

MODAL VERBS: TOWARDS A UNIFIED SEMANTICS

Anna Papafragou

*Department of Linguistics
University College London*

Abstract: This paper offers some arguments in favour of a unitary meaning approach to the English modals. Using *may* as a case study, I argue against ambiguity-based and polysemy-based approaches to the modals. Adopting the relevance-theoretic framework of communication (Sperber and Wilson, 1986/1995), I then show how a unitary semantics for *may* can give rise to the range of root and epistemic modal meanings communicated by the verb; the derivation of root and epistemic interpretations is achieved through general pragmatic considerations during the process of comprehension. I conclude by exploring some implications of my approach for theories of word meaning, polysemy and the semantics/pragmatics distinction.

Keywords: modals, *may*, relevance theory, polysemy, word meaning, semantics, pragmatics.

1. INTRODUCTION

My aim today is to reconsider the debate between monosemous and polysemous approaches to English modals, and to offer some arguments in favour of a unitary meaning approach. I will use *may* as a case study, but I expect my results to carry over to the rest of the English modal set (and possibly to modal items in other languages where the alleged polysemy exists). What I ultimately hope to show is that the discussion of modals has direct implications for the organisation of lexical semantics and the interface with pragmatic mechanisms of utterance comprehension.

It is widely acknowledged that *may* communicates two types of modality: root and epistemic modality. On root readings, some state of affairs is presented as possible given the real-world circumstances, the social and moral laws, and so on; root readings, that is, include the ‘permission’ and ‘real-world possibility’ meanings of the verb which are given in (1). On epistemic readings, a certain proposition is presented as a tentative product of inference from available information, and thus carries low speaker commitment - see (2):

- (1) a. You may go home now.
- b. This mountain may be climbed from other points on this tour.
- (2) a. Given the new evidence, his decision may have been unjustified.
- b. I may have to leave the country.

There have been two types of response to these data. One line of research has considered *may* as semantically ambiguous (Coates, 1983; Palmer, 1990). An interesting variant of this approach is Sweetser's (1990) proposal to consider modals as polysemous between force-dynamically related senses. On her analysis, polysemy in natural language is often motivated by a metaphorical mapping from the concrete, external world of socio-physical experience to the abstract, internal world of reasoning and of mental processes in general. *May* started out as encoding the existence of a potential but absent barrier in the external world - cf. (1), which then became metaphorically projected in the abstract domain of reasoning and gave rise to the epistemic meaning - cf. (2). According to Sweetser, the metaphorical mapping between the two domains of application of the modal concepts has become part of the semantics of English and has given rise to systematic polysemy.

I want to pursue a different line of explanation: I will claim that the semantic content of *may* greatly underdetermines the meaning communicated by the verb, and that independently motivated processes of pragmatic interpretation are drawn upon to flesh out the information semantically encoded by the verb. Unlike most previous approaches adopting a unified semantics for modals (Kratzer, 1981; for an exception, see Groefsema, 1995), I will explore the process of contextual enrichment in some detail; my framework will be placed within a relevance-theoretic account of utterance interpretation (Sperber and Wilson, 1986/1995).

2 THE SEMANTICS OF *MAY*

I propose the following as the semantically encoded content of *may*:

- (3) p is compatible with the set of all propositions in domain x
(where p is the proposition expressed by the rest of the utterance)

The semantic representation of *may* contains a slot (domain x) which is pragmatically saturated: depending on the type of propositional domain selected, *may* will express a different kind of modality. Before proceeding to the details of the pragmatic analysis, let me say what I have in mind by 'domains of propositions'.

According to relevance theory (Sperber and Wilson, 1986/1995), a given proposition can be entertained and stored in memory in several different ways. First, and obviously, it can be entertained as a truth-conditional description of a state of affairs in the actual world (i.e. as a factual assumption). This is involved in what relevance theory calls the "descriptive" use of language, where an utterance is presented as representing a thought entertained as a factual assumption. Factual assumptions are the means whereby we represent reality to ourselves. Since the actual world is the only world which exists, it provides a rich and easily accessible contextual background with respect to which we process and comprehend (i.e. establish the relevance of) external stimuli. Therefore, in the default case for comprehension, the actual world forms the

base world against which ostensive stimuli are processed. Propositions describing the actual world (at any temporal manifestation, present, past, or future) can be said to belong to a single domain - the factual domain.

Factual propositions may be further broken down into sub-domains. Take for instance systems of laws, regulations and other generalisations which describe a sub-type of events in the actual world: laws of biology or chemistry, rules of chess, bank regulations, legal rulings, and so on. The propositions describing these sub-types of event are factual propositions, but belong to a more specific domain (biological, etc.).

Somewhat different are propositions taken to describe stereotypical or ideal states of affairs; again they can be considered to form different domains, and there is some evidence that they are stored and handled by a specific internal module (cf. Premack and Premack, 1994).

Yet other assumptions are handled as descriptions of states of affairs in worlds which differ from the actual one: for instance, worlds desirable from someone or other's point of view. Since the actual world is by definition the only one which is actualised, it follows that alternative worlds do not exist in reality, but form mere thought possibilities, mind-dependent specifications of situations which differ from the way things in fact are.

Finally, assumptions can be entertained and stored as abstract representations, (i.e. hypotheses), or representations of representations (where the initial representation may or may not be attributed to some source). These are cases of what in relevance theory is called the "interpretive" use of language, and will come out as yet a separate domain of propositions (I will further elaborate on this type later on in my paper).

On this picture, then, propositions come with an in-built index, or indication of their domain. In factual assumptions this index is null, since these assumptions are the default case of representing the world; in other types of assumptions, I assume that there is an indication in the language of thought as to their domain, i.e. the sub-type of event or state of affairs they are taken to be descriptions of. Logical relations such as entailment and compatibility apply only among propositions in a single domain.

This last idea will prove crucial for the analysis of modality, since what modal verbs express are generally logical relations of entailment and compatibility holding among propositions in various domains. In the following sections I will illustrate the details of this analysis for various interpretations of *may*.

3 THE PRAGMATICS OF *MAY*

Imagine that (4) is uttered by a bank employee in a conversation with a young customer:

(4) You may convert your account into a student account.

The logical form of the utterance is (roughly) given in (4'):

(4') It is compatible with the set of all propositions in domain x that p [You convert your account into a student account].

This logical form requires some fleshing out before it yields a truth-evaluable representation; it remains to be specified what propositional domain p is considered compatible with. This domain will have to contribute to an interpretation of (4) which is accessible enough for the hearer, and capable of achieving adequate contextual effects in a way compatible with the speaker's abilities and preferences (i.e. it should be an optimally relevant interpretation); furthermore, the resulting interpretation of the utterance should be one that the speaker could reasonably have intended to be optimally relevant for the addressee (thereby satisfying the requirements of consistency with the communicative principle of relevance - Sperber and Wilson, 1986/1995). Assume that (4) is an answer to a customer's query about changing one's account from normal to student: the customer has already made mutually manifest that this change is compatible with his own interests and preferences. The only remaining domain of propositions for which it would be relevant to know whether p is compatible with it involves bank regulations and procedures. What (4), then, will be taken to communicate is that it is compatible with the bank's regulations that the customer's account be converted into a student account.

How exactly would such an utterance achieve relevance? After all, to say that a proposition p is compatible with the set of propositions in a certain domain is to make a very weak claim: how can we narrow down the intended domain of propositions, that is, intuitively, the domain which may serve as evidence for or against a certain proposition? This is a more general problem of what can be considered as "relevant evidence" (in a pretheoretic sense of "relevant"), which has occupied much research in philosophy, psychology and artificial intelligence (cf. the "frame problem"). For my purposes, suffice it to say that, in assessing evidence for/against a certain proposition, a relevance-oriented organism will be geared towards highly accessible domains which are recoverable on the basis of general assumptions about causality and law-like generalisations. In the case of *may*, the hearer should be particularly looking for a domain of propositions with which proposition p was previously contextually assumed to be incompatible; an utterance containing *may* typically achieves relevance by aiming to contradict and eliminate this existing contextual assumption. This way of achieving relevance is actually common to a number of quantifiers, which introduce a "denial of expectation" (see Sanford, Moxey and Paterson, 1994 - cf. also negative constructions).

Similar considerations apply to examples (5) through (7):

- (5) I'll tell you about your trip so that you may make arrangements.
- (6) To make this dish, any sort of pasta may be used.
- (7) During the seminar, you may interrupt as often as is needed.

In (5), the domain of propositions with which the embedded proposition is compatible is a sub-domain of factual propositions describing the hearer's abilities and general circumstances concerning his trip; in (6) this domain is provided by a description of a recipe (a regulatory domain). In (7), the situation is a little more complex. Suppose that the utterance is produced by the person teaching the seminar to the class. Then it is mutually manifest to the interlocutors that students are theoretically in a position to interrupt as long as the teacher does not object, or the structure of the seminar allows

it. It follows that the only domains of propositions that could relevantly contribute to an interpretation of the modal verb in (7) are those concerning the teacher's own preferences and the structure of the seminar. Given two further, mutually manifest contextual assumptions, namely that interrupting the seminar as often as is needed is desirable from the students' point of view, and that the teacher has some authority as far as the seminar is concerned, (7) is felt to convey a "permission" interpretation (see Papafragou, forth. for fuller discussion).

So far I have considered so-called root possibility interpretations of *may*. I now turn to epistemic interpretations. I want to claim that what sets these apart is the fact that the domain of propositions involved in saturating the modal semantics consists of metarepresented propositions.

I have already mentioned the possibility of storing assumptions as abstract representations, and not as truth-conditional descriptions of states of affairs in the actual (or an alternative) world. It seems that humans are capable of entertaining attitudes towards propositions *qua* propositions, and manipulate them inferentially in the context of other abstract representations which are cognitively available. This happens when we consider evidence for and against a given hypothesis: a proposition is "held up", inferentially processed and ultimately rejected or accepted with a certain degree of endorsement (Sperber, 1997). A similar situation, I believe, arises in the case of epistemic modality, which involves the relation between an abstract proposition and the domain provided by a thinking agent's repertory of beliefs (again, handled with an emphasis on their representational properties rather than their representational content). Epistemic modals range over propositions which are not meant to be treated as picking out a state of affairs, but as capturing an "epistemic object" (Sweetser, 1990).

What is crucial for an epistemic interpretation to arise is the presence in context of a mutually manifest assumption to the effect that the speaker lacks some piece of presently available information which could provide evidence for or against the proposition *p*. Consider (8):

(8) Susan may be John's sister.

It is clear that the proposition _p[Susan is John's sister] has a determinate truth value; the speaker, however, lacks conclusive evidence as to whether the proposition is true or false. Knowing that her data-base of beliefs is in this respect incomplete, the speaker cannot take this data-base to be an accurate description of real-world states of affairs - therefore, she should be taken to communicate that *p* is compatible with a domain of propositions available specifically to her. The resulting interpretation is naturally epistemic.

Root and epistemic interpretations are not always easy to distinguish; as relevance theory would predict, indeterminate cases arise when it is not necessary for the hearer to retrieve a description of the domain of propositions used to saturate modal semantics - i.e. when root and epistemic interpretations would achieve a roughly equal (and equally accessible) range of contextual effects. In other cases, it is possible to differentiate between them by a couple of diagnostics. Firstly, modals on their epistemic interpretations cannot be questioned, doubted/rejected, accepted as a fact by the addressee, or form the complement of a factive predicate or a verb of telling

(Lyons 1977, p. 799). The responses in (10) are appropriate only on a root reading of (9):

- (9) More people may join us later.
- (10) a. Is that so?
 b. I don't believe it.
 c. I agree.
 d. It is surprising that more people may join us later.
 e. She told us that more people may join us later. (not reported speech)

Secondly, epistemically interpreted modals cannot occur in antecedents of conditionals: (11) is unacceptable. (12), however, is fine on a root interpretation: what it conveys is that the speaker won't come to the party in case there is an (objective, factual) possibility of Paul's getting drunk:

- (11) ?If that blonde may be Jack's wife, then we should not mention his trip with his secretary to her.
- (12) If Paul may get drunk, I'm not coming to the party.

The differences between root and epistemic interpretations can be explained by the metarepresentational account in the following way. Assume that (9) is interpreted epistemically: in her response, the speaker in (10) can only be taken to express an attitude (of wondering, disbelief, endorsement, etc.) towards the descriptive content of the proposition p [More people join us later], not towards the attitude which the previous speaker has expressed with respect to p . The proposition p will then be "disquoted" from its metarepresentational environment in (9) and used by the next speaker with a certain attitude. No such process need take place in the root reading of (9): in her response, the speaker in (10) can pick out the whole of the conceptual content of (9) and express an attitude to it. Similarly, it is only in (12) - and not in (11) - that *may* conceptually contributes to the claim the utterance makes about the world: in (11) the modal simply indicates the speaker's private stance on the content of the embedded proposition p [That blonde is Jack's wife].

The utterances in (10) and (11) - (12) can be used to illustrate the claim that epistemic uses of modals do not contribute to the main proposition expressed by the utterance (cf. Lyons, 1977; Traugott, 1989; Bybee and Fleischman, 1995). Rather, the function of the epistemic modals seems to be to fine-tune the interpretation of the embedded proposition. For instance, in the case of the epistemic interpretation of (9), the basic truth-conditional content of the utterance is the embedded proposition p [More people join us later]; the hearer has again to disembed this from its metarepresentational environment and treat it as the base explicature of the utterance. What the modal does is indicate the speaker's attitude towards this base proposition, thereby constraining the sort of effects p may give rise to (by altering the range of evidence that can be used for/against p from factual to epistemic, *may* will mainly yield implications of reduced commitment).

4 CONCLUSION: IMPLICATIONS FOR LEXICAL SEMANTICS

The proposed reanalysis of the root-epistemic distinction should have wider implications for the arguments standardly used in favour of polysemy accounts within

lexical semantics. For instance, it is often assumed that root meanings have to be somehow more basic than epistemic ones since they appear earlier in language acquisition (Stephany, 1979/1986; Sweetser, 1990); this is supposed to be a specific case of a broader phenomenon where the basic meaning of a polysemous item is also the one which is acquired first. An account of epistemic modality in terms of metarepresentation could offer a new explanation for the later emergence of epistemic uses of modals in language acquisition without appealing to polysemy. Here, I will give a very brief sketch of how such an explanation might go.

The ability to entertain and use metarepresentations may be seen as a manifestation of a broader metarepresentational capacity which has been called “theory of mind” in psychology (Carruthers and Smith, 1996; Wellman, 1990). On the theory of mind hypothesis, part of the human cognitive mechanism is the ability to know one’s own mind as such, i.e. to reflect on one’s mental contents and processes and to accommodate the results in a coherent commonsense theory about the mental world. The existence of metarepresentations and of a metarepresentational mechanism is obviously independent of specific linguistic data, although the appearance of certain linguistic constructions and items such as mental terms (*know*, *think*, *believe*, etc.) has been instrumental for the construction of the theory of mind hypothesis. Now this ability appears in a fuller form around the fourth or fifth year of age, after which mental representations are increasingly being employed in explanatory accounts of human thought and action.

If my hypothesis about the metarepresentational character of epistemic modality is correct, it follows that there can be no genuine epistemic modals before the threshold of 4;0 or 5;0 years. By contrast, root interpretations are free to appear earlier, since they do not have similar constraints. This prediction is borne out: as the developmental literature demonstrates (Perkins, 1983; Wells, 1985; Stephany, 1979/1986; Shepherd, 1982), not only do epistemic modals appear later than root ones, they also appear around or later than the fourth birthday. The empirical correlation between epistemic modal data and other manifestations of our metarepresentational capacity renders the present account stronger than previous polysemous approaches: recall Sweetser’s (1990) proposal for treating epistemic meanings as metaphoric projections of root meanings. Since the metaphorical analysis entails a specific developmental ordering, one should be able to offer some independent evidence that the onset of the acquisition of metaphor coincides with that of epistemic modality; to say simply that the manifested root/epistemic ordering vindicates a metaphorical analysis of modals is to beg the question (for further discussion, see Papafragou in press/a, b, forth.).

I do not have the time to go into the arguments from historical linguistics often used to support the existence of polysemy. In the case of modals, semantic change has typically yielded a development of root meanings into epistemic ones, a development which has been taken to signal a “rise in subjectification” in language (Traugott, 1989; 1995). The link between subjectivity and epistemic modality is made transparent in the metarepresentational analysis I have suggested; since such an analysis has a concrete psychological basis, it could also make predictions for the comprehension and use of epistemic modals by individuals with an attested deficit in their metarepresentational capacities (as is assumed to be the case in autistic subjects, see Leslie, 1995). With support from such diverse sources, a unitary semantic account of modals together with independently motivated assumptions about the use of metarepresentations in humans,

might serve as a starting point for narrowing down the scope of polysemy in natural language, and for reassessing the shape of lexical semantics.

Acknowledgements

Thanks to Deirdre Wilson, Neil Smith and the members of the Relevance Theory Reading Group in the Department of Linguistics at UCL for comments and suggestions on the topic of modality. Research for this paper was supported by grants from the A. G. Leventis Foundation, the Newby Trust and the Gilchrist Educational Foundation.

REFERENCES

Bybee, J., and S. Fleischman, Eds. (1995). *Modality in grammar and discourse*. Benjamins, Amsterdam.

Carruthers, P., and P. Smith, Eds. (1996). *Theories of theories of mind*. Cambridge University Press, Cambridge.

Coates, J. (1983). *The semantics of the modal auxiliaries*. Croom Helm, London & Canberra.

Groefsema, M. (1995). *Can, may, must and should*: a relevance-theoretic account. *Journal of Linguistics* 31, 53-79.

Kratzer, A. (1981). The notional category of modality. In: *Words, worlds and contexts* (H. - J. Eikmeyer and H. Rieser, (Eds.)). de Gruyter, Berlin.

Leslie, A. (1995). Pretending and believing: issues in the theory of ToMM. In: *Cognition on Cognition* (J. Mehler and S. Frank (Eds.)), 193-220. The MIT Press, Cambridge, Mass. & London.

Lyons, J. (1977). *Semantics*. 2 vols. Cambridge University Press, Cambridge.

Palmer, F. (1990). *Modality and the English modals*. 2nd ed. Longman, London.

Papafragou, A. (forth.). *Modality and the semantics-pragmatics interface*. University College London Ph.D. thesis.

Papafragou, A. (in press, a). Modality and semantic underdeterminacy. In: *Recent developments within relevance theory* (A. Jucker and V. Rouchota, (Eds.)). Benjamins, Amsterdam.

Papafragou, A. (in press, b). The acquisition of modality: Implications for theories of semantic representation. *Mind and Language*.

Perkins, M. (1983). *Modal expressions in English*. Frances Pinter, London.

Premack, D., and A. J. Premack (1994). Moral belief: form versus content. In: *Mapping the mind: domain specificity in cognition and culture* (L. Hirschfeld and S. Gelman (Eds.)), 149-168. Cambridge University Press, Cambridge.

Sanford, A., L. Moxey and K. Paterson (1994). Psychological studies of quantifiers. *Journal of Semantics* 10, 153-170.

Shepherd, S. (1982). From deontic to epistemic: an analysis of modals in the history of English, creoles, and language acquisition. In: *Papers from the 5th International Conference on Historical Linguistics* (A. Ahlqvist (Ed.)), 316-323. Benjamins, Amsterdam.

Sperber, D. (1997). Intuitive and reflexive beliefs. *Mind and Language* 12, 67-83.

Sperber, D., and D. Wilson (1986/1995). *Relevance: communication and cognition*. Blackwell, Oxford.

Stephany, U. (1979/1986). Modality. In: *Language Acquisition* (P. Fletcher and M. Garman (Eds.)), 375-400. Cambridge University Press, Cambridge.

Sweetser, E. (1990). *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure*. Cambridge University Press, Cambridge.

Traugott, E. (1989). On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language* **65/1**, 33-55.

Traugott, E. (1995). Subjectification in grammaticalisation. In: *Subjectivity and subjectivisation: linguistic perspectives* (D. Stein and S. Wright (Eds.)), 31-54. Cambridge University Press, Cambridge.

Wellman, H. (1990). *The child's theory of mind*. The MIT Press, Cambridge, Mass. & London.

Wells, G. (1985). *Language development in the preschool years*. Cambridge University Press, Cambridge.