Degrees of modality

Raphael Salkie

1. Introduction

This paper sets out a new account of modality as a typological category, and applies it to English[1]. Previous studies have argued for a prototype approach to modality, but have not constructed such an account consistently. Section 2 sets out the need for a prototype analysis, and section 3 offers a set of criteria which define the core of the category of modality. We examine several English modal auxiliaries with respect to the criteria, concluding that some of them frequently fail to meet the criteria and are therefore not core members of the category – in other words, they have a low degree of modality. Section 4 locates modality within a wider framework. In section 5 we revisit two problems in the analysis of English modals, arguing that the proposed framework sheds new light on them. Section 6 criticises the alternative analysis put forward by the *Cambridge Grammar of the English Language* (Huddleston and Pullum 2002). We conclude that the framework needs to be tested on other languages, but that it appears to be useful for English.

2. Why modality is elusive

What do linguists mean by the term "modality"? In the literature we find four different approaches:

(a) A broad interpretation. Palmer's extensive survey, for instance, says that "modality is concerned with the status of the proposition that describes the event" (2001: 1). The weakness of this approach is that the description is very vague, so that what it includes and excludes is unclear.

(b) A narrow definition. Van der Auwera and Plungian, in their typological study, say that they "propose to use the term 'modality' for those semantic domains that involve possibility and necessity as paradig-

matic variants", a definition which they describe as "relatively re-stricted" (1998: 80). This strategy has the virtue of starting with do-mains that virtually everyone recognises within modality; unfortu-nately, though, it excludes some semantic domains, such as volition and evidentiality, which many scholars would also want to include.

(c) A list of sub-types of modality. An example is Downing and Locke, who write: "modality is to be understood as a semantic category which covers such notions as possibility, probability, necessity, volition, ob-ligation, and permission" (2002: 382). This gives an indication of the domain, but fails to set clear boundaries.

(d) Avoidance of the problem by focusing on a set of expressions, such as modal verbs in English; modality is then taken to include whatever these expressions mean. Coates (1983), for instance, analyses the meaning and use of English modals in corpus data, without attempting any definition of the term *modal*.

The difficulty with the domain of modality is that there is broad agreement about some central members of the class, but disagreement about some of the candidates for inclusion. It is not possible to provide an acceptable definition with clear boundaries: instead, we need a clearly defined core but fuzzy boundaries. This suggests that a prototype approach is more promising: we should specify the criteria which demarcate the core, and allow for peripheral members of the category which do not meet all the criteria. Items that meet all or most of the criteria will be said to have a high degree of modality, while those in the periphery will have a low de-gree.

It is common practice in typological studies to treat grammatical catego-ries as having a core and a periphery, as noted by standard handbooks such as Comrie (1989: 37–38 and 106–110) and Croft (1990: 66–67 and 124–127). The prototype strategy was explicitly proposed for English grammar by Huddleston (1984: 72), and is adopted in a number of recent grammars of English, such as Greenbaum (1996: 92) and Biber et al. (1999: 59). It is applied to modality in English by Huddleston and Pullum (2002), who say that "modality … is not sharply delimited or subdivided, so that we shall need to make reference to the concept of prototypical features and to allow for indeterminacy at the boundaries of the categories" (2002: 172). We shall argue below that Huddleston and Pullum's specific analysis is prob-lematic, but their general strategy seems to be correct.

The prototype approach to grammatical categories is sometimes linked with work in cognitive psychology by Rosch (cf. Rosch and Lloyd 1978). More commonly, though, a commitment to prototypes is primarily a recognition of the need to allow for different degrees of category membership, and this is the situation with the work in typology and English grammar just cited. The consequences of adopting a prototype approach will therefore vary from category to category, depending on the criteria for membership. There is no reason, for example, to expect core members of a category to be more frequent, acquired earlier by children, or less liable to change over time. What we would expect is for the criteria to have implications for the analysis of individual languages, and for cross-language comparison.

3. Criteria for modality

We propose four criteria for the typological category of modality, the first two quite traditional, the second two more problematic. The criteria are semantic and pragmatic, the assumption being that cross-linguistic categories have to be characterised in this way. Other properties of modal expressions, particularly their morphosyntactic behaviour, will vary from language to language.

3.1. Possibility and necessity

We shall adopt the proposal cited above from Van der Auwera and Plungian as a starting point, treating it as just one criterion among others, rather than the only one. As they note, the view that possibility and necessity are central to modality is widespread and traditional. Their position is supported by Kiefer, who claims that "the relativization of the validity of sentence meaning to a set of possible worlds" is "the essence of modality" (1994: 2515). If we take this criterion seriously, then it excludes from the core the category of dynamic modality, proposed by some analysts to cover ability *can* and various uses of *will* in English (Palmer 2001: 9). These uses of *can* and *will* do not invoke other possible worlds: we shall look at some examples in sections 3.6 and 5.1. Since several scholars have suggested that dynamic modality may not be a type of modality at all (e.g. Palmer

1990: 37; Papafragou 1998: 2), assigning dynamic uses of modals to the periphery seems reasonable.

3.2. Epistemic and deontic

At the heart of many analyses is the distinction between epistemic modality, which refers to "judgements about the factual status of the proposition" (Palmer 2001: 8), and deontic modality, involving attitudes to "acts performed by morally responsible agents, e.g. obligation and permission" (Papafragou 2000: 3). It is natural to take these two domains as our second criterion: items which express either of them will have a high degree of modality, while items which do not will be nearer the periphery. This criterion will assign to the periphery dynamic modality (which already failed the first criterion) and alethic or logical modality (Lyons 1977: 791). Future time uses of *will* also fail this criterion: we will look at some examples in section 3.6.

3.3. Subjectivity

This term, introduced by Lyons (1977: 797), has been used in several different ways in the analysis of modality, and remains controversial (cf. Verstraete 2001, Timotijevic, this volume). Verstraete conceives of subjective modality as bringing into existence "a particular position of commitment [by the speaker – RS] with respect to the propositional content of the utterance" (2001: 1517). He calls this type of commitment "modal performativity", distinguishing it from the traditional use of the term *performativity* which links it with illocutionary force (and for which he proposes the term "interactive performativity"). Let us suppose that core instances of modality are subjective in this sense, making this our third criterion for modality.

In a recent paper, Narrog (2005: 169-171) argues that subjectivity is not a useful basis for defining modality because it is difficult to define precisely, and because it does not apply to all uses of items which are normally taken as core instances of modality. If subjectivity was the sole basis for characterizing the category of modality the latter criticism would be a reasonable one, but in the prototype framework proposed here it is just one criterion among others. Including subjectivity reflects the views of many

scholars who have studied modality; using it alongside other criteria avoids giving it too much weight. We return below to Narrog's first criticism.

Verstraete is one scholar who explicitly argues that uses of English modal verbs vary with respect to subjectivity: he argues that epistemic modals are always subjective, dynamic modals are never subjective, while deontic modals sometimes are and sometimes are not (2001: 1525). If we take subjectivity as a criterion for modality, then epistemic modals in general will have a higher degree of modality than dynamic ones, with deontic modals in between. Consider these authentic examples, of which the first two and the fifth are taken from Verstraete (2001):

(1)   …on my return from Mogadishu last week Amnesty confirmed that the preacher for whom it had campaigned had actually been released some time in nineteen eighty-nine. The flood of letters **must** have had some impact after all.

(2)   You seem to be seeking to destroy yourself in some way, but you **must** not include me in your plan of action.

(3)   Sometimes the hills were so steep that, despite our driver's haste, the horses could only go slowly. I wished to get down and walk up them, as we do at home, but the driver would not hear of it.
"No, no," he said. "You **must** not walk here. The dogs are too fierce." (Bram Stoker, *Dracula*)

(4)   Macmillan English Dictionary for Advanced Learners
The most up-to-date dictionary CD-ROM for learners of English
(…) I've been unable to install the CD-ROM on my computer. What have I done wrong?
(…) If you are using Windows® 2000 or Windows® XP, only users in particular groups can install programs. For Windows® 2000, you **must** be a member of the Administrators group or the Power User group to install the CD. For Windows® XP, you **must** be a member of the Computer Administrators group to install the CD. (*User Guide*)

(5)   But to reach orbit an object **must** accelerate to a speed of about 17,500 miles per hour (28,000 kilometers per hour, called satellite speed or orbital velocity) in a horizontal direction; and it **must** reach an altitude of more than 100 miles (160 kilometers), in order to be clear of the atmosphere.

Verstraete argues that epistemic uses of *must*, as in (1), always involve the speaker's commitment to "the propositional content of the utterance", while for deontic uses this varies. It appears that the speaker is the source of the obligation in (2), but this is only partially the case in (3), where the constraint is a result of the circumstances, and the driver's words are somewhere between reporting an obligation and giving advice. In (4) the *User Guide* is endorsing an obligation imposed by a third party, rather than by a social or ethical norm. Cases like (5) are more peripheral still: the necessity here derives from the laws of physics and the speaker is simply reporting it. We represent the four instances in table 1.

Table 1 Different uses of *must*

Example	Deontic (social or ethical norm)	Subjective
2	✓	✓
3	✓	X
4	X	✓
5	X	X

We now take a closer look at subjectivity, responding to Narrog's point that it needs to be made more precise. As defined by Verstraete it is clearly a pragmatic notion, since it crucially involves the speaker. A possible refinement of our third criterion would be to add to it as a statement about pragmatics: modal expressions which involve pragmatics will have a higher degree of modality than those which do not. Specifically, we propose that core instances of modality involve "primary pragmatic processes" in the sense of Recanati (2004: 21). These are processes which are required for a sentence to have a complete propositional meaning in a particular context: the clearest examples involve indexical expressions such as personal pronouns and tenses, which must have their reference assigned from the context. A sentence such as *I live in Belgium* only has a full propositional meaning after the application of the pragmatic processes of assigning a reference to the subject pronoun and the time of speaking. Recanati calls this type of pragmatic process *saturation*, because it involves filling in a

slot in the meaning of a sentence which would otherwise be empty and therefore uninterpretable. Any process which involves identifying the speaker will be a primary pragmatic process in this sense. Secondary pragmatic processes, in contrast, involve inferences or other extensions from the propositional content: they have a much wider range and are typically optional.[2]

It is evident that core instances of modality, such as (1) and (2) above, involve a similar process of saturation, because they require the speaker to be identified in order to be interpretable. Peripheral instances like (5) do not require saturation in order to interpret the modal. All three examples involve the modal *must*, however, which highlights a key difference between our analysis and the framework put forward by Papafragou (1998, 2000). She makes a distinction between *may* and *must*, which in her system always involve a primary pragmatic process (specifically, saturation), and *can* and *should*, which never do (1998: 14; 2000: 43). To analyse example (5), Papafragou would have to argue that it is analysed in a parallel way to (2): just as in (2) we need to invoke an (unspecified) social or ethical norm, in (5) we need to bring into play the domain of laws in physics. So far, we agree. Papafragou would also have to argue that invoking the domain of physics involves a primary pragmatic process of saturation which includes assigning a referent to the speaker. There is no reason to accept this, in my view: I would argue that sentence (5) fails the pragmatics part of the subjectivity test. In the framework proposed here, unlike that of Papafragou, modals are not tested on the criteria as isolated words, but in specific sentences. A modal may pass the subjectivity test in some uses, but fail it in other instances.

If we refine our third criterion in this way, the question arises of the relation between the modal expression and the rest of the sentence that contains it. Many studies of modality simply take for granted that a distinction has to be made between the (pragmatic) modal part of an utterance and the (semantic) residue. Various ways of representing this distinction have been proposed.. In generative studies, the distinction is often made in the syntax (cf. Wurmbrand 1999, Butler 2003), while other work distinguishes between the modal expression and "the propositional content of the utterance" (Verstraete 2001: 1517, cited above), or assumes a distinction between the modal expression and the "embedded proposition" Papafragou (2000: 43) as part of a semantic and pragmatic analysis. In some cases it is easy to recognise this division, particularly where negation can apply either to the modal or the proposition:

(6)   You must not go (p is negated – "it is necessary that you not go")
(7)   You need not go (modal is negated – "it is not necessary that you go")

In other instances, the distinction is less clear:

(8)   John can't swim
(9)   There won't be any trouble at the next match

Here the negation seems to have scope over the whole sentence, since we can paraphrase (8) and (9) as:

(8')   It is not the case that John can swim
(9')   It is not the case that there will be trouble at the next match

Another aspect of our third criterion, then, is whether there is a sharp division between the meaning of the modal expression and the meaning of the rest of the sentence; that is, whether or not there is a complete proposition which can be identified separately from the modal.

Summarising, the third criterion has three components:

    Subjectivity
    (i) commitment by the speaker
    (ii) primary pragmatic processes.
    (iii) a sharp distinction between the modal expression and the propositional content.

It is hard to decide whether these are three separate criteria or so closely connected that we should take them together. For ability uses of *can*, the three criteria yield the same result. We have already noted that "dynamic" *can* ranks low on part (iii); for this use of *can* we can accept Papafragou's claim that no primary pragmatic process is involved, so it fails (ii); and it is easy to see that speaker commitment is not involved in interpreting ability *can*. In this case, then, the three criteria all work in the same direction, consigning ability *can* to the periphery. We return to *can* in sections 3.6 and 5.1 below.

For *will*, part (ii) of the subjectivity criterion has to be treated separately from the other two. Papafragou does not discuss *will*, but in its straightforward future time uses like (9) it lacks speaker commitment and forms part

of the proposition (on the reasonable assumption that time reference is part of the truth-conditional content of a sentence, since *John will arrive* has different truth conditions from *John arrived*); hence it fails parts (i) and (iii). Future time *will* passes part (ii), however, since time reference is indexical, but this is not the reason why core instances of modality pass part (ii): it just so happens that primary pragmatic processes are involved in time reference. Thus future time *will* meets one of the parts of the third criterion but fails the other two, which suggests that they should in fact be regarded as separate.

3.4. Extremes of the modality scale

Our final criterion takes account of the fact that modal expressions can often be located on a scale. This is clearest for epistemic examples, where the factual status of a proposition may be judged to be just possible at one extreme, or highly certain at the other. In between we find expressions of probability or moderate certainty, for example:

(10)   John has possibly arrived by now.
(11)   John has probably arrived by now.
(12)   John has almost certainly arrived by now.

The two extremes have well-known logical properties with respect to negation. They also have pragmatic properties which are shared with other scalar expressions (cf. Levinson 2000: 83). The logical and pragmatic properties of modal expressions at intermediate points on the scale, such as *probably* in (11), are more variable. Let us tentatively take location near one of the extremes of the scale as a criterion for modality. We thus treat *may* and *must* as near the core. The English auxiliaries *can*, *could* and *might* also meet this criterion, as does *need* at the opposite extreme. As for *will*, it is close to the extreme occupied by *must* in its epistemic use, as in this example:

(13)   John will have arrived by now.

In its other main uses, however, such as future time (9), propensity (14) and volition (15), there is no relevant scale on which to locate the meaning of *will*:

(14)   They'll often go for days without speaking to each other
(15)   Will you please open the door?

In these uses, *will* fails to meet the fourth criterion.

It could be argued that the fourth criterion follows from the first one, since possibility and necessity are extreme points on a scale. We shall keep it separate here, however, in the hope that further research may clarify this question.

3.5. Summary of the criteria

Core members of the modal category meet these four criteria:

A. They express possibility or necessity.
B. They are epistemic or deontic.
C. They are subjective, involving
     (i)   commitment by the speaker.
     (ii) primary pragmatic processes.
     (iii) a sharp distinction between the modal expression and the propositional content.
D. They are located at one of the extremes of a modal scale.

3.6. Applying the criteria

Consider now some examples in relation to the criteria (those labelled *BNC* are from the British National Corpus):

(16)   I don't know for sure but there may be milk in the fridge.
(17)   Nursing and medical staff may also need psychological support to cope with the intense nature of the treatment and the uncertainties in outcome. (BNC EE8 800)

(18)   I can track him, and he won't know. (BNC CM4 2289)
(19)   It can be cold in Stockholm.

(20)   An analysis of the results should shed light on the workings of the Northern Ireland labour market. (BNC HJ0 18810)

(21)   Your essay is an argument, and you should present different sides of that argument. (BNC HXH 1408)

Example (16) meets all the criteria, whereas (17), which does not express speaker uncertainty but generalises over the members of the class of nursing and medical staff, fails criterion (Ci) because there is no special speaker commitment. It also fails criterion (Cii), because unlike (16) there is no need here to invoke the knowledge of the speaker in order to interpret *may*: to interpret (17) we simply have to assume certain properties of nursing and medical staff, exactly as we do in interpreting the non-modal *Nursing and medical staff also need psychological support*. However (17) passes (Ciii), since the proposition can be negated separately from the modality (that is, *Nursing and medical staff may also not need psychological support* means "It is possible that nursing and medical staff do not need psychological support"). We conclude that (17) has a slightly lower degree of modality than (16).

   Turning to (18), *can* here has a low degree of modality because it fails criteria (A), (B) and all of (C) (admittedly this example involves reference to the speaker, but that is due to the first person pronoun, not the modal). Example (18) fails criterion (A) because it does not involve reference to other possible worlds. Some analysts have captured this insight by including "factuality" in their analysis of *can* (cf. Papafragou 2002: 43; Larreya and Rivière 2005: 89), but thereby they are effectively conceding that this use of *can* fails criterion (A), since factuality is the very antithesis of our first (and least controversial) criterion for modality. In short, this common use of *can* fails the first three criteria for modality, and thus seems to have a very low degree of modality.

   Against this claim one might give these arguments:

(a)   *can* shares the morphosyntactic properties of auxiliaries like *may*, *must*, etc.
(b)   there are semantic similarities between (non-modal) ability and (modal) possibility.
(c)   *can* and *may* are similar in other uses; for example they are both used for permission.
(d)   In negative and interrogative contexts, *can* has an epistemic use (e.g. *That can't be true* and *Can that really be my brother?*)

Argument (a) is only relevant if it can be demonstrated that members of this morphosyntactic class of items share semantic properties – and much of this paper has argued that this is not the case. Argument (b) is more important, and it is clear that in many instances, *can* is used in English with a more general sense of possibility rather than just ability. My argument is that these more general uses are pragmatic inferences from the ability sense, and do not indicate that *can* meets criterion (A). On argument (c), Papafragou (2002: 58) gives a subtle analysis of differences between *can* and *may* in their permission uses which follow from the factual nature of *can* versus the wider domain of evidence invoked by *may*: as just noted, this effectively concedes that permission *can* is not modal (in our terms, it fails criterion A). As for the fourth argument, we can again cite Papafragou (2002: 78–79), who gives reasons against treating such examples as epistemic.

Example (19) shares all these properties of (18), except in respect of criterion (Ci). There is an element of subjectivity in (19): by choosing not to say *It is cold in Stockholm*, the speaker indicates that her utterance is based on previous knowledge about that city, not current knowledge of its temperature at the time of speaking. Compare *It may be cold in Stockholm*, a prototypical epistemic judgement like (16) which is a statement about the likelihood of the actual temperature being low, either now or in the future. Hence (19) is a weak assertion which stays in the realm of potentiality, and expresses a low degree of speaker commitment to the proposition *It is cold in Stockholm*. This gives (19) a higher degree of modality than (18), but a lower degree than core instances of epistemic *may*.

Turning now to *should*, most analysts would treat (20) as epistemic. Note, however, that in prototypical instances of epistemic modality like (16), a body of evidence is brought to bear on the likelihood of a proposition being true. The evidence normally includes general facts, not just the facts of the specific situation expressed by the proposition: in (16), for instance, the evidence might include knowledge that fridges often contain milk, or that on previous occasions milk was located in this fridge. In (20), the body of evidence does not include such general facts, but simply the facts of the situation in question. The effect of using *should* in (20) is to make a prediction about real future events which is less confident than a prediction using *will* in its peripheral future time use. This example does not have the crucial generalised context that is characteristic of core epistemic modality.

Papafragou claims that *should* in all its uses "relies quite heavily on the sort of structured knowledge humans typically possess about the normal

course of events" (2000: 62). She proposes that this is a restriction on the body of evidence that can be brought to bear on sentences containing *should*, and that the semantics of *should* is otherwise identical to that of *must*. While the claim about the normal course of events is correct, I would argue that this is not a small difference between *should* and *must* but a fundamental one. If so-called epistemic *should* in fact stays firmly in the real world, then we are not dealing with the domain of possibility and necessity at all: as the quotation from Kiefer in section 2.1 makes clear, these domains invoke other possible worlds. Thus (20) fails criterion (A), and we analyse it as saying "In the normal course of events, an analysis of the results will shed light on the workings of the Northern Ireland labour market", yielding the implication of a moderately confident assertion.

Turning now to (21), we can argue that this example too fails criterion (A). Its meaning too refers to "the normal course of events": its construal as imposing a moderate obligation comes from the interaction of that meaning with the context. Many of the relevant aspects of the context will be of the same kind that results in a deontic sense of sentences with *must*; but *should* arrives at its obligation interpretation via a different route from *must*. We revisit these matters in 5.2 below.

4. Modality and Irrealis

A characterisation of modality should situate it in relation to other grammatical categories. In this section, we make some suggestions about this wider picture, drawing on Palmer (2001). Suppose we take as the default type of sentence a factual description of an event or state of affairs in front of the speaker's eyes. We can use the general term *realis* for this type of sentence, and any sentence which departs from the default will be called *irrealis*. Languages which make this distinction in their grammar will be said to instantiate the category *Reality Status*. For languages such as Manam (Papuan – Palmer 2001: 147–148) this top-level distinction may be the only expression of Reality Status: every finite verb has to be specified morphologically as realis or irrealis. Exactly where a language draws the line seems to vary: Manam treats past, present and habitual as realis, while future, imperatives, counterfactual and sequences of habitual are marked as irrealis. Palmer states that simple past and present time are always realis (2001: 168), but that some semantic domains appear as realis in some languages and irrealis in others. An example is future, which is irrealis in

Manam but comes under realis in Caddo (Caddoan, Midwestern US – Palmer 2001: 169). This combination of an invariant core and a fluctuating periphery argues for a prototype approach to Reality Status, like the one we have adopted here for modality.

Within the irrealis category, we propose that a language can choose one or more of three subcategories.

A. The sentence is asserted or not asserted. We shall call this category *mood*, with a binary division between *indicative* and *subjunctive* (cf. Palmer 2001: 3–4)
B. The sentence can be marked as one whose truth is not in doubt, or marked to indicate judgements based on hearsay, personal experience, etc. The usual term for this category is *evidentiality*.
C. A language can refer only to the real world, or can invoke other possible worlds. This is the (first criterion for the) category of *modality*.

We thus treat modality as a sub-category of irrealis, within the top-level category of Reality Status. It is useful to contrast this with the different taxonomy proposed by Narrog (2005). He first proposes that terms used in the literature, such as "(realis vs.) irrealis", "factuality" and "validity", are essentially equivalent, and indicate "whether a state of affairs is 'actual' or 'existent' at a specific point or interval of time" (2005: 183). He gives reasons to choose "factuality" as the best term, and defines modality as follows:

> Modality is a linguistic category referring to the factual status of a state of affairs. The expression of a state of affairs is modalized if it is marked for being undetermined with respect to its factual status, i.e. is neither positively nor negatively factual. (2005: 184)

Narrog thus equates our category of Reality Status with modality. I think that there are two problems with this system. Firstly, it takes one attribute of modality and makes it the only defining criterion, whereas there are good reasons for taking more than one criterion. Secondly, it fails to capture the importance of real versus possible worlds as distinctive of modality, as opposed to the wider realis-irrealis distinction in languages like Manam.

The notion "possible worlds" is crucial to another difference between Narrog's system and the one proposed here. He argues that evidentiality is

a type of modality, giving as an example the evidential use of the German modal verb *sollen*:

(22) Er soll bei einem Unfall umgekommen sein.
      "He allegedly died in an accident"

Narrog comments:

> The factuality of the proposition (He died in an accident) is undetermined, just as it is in the case of epistemic modality. The difference is that while in epistemic modality the proposition is undetermined with respect to its factuality relative to the world of knowledge and beliefs of the speaker, with evidentiality it is undetermined relative to sources of information other than the speaker. (2005: 188)

My argument against this is that examples like (22) do not invoke other possible worlds, whereas epistemic modals do. If I say *He may have died in an accident*, I am invoking other worlds in which *He died in an accident* is not true. The evidential in (22) does not invoke other possible worlds, but takes a view about the real one. Narrog is right to point out similarities between evidentiality and modality, but conflating them goes too far.

5. Some consequences of the prototype approach

We have argued that the English auxiliaries *can*, *should* and *will* often have low degrees of modality. This section revisits two controversies in the analysis of English modals and attempts to show that the approach proposed here sheds new light on them. The first involves *can* in contrast with *may*, and the second compares *should* and *must*. We intend to deal with *will* in a separate paper.

5.1. *Can*, *may* and the real world

Leech (2004: 82–83) examines these two sentences:

(23) The road may be blocked.
(24) The road can be blocked.

Leech says that (23) expresses "factual possibility" while (24) is an instance of "theoretical possibility": example (24), he claims, "describes a theoretically conceivable happening", whereas (23) "feels more immediate, because the actual likelihood of an event's taking place is being considered". He contrasts two likely contexts:

(23')   The road may be blocked by floodwater ("that possibly explains why our guests haven't arrived" – dialogue between husband and wife expecting visitors).

(24')   The road can be blocked by police ("and if we do this, we might intercept the criminals" – said by one detective to another).

What Leech says about the different contexts is exactly right, but his explanation is wide of the mark. We have argued above that "factual modality" is a contradiction in terms; and we have also accepted the claim in earlier work that *can* has factuality as part of its meaning. How does Leech come to put things the other way round, treating *may* as factual, rather than *can*?

   The reason is that epistemic *may* takes a proposition (*The road is blocked*) and expresses a judgement about the likelihood of this proposition being true in the real world, as compared to other possible worlds. With *can*, on the other hand, no other possible world is invoked: example (24) (in a context like 24') contains a single proposition about a current capability of the police. Leech seems to consider that the capability in (24) is further removed from the real world than the comparison with other possible worlds in (22). That is not, however, the relevant point.[3] In (24) the capacity exists in the real world and no other world is invoked, whereas (23) compares the real world with other possible worlds. Thus (24) has a high degree of modality (it meets all the criteria) while (23) is marginal (it fails all of them). (For more on Leech's methodology, and a comparison with other analyses of *can* and *may*, see Salkie 2001: 578–580).

5.2. Should and *must* – Rivière's problem

Rivière (1981) discusses these examples:

(25)   You live in Oxford, you must know Professor Fen then.

(26)   You live in Oxford, you should know Professor Fen then.

(27)   You know Professor Fen, you must live in Oxford then.
(28)   *You know Professor Fen, you should live in Oxford then.
        [OK in obligation sense]

(29)   *John is a brilliant student, he must pass his exam easily.
(30)   John is a brilliant student, he should pass his exam easily.

The asterisk on (28) indicates that *should* here cannot have the sense it has in (26), where it appears to be a weaker version of epistemic *must*, but only the sense of "you are obliged to live in Oxford". The main part of Rivière's problem is to explain why (28) is not acceptable in this sense, while (26) is; a secondary part is to explain why (29) is unacceptable in contrast to (30).

I have argued in earlier work that Rivière's proposed solution is not workable (Salkie 2002: 85–87). In the framework put forward here, *should* has the sense of "in the normal course of events", and has a low degree of modality since it fails criteria (A), (B) and (D). In the right pragmatic context, we interpret "normal course of events" as carrying the stronger sense that one event doesn't just follow another but is the normal consequence of another. The word *then* in (26) invites us to apply this stronger interpretation here. Thus (26) does not express an epistemic judgement but a statement about a normal consequence of living in Oxford. This is very different from (25), which expresses a judgement of the likelihood of the proposition *You know Professor Fen* being true, given the evidence in the context (which includes the first clause). Hence *should* and *must* are both possible in this context, but for different reasons.

In (27), the speaker once again makes an epistemic judgement, but the crucial example (28) cannot express the notion "normal consequence", because living somewhere is not normally a consequence of knowing someone. Hence (28) cannot be interpreted as parallel to (26), and another sense of *should* is required for the sentence to be acceptable.

Turning to the secondary part of Rivière's problem, we can easily interpret (30) in terms of "normal course of events". As for (29), this seems to be unacceptable because of a lexical restriction on epistemic uses of *must* in English, which cannot normally refer to future time. (The fact that epistemic *may*, and translation equivalents of *must* in French and German, are

not so restricted suggests that we are dealing here with an idiosyncrasy of *must* in English).

The framework set out in this paper thus enables us to find an alternative solution to Rivière's problem.

6.   Degrees of modality in other work

In his paper on the future tense controversy, Huddleston (1995: 421) introduces the notion that some uses of English modals have a low degree of modality, because the additional element of meaning which the modal adds to the sentence is small. The notion is amplified in the *Cambridge Grammar of the English Language* (Huddleston and Pullum 2002 – henceforth *CGEL*). This section tries to show that the notion as they use it is incoherent, because it groups phenomena which have nothing in common, and redundant, adding nothing to an analysis of modality.

CGEL proposes that degree of modality "has to do with the extent to which there is a clearly identifiable and separable element of modal meaning" (179). This sounds similar to our criterion (Ciii), but in developing the idea, *CGEL* uses only one test: whether there is an appreciable difference in meaning between pairs of clauses where one member is unmodalised, the other modalised. They give four types of example, the first one involving *may*, which they claim adds a distinct element of meaning in (32) compared with (31):

(31)   They know her.
(32)   They may know her.

In examples like (33) and (34), however, the modal adds little to the meaning:

(33)   Strange as it seems, I know you.
(34)   Strange as it may seem, I know you. (*CGEL*, 179)

There is hardly any difference between the first clause of (34) and its non-modal counterpart in (33), so *may* is said to have a low degree of modality in (34). In their second type of example, *CGEL* applies the same reasoning to *should*, which in their view has a high degree of modality in (36), where

it is clearly distinct from (35), but a low degree in (38), which barely differs from (37):

(35)   Your passport is in the drawer.
(36)   Your passport should be in the drawer.

(37)   It's odd that he is so late.
(38)   It's odd that he should be so late.

Before commenting on these examples, we should note that similar data with *can* are analysed in Kjellmer (2003), who observes that with performative verbs, verbs of perception and verbs of mental processes, *can* appears to add little to the meaning:

(39)   Detective Superintendent Mike Morgan, leading the inquiry, said: 'I can assure you he is not a suspect.' [cf. *I assure you*]
(40)   I can just see a soldier beating a man with a stick. [cf. *I just see*]
(41)   I don't think they particularly enjoy the reading log [and] to a certain extent I can empathize with that. (Kjellmer 2003: 164-6) [cf. *I empathize*]

Kjellmer argues that *can* does in fact add to the meaning here, but the important point for our purposes is his demonstration that this use of *can* is mostly restricted to first person singular constructions and to particular verbs (2003: 149–153). What we have here is a type of partially fixed idiomatic construction with a slot that can only be filled by a limited class of lexical items (compare "It was a(n) X's dream", where X usually has to be an agent noun like *promoter* or *scientist*).

   Returning to the "X as it may seem" construction in (34), there is little doubt that it is idiomatic in the same way. Evidence for this comes from the 48 instances of "X as it may seem" in the BNC. In nine of these examples X is the word *strange*, and a further 21 contain an adjective which is similar in meaning, most commonly *surprising*, *incredible* and *unlikely*. In 11 of the remaining examples the adjective has a negative connotation (*daft*, *arduous*, *haphazard*, etc), or is in a negative context (e.g. *not so easy as it may seem*), leaving only three instances with an adjective such as *sensible* in a neutral or positive context.

   Examples like (38) are not as fixed as this construction with *may*, though this near-vacuous use of *should* is restricted to subordinate clauses

which complement evaluative predicates like *odd*, *strange*, etc. and their positive counterparts like *natural* and *interesting*. Furthermore, both the examples with supposedly low-degree *may* and those with *should* are marginal in other ways: they are formal, rather archaic, and rare except in certain genres. It would not be worth inventing a special term for these marginal examples unless there was a reason to apply the term also to central uses of modals. These uses of *may* and *should* are similar to the reduction of meaning sometimes called "semantic bleaching" which often occurs when a content word evolves into a function word. They should be dealt with under this heading, rather than as a reduction in meaning that is specific to modal expressions.

   CGEL gives a third type of example, arguing that low degree modality also applies to some core uses of *will*:

(42)   She is one year old tomorrow.
(43)   She will be one year old tomorrow. (*CGEL*: 179)

They note that "there is no effective difference" between (42) and (43) (*CGEL*: 190). These examples are very different, however, from those involving *may* and *should*. In the first place, (43) is not marginal or idiomatic. More importantly, in (34) and (38) the modals have lost some of the meaning that they have in their central uses, but here the point is surely not that *will* has lost some of its usual meaning: rather, it is the present tense in (42) which under certain conditions has acquired a future interpretation. The small difference between (42) and (43) arises because the non-modal (42) has taken on some of the futurity meaning which is usually expressed by *will*. If we wanted a term for this we should call it "acquired modality" of the present tense, not "low degree modality of *will*".

   The last type of example that *CGEL* uses to illustrate low degree modality again involves *will*, and is illustrated in (44):

(44)   Ed is Tom's father and Tom is Bill's father, so Ed will be Bill's
       grandfather.

CGEL notes, correctly, that "the evidence for [the conclusion] is so strong that I could equally well have used unmodalised *is*" (2002: 189). They give a parallel example in the section on epistemic *must*, and make a similar remark about the small difference between the modalised and unmodalised version (2002: 181). Here again, though, we surely have a very different

type of phenomenon from the other three examples. When there is a small difference between epistemic *must / will* on the one hand, and unmodalised assertions on the other, the reason is simply that in contexts like (44), claiming "p is almost certainly true" is not very different from just saying "p". In other words, what *CGEL* calls "high strength" modality (near the necessity end of the modality scale) is often close to making an unmodalised assertion. That is a rather banal fact about epistemic necessity, and does not motivate introducing a separate notion of "degree of modality".

To sum up, cases where there is a small difference in meaning between a sentence with a modal and its unmodalised counterpart occur for three completely different reasons:

A. The modal loses some of its meaning (*may* and *should* in some marginal uses).
B. The unmodalised sentence has acquired some of the same meaning as the sentence with the modal (futurate use of the present tense vs. *will*).
C. Deductions based on strong evidence, and expressed using epistemic *must* or *will*, do not differ greatly in strength from unmodalised assertions.

The notion of "degree of modality" in *CGEL* thus appears not to cover a single phenomenon but three unrelated ones. Furthermore, these phenomena can each be explicated by analytic notions which are independently needed, so the proposed new notion adds nothing useful to the analysis of modality. In this paper we have treated the notion "degree of modality" as a more complex one, involving several different criteria instead of the single one used in CGEL. We might add that a key motivation for introducing the notion in *CGEL* and in Huddleston (1995) was to support the claim that *will* is a marker of modality, not of tense. In subsequent work we hope to show that *will* should be treated as both a non-prototypical modal and a non-prototypical tense, thereby resolving the controversy about its category status.

7. Conclusion

We have set out an account of modality using the prototype approach to grammatical categories, a strategy which is widely accepted by typologists and by English grammarians. Using four criteria for membership of the category, we claimed that some auxiliaries in English (e.g. *must* and *may*) have core uses which meet all the criteria and peripheral ones which do not. Other auxiliaries such as *can*, *should* and *will* usually fail one or more of the criteria. We have tried to show that the notion "degree of modality" can be broken down into a small set of criteria, and that the proposed framework sheds light on a range of English data. We argued that the alternative prototype approach to the notion "degree of modality" in *CGEL* is misconceived.

Any account of a typological category needs to show its value in more than one language. In future work we plan to apply the prototype framework to modal expressions in French and German. For English, at least, treating different instances of modal expressions as varying in the degree to which they match the prototype seems to be a useful grammatical tool.

Notes

1. I am grateful to Jelena Timotijevic, Heiko Narrog and an anonymous Mouton reviewer for comments on an earlier version. None of them is responsible for any errors

2. Recanati argues that several processes which are often taken as secondary are in fact primary, which sets him at odds with other theorists of the semantics-pragmatics interface. Processes like saturation, however, are recognised by many schools of thought, even if they disagree about other aspects of pragmatics (cf. Levinson 2000: 177; Carston 2002: 185ff.).

3. The passive voice probably contributes to Leech's faulty analysis. Example (23) describes a property of the police (in a context such as [23']), while (22) makes a judgement about a property of the road. If the examples had been *We can block the road* versus *We may block the road*, the example with *can* would perhaps have felt "more immediate" and factual than the tentative modalised one with *may*.

References

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and E. Finegan
  1999   *Longman Grammar of Spoken and Written English.* London:
            Longman.
Butler, Johnny
  2003   A minimalist treatment of modality. *Lingua* 113: 967−996.
Carston, Robyn
  2002   *Thoughts and Utterances.* Oxford: Blackwell.
Coates, Jennifer
  1983   *The Semantics of the Modal Auxiliaries.* London: Croom Helm.
Comrie, Bernard
  1989   *Language Universals and Linguistic Typology.* 2nd edn. Oxford:
            Blackwell.
Croft, William
  1990.   *Typology and Universals.* Cambridge: Cambridge University
            Press.
Downing, Angela and Philip Locke.
  2002   *A University Course in English Grammar.* London: Routledge.
Greenbaum, Sidney
  1996   *The Oxford English Grammar.* Oxford: Oxford University Press.
Huddleston, Rodney
  1984.   *Introduction to the Grammar of English.* Cambridge: Cambridge
            University Press.
  1995   The case against a future tense in English. *Studies in Language*
            19: 399−446.
Huddleston, Rodney and Geoffrey Pullum
  2002.   *The Cambridge Grammar of the English Language.* Cambridge:
            Cambridge University Press.
Kiefer, Ferenc
  1994.   Modality. In *The Encyclopedia of Language and Linguistics*,
            Volume 5, Ronald Asher (ed.), 2515−2520. Oxford: Pergamon
            Press.
Larreya, Paul and Claude Rivière
  2005   *Grammaire Explicative de L'anglais*. 3e édition. Paris: Pearson
            Longman France.
Leech, Geoffrey
  2004   *Meaning and the English Verb.* 3rd Edn. London: Longman.
Levinson, Stephen
  2000   *Presumptive Meanings.* Cambridge, MA.: MIT Press.

Lyons, John
    1977   *Semantics*. Cambridge: Cambridge University Press.
Narrog, Heiko
    2005.   On defining modality again. *Language Sciences* 27: 165-192.
Palmer, Frank
    1990   *Modality and the English Modals.* 2ⁿᵈ edn. London: Longman.
    2001   *Mood and Modality.* 2ⁿᵈ edn. Cambridge: Cambridge University Press.
Papafragou, Anna
    1998   Inference and word meaning: the case of modal auxiliaries. *Lingua* 105: 1–47.
    2000   *Modality: Issues in the Semantics-Pragmatics Interface.* Amsterdam: Elsevier.
Recanati, François
    2004   *Literal Meaning*. Cambridge: Cambridge University Press.
Rivière, Claude
    1981   Is *should* a weaker *must*? *Journal of Linguistics* 17: 179–195.
Rosch, Eleanor and Barbara B. Lloyd (eds.)
    1978   *Cognition and Categorization*. Hillsdale, MI: Lawrence Erlbaum.
Salkie, Raphael
    2001   Grammatical explanation in Britain and France. In *Langage et Référence:Mélanges offerts à Kerstin Jonasson à l'occasion de ses soixante ans,* Hans Kronning, Coco Norén, Bengt Novén, Gunilla Ransbo, Lars-Gören Sundell and Brynja Svane (eds.), 573-582. (Studia Romanica Upsaliensia 63.) Uppsala: University of Uppsala.
    2002.   Probability and necessity in English and German. In *Information Structure in a Cross-linguistic Perspective,* Hilde Hasselgård, Stig Johansson, Bergljot Behrens and Cathrine Fabricius-Hansen (eds.), 81-96. Amsterdam: Rodopi.
Timotijevic, Jelena
              Another look at modals and subjectivity. (In this volume).
Van der Auwera, Johan and Vladimir Plungian.
    1998.   Modality's semantic map. *Linguistic Typology* 2: 79–124.
Verstraete, Jean-Christophe
    2001.   Subjective and objective modality: Interpersonal and ideational functions in the English modal auxiliary system. *Journal of Pragmatics* 33: 1505–1528.

Wurmbrand, Susi

    1999.   Modal verbs must be raising verbs. In *WCCFL 18: Proceedings of the 18th West Coast Conference on Formal Linguistics*, Sonya Bird, Andrew Carnie, Jason D. Haugen, and Peter Norquest (eds.), 599–612. Somerville, MA.: Cascadilla Press.