

- (...) I'm hearing a siren outside. I think my, my neighbor's car must have got bumped, you know, he has one of those alarms on his car. [S12]
- (...) So they started going around and, and looking at paint colors and they must have looked at five hundred different paint chips ... [S349]
- (...) - Uh-huh, uh-huh. I'm not sure what the name, maybe it was another woman, so.
 - Must have been from, uh, DISD.
 - Maybe so, yeah, maybe so, [S243]

The conclusions from the Switchboard data are that some constructions are genuinely ambiguous, most notably the *must V* construction. The only construction that is unambiguously deontic is the *must be V-ed* construction, but there are only a handful of examples. There are a number of constructions which are epistemic only, namely those involving Progressive and Perfect constructions.

The situation in the Brown corpus is shown in Table F. In some cases, the data agree with the data from the Switchboard corpus, but in other cases there are striking differences.

Table F
Occurrences of *must* with verbal complements in the Brown corpus

	Deontic	Epistemic	Indeterminate
<i>Be + V-part.</i>	236	2	8
<i>V</i>	392	14	19
<i>Have</i>	34	3	8
<i>Have been</i>		36	
<i>Be</i>	92	37	20
<i>Be + V-ing</i>	1	12	2
<i>Have got + V-ed</i>			
<i>No V</i>	5	1	1
<i>Have + V-ed</i>		50	2
<i>Have been + V-ing</i>		4	
<i>Needs + V</i>	1		
<i>Get + V-part</i>	1		

A comparison with Table E reveals that the most important differences between the two corpora are in the relative frequencies of the constructions *must V*, *must have* and *must be*. These three constructions have a preferred interpretation which is different in each corpus. So is the case of *must* without a main verb, but the numbers are much smaller here than in the Switchboard corpus. The number of constructions in which a deontic interpretation is possible has also increased, although it only involves one instance in each case. There is one instance of the construction *must needs V (...)* and one a passive construction with *get* instead of *be (...)*:

(...) In some cities games were broadcast throughout the week and then on weekends the announcer was silenced, and fans must needs drive to the city from all the broadcast area to discover how their heroes were faring. [B339]

(...) get example B86 in Passive section

There is an instance of a Progressive construction with a deontic interpretation, namely:

(...) The article also said that a person had to be 18 years old or over, and must not be going to high school to attend these classes. [B79]

The most common deontic construction in both corpora is the construction *must V*, which accounts for 83% of all cases in the Switchboard corpus, and 52% of all cases in the Brown corpus.

PASSIVE

Another category which has been thought to play a role in determining the choice of modality is the correlation with passive constructions. In Coates' study, 93% of all passive constructions with *must* are deontic in the Lancaster corpus, but only 60% in de SEU corpus (1983:37).

The situation in the Switchboard corpora is similar; in the spoken Switchboard corpus, there is a roughly even split between epistemic and deontic modality, but in the Brown corpus, the overwhelming majority of passives is deontic (Table ...), 97.9% of all passives is deontic. The pattern is reliable for the Brown corpus due to the large percentage of passive sentences with *must* in the corpus, as opposed to the handful of occurrences in the Switchboard.

Table ..
Correlation of modality and voice

	Switchboard		Brown	
	Active	Passive	Active	Passive
<i>Deontic</i>	59	4	521	237
<i>Epistemic</i>	402	3	152	6

Examples of epistemic passives are:

(...) But contest definition – that dramatic muscular separation of every muscle group that seems as though it must have been carved by a sculptor's chisel – is something quite different. [B177]

(...) It may, however, be noted that his gift for color and imagery must have been greatly stimulated by his stay in Paris. [B392]

(...) Somehow he had forgotten what he must have been told, that combat was an

intermittent activity. [B830]

epi switchboard

0012 [see above]

(...) They must have been hidden somewhere, because they came out. [S381]

(...) Obviously if we can do high class research, we must be well trained ... [S413]

deontic switchboard

(...) Course they have a requirement that, uh, that every single appeal must be permitted. [S199]

(...) whether, if it, if it's, uh, a toddler, uh, diapers have to be changed, uh, must be tended to constantly for fear that they might get into something. [S468]

0475 [see above]

(...) I think that any money that is given must be given carefully just to something that is a proven point ... [S511]

All examples but one of deontic passives in the Brown corpus have *be* as auxiliary. The sole exception is (...), where the auxiliary is *get*:

(...) How many times must we get burned before we learn? [B86]

From the point of view of predictability, the passive is most predictive in the written corpus. In the spoken corpus the passive is equally divided between the two interpretations. The overwhelming deontic interpretation of passives in the Brown corpus is probably due to the already large preponderance of deontic interpretations in written corpora.

REFERENCES

Barnes 1994

Biber et al 1999

Bybee et al. 1994 Chafe and Nichols 1986

Coates 1983

Collins 1991

De Haan 1997

De Haan 1999

De Haan 2001

Diewald 2001

Drubig 2001

Ehrman 1966

Floyd 1999

Friedman 1986

Friedman 1999

Guentchéva 1996

Johanson and Utas 2000
Kennedy 2002
Kucera
LDC 1999
Nuyts 2001
Palmer 1986
Palmer 1990
Van der Auwera and Plungian 1997
Westmoreland 1998
Willett 1988

¹ The lone example that does not fall into any of these groups is:

“Until such work is done, there must remain the nagging suspicion that alienation may be little more than an expression of the malaise of the individual, ...”

² As can be seen from the data in Table D, there are three occurrences of *it must be + N/Adj* in the Brown corpus. Interestingly enough, these three occur within the less formal registers, i.e. closest to the spoken register, of the Brown corpus. One occurs quoted within a letter, one, example (...), in fiction in an interior monologue, and one in fictional dialogue.

³ Similarly, it is not the case that first person singular does not occur with indirect evidentials at all. It can happen when the speaker is not in full control of his faculties, e.g. when he is drunk or unconscious.

⁴ This is in agreement with the figures from Biber *et al.* (1999:497-501).