

THE ANALYSIS OF THE ENGLISH "MODAL AUXILIARIES" AS MAIN VERBS

P. Collins

University of Sydney

This paper contains two sections. In the first, various arguments will be presented in favour of treating the "modal auxiliaries"¹ of English as main verbs taking sentential complements in underlying structure. In the second part, several grammatical analyses which have been proposed for the modals will be examined and revisions which may be necessary for these suggested.

[.] The status of the modals in the grammar of English has never been satisfactorily established. I wish to show that their traditional description and formalization as "auxiliary" or "helping" verbs gives us little understanding of their role in a grammar.

Within the framework of transformational grammar the first proposal for handling the modals was Chomsky's (1957) phrase-structure rule :

1) Aux \rightarrow Tense (modal)(have+en)(be+ing).

This analysis of the auxiliary, which appears essentially unchanged in Chomsky (1965) has been accepted in much of the transformational literature and commonly appears as a standard example in textbooks. What it does is capture the regularity in English verb and makes possible the formalization of certain syntactic processes.²

However what was attempted in this, admittedly elegant, formulation was merely a characterization of morphological and syntactic observations. In fact, the Chomsky analysis assumes a large degree of independence between semantics and syntax, e.g. it claims that when 'past' is chosen in the auxiliary, the modal 'may' should be realized as "might." However "might" rarely has a semantic past tense interpretation :

- 2) Sue might try to catch the 6.30 train.
- 3) Carol might study French.

Furthermore, Chomsky expresses the belief in Aspects (p.17), that ambiguity is not introduced in the base. However, we are frequently faced with sentence ambiguities directly

dependent on the interpretation of individual modals. Hence the sentence:

4) Fred can carry the basket.

has three distinct interpretations:

5) (a) Fred is able to carry the basket.

(b) It is possible for Fred to carry the basket.

(c) Fred is permitted to carry the basket.

Each interpretation, it may be noted, implies a different set of grammatical relations. In (a), (the "ability" sense), Fred is subject of both "can" and "carry". In (b), (the "possibility" sense), Fred is subject of "carry", but neither subject nor object of "can", which has a sentential subject, "Fred's carrying the basket". In (c), (the "permission" sense), Fred is object of "can" and subject of "carry".

Several recent papers have questioned the assumption that Chomsky's rule is part of English grammar. It is argued, and I will present evidence to support this view, that the modals are not generated in the phrase-structure, but should be removed from their role as verbal auxiliaries and analysed as predicates of higher sentences.

1.2 The first argument is based on the behaviour of negation in simple modal sentences. A negative particle that occurs directly after Tense or Modal in a simple surface sentence in English is normally an instance of sentence negation. However a negative occurring after a modal may not be associated with the whole sentence. Thus, although sentential negation is exhibited in (6),

6) Kathy cannot talk to her mother.

in (7) the negative, which is also positioned after the modal, is associated not with the whole sentence but with the remainder of the predicate phrase to the right of the modal:

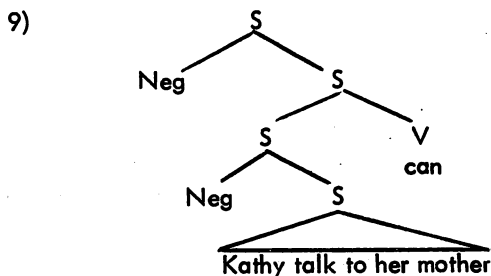
7) Kathy can not talk to her mother, if she wishes.

Both negatives appear in a simple surface sentence such as (8):

8) Kathy can't not talk to her mother, and stay on friendly terms with her.

where one negative intuitively negates "can" and the other negates "talk".

By the Chomskyan analysis the first clause in (8) would be generated with one underlying sentence since there is, in this analysis, only one main verb, "talk". Specific rules would be needed to optionally negate the nodes M and V. This unnecessary complication, which would thus have to be dealt with in an 'ad hoc' manner by an analysis based on rule 1, would follow naturally from a higher verb analysis, (the underlying structure would consist of two separate sentences, both of which are separately negated).



1.3 A further argument which indicates that modals should operate as independent constituents in underlying structure, is their capacity to be modified by various types of adverbial clauses.

- 10) John could, because of his exceptional strength, lift huge weights.
- 11) I may play if my mother says so.

In a Chomskyan framework an ad hoc rule would be needed to associate the adverbial element specifically with the modal in (10) and (11). This significant loss of generality could be obviated if the hypothesis that M and V are under the same S node were rejected and M given full verbal status.

1.4 The final argument is based on the facts of language change.³ The Chomsky analysis and the main verb analysis imply very different things about the nature of syntactic change in the history of the English language. In Old English the modals were independent verbs, behaving similarly enough to form a special sub class of 12 irregular verbs (the "preterite-presents"), but undoubtedly members of the category, verb. These include "cunnan" ("to know how to", "be acquainted with"), "magan" ("be able to", "have the power to"), "sculan" ("to be obliged to", "have to"), "motan" ("be allowed to", "have to") from the preterite of which we have "must", and "agan" ("to own", "possess") from the preterite of which we have "ought". Several examples from the Old English heroic poem Beowulf follow:

- 12) ; men ne cunnon
 hwyder hel-runan hwyrftum scripað (1. 162)
 ("it is not known whither such demons go wandering.")
- 13) ; a mæg God wyrcan
 wunder æfter wundre, wuldres Hyrde (1. 930)
 ("God always has power to work wonder upon wonder,
 (God) the guardian of glory")

In Middle English, too, we frequently find the modals used as independent verbs. The following are taken from Chaucer:

- 14) In alle the ordres foure is noon that kan
 So muchel of dalliaunce and fair language.
 (General Prologue, l. 210)
 ("In all the four orders there is noone who knows so much
 about dalliance and sweet talk").
- 15) Oure Lord hath yeve it me; our Lord hath biraft
 It me; right as oure Lord hath wold, right so it
 Is doon; blessed be the name of oure Lord!
 (B.Mel., l. 2190-2195)
 ("Our Lord has given it to me; our Lord has taken it from
 me; just as our Lord has willed, so it is done; ...").

The Chomsky analysis implies that the preterite-present verbs disappeared from the grammar and were regenerated as auxiliary elements to the main verb. That language change could be so loosely constrained is unlikely.

Thus, although the modals in English unquestionably form a class with certain idiosyncratic properties, their traditional description as "auxiliary" or "helping" verbs gives us little understanding as to their role in a grammar.

2.1 Before proposing any new analysis of the modals, it may prove useful to review some basic facts about them. The modals in English fall into two basic classes, which are usually termed "epistemic" and "root". In their epistemic meanings the modals express the speaker's state of knowledge or belief or opinion about the proposition. In their root senses the modals modify the surface structure subject of the sentence, indicating his volition, obligation, ability etc.

The meaning and use of two modals in their typical epistemic and root senses are illustrated in the sentences in (16) and (17) respectively:

- 16) (a) They may come on Tuesday (= "It is possible that...")
 (b) The solution must be true (= "It is necessarily the case that...")
- 17) (a) Peter may come if he wants ("Peter is permitted...")
 (b) You must set the example ("You are obliged to...")

There is a certain amount of syntactic evidence for this root-epistemic distinction.

While use of the perfect aspect in the complements of the epistemic modals indicates their past tense, suppletive forms are generally used to indicate past tense for the root modals:

- 18) The solution must have been true.
 19) You had to set the example.

Progressive forms are normal after epistemic, but not root modals. (20) would thus be unacceptable if interpreted in a root sense:

- 20) * They may be coming on Tuesday.

Root modals cannot be followed by stative verbs. Hence (21) (a) and (b) are unacceptable if interpreted in a root sense:

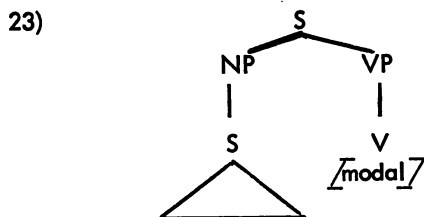
- 21) (a)* Fred may seem to be a fool.
 (b)* Fred must resemble a gorilla.

The subject of root, but not epistemic modals is generally required to be animate. Hence the unacceptability of (22) if interpreted in the root sense:

- 22)* The earth must circle the sun.

I will later argue that these apparently idiosyncratic and haphazard facts are related and can be accounted for in an analysis within the theoretical framework of generative semantics. First, let us look at some of the deep-structure analyses proposed for the root and epistemic modals.

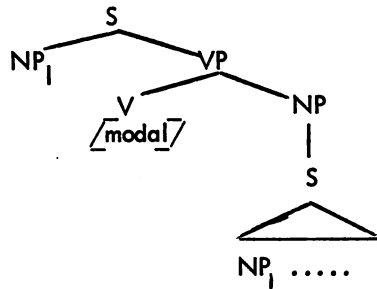
2.2 The following underlying structure has been proposed for the epistemic modals:



This structure is that generally accepted for the class of intransitive, subject-embedding verbs including 'seem', 'happen', 'appear', as well as adjectives such as "certain", "probable", "possible" etc.

The root modals have been regarded by most proponents of the main verb analysis as transitive verbs taking object-complementation, (like "refuse" and "try"), in the underlying structures of which the subjects of the higher sentence and its complement are identical.

24)



The choice of these two structures appears to have been based primarily on the behaviour of roots and epistemics with respect to the passive transformation. It was observed that (25) (a) and (b) will be paraphrases if epistemic "may" is understood, but if the root ("obligative") sense is understood, then the meanings differ:

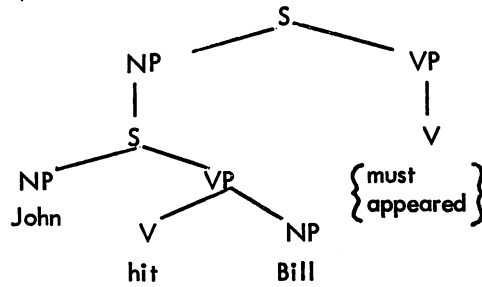
- 25) (a) John must hit Bill.
- (b) Bill must be hit by John.

These facts follow from the proposed deep structures for the same reason that (26)(a) and (b) are paraphrases, while (27)(a) and (b) are not:

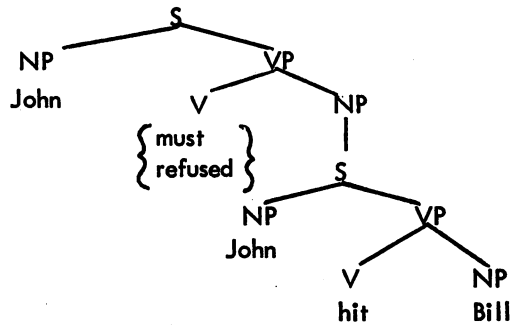
- 26) (a) John appeared to hit Bill.
- (b) Bill appeared to be hit by John.
- 27) (a) John refused to hit Bill.
- (b) Bill refused to be hit by John.

The synonymous active/passive pairs containing epistemic modals and verbs like "appear" arise from the same deep structure, (28). The non-synonymous active/passive pairs containing root modals and verbs like "refuse" arise from the different deep structures, (29) and (30):

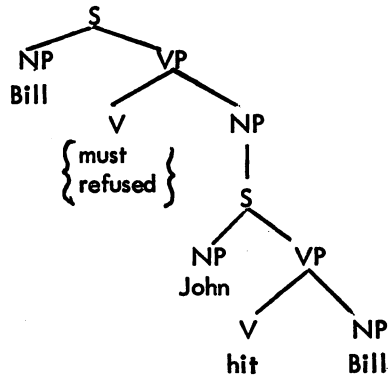
28)



29)



30)



However, while (23) appears to be well motivated as the deep structure for epistemic modals, the principal argument for (24) is unsatisfactory. Newmeyer (1969) notes that there are cases where sentences containing root modals can be passivized with no change of meaning. The synonymy of (31) (a) and (b) cannot be accounted for by the transitive analysis.

31) (a) Visitors may pick flowers .

(b) Flowers may be picked by visitors .

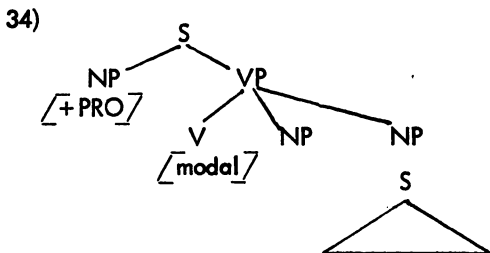
Newmeyer furthermore observes that (24) fails to capture grammatical relations adequately. For example it claims that in (29) above, John is the subject of "must" as well as "hit". But this counters our intuitive feeling that in (25)(a) John is a patient rather than an agent. Sentence (32), a paraphrase for (25)(a), supports this intuitive feeling:

32) John is obliged to hit Bill.

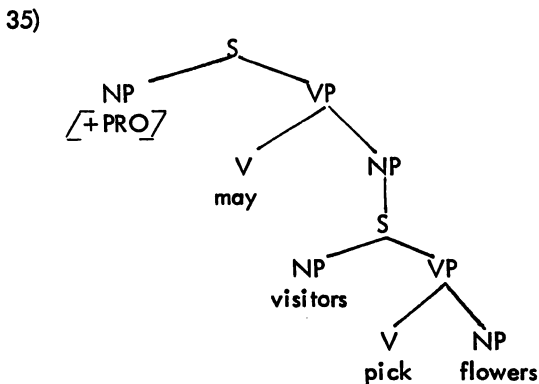
(33), in turn, is the active version of (32):

33) Someone obliges John to hit Bill.

Newmeyer proposes a slightly altered version of (24) in which the surface subject becomes the logical object of the modal, and the content of the subject NP is left partially semantically open:

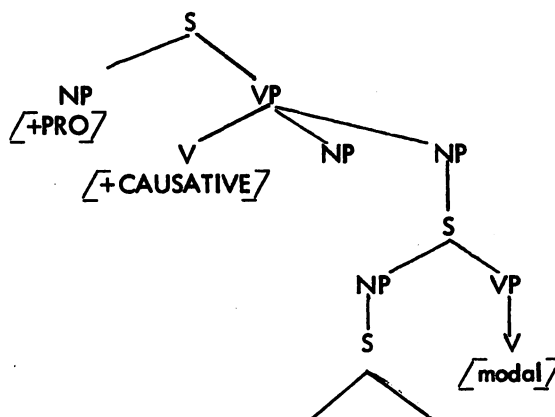


This suggestion appears both to capture logical and grammatical relationships, and account for the fact that in some cases passivising a root modal changes the meaning and in some cases it does not. Thus (25)(a) and (25)(b) would occur in underlying structures like (34) with indirect objects John and Bill respectively. (In the first case EQ UI alone would apply, in the second PASSIVE followed by EQ UI). On the other hand (31)(a) and (31)(b) would both have the underlying structure, (35), where there is no indirect object and different surface structure alternatives arise through the application/non application of PASSIVE prior to SUBJECT RAISING.



However neither of the analyses thus far presented, argues Newmeyer, can account adequately for all the semantic relationships. There is a certain semantic relationship between root and epistemic modals which would be difficult to capture by either (24) or (34). The root modal can often be interpreted as the epistemic modal with an added causative or affective sense. Epistemic "may", for instance, means approximately "possibility", root "may" means "to cause to be possible in a characteristic way" or "to allow". Epistemic "must" expresses "logical necessity", while root "must" means "to cause to be necessary" or "to require". Newmeyer proposes the following underlying structure for root modals:

36)



Newmeyer's defence of his proposed structure is somewhat inconclusive and he closes his book with the statement that: "since there has never been a moderately satisfactory analysis of the root modals proposed, the nature of their structure is still an open question."

Newmeyer's suggestions are indeed worth following up. We will argue that the semantic unit "cause", which unarguably plays an important role in the formation of the English verbal system as a whole, also figures in the modal system of English. We may establish with a fair degree of certainty that a causative process underlies the use of the root modals. Consider the following:

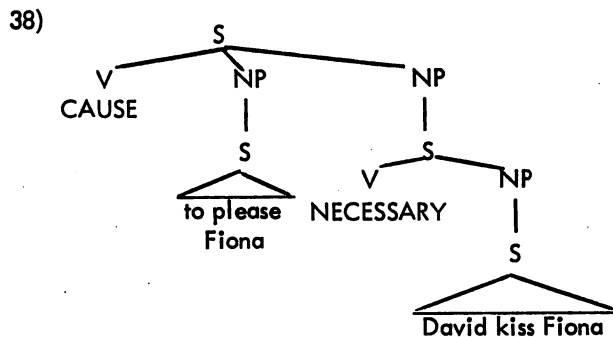
37)(a) David kisses Fiona to please her.

(b) David must kiss Fiona to please her.

While there is no explicit connection between the principal and subordinate clauses in (37) (a), in (37) (b) the root modal implies a causation (the first clause communicates that David is obliged to kiss Fiona and the second makes explicit what is causing the existence of this obligation).

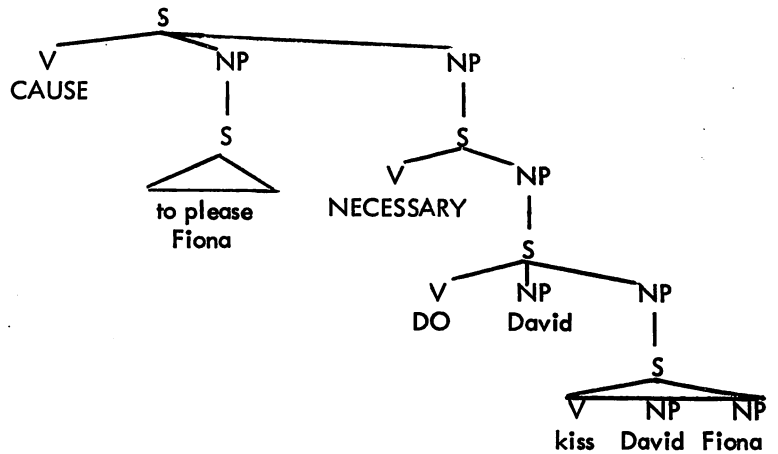
Given that root "must" expresses that there is a requirement for a certain state of affairs to be actualized and that it implies that there is something which causes the existence of this obligation,⁴ let us propose an analysis of root "must" within the theoretical framework of generative semantics. We will assume that the meaning of a verb, and of the sentence containing it, is decomposable into a structured set of semantic units - these, not morphemes, will constitute the ultimate units of semantic structure.

To formalize the "obligation" aspect of the meaning of root "must" let us introduce the semantic element NECESSARY, and for the causal aspect let us introduce the element CAUSE. (CAUSE represents a relation between two events and not, as it has often been assumed, for example in (36), a relation between a person and an event). The underlying structure for (37) (b) will be:⁵



This analysis of "must" is still unsatisfactory for two reasons. It was mentioned previously that the verb in the complement of a root modal cannot normally be stative and that root modals normally have a human (surface structure) subject. These facts are related and may be formalized by the introduction of the single semantic unit DO, whose subject is normally human and whose object expresses the result of his intentional action.⁶ Our analysis of (37) (b) will accordingly be modified as:

39)



After EQUI has applied, successive applications of the transformation of PREDICATE-RAISING result in a tree in which we have a configuration of lexical predicates CAUSE-NECESSARY-DO for which the surface lexical item "must" is inserted.

This structure accounts, more adequately than (38) above, for the problems associated with the active and passive forms of sentences containing a root modal. Now (25) (a) and (b), in their root readings, will have underlying structures similar to (39), in which John and Bill respectively are the subjects of DO.

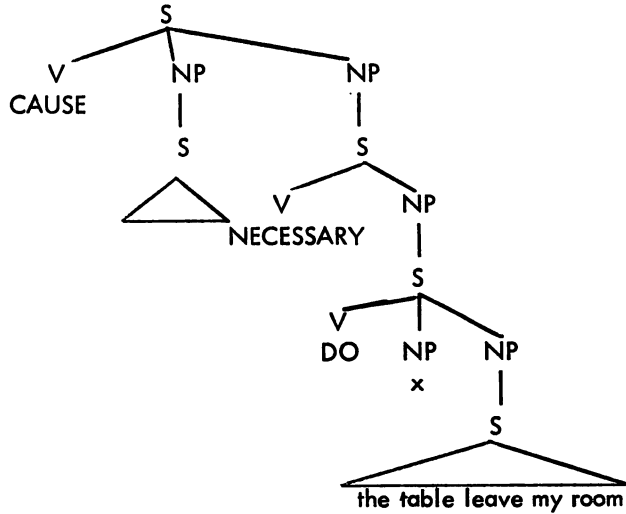
(31) (a) and (b) above will both derive from a structure in which the sentential subject of CAUSE is unspecified and the subject of DO is "visitors".

Consider also the following sentence:

40) There must be peace and quiet.⁷

In (40) the surface structure subject of "must" is neither the logical subject nor object in the verbal reading involving the modal. The subject of DO will thus be unspecified:

41)



I admit that sentences like :

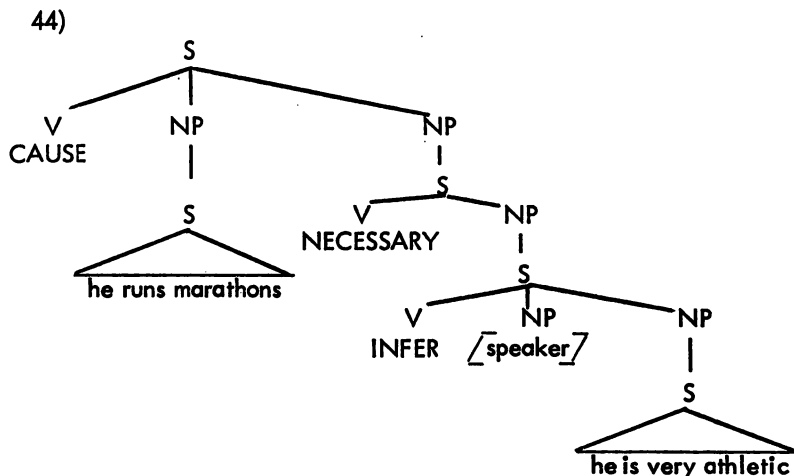
- 42) A river must be very deep, before I will sail my yacht in it.

pose difficulties for this analysis - nobody can do anything under normal circumstances, intentionally or otherwise, such that a river is deep. (I suspect however that the surface structure of this sentence conceals a quite elaborate underlying structure).

Let us now re-examine the analysis of epistemic modals given earlier, in (23). We have seen that sentences containing such modals express an inference or belief on the part of the speaker concerning the state of affairs referred to by the simple sentence without the modal. To account for this epistemic sense we introduce the semantic unit INFER, whose subject will be human, and whose object will be a state of affairs.

This inference, it would seem, results from a process of causation aimed not at the actualization of a state of affairs (as with the root modals), but the belief of the speaker.⁸ As with root modals the subject of causation (now representing the grounds upon which the speaker bases his inference) may be expressed in a subordinate clause. The underlying structure for (43) will thus be (44):

- 43) He must be very athletic, for he runs marathons.



Thus with epistemic "must" strong circumstantial evidence imposes an "obligation" on the speaker to infer something.⁹

Let us return, finally, to the root modals and reconsider our analysis in the light of the "performative analysis". It was the Oxford philosopher, J.L. Austin, who established the idea of "performative" utterances (which contain verbs which are such that the utterance of the sentence itself constitutes a certain action), which he wished to separate from "constative" utterances (merely descriptive utterances subject to criteria of truth-falsehood).¹⁰ Sentences such as:

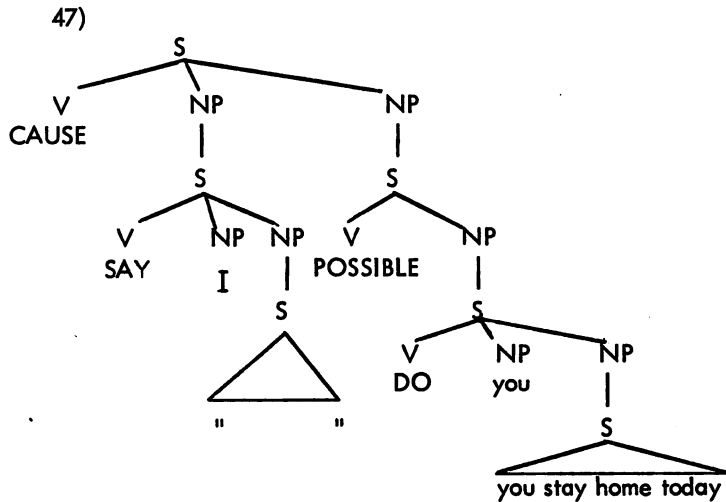
45) I permit you to stay home today.

were called "explicit" performatives by Austin, since the main verb corresponds directly with the name of the action being performed (in this case, "permitting"). However, there are also performative utterances where the action is not named, but is inferred. These are labelled "primary" performatives by Austin - their intrinsic vagueness leads him to say they "resemble" explicit performatives. Thus (46) "resembles" (45):

46) You may stay home today.

The justification for relating "may" to "permit" rather than to another of the great number of performatives resides in the applicability of the following paraphrase to both (45) and (46): "I make it possible for you to stay home today (by saying it is possible for you to stay home today)". The essence of this paraphrase is an implicit claim that "permit" and root "may", when used performatively, are causatives of "possibility", where the means of causation is speaking.¹¹

These observations can be accommodated into the analysis we have presented above in (39). Now we may suggest a more exact specification, at least for the root modals in their performative uses, of the nature of the subject NP of CAUSE. This NP will dominate a semantic unit which we shall call SAY, and which indicates that the issuing of the utterance constitutes the performance of the action, an action which one could scarcely perform, at least with such precision, in any other way. We allow the surface structure of the quoted sentence to be embedded as the sentential object complement of SAY.¹² The underlying structure for (46) will be:

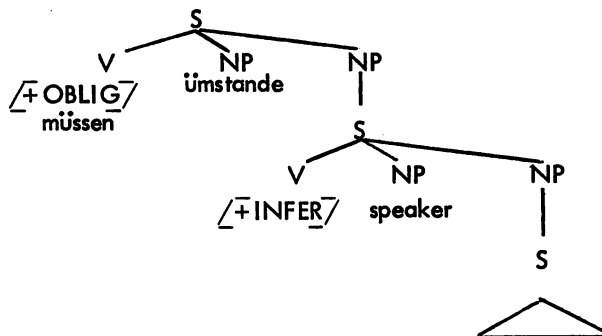


Thus, operating within a generative semantic framework with a small set of semantic units we can explain apparently diverse syntactic peculiarities of the English modals and capture some of their deep semantic regularities.

FOOTNOTES

1. By "modal auxiliaries", hereafter referred to simply as "modals", I mean that closed class of verbs including the four "paired" modals ("can", "may", "shall" and "will"), which have accompanying past tense forms ("could", "might", "should" and "would"), and the four "unpaired" modals ("must", "ought", "need" and "dare").
2. One syntactic process whose formalization is made possible by the above rule is that of "question-inversion", which has the effect of interchanging the subject NP with the segment consisting of tense plus the first auxiliary.

3. While we cannot construct a synchronic rule solely on the basis of such facts it is instructive to compare constructions at an earlier stage of the language analogous to those we are examining in the present language. Hypotheses of syntactic change are formulated from such comparison. We may suspect the validity of any hypothesised change which appears to be at variance with most of the others which have been proposed.
4. Dakin (1970) observes that modals always imply a causation.
5. In the analysis given in (38), and those following we accept the proposal made by McCawley in a number of recent papers that an S is analysable as V, followed by one or two arguments, and that no VP nodes occur in deep structure.
6. See, in particular, the arguments in Ross (1971).
7. This sentence is discussed in Lakoff (1972).
8. In a pioneering article in the semantics of the modals Boyd and Thorne, (1969), analyse the difference between root and epistemic "must" in terms of illocutionary force. In their view, root "must" has the illocutionary force of "a statement of a necessary demand", epistemic "must", that of "a necessary statement". But surely, as we have argued, the obligation or necessity is incumbent on the speaker to infer, rather than state, something.
9. In an article which proposes a case grammar analysis of the German modals, Calbert (1971) considers a structure rather similar to (44):



Although Calbert analyses "müssen" simply as a transitive verb, it is clear from his paraphrase of the sentence, "He must be sick", as "circumstances force one to think that he is sick" (p.111), that causation is here involved, in his view.

10. See, in particular, Austin (1962).

11. It would appear from the potential insertability of the adverb "hereby" into performative utterances that they all involve a causation where the means of causation is speaking. Interpreted literally, "hereby" indicates that the utterance of the sentence is, as it is said, the instrument effecting the action denoted by the performative verb.
12. Kiparsky and Kiparsky (1971) discuss the possible necessity of allowing quoted sentences to be embeddable in a generative grammar.

BIBLIOGRAPHY

- Austin, J.L. (1962) How to do Things with Words. (Ed) Urmson, J.O. Oxford. O.U.P.
- Boyd, J. and Thome, J.P. (1969) "The Semantics of Modal Verbs". Journal of Linguistics. No. 5 : 57-74.
- Calbert, J.P. (1971) "Modality and Case Grammar". Working Papers in Linguistics, No. 10 : 85-132.
- Chomsky, N. (1957) Syntactic Structures. Mouton and Co., The Hague.
- Chomsky, N. (1965) Aspects of the Theory of Syntax. M.I.T. Cambridge, Mass.
- Dakin, J. (1970) "Explanations". Journal of Linguistics. No. 6 : 199 - 214.
- Kiparsky, P. and Kiparsky, C. (1971) "Fact", in Steinberg, D. and Jakobovits, L. (Eds) Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, and Psychology - New York : C.U.P.
- Lakoff, R. (1972) "The Pragmatics of Modality". (Ed) Peranteau, P.M. et al. Papers from the 8th Regional Meeting of the Chicago Linguistic Society.
- Newmeyer, F.J. (1969) English Aspectual Verbs. Studies in Linguistics and Language Learning No. 6. Seattle : University of Washington.
- Palacas, A.L. (1970) The Place of Modal Verbs in the Grammar of English with special emphasis on 'can', 'may' and 'must'. Indiana University Ph.D. University Microfilms, A Xerox Co., Ann Arbor, Michigan.
- Ross, J.R. (1967) "Auxiliaries as Main Verbs". Studies in Philosophical Linguistics. Series I : 77 - 102.
- Ross, J.R. (1971) "Act" in Davidson, D. and Harman, G. (Eds) Semantics of Natural Languages. Reidel.