A Procedural Account of Some English Modals

Maija Leena McKinnon

Ph.D.
University of Edinburgh
1985

Declaration

I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself unless otherwise indicated.

Edinburgh, 7 June 1985

Maija McKinnon

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Keith Brown, Barry Richards, and Jim Miller, for their advice and encouragement during the preparation of this thesis.

I am also grateful to many other people connected with the School of Epistemics: to Chris Mellish and Henry Thompson for helpful suggestions, to Mark Steedman for comments and criticism, and to Ken Beesley and Lee Humphreys for many elucidating discussions.

I wish to thank my husband for his support and his interest in my work.

ABSTRACT

This thesis investigates modality within a procedural framework. It is mainly concerned with dynamic modality; procedures are defined for the dynamic senses of can and must. This is done within a small query program for a limited domain, but the procedures for the modals are general and would be valid in other domains.

The program uses a planner and in addition employs two pragmatic concepts, the focus of a question, which is related to the conversational goal of the questioner, and the possible higher level goals that the interlocutors may have.

Dynamically modalised questions can be used to ask about the possibility or necessity of an action with respect to such an (explicit or implicit) goal. In determining the degree of modality the procedures proposed here test whether carrying out the action is compatible with achieving the goal.

A proposal is made to extend the system to handle deontic modality.

Epistemic modality and its relation to root modality are discussed in a more general way. Finally, some theoretical consequences of handling modals in a procedural framework are discussed.

CONTENTS

1. INTRODUCTION

1.1. The aim and the scope of the study

The aim of this study is to investigate how some aspects of modality could be handled in a language understanding system. It deals with the semantics and pragmatics of some English modals, and attempts to give a procedural definition of them. Root modality will be the main concern, although epistemic modality is also discussed briefly.

The main part of the study proposes a procedural treatment for the root senses of modals like must and can and the related semi-modals HAVE TO, and BE ABLE TO. In the discussion of epistemic modality and epistemic senses of modals like must and may, consideration is given to how far the procedural treatment proposed for the root modals is relevant or applicable to the treatment of epistemic modals.

The method of investigation is to work out how these modals could be defined in a natural language question answering system. In the system the domain of discourse is limited and the form of questions is restricted to ones that expect a modalised answer. The modality expressed in the questions and answers is dynamic modality. An example program deals with such a restricted fragment.

Although many of the proposals and conclusions in this thesis relate to the construction of the example program, the aim is to keep the discussion on a broader basis and consider other aspects of modality as well as those that come up in the limited fragment of English. Following the example of Isard (1974: 233), the program is intended as a "small working model" to illustrate the proposals made. The program is not to be understood, however, as a model of how actual speakers behave, but as an example of one way of dealing with the problems that are encountered in trying to represent the meanings of modals. In other words, no psycholinguistic claims are made.

The domain of discourse chosen for the study was that of house construction. The task of the system is to answer questions about the timing of the different jobs within the construction schedule. It is assumed that it is important to finish the construction of the house by a certain date. The timing of a given job is restricted by constraints such as which jobs must have been completed before that job can be started, and the starting or completing of some other jobs may be dependent on that one being completed first. The timing of all jobs may thus affect the completion time of the house. The task of providing information about the optimal schedule can thus be seen as a useful one. Questions like the following might occur in this domain:

(1) Can the painters start on the 29th?

(2) Must the painters start on the 29th?

(3) When can the electricians start?

(4) When should the rough plumbing be finished?

Details of the domain and an outline of the system will be given in chapter 2.

The scope of the study is narrow in the sense that only present tense forms of the modal verbs are considered; past tense forms or expressions in which the main predication refers to past time are not discussed. The modals relating to the expression of futurity, will and shall, are also excluded. The question of the relation of modality to time and tense is therefore outside the scope of this study. Although the scope is limited, it has the advantage that it was possible to investigate the problems of dynamic and deontic modality in much more detail.

This is a study about the semantics and pragmatics of modal verbs. Except for the fragment of syntax that is needed for the parser of the example program, it will not deal with the grammar of modal verbs or the details of how they are used in speech. These areas have been extensively covered in the literature[1].

1.2. Some kinds of modality and some issues

The issues, categories and terms that are relevant to the study of modality and modal verbs will be introduced relative to one important investigation of the English

modals, Palmer's (1979) 'Modality and the English modals'. In this section, references will be made to other studies as well, but other research will be more fully discussed in later sections.

Palmer's study is a detailed taxonomy of the English modals and a comprehensive, corpus-based description of their uses. The starting-point for Palmer's discussion are the modes distinguished by von Wright (1951: 1-2); these include the alethic, the epistemic, and the deontic mode. Of these, the epistemic and the deontic mode are the ones that have received most attention in studies on modality in natural language.

Epistemic modality expresses the speaker's assessment of the truth of a proposition on the basis of what he believes or knows, as in

(5) He may be in his study.
(6) He must be in his study.

Deontic modality has to do with whether an action or an event is permitted, obligatory, or prohibited:

(7) May I come in?
(8) You must close the door.
(9) You mustn't leave the door open.
(10) All books must be returned in June.

In deontic modality, the permission, obligation, etc., may be imposed either by the speaker (cf. examples 7-9), or it may be imposed by some external source and only stated or reported by the speaker (cf. example 10). The 'source' who or which imposes the obligation, etc., is usually called the

deontic source, although Palmer does not use this term.

The modality that occurs in logic, in the form of the logical operators POSSIBLE (the proposition is true in some possible world) and NECESSARY (the proposition is true in all possible worlds) can also be expressed in natural language; this is the alethic mode. It is expressed by the same modals as epistemic modality: for example

(11) If a prime p is a factor of the product ab, then p must be a factor of either a or b.

but it is not as common as epistemic modality. Palmer leaves alethic modality entirely out of his study. It will not be discussed in this thesis either, but it is mentioned here as some references will be made to it.

In addition to these categories, Palmer introduces a new category, called dynamic modality. Dynamic modality has to do with whether an action or an event is possible in the sense of whether it is feasible: whether it can be done, or whether it is necessary in order to achieve something. Dynamic modality and the justification for treating it as a (to some extent) separate category from deontic modality will be dealt with in detail below.

The term 'degree of modality' is a central notion in this thesis. It will be used to refer to the degree of commitment to the truth of the proposition, or to the desirability or otherwise of the proposed action; i.e. necessity, possibility, and impossibility, as well as obligation, permission and prohibition are degrees of

modality. In natural language there are usually considered to be intermediate degrees of modality as well, as in

(12) It will probably rain tonight.
(13) You should do your homework.

In these, the degree of modality is somewhere between possibility and necessity, or between permission and (strict) obligation. (Palmer does not discuss this point, but cf. Halliday 1970, Lyons 1977, Hakulinen & Karlsson 1979.)

Let us clarify the concept of dynamic modality, and the extent to which it differs from the other modalities. Dynamic possibility is usually expressed by CAN or BE ABLE TO. Two kinds of dynamic possibility can be distinguished: circumstantial possibility, and possibility in the sense of ability (Palmer 1979). Examples of the ability sense are found in (14) and (15), where the possibility or feasibility of the action depends on the subject's ability to perform the action:

(14) I can swim.
(15) They can't speak English.

In contrast, (16) and (17) are examples of circumstantial possibility, where the possibility or the feasibility of the action depends on circumstances:

(16) We can't go swimming tomorrow.
(17) Well, I'll see what can be done and give you a ring. (Palmer 1979: 72)

The following situation, taken from the domain of house construction used in this thesis, exemplifies circumstantial

dynamic modality, possibility in (a) and necessity in (b):

(18) Can the painters start on the 29th?
a. Yes, they can.
b. They must start then if the house is to be finished on time.

Possibility or necessity of the painters' starting at the given time depends on some circumstance, for example, the goal of completing the house by an agreed date. Circumstantial dynamic modality will be a central concern in this thesis.

These examples have intuitively different meanings from examples of epistemic modality. Compare example (19) (dynamic) to (20) (epistemic):

(19) The painters can start on the 30th.

(20) The painters may start on the 30th.

(19) says that a course of action is feasible; (20) says that something may happen.

In this thesis a distinction will be made between deontic and dynamic modality. The term 'dynamic modality' will normally be used in a sense close to Palmer's 'circumstantial dynamic modality'. Making a distinction between dynamic modality and deontic modality can be justified on the following grounds.

The term 'deontic' is traditionally used to refer to modalities that express duty, obligation and permission, both by linguists and by philosophers. In examples of dynamic modality, such as (18) and (19), duty, obligation and

permission play no part; the degree of modality depends on circumstances.

In the kind of deontic modality where the deontic source is something other than the speaker (of the modalised utterance), it may be another person, whose deontic statement is being reported, or it may be rules, regulations or a moral or ethical code. According to Lyons (1977: 834, 840) a deontically modalised sentence contains the sense "X obligates that p", or "X permits p", where X is the deontic source.

In dynamic modality, too, there may be something equivalent to a 'source'. X is the circumstance or cause that makes an action possible or necessary or impossible, but the circumstance is different in nature from a deontic source, in that in deontic modality the source is assumed to have authority over whether the action is to be carried out whereas with dynamic modality the source or circumstance is not linked to authority. However, in dynamic possibility there are also examples in which there is no such X: an action may be possible without the possibility being linked to some single circumstance (see section 4.1.1.).

In this thesis dynamic possibility will be seen in terms of plans for actions. For example, in order to answer a question like

(18) Can the painters start on the 29th?

the respondent has to find out whether there is a plan for

the painters starting on that date and whether the preconditions for carrying out that plan are satisfied. Dynamic necessity will be shown to be linked to a goal or goals of one of the participants of the discourse. Ability will be regarded as something which can be handled under dynamic modality, since it can be seen as one precondition for a plan.

While a distinction will be made between deontic and dynamic modalities, it is accepted that they are related, and that they differ from epistemic modality in a systematic way: deontic and dynamic modalities have restrictions on the time reference and the nature of the predicates; no such restrictions apply to epistemic modality. The term 'root modality' will be used as a superordinate term to cover both deontic and dynamic modality. The main distinction is thus between epistemic on the one hand, and root modality (deontic and dynamic modality), on the other. Some main differences between the two categories are listed below; the links and differences between them, and whether the procedures for the two could be combined will be discussed in detail in chapter 6.

The following are generally considered to be the main differences between epistemic and deontic modality (cf. Lyons 1977, Palmer 1979). Epistemic modality has to do with propositions; it expresses whether the truth of the proposition is possible or necessary. Deontic modality involves actions: it expresses whether an obligation, permission, prohibition or exemption to perform an action

exists. Moreover, the action must be (in relation to the modality) in the future: there cannot be an obligation, etc., to perform something in the past. This restriction does not apply to the propositions of epistemic modality which may be past, present or future. Since dynamic modality has to do with events and actions, it differs from epistemic modality in just the way that deontic modality does.

It has also been claimed that epistemic and deontic modality differ in that deontic necessity "typically proceeds, or derives, from some source or cause" (Lyons 1977: 824); i.e. from the so-called deontic source. Epistemic necessity is thought not to be linked to a source in the same way. However, epistemic necessity also can be thought of as deriving from a cause in a sense: an utterance like 'It must be raining' is based on some argument that the speaker has in mind; this argument is the 'cause' of the conclusion.

Leech and Coates (1980) add to these differences the following points as evidence for the claim that there is a clear-cut boundary between the epistemic and root senses of the modals <u>may</u> and <u>must</u>. Firstly, they mention syntactic/semantic criteria such as the differences in the scope of negation between the different senses (e.g. 'He may not go' has the senses (i) "He is not allowed to go" and (ii) "It is possible that he will not go" - i.e. <u>may</u> has wide scope with the deontic sense and narrow scope with the epistemic sense); secondly, they claim that within root modality, there are continua from one root sense to

another; epistemic modality has no such continua between senses.

A central theoretical question that is addressed in many studies on modals is whether they are monosemantic or polysemantic. Can a modal can be described as having a basic meaning, of which other senses are variations, or must the different senses be regarded as independent? Ehrman (1966) and Perkins (1980) take the monosemantic view. For example, the root and epistemic senses of the modal must, as in

(21) You must close the door.
(22) It must be raining.

can in their view both be expressed in one definition, such as "X requires that p", with the difference between the senses lying in the nature of X (cf. Ehrman 1966: 67, Perkins 1980: 64).

Most other studies implicitly accept the view that modals are polysemantic (e.g. Leech 1971, Quirk et al. 1972, Palmer 1979), although Palmer (1979) notes that the basic meaning hypothesis has more justification with some modals than with others. While MAY is a modal which appears to have two distinct senses (op.cit.:10), CAN seems to have a basic meaning. The basic meaning of CAN would be dynamic possibility (op.cit.: 61), given that it is possible to combine the permission sense of CAN with its possibility sense.

Coates (1980) explains the meanings of modals in terms of fuzzy set theory: each modal has one or more core

meanings, and the core meanings may have a 'skirt' and a peripheral area around them. Those uses of a modal which do not belong to the core sense but are related to it belong to the skirt or the periphery. For example, the core meanings of can are "ability" and "permission" and both of these have a skirt where the sense is that of "possibility". Here the two skirts overlap. The aim of this approach is to explain the "messiness" of modals; there is frequently indeterminacy between the different senses (cf. chapter 7).

The only studies to my knowledge that take a procedural approach to the English modals and to modality are Isard and Longuet-Higgins (1973) and Isard (1974). The studies deal with conditionality and the modal verbs in the context of playing the game of tic-tac-toe, and they give procedural definitions of the (senses) of modal verbs that are used in that domain. The modals included in their program are may, must, will and can, and they are defined in terms of possible games which are continuations from the reference situation.

Although the definitions are both insightful and successful in the domain of the game, they are limited in that, with the exception of can, the sense that occurs in this context is that of objective epistemic, or possibly alethic modality. The definitions of may and must are given in terms of possible games, that is, whether p is true in some or every possible game, which is exactly parallel to the definition of the logical modal operators

POSSIBLE and NECESSARY. Will and can obviously have no equivalent in logical modality; however, they occur in a very neutral or objective sense in the context. Can is the only one that expresses root modality. Their procedure for can involves looking for a strategy for carrying out the action given in the proposition (their definition of can will be discussed in chapter 4).

The research done for this thesis can be seen as a continuation of the approach taken by Isard and Longuet-Higgins. The proposals that will be made here for dealing with can are built on their work: defining dynamic possibility of actions in terms of plans for those actions is close to Isard and Longuet-Higgins' definition of can in terms of strategies. However, a different range of modality will be covered here. Their program does not cover the kind of circumstantial modality that may depend on the interlocutors' goals. The definitions of can and must proposed here will therefore be different from theirs.

1.3. Lyons (1977): Semantics (Chapter 17)

Lyons (1977: Chapter 17) represents a different approach to modality from what has been adopted in this thesis. As it is an important contribution, it will be reviewed here in some detail in order to see how it is located relative to the framework chosen here, and also because Lyons explores areas like the contrast between root and epistemic modality, which will be discussed in the later chapters of

this thesis.

Lyons' discussion of modality is based on his tripartite analysis of the logical structure of utterances in general. This system is adopted from Hare (1970): the three components are the phrastic, the tropic and the neustic. The phrastic is the propositional content of the sentence. The tropic expresses what kind of speech act the sentence is characteristically used to perform. The neustic expresses to what extent the speaker is committed to the factuality or desirability of what is expressed in the propositional content of the sentence (Lyons 1977: 749). For example, statements are described as follows: "I say so (neustic) - it is so (tropic) - that p (phrastic)"; yes-no questions as "I don't know - it is so - that p"; and commands as "I say so - let it be so - that p". Lyons uses symbols to express the values that the neustic and the tropic may take: . stands for unmarked neustic and tropic, ? for questions, and ! for mands. By using these symbols, the structure of a statement is ". . p", that of a yes-no question "? . p" and that of a mand ". ! p". A request is analysed as "? ! p".

In connection with modality this tripartite system is used to explain the differences between subjective and objective modality, and the affinity of subjective deontic modality to mands.

Both epistemic and deontic modality are divided into subjective and objective categories. If the utterance (e.g. 'Alfred must be unmarried') contains subjective epistemic

modality, the assessment of the truth of p is the speaker's subjective view; if objective epistemic, the possibility or necessity of p is an objective fact. Lyons claims that this distinction can be made in theory, even if not always in practice.

But if it is intended that the objective sense of for example 'Alfred must be unmarried' is the result of objective, logical reasoning, (cf. Lyons 1977: 798), it could equally well be used as an example of alethic modality, and it is not quite clear what the difference between alethic and objective epistemic is. The problematic status of objective epistemic modality is acknowledged by Lyons when he admits that it lies between alethic modality and subjective epistemic modality, "and might be assimilated to either" (op.cit.: 798).

Subjective epistemic modality is analysed as having a qualified neustic, i.e. the speaker's commitment to the truth of p is qualified: the neustic is "I think it possible" instead of "I say so"; in objective epistemic modality the neustic expresses commitment but the tropic is qualified. Subjective is thus "poss . p" and objective ". poss p", although Lyons also allows the possibility of having the modality in the phrastic, as part of the propositional content: ". . poss(p)".

It would seem that, certainly from the hearer's point of view (Lyons' analysis seems to be from the speaker's viewpoint), there is perhaps a scale in epistemic modality in that the hearer may assess the comment of the speaker

about the factuality of p to be more or less subjective. This assessment may be made on such grounds as the role of the speaker and how much the hearer thinks the speaker knows about the subject (cf. Lyons' (1977: 799) example 'It may be raining in London' as said by a meteorologist or by a layman). But there is a clear boundary between epistemic and logical modality, as Karttunen (1972) has shown (see section 4.3.1.).

The difference between objective and subjective deontic modality is that in the subjective case the deontic source is the speaker; in the objective case it is something or somebody other than the speaker. Lyons considers the possibility of analysing subjective deontic necessity as ". ! p". This would make the semantic structure of modalised statements like 'You must open the door.' identical to that of commands, e.g. 'Open the door!', and, inevitably, Lyons concludes that these are not identical semantically and that in modalised statements the obligation has to be part of the phrastic. He proposes analysing the modalised statement as ". ! that !p exists" (Lyons 1977: 840). In other words, 'You must open the door' is an assertion which expresses the existence of an obligation, but it also expresses the illocutionary force of a mand.

Permission, prohibition, and exemption are analysed in the same way, as assertions expressing the existence of a permission, etc., using negation of either the modality or the main predication to achieve the desired description

(these are based on the relationship of possibility and necessity through negation: nec p = ∽ poss ∽ p; poss p = ∽ nec ∽ p; (Lyons 1977: 789, 840):

You may open the door. ". ! that ∽!p not exist"
You needn't open the door. ". ! that !p not exist"
You must not open the door. ". ! that ∽!p exist"

The difference between objective and subjective deontic modality is that in the objective case the tropic is unqualified, "it is so" rather than "it be so". Lyons at first gives the structure "I say so - it is so - that X obligates (permits) p" (p. 834). However, he later expresses doubt as to whether the ambiguity between subjective and objective deontic modality really is a linguistic ambiguity at all; and whether all deontic statements should in fact be analysed as having the primary illocutionary force of precisely statements (p. 841). But, to sum up the theory as it is presented, the difference between subjective and objective deontic modality is whether the tropic is . or !. The primary structures in natural language, at least in English, are "poss . p" (subjective epistemic possibility) and ". ! p" (subjective deontic necessity) (p. 845).[2]

It is not clear whether he intends the subjective/objective modality dichotomy to be applied to modalised questions, or whether questions are neutral as to subjectivity/ objectivity. It is difficult to see how this distinction could apply to questions. The first difficulty is that the subjectivity or otherwise cannot apply to the

speaker's (= questioner's) utterance (he is asking for the hearer's response). The second difficulty is that, if the subjectivity or objectivity is thought to apply to the expected answer, the questioner may be asking the question without any presumption as to who or what the deontic source is, or whether something is the respondent's view or an objective view.

The central feature of Lyons' analysis is that it attributes the differences between modalities partly to the propositional content and partly to the illocutionary component of utterances. By contrast, the procedures for modality proposed in this study do not include illocutionary force at all, although they do make use of both semantic and pragmatic concepts. The pragmatic notions that have to be accounted for, e.g. the intentions behind utterances, are richer than what can be included in the tripartite system proposed by Lyons. The system proposed here includes one of these pragmatic concepts, the goals of the interlocutors, including both the general, non-linguistic goals as well as conversational goals.

It seems clear that any utterance has to be labelled as to its function, that is, whether it is a statement, a question or a mand. This labelling can be seen as part of the semantic level and as the result of a direct mapping from the syntactic structure (declarative, interrogative and imperative sentences) (cf. Leech 1983: 114). When an utterance is used indirectly to convey some other illocutionary force in addition to these, pragmatic

principles (principles of conversation, general problem-solving methods, knowledge about the world and about people's behaviour) have to be applied.

This study differs form Lyons' in its approach and to some extent in the range of modalities it covers. Lyons gives a logical analysis of utterances including propositional and speech act elements; this study attempts to find procedures that would deal with modalised questions and answers. Lyons' treatment covers alethic, epistemic and deontic modalities and draws a distinction between subjective and objective modality in the latter two. This thesis will cover subjective epistemic modality (but not objective epistemic or alethic modalities) and subjective and objective deontic modality, in the senses that Lyons uses these terms. The main concern here will be dynamic modality (not discussed by Lyons).

1.4. Procedural semantics

As the approach taken in this study, procedural semantics, differs from that of most other studies on the semantics of modals, this section aims to clarify what it is and what its methods and advantages are.

Procedural semantics defines meanings in the form of algorithms. The method typically used by advocates of this form of semantics is to incorporate these algorithms in a natural language question-answering program; in this way it is possible to test whether they work in practice and so

establish whether they are at least possible candidates for defining the meanings of given expressions. This approach offers several other advantages for semantic investigation. It forces the investigator to develop the proposals in detail and prevents possible oversights in the application of rules. It prevents one from developing over-complicated theories or theories which shift difficulties into other areas, which are themselves often unsolved. It also forces one to consider the analysis of utterances both from the speaker's and the hearer's point of view, rather than purely abstractly; for example, phenomena like ambiguity, or determining what goals are relevant in the context, are different from the two view points. At the same time it shows what is common in production and comprehension. All these advantages derive from the concreteness and explicitness of the approach.

The procedural approach sees language as goal-oriented, functional behaviour. Both semantics and pragmatics have to be brought in, but it is still possible to make a distinction between what belongs to each sphere, e.g. what part of the meaning of a modal has to be general, applicable to all contexts, and what has to be derived form the context in which the utterance is spoken. For example, the procedure that answers dynamically modalised questions takes into account the goals that the interlocutors may have, since the degree of modality often depends on such goals. This is part of the general procedure for the semantics of dynamic modality. However, finding what the

goal or goals are in a particular context has to be done by using pragmatic principles.

The most important feature of the procedural approach is of course that it enables one to test whether the theory developed can be made to work in practice. It also gives new perspectives on other people's theories: in this case, on theories about the semantics of modals, such as the question of whether modals are monosemantic or polysemantic.

One of the conclusions drawn in this thesis is that neither the strictly monosemantic nor the strictly polysemantic view of the modals is helpful as a basis for a procedural account of the modals. It will be shown, on the one hand, that there is a high-level similarity between dynamic and epistemic modality (a similarity which may even be shared by alethic modality). On the other hand, there are important distinctions between epistemic modality and root modality (including dynamic modality). These differences emerge with clarity because of the explicitness required for the procedural treatment. They lead to the conclusion that separate procedures are needed for dealing with epistemic modality and with root modality; however, it will be proposed that the epistemic procedure should have access to and be able to use the root modality procedure as part of its inference making process when appropriate (but not vice versa).

1.5. Organisation of thesis

The organisation of the rest of this thesis is as follows: Chapter 2 is an outline of the system proposed for handling modalised questions. Chapter 3 introduces the pragmatic concepts that will be employed in the proposed system. Chapters 4 and 5 give a detailed account of the dynamic and deontic modalities and the procedures developed for handling them. Chapters 6 and 7 discuss the relation of epistemic modality to root modalities and the question of to what extent modals can be said to be mono- or polysemantic. A discussion of the relationship of semantics and pragmatics in the procedures for modals is also included in chapter 7.

2. THE DOMAIN OF DISCOURSE AND AN OUTLINE OF THE PROPOSED SYSTEM

2.1. The domain of house construction

In order to be able to investigate what the procedures for the modals should be, a restricted domain of discourse was chosen for the study.

The domain is that of a schedule for the construction of a house. The construction process involves a number of jobs, which have to be done in a certain order. The duration time of each job is known beforehand. The jobs, their immediate predecessors and their duration times chosen as an example are given in Figure 1.[1]

On the basis of the information given in Figure 1 it is possible to construct a network (Figure 2), which shows the sequence of jobs, and given that the goal is to finish the house in the shortest possible time, the time for the earliest and latest possible starting times for each job can be calculated.

For some jobs there is a margin; for instance job k could be started any time between day 25, which is the shortest time required to complete all the jobs that have to precede job k (day 25 = its earliest starting time), and day 27, which is got by subtracting the durations of all the jobs that follow job k from the finishing time (day 27 = the latest starting time of job k).

Job Name	Description	Immediate Predecessors	Time (days)
a	Excavate	-	4
b	Pour concrete foundations	a	2
c	Erect frame and roof	b	4
d	Lay brick work	c	6
e	Install drains	b	1
f	Pour basement floor	e	2
g	Install rough plumbing	e	3
h	Install rough wiring	c	2
i	Fasten plasterboards and plaster	g,h	10
j	Lay flooring	i,f	3
k	Finish plumbing	j	2
l	Joinery	j	3
m	Finish roofing	d	2
n	Paint	k,l	3
o	Finish electrical work	n	1

Figure 1 : House construction

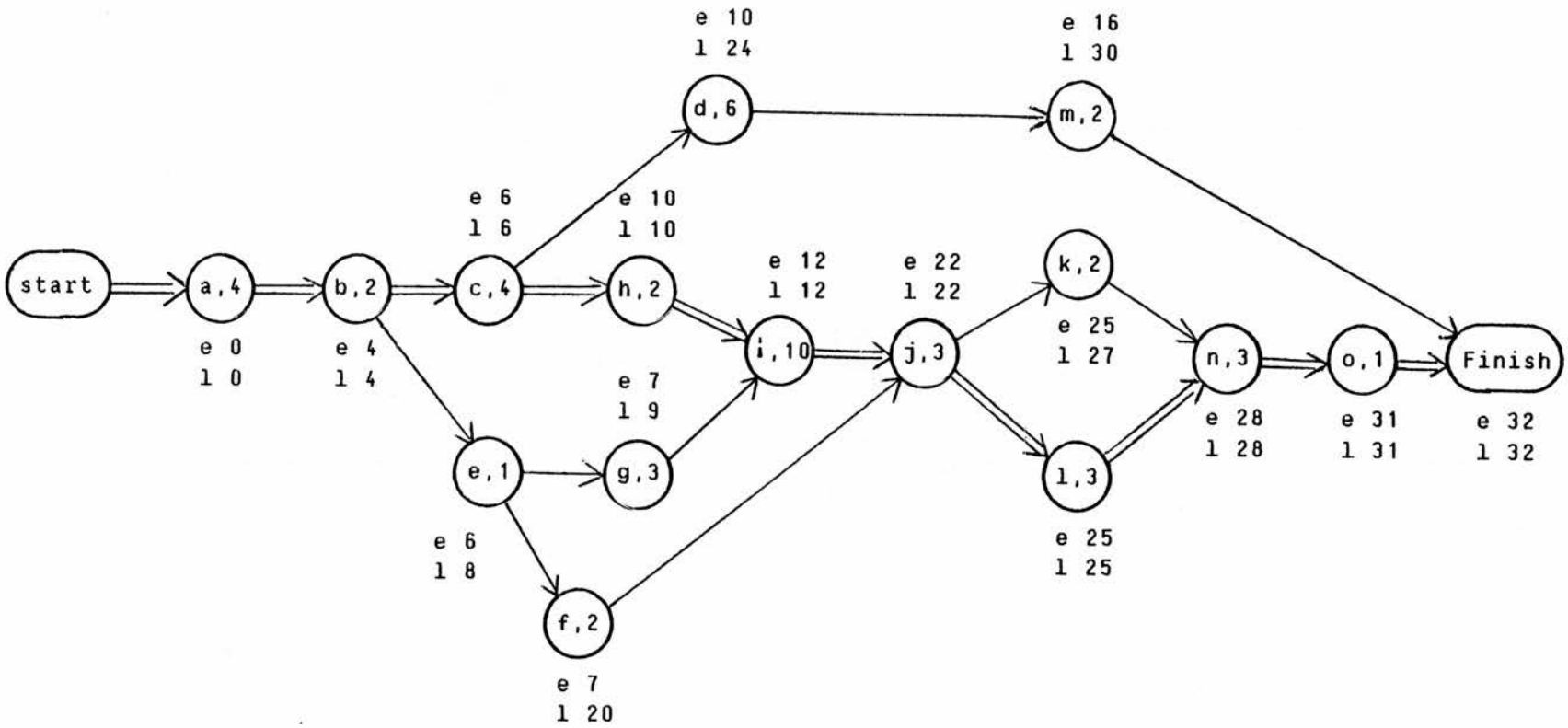


Figure 2 : House construction: sequence of jobs

The numbers within the nodes refer to the duration times of the job in question
e = earliest starting time
l = latest starting time
Double line arcs mark the critical path

For some jobs the earliest and latest starting times are the same, that is, there is only one possible starting time. These jobs form a path in the network, which is called the critical path. No job which is on the critical path can be delayed without delaying the completion of the house. (Critical path problems are a well-known type of problem to computer and management scientists. There are programming methods to find out the earliest and latest starting times for each job and to find the critical path, given the information about immediate predecessors and duration times for each job.)

Assuming that we had an 'expert' or a system that had all the information given in the network described above and which could answer questions about scheduling the construction work, one could ask the system questions like the following:

(1) When can the plumbers start the rough plumbing?

(2) Can the plasterers start on the 9th?

(3) When should the painters start?

(4) Must the painters start on the 28th?

Examples of the system's answers would be[2]:

(A1) a. They can start on the 7th.
b. They can start between the 7th and the 9th.

(A2) a. Yes.
b. No, because the plumbers haven't finished by then.
c. No, if the house is to be finished by day 32.
d. Yes, in fact they must start then.

(A3) a. They should start on day 28.

(A4) a. No, they can start between the 26th and the 28th.
b. No, they can also start on the 27th.
c. Yes, they must.

The only constraint on the schedule in this simplified network are the order of jobs and the time constraints. Other parameters one might want to add are for example availability of workmen at certain times, supplies of materials, and so on. If constraints like this are taken into account, a question like

(5) Can the bricklayers start on the 12th?

could be understood as a query about whether they can from the point of view of the time network for the completion of the house, or whether they have a free slot in their diary, or whether the bricks will have been delivered by then.

The questions and answers one gets in this domain illustrate dynamic modality. They also illustrate the way in which dynamically modalised questions and responses are often interpreted as being about possibility or necessity in relation to some specific point of view (cf. Austin 1961, Kratzer 1977, Cresswell 1979). This and other pragmatic problems will be discussed in the next chapter.

2.2. The example program

The example program can answer questions like the following:

(6) Can the painters start on day 6?

(7) Must the painters start on day 6?

(8) When can the painters start?

(9) When must the painters start?

The program is written in Prolog and consists of a parser, a planner, a database for the planner, and routines for determining the possibility or necessity of an action and for giving the result as an appropriately formulated answer to the question asked by the user. The last mentioned component, the analyser of the degree of modality for the proposed action, is the most important part of the program. The function of both the parser and the planner is to serve as tools for the analyser, and they are not necessarily the ideal ones for this domain.

This section gives an overall view of how the example program and its components work. There is a listing of the program and more detailed notes in the Appendix.

2.2.1. The parser

The parser can deal with sentences like

(6) Can the painters start on day 6?

(7) Must the painters start on day 6?

(8) When can the painters start?

(9) When must the painters start?

(10) The painters can start on day 6.

(11) The painters must start on day 6.

(12) Can't the painters start on day 6?

The parser takes the input sentence and produces a simple semantic representation (SR); for example,

Sentence: Can the painters start on day 6?

SR: question(yesno, can(starttime(painters,6)))

(The sentence is yes-no question and the propositional content is expressed in a simplified form, "can(starttime(painters,6))", which is adequate for the present purposes. The meaning of can is represented as "can" and the meaning of must as "must" in the SRs; they could be represented as "feasible" and "necessary"; cf. section 7.2.)

The aim in writing the parser was to analyse as accurately as possible verb phrases containing present tense modal auxiliaries. All other areas of syntax and semantics are dealt with as simply as possible, extracting only the minimum information necessary in the specific domain.

The framework for the parser is that of definite clause grammars, and it is written using the method described in Pereira and Warren (1980: 252-253), which produces a semantic representation as its output. The semantic

representation is built up from the SRs of the constituents: as the parser processes each constituent it assigns it a SR; these SRs are passed on to the left hand sides of rules and thus eventually to the top level.

The present parser also produces other information about the input sentence: it has a rudimentary system of assigning some constituent as a discourse focus of the sentence. This facility is needed especially when answering questions. For example, in the question

(13) Can the painters come on day 6?

the focus might be either 'the painters' or 'on day 6', and the respective answers would accordingly be different:

(A13) a. No, but the electricians can come then.
b. No, but they can come on day 7.

In reality, determining what is in focus depends to a large extent on the pragmatic context and the intentions of the interlocutors. The problem of focus will be discussed in chapter 3.

In the present program, the focus is always given; the system does not in any sense deduce it from the linguistic or the extralinguistic context. The focus is assumed to be the time adverbial unless otherwise indicated (in which case it must be indicated by a special symbol in the input question).

The output of the parser is then the semantic representation together with an indication of what is in

focus; and the focus is either the time or a marked element.

The grammar used in the parser is a phrase structure grammar based on ideas drawn from Gazdar, Pullum and Sag (1982). In the process of adapting the fragment of grammar which was needed to process the modals and the sentence types relevant in the restricted domain, it was expanded and modified considerably. In order to make the grammar more concise and the parser more efficient, the grammar was changed from a context-free to a context-sensitive grammar.

The parser can handle both narrow and wide scope negation in declarative as well as interrogative sentences. If requested, it will output all possible readings of a negated sentence. However, the system as a whole cannot answer negated questions; because of restrictions of the Prolog language and of the planner, this was not attempted.

Details of the parser are explained in the Appendix.

2.2.2. The planner

A program that attempts to answer questions containing modals needs a planner. The general purpose planner WARPLAN, written by Warren (1974) was used here to demonstrate the responses that are needed from the planner. However, it is the interface between the output (the SR of the proposition) given by the parser, and the planner that defines the meaning of modals like <u>can</u> and

must. If this interface is not independent from the domain and from the planner, it loses its generality: it is no longer a general definition of can or must but of can or must in a specific domain.

WARPLAN is a STRIPS-like planner[3], which has operators, preconditions, delete lists and add lists. The operators correspond to actions and the preconditions are conditions which have to be true before the action can take place. The delete and add lists represent the effects of the operators when they are applied to states of the world. Operators may require the action of other operators to achieve their preconditions.

The main difference between STRIPS and WARPLAN is that, unlike STRIPS, WARPLAN is complete in the sense that in theory it will find a solution to any solvable problem. But WARPLAN is not very efficient, which limits its application in many practical cases. (For differences between STRIPS and WARPLAN see Warren 1974.) From the point of view of efficiency WARPLAN is not the ideal planner for the house construction domain. Better efficiency would be achieved by incorporating into the planner a special purpose critical path algorithm. Such a procedure was used at an early stage of this research but it was not incorporated in the general planner as this is a technical problem which does not affect the linguistic issues. The example program uses the planner for demonstration purposes rather than to find efficient solutions to particular problems.

The example program was run using the example below,

which is much simpler than the one given in section 2.1. for the critical path network.

2.2.3. The database

The database contains the facts about the world that the planner needs, in this case facts about house construction, and the operators and preconditions that are relevant.

As all the questions that are allowed have to do with finding a time when particular tradesmen can start a particular job, the only operator that is needed is 'fixstart', which tries to find a starting time for given job.

The example that is given in the Appendix and which was used for demonstration has the simplest possible configuration in that we have only a few jobs and the jobs follow each other (i.e. each job has only one predecessor). The time for the beginning of the work is given; the starting times for the actual jobs are not fixed. The predecessor and the duration of each job is defined in the database.

The precondition for the operator 'fixstart' defines the computation for finding out whether a given time T is a possible starting time for a job J. This involves looking at the starting time of the predecessor of J and checking what the duration of the predecessor is. If the starting time of the predecessor of J is not known, the planner has to find that out first.

2.2.4. Determining whether an action is feasible or necessary

The example program has a routine for analysing the feasibility or necessity of an action. This routine is called when an answer is needed for questions containing <u>can</u> or <u>must</u>, e.g.

(6) Can the painters start on day 6?

(7) Must the painters start on day 6?

It takes as input the semantic representation of the question, uses the planner to test the feasibility of the action, and produces an appropriate response. The feasibility of the action is tested by consulting the schedule for the construction, and it is related to the goal of finishing the house by an agreed date. Thus examples (6) and (7) might get answers like

(A6) a. Yes, they can.
b. No, they can't if the goal is to be met.

(A7) Yes, they must if the goal is to be met.

This routine is the most important part of the program as well as the central part of this study, and it is described in detail in chapter four.

This program produces answers to dynamically modalised questions. It illustrates the process of interpreting the question and of producing a relevant answer. In this process the pragmatic concepts of the focus of a question

and the understood higher level goal of the interlocutors (finishing the house by a certain date), to which the modality relates, are taken into account. To some extent the program illustrates both the understanding and the production of modalised utterances.[4]

3. PRAGMATIC PROBLEMS

The purpose of this chapter is to introduce and to discuss two pragmatic problems that are encountered in and highlighted by a natural language question-answering system. These are the focus of the question, on the one hand, and the goals and intentions of the interlocutors, on the other, and the effects that these have on the production of answers to questions. It should be emphasized that both of these are general phenomena that have to be tackled in order to get a question-answer program to function satisfactorily; they are not restricted to the study of modals nor to the domain or framework chosen here.

Both of these notions will be used in the program developed here for dealing with modalised questions not only to make an answer more relevant to the interlocutor by recognising his conversational and his ulterior goals; but they will also be used in the actual procedure that determines the degree of modality. The focus of the question will be used to define a set of alternative worlds: this is necessary for establishing whether the proposed action is necessary or not (see chapter 4). The general, non-linguistic, goals of the interlocutors are important for dealing with dynamic modality and it will be shown that dynamic necessity is always and that other degrees of dynamic modality are sometimes linked to such a goal.

The focus of the question will be dealt with first: there is a preliminary section to set out the problems, and after that, a description and a discussion of some views of focus in questions in the literature. This is followed by a review of a discussion of how the interlocutors' general goals affect the answering process. We will then return to discuss these phenomena in the domain of this study.

3.1. The focus of the question

A direct can-question is very often a preliminary for arranging something, as it is in the following example.

(1) User: Can the painters start on the 29th?
System: Yes.
Yes, in fact they must if the house is to be finished on time.
No, they cannot if the house is to be finished on time.

When the answer to this type of question is affirmative, it is followed by the interlocutors agreeing to the arrangement.

In example (1) the role of the system is to give expert advice on scheduling. The system knows that the user has an overall goal of finishing the construction of the house by some time X and the question is interpreted in the sense "Is p possible in view of the goal of completing the house by X?". (This is of course only one possible interpretation of the question. In another situation, it may be intended/interpreted as relating to some other

goal.) In order to give a meaningful, relevant answer the goal has to be taken into account in the procedure that produces the answer.

We will assume that the illocutionary force of the user's utterance in (1) is that of question, i.e. a request for information (rather than a request for action). In asking the question, the user has an immediate goal of either (i) fixing a time for the painters to start work, or (ii) fixing for somebody (not necessarily the painters) to start work on the 29th.

We will use the term 'focus of the question' to refer to the element in the question that the questioner's interest is focused upon. If the questioner's immediate goal is (i) fixing a time for the painters, then the time (the 29th) is the focus of the question. If the questioner's immediate goal is (ii) fixing for (any) job to be started on the 29th, then the focus of the question is the description of the job (here described by referring to the tradesmen in question, the painters). Possible helpful answers would be different depending on what is taken to be in focus:

(i) No, but they can start on the 30th.

(ii) No, but the electricians can come then.

This kind of helpful answers are based on the conversational principle of cooperation (Grice 1975). The concept of the focus of a question is related to the conversational goal of the questioner, and in trying to find it and answer appropriately the respondent is trying to

make his contribution relevant. Leech (1983) indeed redefines Grice's "Be relevant" maxim in terms of the interlocutors' conversational goals:

> "An utterance U is relevant to a speech situation to the extent that U can be interpreted as contributing to the conversational goal(s) of [the speaker] or [the hearer]." (Leech 1983: 99)

The program for modalised questions uses both the overall goal of the user (= finishing the house by a given time), and the focus of the question, which is related to the immediate (conversational) goal of the user in his asking the question. The immediate goal is a subgoal of the overall goal.

To consider a different example, in

(2) A: I've got to go to Glasgow tomorrow. Can you drive me to the station tomorrow morning?

B: a. Yes, I can do that.
b. No, I can't because [...] . You'll have to go by bus.
c. I suppose I'll have to (because the busdrivers are on strike).

the roles of A and B might be husband and wife; A's goal is to get to Glasgow the following day, and a subgoal of that is for him to get to the station in the morning. The former is spelled out and the latter can be inferred by the hearer, the inference being based on both the information given in the utterance and the hearer's knowledge about ways of travelling and what they involve. The illocutionary force in this case is that of a request (for action) , and

the immediate goal is of course the action requested, namely to get B to drive A to the station.

This example shows that the focus of the question need not be one single constituent, but may be a larger part of the proposition. If B cannot drive A to the station and wants to give a helpful, suggestive answer, B may check 'B's driving A to the station' against some other way of getting A to the station (cf. answer (2b)). So the focus would seem to be 'you drive me'.

In a different context, however, 'Can you drive me to the station tomorrow morning?' could have the focus on some other part of the proposition. For example, if A wants to go to the station for train spotting some time (but it does not have to be tomorrow), the focus is on the time, and it would be appropriate to answer 'No, but I could take you on Friday afternoon'; but this would not be an appropriate answer to the question in (2).

In this example, knowing the questioner's goal may help to restrict the set of what might be in focus. However, the focus cannot always be predicted even when the goal is known. In example (1), where the goal is the finishing of the house at time X, the focus may still be either the job or the time.

3.2. Focus in questions: related work

This section reviews how the 'focus of a question' has been understood and treated by different scholars.

According to Quirk et al. (1972: 388) focussing is marked by placing the nuclear stress[1] on the chosen item in the sentence. In the case of yes-no questions the constituent focussed upon refers to an item of information which is unknown to the questioner (unlike the rest of the sentence). They give these examples:

(3) a. Was he a famous actor in 'those days?
("I know he was a famous actor -
but was it then or later?")

b. Was he a 'famous actor in those days?
("I know he was an actor in those days -
but was he a famous one?")

These questions are identical except that the focus falls on different elements, but this difference means that the addressee is asked to consider different aspects in each case.

Quirk et al. also claim that in order to avoid the difficulty in locating the focus of the question, one of the following strategies is generally used. Yes-no questions are either (a) kept short, introducing one factor at a time, e.g.

(4) Is this the room?
(5) Did anyone search it?
(6) Was it John?
(7) Did he do it carefully?

or (b) they are made unambiguous by grammatical and/or prosodic focussing, e.g.

(8) Is ´this the room that ´John searched carefully?
(9) Did John search ´this room carefully?

or (c) they are presumed to question the predication as a whole:

(10) Did John search the room?

or (d) they are presumed to question the last element of the sentence:

(11) Did John search the room carefully?

(Quirk et al. 1972: 53). (The viewpoint in this description seems to shift from the speaker (cases a-b) to the hearer (cases c-d).)

Except for the case in which the whole predication is questioned, the claim is that in questions containing several elements the focus is on the last element, unless some other element is marked as focus by grammatical means (example 8) or by stress and pitch (example 9).

Although Quirk et al. refer to ´nuclear´ stress being placed on the focussed item, the phenomenon in examples (3a-3b) and (8-9) seems to come close to contrastive stress. These examples contain devices that are often used for expressing contrastiveness: marked stress/pitch and cleft sentences (cf. Chafe 1976: 34-37).

One could perhaps summarise the view of Quirk et al. by saying that unless there is contrastive focus on some element in the question, then the question is either about the predication as a whole, or about the last element in the question.

It is true that there are modalised questions as well that question the whole predication, e.g.

(12) Can John search the room?
(13) Can the police search the house? (permission sense)
(14) Need the police search the house?

But with examples like

(1) Can the painters start on the 29th?

Quirk et al. would argue that if the speaker is interested in the painters, a contrastive stress will be placed on the word painters.

It is noteworthy about the examples given by Quirk et al. that (5) differs from all the rest in the following way: an affirmative answer to all the other questions (3-4, 6-11) would be 'Yes' only; an answer to

(5') Did anyone search the room?

could also give the information who it was:

Yes, John did.

Winograd (1972: 139-140) deals with this type of

question, in which the item in focus is an indefinite noun phrase. In his example

(15) Does the box contain a block?

'a block' is the element is focus. He claims that it is the grammatical fact that 'a block' is an indefinite NP that marks it as the focus of the question and which the hearer uses as a clue to find the focus. An indefinite NP represents an unknown element, and even if a mere 'Yes' or 'No' would be an adequate answer, Winograd's program operates on the assumption that since the questioner is interested in 'a block' or 'blocks' which are in the box, he is probably interested in being able to identify the unknown block or blocks if such are found. Thus the program might answer

(16) Yes, two of them: a red one and a green one.

(Winograd 1972: 140). This strategy of identifying the focus with an indefinite NP (if there is one in the sentence) and construing the question as asking for a value for the indefinite element, works well in the domain of the 'blocks world', and if there is no special emphasis on any element.

However, a question with an indefinite NP but also with a root modal, such as

(17) Can you put a block in the box?

(18) Can you get a painter for the 29th?

turns out to be less straightforward. Although these too

can be answered by 'Yes, the red one.' or 'Yes, I can get Mr Duncan.', it is possible to construe these examples in such a way that the questioner is not asking for a value to be supplied for the indefinite NP, but is only asking whether the addressee is capable of performing the action (or, in the impersonal interpretation of (17), whether the action is feasible, performed by any human agent). (17) and (18) can also be interpreted as requests for action, in which case the questioner is asking the addressee to perform the action, but again is not necessarily interested in finding an identity for the NP. If this is true, then Winograd's procedure of interpreting the indefinite NP as focus does not work in all cases.

Winograd claims that in question (15) 'the box' cannot be in focus because 'the box' would not be a meaningful answer to question (15). This seems to be saying that an element can only be in focus if it can occur in the response. This need not be the case. 'The box' may be in focus, but the reason why it cannot occur in the answer is that it refers to a known entity (repeating it would not add any information), whereas if 'a block' is in focus, the identity of the entity is not known and can be given in the response.

It is certainly possible to place contrastive focus on any other element in question (15), e.g.

(15') Does the 'box contain a block?
(as opposed to the basket)

(15´´) Does the box con´tain a block?
(as opposed to there being a block next to it)

Winograd also discusses some other syntactic devices as clues to what is in focus.

It seems that Winograd´s concept of focus is different both from what Quirk et al. refer to when they talk about the focus of a question (they are primarily interested in the functions of syntactic and intonational phenomena), and from what the focus of the question is taken to be in this study. Winograd´s view of focus is probably determined by the limited domain and language but it is too restricted since in that view focus can only fall on indefinite NPs. It restricts focus only to questions asking about the existence of an entity with a particular property. These questions are different from other yes-no questions in that they can be construed as wh-questions: they sometimes expect the addressee not only to respond with ´Yes´ or ´No´ but also to supply a value for the indefinite expression (Lyons 1977: 759). This is what Winograd´s program does: it treats these questions in the same way as it treats wh-questions.

The term ´focus´ has also been used by researchers in artificial intelligence to describe a discourse phenomenon.

Grosz (1978, 1979) uses the term ´focussing´ in the following sense: participants of a discourse focus or concentrate their attention on a small portion of what each of them knows or believes; entities that are central to the dialogue at a given point are the ones that are

focussed upon. Focussing is an active process both on the part of the speaker and on the part of the hearer. The speaker provides clues as to what he is going to focus on next; these clues may be linguistic or they may be derivable from the non-linguistic context. For communication to succeed, one has to assume that at any given time both the speaker and the hearer are focussing on the same thing(s) (Grosz 1979: 84).

Grosz' studies examine the relationship between focussing and definite descriptions. Focussing is shown to be an important factor both in interpreting and in generating definite descriptions. As a mechanism that separates those items that are currently highlighted (from all other possible ones), focussing limits the set of entities from which the entity referred to must be identified (Grosz 1979: 91).

The data for Grosz' study is task-oriented dialogues. One of the participants of the dialogue is an apprentice whose task it is, for instance, to assemble or disassemble a piece of machinery, and the other one is an expert on that subject, giving advice to the apprentice. Grosz shows how the task at hand affects focussing. She shows how the structure of the mutually known part of the plan tree (for the task) determines what is in focus, and she relates movement to different subproblems in this tree to shifts in focus.

Grosz has shown that focussing is one of the mechanisms that are needed for finding referents for definite

descriptions and for generating definite descriptions in such a way that the referents are identifiable by the hearer. However, the important question of how a shift in focus is indicated or understood in general, or for example in other, less well structured domains, remains largely open. The only clue discussed in her work is the structure of the task. It is possible that there are other means that are used in determining focus (Grosz 1979: 98).

Kaplan (1979) discusses focus in connection with 'suggestive responses': these are responses that give some additional information (in addition to the minimum yes/no response) that the respondent thinks may be relevant to the questioner. For example,

(19) A: Is there a mailbox on this block?
B: No, but there is one down the street.

The suggestive response is usually a response to a related, modified question (Kaplan 1979: 35; cf.also Allen & Perrault 1980, discussed below). The modified question is a result of varying or eliminating the focus of the original question. For Kaplan, focus is

> "roughly speaking, [...] that aspect of the question that is most likely to shift in a follow-up question" (Kaplan 1979: 35).

Determining what is in focus in a question involves plan inference (Kaplan 1979: 39). (Plan inference has been investigated by Allen & Perrault (1980), which is discussed below.)

Discourse focus in this sense has also been investigated by Sidner (1979), who discusses its role in definite anaphora comprehension.

3.3. The goals of the interlocutors

It must first be made clear in what sense the phrase 'the speaker's intentions' or 'the speaker's goals' is used here.

Intentions are often discussed in the context of speech act theory. In uttering a sentence, it is the speaker's intention to get the hearer to recognise that a particular speech act is intended (i.e. its illocutionary force), or, to quote Searle (1969: 45), "the speaker intends to produce a certain effect by means of getting the hearer to recognise his intention to produce that effect".

By his analysis, in saying

(20) Are you coming to the party tomorrow?

the speaker intends the hearer to recognise that the speaker has the intention of getting the hearer to supply the information whether the hearer is coming to the party the following day.

However, even if our example

(1) Can the painters start on the 29th?

can obviously be analysed in the same way as (20), the intention or the goal of the speaker that the hearer needs

to know in order to provide the answer for the question (e.g. finishing the house by X) has nothing to do with the illocutionary force of the utterance or the intention to make the hearer recognise that illocutionary force. Rather, we are talking about some general, non-linguistic goal that the speaker may have and which the hearer may either know or believe or infer that the speaker has or may have. This is the sense in which the word 'goal' will be used in this study.

Allen and Perrault (1980) describe a natural language program that takes external goals into account. They show how recognising the interlocutor's plan is used in giving a helpful response in question-answer dialogues.

In their example situation the participants of the dialogue are the clerk at the information desk at a railway station, and the customer who is asking for information. The approach is that of an artificial intelligence natural language question-answering system; the system takes the part of the information clerk.

When answering a question like

(21) When does the train to Windsor leave?
(uttered at the information desk in Toronto)

the system responds helpfully by giving both the time and the platform:

(22) 1600 at gate 7.

This answer is based on inferences made from the question

itself, (implicitly) from the role of the speaker and the hearer, and about the speaker's (user's) possible intentions. The system tries to reconstruct a plan that the user is likely to want to realise, and then see if there are any obstacles to the carrying out of this plan, and if so, it tries to take these into account in its response. With the above question it proceeds in the following way:

It first discovers that this is more likely to be a question linked to boarding a train rather than meeting a train. These are the only two actions the program knows about; the choice of these particular actions as possible actions to be considered is of course linked to the role of the questioner, i.e. the customer requesting information, which at a railway information desk is likely to be about boarding or meeting trains (but not, for example, about arranging drivers' shifts or about putting on extra trains); thus the roles of the clerk and the customer limits the search for the possible intentions of the user. This is implicit in the program but not discussed in the article by Allen and Perrault.

The assumption that (21) is more likely to be about boarding a train than meeting one is made by matching the question with the plans for boarding and for meeting trains and discovering how much it has in common with each of these. To describe the process in a very much simplified way, it is done on the basis that the question is about a (remote) destination and a departure time for that destination, rather than about arrival time at the station

that the interlocutors are at. (The hypothesis as to which one is the intended plan in fact goes through several stages, with various rating heuristics applied to it.)

'The train' in example (21) is taken to mean "the next train" (although, depending on the context, it could obviously refer to other things), and such a train is searched for and found, and its departure time recorded.

The program now works on the assumption that the hearer's plan is to board the next train to Windsor and that he wants the information as to when it leaves.

The next step that the program takes, however, is to check that there are no (other) obstacles for the customer in carrying out his plan. It discovers that the plan for boarding the train has the precondition that the agent be at the departure location (platform) of the train at the right time; unless the customer knows which platform to go to, this will be an obstacle to his plan. Once this obstacle has been detected, the program therefore informs the customer about the platform as well as the time, so removing the obstacle.

Allen and Perrault (1980) also show how the plan inference and obstacle detection mechanism can deal with providing helpful answers to yes-no questions. To a yes-no question answered 'No', e.g.

(23) Does the Windsor train leave at 4?

a mere 'No' would be considered an unhelpful answer. What the customer would expect would be the time the train

does leave, e.g.

(24) No, it leaves at 4.30.

When a person asks about the truth of some proposition,in the case that that proposition is false, the person is often interested in a related, true proposition (Allen & Perrault 1980: 168: cf. also Kaplan 1979: 35).

3.4. Focus and goals: tentative conclusions and open questions

In this section the points made so far about focus and goals are summarised and some examples re-examined in order to see if any conclusions can be drawn.

There are first of all, the cases in which the question has no focus, but the speaker simply wants to know the truth value of the (whole) proposition, as in

(25) Did John search the room?
(with no marked stress)

or, in the case of modalised questions, whether something is epistemically, deontically or dynamically possible, impossible or necessary:

(26) Can the police search the house?
(deontic or dynamic)

(27) Need I come?
(deontic or dynamic)

(28) Could they be on holiday?
(epistemic)

Otherwise the speaker focuses attention on a particular element in the question, as in

(29) Can you come on ´Wednesday?

There are linguistic markers of focus of this kind, such as stress and pitch and cleft sentences (Quirk et al. 1972, Chafe 1976).

However, the focus of the question in our sense (= the most likely element to be modified in a subsequent question) does not always coincide with the element that has the nuclear stress. If the modal verb itself is stressed,there may also be (another) focus within the main predication. In any case, in the system presented here, the modal verb itself cannot serve as the focus that is needed in the procedure for modals. A focus has to be found within the main predication. This is because in order to determine the degree of the modality one has to look at alternative states of the world; this is done by altering the value of the focus (see chapter 4). In

(30) Do you ´have to go tomorrow?

´have to´ is stressed and has the emphasis of the question on it, but the process for answering the question would have to look at whether ´tomorrow´ is the only possibility or not (to determine whether ´going tomorrow´ is necessary or possible). Notice also that a follow-up question to this question would be something like ´Couldn´t you stay until Wednesday?´. It is not clear whether the case in which the

emphatic stress falls on the modal verb itself is the only exception, i.e. whether in all cases in which the emphatic stress does not fall on the modal but on another element, this other element would also be the focus of the question. But since this is not known, perhaps one should keep the two concepts separate, so that we may have elements that are stressed and elements that are in focus, and these do not have to coincide.

In cases in which there is no focus (in our sense), the whole of the main predication has to be changed in order to see whether an alternative state of the world is possible or necessary. For example, with

(31) Do you have to go?

one would see whether, from the point of view of some goal, 'going (now)' is necessary, or whether staying is possible as well. (Predications without focus are a problem for the system presented here, since it can only look at an alternative state of the world by changing the focus.)

It is possible that there are other linguistic rules, in addition to the ones mentioned above, for finding the focus. For example, any syntactically optional element at the end of the sentence could be taken to be a 'default' focus, if there is no other marked focus (cf. Quirk et al. 1972: 388).

Sidner (1979) has proposed that deep case roles of NPs may play a role and that there is a hierarchy in which the case roles are considered (from the hearer's point of view) as candidates for focus.[2]

In wh-questions the focus is of course the wh-word. As to yes-no questions, one could say that trying to determine the focus of a yes-no question is comparable to trying to find a related wh-question. If the yes-no question can be turned to a wh-question it is the element that is replaced by a wh-word which is the focus of the question. Interpretations (i) and (ii) of example (1) ('Can the painters start on the 29th?') can be related to the wh-questions

(i') When can the painters start?

(ii') Who can start on the 29th?

respectively.

It is not claimed, however, that focussed questions are another form of wh-questions. As was pointed out above, there are some yes-no questions, especially ones asking about the existence of an entity/entities that can be construed as wh-questions. But the kind of modalised, focussed questions that occur for example in the scheduling domain are different from wh-questions in that a value for whatever is in focus is given in the question, and the point of the question is whether that value is feasible with respect to the goal.

In this study 'focus of the question' refers to that element or part of the question that the questioner's interest is focused upon. Following Kaplan (1979), it may be described as that element which is most likely to be changed in a follow-up question. What is in focus depends on what immediate (conversational) goal the speaker wants

to achieve by uttering the question.

In the house construction example

(1) Can the painters start on the 29th?

we have the following context: the overall goal, shared by the participants, is to finish the house at the earliest possible time or by some specified time X. The overall goal is known to the participants either because they recognise each other's roles or because it has been established in (some) previous discourse.

The immediate goal of the speaker in uttering the question is either to arrange a time for job J or a job for a time T (i.e. either the job or the time is in focus). Both of these alternatives are subgoals of the overall goal of finishing the house by time X. This means that if (as is assumed here) the overall goal is known to the interlocutors, the search for the immediate goal is limited to the subgoals of the overall goal; and vice versa, if the immediate goal is deducible from the utterance itself, but the overall goal is not known to the hearer, the search for the overall goal is limited to those goals of which the immediate goal is a subgoal.

However, in the present example, knowing what the overall goal is does not solve the problem of which one of the two possible subgoals the speaker wants to solve when he asks the question.

What is in focus depends on which subgoal is being considered. The hearer knows that the goal and the

subgoal have to be part of the same planning tree or planning network, but knowing what the goal is does not necessarily help to choose the subgoal if there are several candidates, and vice versa.

What has here been described as determining the speaker's immediate goal in uttering the question corresponds closely to what Allen and Perrault (1980) call plan inference (cf. discussion above), i.e. the hearer trying to infer the plan of the speaker, e.g. inferring whether the speaker wants to board a train or meet a train when asking 'When does the train to Windsor leave?'

Our domain is different from theirs in that in their (non-modal) context the overall goals of the interlocutors do not matter, whereas in the modal context they do.

Allen and Perrault use semantic clues, (pragmatic) knowledge about the plans relevant in the domain, and the interlocutors' knowledge of each other's roles in their plan inference mechanism. There may also be phonological and syntactic clues, such as the ones discussed above, used by the speaker and the hearer to mark/find the focus of the question, which in turn may help to infer the plan of the speaker. Which of the clues discussed above are important, and how they work and interact, are open questions at the moment.

3.5. Root modals and goal orientation

Deontic modality, especially deontic necessity is obviously related to the goals of the speaker, if he is also the deontic source, since it is the primary function of this modality to issue mands and directives (Lyons 1977: 840), in other words, to get people to do things or to prevent them from doing something. For example,

(32) You must bring that book back tomorrow.

(33) You must not pick these flowers.

seem to be related to some ulterior goal or reason; the goal may be expressed as well:

(32′) You must bring that book back tomorrow, because I´ve promised to give it to Nigel.

(33′) You must not pick these flowers, because I want the garden to look nice.

Deontic possibility (permission), on the other hand, does not have to be related to a goal; A´s saying to B

(34) You may go.

may be the result of establishing that A has no goal that conflicts with B going. But establishing this does involve searching for goals that might conflict with B going.

Linguists who are in favour of a monosemantic view of the modals have argued that necessity modals can be analysed in terms of "something requires p" (where p is

the proposed action) (Ehrman 1966:67,73) or "something entails p" (Perkins 1980: 64).

Kratzer (1977) goes further in claiming that not only assertions containing necessity modals but all modalised assertions should be analysed as having an implicit "in view of" phrase attached to them. How the "in view of" phrase is completed determines what kind of modality the assertion contains: whether it is epistemic, deontic or dynamic. For example, consider the interpretations (a-c) of (35) (modified from Kratzer 1977):

(35) Bear hunters must be brave.

a. In view of what is known, bear hunters must be brave. (epistemic)
b. In view of their duties, bear hunters must be brave. (deontic)
c. In view of their aims, bear hunters must be brave. (dynamic)

Dakin (1970) discusses the way that the use of modals or related verbs (the so-called semi-modals) implies that there is a reason for the event that took place (all his examples are in the past tense), even if the reason is not given explicitly. For example,

(36) John had to stop.

implies that there was something that caused him to stop. The explanation as to what caused the event can be given, e.g.

(37) John had to stop because his brakes jammed.

but it does not have to be expressed for the implication to

be there. There are various mechanisms for giving the explanation (because ..., explanation followed by so, order of the clauses, etc) (see Dakin 1970: 199-200).

Dakin employs a speech act analysis: 'John stopped' is represented as "I state S" and 'John had to stop' as

I state S
I state X caused S

The 'something' or X that caused John to stop can according to Dakin be either an antecedent cause or a state of affairs which required him to stop. If John had to stop because his brakes jammed, that is a physical cause; if John had to stop because the traffic lights changed to red, he was 'demanded' to stop. The difference between causation and demand in the sense Dakin uses the terms is based on time relations: a causing event precedes the event which constitutes the effect; an existing state of affairs demands an action (Dakin 1970: 200, 203). It seems that the demands have to do with laws (of the society), regulations, and other comparable systems. There is a similarity between demands in this sense and goals in that the action demanded by a state of affairs and an action required in order to reach a goal are voluntary actions. Dakin's positive past tense assertions are factive, i.e. the action is asserted to have taken place, but if we take an assertion referring to the future, such as

(38) You must stop when the light is red.

the action does not take place inevitably; one can choose

not to stop (even if it may be at a cost). And in fact even in the past tense assertions, the action did not inevitably follow from the demand; the agent could have chosen not to stop.

(39) John had to stop because the lights turned red.

is an example of deontic modality. It is clear that in deontically modalised sentences there has to be an agent and a voluntary verb. But in other cases, if the action is involuntary, if there is no choice, or if there is no agent, it is odd to use MUST or HAVE TO:

(40) John slipped on a banana skin and had to fall.

The verb FALL normally denotes an involuntary action; one would normally say 'John slipped on a banana skin and fell'.

In Dakin's example

(41) John had to stop because his brakes jammed.

although the cause is a physical cause (Dakin 1970: 213), the verb STOP is still a voluntary verb, and there is an agent performing the action, and so one can still construe the main clause as depending to some extent on John's decision. If there is no agent, one cannot use HAVE TO in a root sense:

(42) ?The car without a driver had to stop.

Dakin's analysis in terms of demands is one way of

explaining objective deontic modality. Perhaps causes in his sense could be interpreted as circumstances for analysing dynamic modality. However, neither are obviously the same as the interlocutors' goals; but they are linked to these goals in a systematic way. Nor are the goals identical to the deontic source; in example (38), in fact, the laws or regulations that the demand depends on are the deontic source.

Root modality always refers to the future. Even if the assertion is in the past tense, the proposed action is in the future as looked at from the point of reference. Goals are always in the future as well: with root necessity, for instance, whatever brings about the necessity is in the future. Even with objective deontic necessity (where Dakin claims the 'demand' is by the present state of affairs) it can be claimed that the demand is also linked to the future consequences of actions. In contrast to root modality, epistemic modality can refer to past, present or future, and the 'reason' for the necessity derives from what the speaker knows or believes at the moment of speaking.

It will be argued in this thesis that the interlocutors goals have to be taken into account in interpreting dynamically and deontically modalised utterances. It will be claimed that necessity is always goal-oriented and that possibility often is but does not have to be goal-oriented.

4. DYNAMIC MODALITY

4.1. Can

4.1.1. The senses of can

The following senses are usually given for can: it expresses ability, permission and possibility (Leech 1971: 667-71; Quirk et al. 1972: 97; Palmer 1979; Coates 1980: 153).[1] Examples of these are:

(1) John can run faster than Mike.
(ability)

(2) I can't go swimming tomorrow.
(possibility)

(3) Mr Duncan can start painting the windows on Monday.
(possibility or permission)

(4) You can go in now.
John can go in now.
(permission or possibility)

These senses are sometimes treated as different meanings of can; linguists who are in favour of a univocal view of the modals see them as variations of one basic meaning. The basic meaning is often given as "nihil obstat", or there is nothing that precludes the event from taking place (Ehrman 1966: 12). Developments of this view are presented in Wertheimer (1972) and Perkins (1980). Both explain the differences between the senses by relating them to different systems of laws.

In the case of can the relevant systems of laws would

be something like laws of nature, physical laws or a scientific theory on the one hand, and laws of the society or a moral code on the other hand. These are relevant to the possibility sense and the permission sense respectively.

Perkins, whose analysis is based on that of Wertheimer (1972), then gives the following formula as the definition of <u>can</u>:

K(c does not preclude that e occur)

where K stands for natural laws or laws of the society, c is either a circumstance or a deontic source and e is the event (Perkins 1980: 61). The formula can be paraphrased as follows:

> "within the system of natural laws, there is no circumstance that precludes the event e from occurring"

This is the possibility sense of <u>can</u>, and the permission sense can be paraphrased as follows:

> "within the laws/rules of society, nothing/ nobody precludes the event e from occurring"

A different view is given by Isard and Longuet-Higgins (1973). In their domain of the game of tic-tac-toe, the relevant sense of <u>can</u> is the possibility sense, and it is defined as "there is a strategy that will necessarily lead to [the proposed event]". So

(5) I can win this game.

can be paraphrased roughly as

"I have a strategy which will necessarily win this game (that is, whatever my opponent does)"

and the definition also includes the sense of choice, i.e. "if I choose to use that strategy" (Isard and Longuet-Higgins 1973: 194-5).

It is unclear why the qualification 'necessary' is included in this definition.

In their program can is construed in the sense "what is possible from the point of view of the rules of the game". The question 'Can you win?' could be interpreted in other ways, for example, as "Have you got the ability to win?". But in this example the possibility is not related to any goals.

The sentence

(6) John: Mr Duncan can start painting the windows on Monday.

may be interpreted in slightly different ways depending on the context and who is uttering the sentence to whom. If John has been trying to arrange for somebody to come and paint the windows of his house, he might say (6) to his wife to tell her that the painter (Mr Duncan) can fit their house into his schedule next week starting on Monday. However, if he says (6) to the foreman of the painters or builders he probably means that that time of starting the painting job suits him. (6) can also be interpreted as expressing permission (deriving from some factual reason) or even as a command, if uttered by a superior of Mr Duncan's.

One might argue that from the hearer's point of view these differences in the interpretation are perhaps not relevant or important, and that all he needs to know is whether in general p is possible. This is probably often the case (cf. Coates 1980 for a discussion of indeterminacy and merger). However, when the sentence is a question, e.g.

(7) Can Mr Duncan start painting the windows on Monday?

the hearer has to decide, before he can answer, what he is being asked about - possibility in relation to what, in view of what?

If we look at the definitions given above in the review of the senses of <u>can</u>, it becomes apparent that they cannot explain all the variations in the interpretation of it.

According to Perkins, the interpretation of

(8) Mr Duncan can/cannot start painting the windows on Monday.

is: "within the system of natural laws, there is no (there is some) circumstance which precludes Mr Duncan starting to paint the windows on Monday". The corresponding question would have the interpretation "is there any circumstance which precludes Mr Duncan from starting painting the windows on Monday?". There are two criticisms of this. One is that although the paraphrase sounds intuitively close to the meaning of <u>can</u>, it is not correct. The impossibility of a proposition or event does not <u>have</u> to depend on one or even a limited number of

circumstances (although it often does) (cf. below). The second is that this definition does not say how one comes to know that something prevents the event, or what sort of circumstances to look for, or how to decide what sort of circumstances to look for; in short, what the procedures are for understanding and answering a question with can.

In Isard and Longuet-Higgins' (1973) program, when it is asked a question like

(9) Can I win?

the system goes into a hypothetical mode where it runs through the rest of the game, checking all variations of it that are still possible, and if it finds a strategy that wins the game, it replies 'yes'. If it cannot find such a strategy, it replies 'no'.

The context of a game shows that it is not always the case that there is one circumstance or a limited number of circumstances that make an event impossible. At a particular moment in a game of tic-tac-toe or chess, it may be the case that it is impossible to win the game, not because of one single fact, circumstance or reason, but because all the available strategies lead to a defeat. In other words, it would not be possible to give any other 'reason' for the impossibility than that there is no winning strategy.

Can does however often imply conditionality, and impossibility is often explained by speakers by giving the reason why the event is impossible. Coates (1980: 167)

gives some interesting statistics about sentences containing root cannot or can't: in most cases the speakers gave the reason of why the event was impossible. But even if the possibility often depends on one condition, this is not necessarily the case, and a more general definition than the "nihil obstat" view in this form must be found for can.

When answering a question like

(10) Mr Duncan: Can I start painting the windows on Monday?

the hearer has to go through considerations like the following:

- is it possible physically (are the windows ready for painting)?
- is his starting on Monday realizable from my point of view (e.g. will there be someone in to open the door)?
- does it have any consequences which are incompatible with goals I have (e.g. getting the windows done before the winter)?

The hearer perhaps also realizes that there are factors that he need not worry about; for example, is the painters' work schedule or diary free for that time, whether the painter can get the paint by then, etc., which he knows are the business of the speaker (the painter), not the hearer in this case.

The procedure will then have to see whether there

exists a plan which realizes the proposition ("Mr Duncan start painting the windows on Monday") and which incorporates the relevant features, including the goals that the speaker, the hearer or both of them hold.

If it is assumed that people generally try to be helpful and co-operative when answering questions (cf. Grice 1975), and that they recognise the interlocutor's intentions (cf. Allen and Perrault 1980), it must also be able to give appropriate (different) answers depending on what the outcome of the procedures mentioned above are. For example,

(11) Mary: Can we have the windows painted this month?

could elicit the following answers:

(12) John:

a. Yes.

b. Yes, in fact we must do it this month, if we want to get it done before the winter.

c. No, because they have to be repaired first.

d. No, we can't afford it, if you want to have a new carpet for the living-room as well.

In (a) the event is possible but not necessary, in (b) it is necessary if a (mutual) goal is to be met, in (c) it is not possible, and in (d) it is not possible if a goal (of the speaker of the question) is to be met.

4.1.2. A procedure for dynamic can for direct questions

This section will describe the procedure that deals with direct questions containing can in the scheduling domain, such as

(13) Can the painters start on the 29th?

This is interpreted in the sense "Is it feasible that the painters start on the 29th, if the completion time for the house is day X?" For simplicity, other factors that might influence the feasibility (and which would have to be taken into account in a real situation) are not considered. The time, e.g. 'the 29th', is taken to be the focus of the question.

The goal is also simply given in the database, in the form 'goal(G)' e.g. 'goal(finish(32)). Although it is possible to have several goals in the database, the example below assumes the single goal of finishing the house on a given date.

The interesting question of how a relevant goal (or possibly goals, in a real situation) are chosen, that is, how the respondent finds and relates a particular goal to a particular question, is an unsolved problem. The algorithm for it would probably use several approaches: a goal may have been established in the preceding context; it may be implicit in the non-linguistic context; it may be deducible from the interlocutors' roles; or it may be explicitly expressed in or deducible from the utterance (the question)

itself.

The system can give any of the following responses to (13):

(14) a. Yes.

b. Yes, indeed they must if the goal is to be kept.

c. No, they cannot if the goal is to be kept.

d. No, it is impossible.

e. Yes, they will start on that day.

The first answer is given if, with respect to the goal, it is possible that the painters start on the 29th, but it is not necessary that they start then; it would be possible for them to start on some other day, say the 30th, without altering the completion time. Answer (14b) is given if it is not only possible but necessary for the painters to start on the 29th: that time is the only starting time which is compatible with achieving the goal (neither any earlier nor any later time is feasible; in this example this is because the job is on the critical path).

The system embodies the narrow meaning of possibility, in which possibility and necessity do not overlap. See below for discussion.

The difference between (14c) and (14d) is that in (14c) the impossibility is caused by the goal, whereas in (14d) the impossibility is absolute. In (14b) and (14c) the if-clause also expresses the fact that the goal is something which may be regarded as flexible. For example,

in (14b), if the goal is altered, the necessity may cease to exist.

The system gives the answer (14e) if it has already been arranged that the painters come on the day proposed in the question (see below for discussion of this point).

When the system is asked

Can the painters start on the 29th?

the parser processes this sentence and returns a simple semantic representation (SR) in the form

question(yesno,can(p))

In this case p = starttime(painters, 29), so the full SR of (13) is

question(yesno,can(starttime(painters,29))).

(The first parameter, 'yesno', indicates the type of question.)

This SR together with the focus f of the question (here f = 29) is the input to the procedure for answering modalised questions. The output of the modal questions procedure is one of the following results:

(R1) can(p)

(R2) must(p) if g

(R3) not(can(p)) if g

(R4) not(can(p))

(R5) will(p)

These results correspond to the responses (14a - 14e).

The actual response is formulated on the basis of both the result and the form of the question.

In the following description of the procedure for modalised questions the notation ´plan(p)´ refers to a routine that calls the planner to find a plan for p. The argument may be a conjunct: ´plan(p & g)´, where g itself may be a conjunct of goals. The procedure alters the focus within p to test whether there are alternatives for p.

The proposition expressed in the question will be denoted by p and where it is important to show the dependency of p on the focus I will use the notation p(f). The notation Foc will be used for the case when the focus is a variable.

I will indicate whether the focus variable is instantiated at the original or at an altered value by the notation Foc=f or Foc=/=f, where f = the value of the focus in the (original) proposition. Taking as an example the question in (13), with the time as focus and the goal being ´finish house by day 32´, then ´plan(p(f) & g)´ means
"plan(starttime(painters,29)) & finish-house(32))"
and ´plan(p(Foc) & g) & Foc=/=f´ means
"plan(starttime(painters,Foc)) & finish-house(32)),
where Foc=/=29", i.e. "find a plan for the painters to start at some time other than day 29 which allows the house to be completed on day 32".

Given the SR as its input, the modal questions procedure first sees whether there is a goal in the database. We will describe first the situation where there is a goal,

as in the scheduling domain. The task then is to find out whether there exists a plan that realizes the conjunction of the proposed action and the goal g, i.e. whether the action and the goal are feasible and compatible.

Having found the goal g the system calls the planner to see whether there is a plan that realizes (p(f) & g). If this call of plan(p(f) & g) succeeds, it then checks whether the time given is the only time, i.e. whether p(f) is necessary with regard to g. This is done by attempting to find some other value for the focus than the one given in the original proposition. The planner is called to plan(p(Foc) & g) & Foc=/=f. If this call succeeds, the result is "can(p)": p is possible but not necessary. The system can give the user the answer (14a).

If plan(p(Foc) & g) & Foc=/=f fails, i.e. no alternative value can be found for the focus variable with which the call plan(p(Foc) & g) succeeds, the conclusion is that p is necessary. But it must be checked whether p is necessary only in conjunction with g, or is also necessary without g. The procedure therefore calls plan(p(Foc)) & Foc=/=f. If this succeeds, p is not necessary on its own; the result is "must(p) if g", and the appropriate answer is (14b).

The check for whether p is necessary on its own, without regard to the goal, is included in the procedure for the sake of completeness. One cannot claim "must(p) if g" without checking that p is only necessary in relation to the goal. (In the scheduling domain, p can only be said to be necessary in relation to a goal. This is in

accordance with the claim that dynamic necessity is always linked to a goal.)

If the call plan(p(Foc)) & Foc=/=f fails, the situation appears to be that p is necessary irrespective of the goal: the planner can find a plan for p, but no plan for any alternative for p. In the example program this only happens if the procedure finds that a starting time for the job in question has already been agreed upon. The plan in this case is an 'empty plan' corresponding to doing nothing. It is a precondition for a plan for a starting time for a job that the starting time has not been fixed yet. Therefore if the starting time has been fixed and is seen as a fact, it is impossible to talk in terms of a plan for it (i.e. to have anything other than the empty plan). The user should be informed of the fact. In this case the answer (14e) is formulated in terms of what <u>will</u> happen, not in terms of what <u>must</u> happen.[2]

If the first call, plan(p(f) & g), fails, then p is not possible with respect to the goal, and the procedure then checks whether p is possible by itself by calling plan(p(f)). If this succeeds, p is possible on its own but not possible when conjoined to the goal. The result is "not(can(p)) if g", and the appropriate answer is (14c). If plan(p(f)) fails, p is (absolutely) impossible: the result is the unqualified "not(can(p))", and the answer (14d).

In this domain it is difficult to envisage a situation in which there is no goal. Even if the goal of finishing the house by a certain date is dropped, the general goal of

finishing the house still remains. However, dynamic modality does not always have to be linked to goals, if the modality in question is possibility rather than necessity. The problems with cases in which there are no relevant goals will be dealt with below; they have not been incorporated in the example program.

The procedure outlined above can be represented in a simple network, which is given in Figure 3. It must be stressed that the ordering of the operations in the network is not the only possible one. The important thing are the results and what operations are necessary to arrive at each result; not in what order the operations are done.

Two points about the procedure that have already been mentioned require further comment.

The first is the case where it is found that an agreement about the starting time for a job has been made. The program discovers this in the process of trying to determine the degree of modality. If p is taken as arranged, then the question of the feasibility or necessity of p ceases to be relevant. The information that an agreement has been made should be given to the user both in the case in which the answer is 'Yes, they will come on that day' and when the answer is 'No, they will come on day t'. In the latter case an agreement having been made for another date is one of the many possible reasons for the proposed action being impossible.

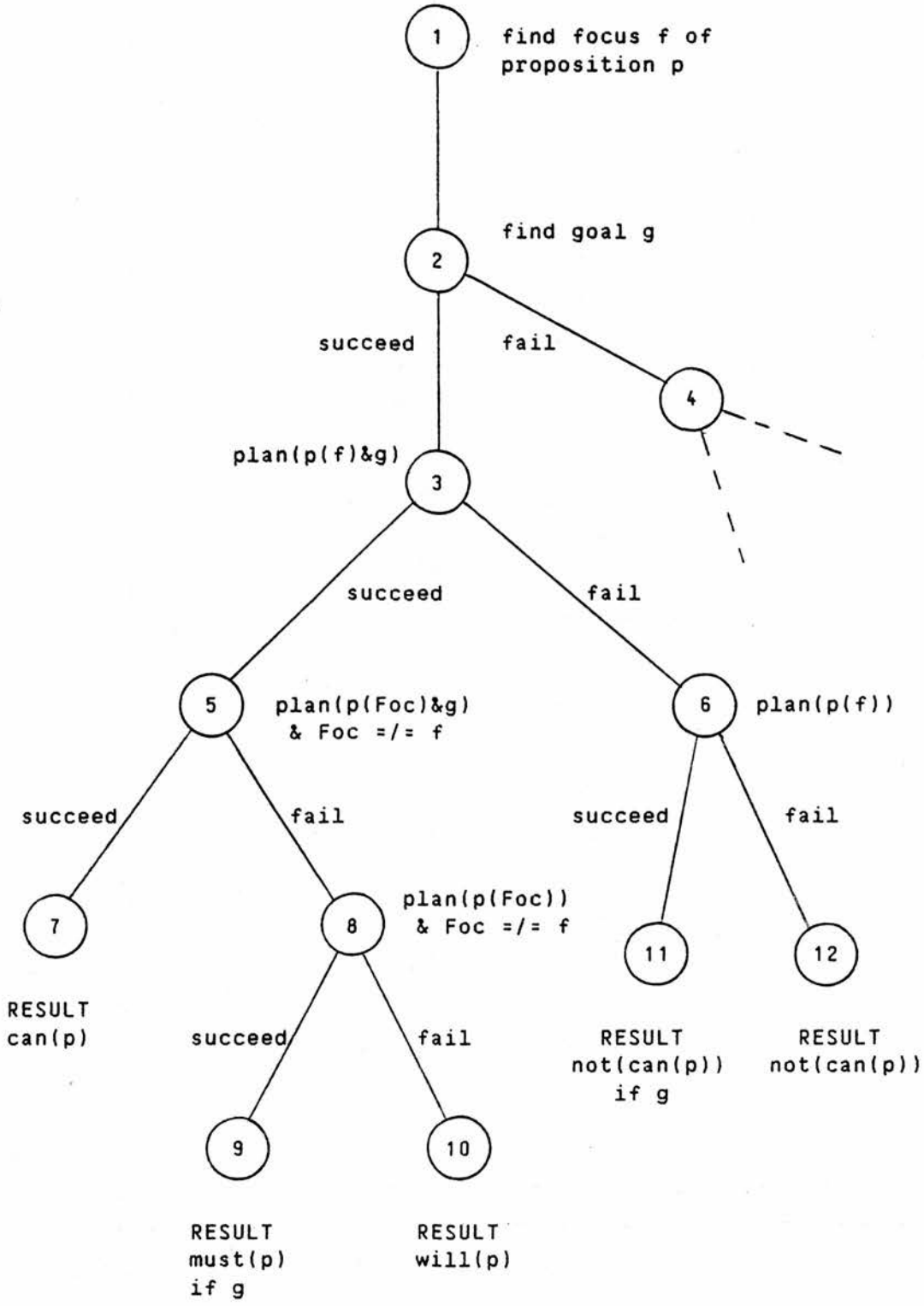


Figure 3 : Decision tree for answering direct questions containing dynamic can or must

People do tend to give reasons as to why an action is impossible (Coates 1980: 167). It would be desirable to have the program give a reason for the impossibility, whatever it is caused by, unless it is caused by a combination of several factors rather than a single factor.

A question containing can is commonly a preliminary for arranging something. In the scheduling domain an exchange like

A: Can the painters start on the 29th?
B: Yes, they can.

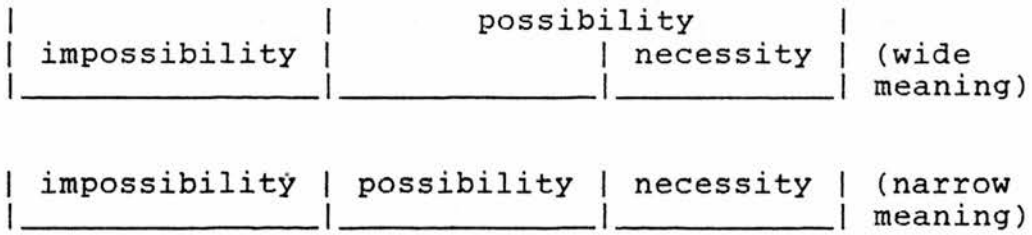
could be followed by something like

A: O.K., lets put them down for the 29th then.
B: O.K.

However, if B knows that the arrangement has already been made, he is likely to point this out to A as an immediate response to A's question. The procedure for modal questions finds out and gives this fact because it cannot satisfy the precondition that the starting time must be 'unfixed'. Can is typically only used if a choice is involved. Another way of doing it would be for the system to check, before exploring the possibility or necessity, whether p has been arranged, and only to go on to the modality routine if no arrangement has been made. The latter method is based on the speaker's attempt to anticipate the interlocutor's next move in the conversation.

The second, more important point about the procedure outlined above is the way it defines "can" by excluding both necessity and impossibility.

Logicians distinguish between a wide and a narrow meaning of possibility. The wide meaning does not exclude the meaning of necessity: if p is necessary it is also possible. The narrow meaning of possibility excludes the meaning of necessity: what is possible is not necessary (Reichenbach 1947: 128; cf. also Jespersen 1924: 324-5; Hintikka 1960, Burton-Roberts 1984). The two meanings of possibility are often represented as follows:



According to Reichenbach (1947: 128) the narrow interpretation of possibility is used in conversational language. People do not say that an event is possible if they know that it is necessary. This is also the view embodied in the system presented here: can is interpreted as representing narrow possibility. Therefore, if the proposed action is necessary, answer (14b) is given to question (13), not answer (14a).

An explanation in terms of the co-operative principle is also possible here: it could be argued that it is likely to be relevant for the hearer to know that something is

necessary, if he is interested in whether it is possible.

If the narrow meaning is accepted as the meaning of dynamic possibility, it could be defined as "not(impossible(p)) & not(necessary(p))".

Our system defines dynamic possibility by verifying that p is possible but not necessary: p is possible if there is a plan which realizes p; p is not necessary if also not(p) is possible and compatible with achieving the goal. However, the system does not ask the planner to find a plan for 'not(p)' or 'not(p) & g'. The planner cannot handle a call to find a plan for a negated action. Instead, the procedure changes the focus element, f, in p(f), giving the proposition p(Foc) and then calls the planner (perhaps several times) to find a value of Foc such that Foc=/=f.

4.1.3. Discussion and extensions within dynamic modality

Can is sometimes associated with a goal, sometimes not. The goal to which the feasibility of the action or event is related may be expressed or it may be implicit. The following are examples of goal-oriented can:

(15) What time can you get away? (Palmer 1979:83)

(16) I can't take any time off this week.

(17) We can't stop now if we want to get back by 8 o'clock.

(18) Can we stop now if we want to get back by 8?

(19) Olive oil can be used to protect the surface of a wooden salad bowl.

Notice that an if-clause that expresses the goal (as in 17 and 18) can only be attached to negative or interrogative (i.e. non-assertive) utterances; it is not normally attached to positive assertions:

(20) ?We can stop now if we want to get back by 8.

although

(21) We can stop now even if we want to get back by 8.

seems acceptable. (Other kinds of if-clauses can of course be attached to positive assertions with can, e.g. 'We can stop here if you want a cup of coffee'.)

Examples of can without a goal include cases in which it expresses ability (ability as, or dependent on, a property that an entity has):

(22) Can you speak English?

(23) John can't lift that weight.
(Palmer 1979: 87)

as well as cases of the so-called existential modality:

(24) Welshmen can be tall.

and cases of non-modal uses of can, as in

(25) I can see the moon.

But the can of feasibility does not have to be linked to a goal either; there are a lot of examples of a purely circumstantial can of feasibility, without any association to

a goal; for example,

(26) You can travel from Belgium to France with less palaver than you can travel from the North to the South of Ireland. (Palmer 1979: 71)

(27) Most house plants can be propagated by this method.

(28) Can I win (this game)?

With dynamic modality, the goal that the modality is related to may be anybody's goal: it may be the speaker's or the hearer's goal, it may be shared, or it may be a third party's goal.

In the exchange

(29) User: Can the plumbers come on the 15th?
System: Yes, they can.

the goal of finishing the house by time x can be thought of as either the user's goal or as a shared goal.

The same question could be used to refer to the plumbers' intentions, that is, whether the plumbers can come on the 15th with regard to their other commitments, e.g. jobs that they have already undertaken to do at that time; roughly, this would be equivalent to asking "Do you know if the plumbers can come on the 15th?". Here the question relates to neither the speaker's nor the hearer's goal.

An example in which the feasibility depends on the respondent's intentions would be

(30) A: Can you drive me to the airport tomorrow morning?

B(1): Yes, I can.
B(2): No, I can't because I have a meeting in the morning.

There has to be a mechanism which finds whose goals are relevant in a given situation. This presumably has to do, among other things, with how the question is interpreted: whether as a request, asking for permission, etc. (cf. discussion of these below).

With (30) the respondent has to look at his own goals and see whether the proposed action and his goals are compatible. One of the preconditions of the action proposed in (30) is (assuming that the starting point for the journey to the airport is 'home')

at(home, B, morning M)

but B's goal, attending the meeting, has the precondition that he be at the meeting place in the morning; so the plans clash, and the result is "not(can(p)) if g".

We now have to look at <u>can</u> in contexts in which we have no goals.

In example (30), if B has no intentions for the following morning, it would seem that only the preconditions for the proposed action need to be checked. The results are either "can(p)" (if there is a plan) or "not(can(p))" (if some condition(s) are not satisfied).

If there are no goals, there can be no dynamic necessity, and the scale of modality does not include necessity modals. But there is still the possibility that the event not only <u>can</u> but <u>will</u> happen. As was discussed above, one context

where we assume that the event will occur is the case of an arrangement having been made. The two possible ways to deal with this are: (1) when encountered with a modalised question the system first checks whether the event has been arranged (or is already actual), and (2) the system finds this out in the process of trying to build a plan for the action, as this fact constitutes an unsatisfied precondition of the action. Method (2) has been adopted here.

There are other cases in which it may be pointed out in the answer that the event will happen, for example:

(31) A: Can caustic soda burn your hands?
B: It will burn your hands if you put your hands in it.

Consider also

(32) You can turn the central heating on by pressing the red button.

which is equivalent to

(33) If you press the red button the central heating will come on.

In the type of sentence exemplified by (32), the by-clause gives the essential condition for the event of the main clause to occur; the meaning of (32) can be expressed as an if-then condition.

Taking into account the fact that dynamic possibility may be goal-oriented or not, we can present a modified algorithm for dealing with questions containing dynamic

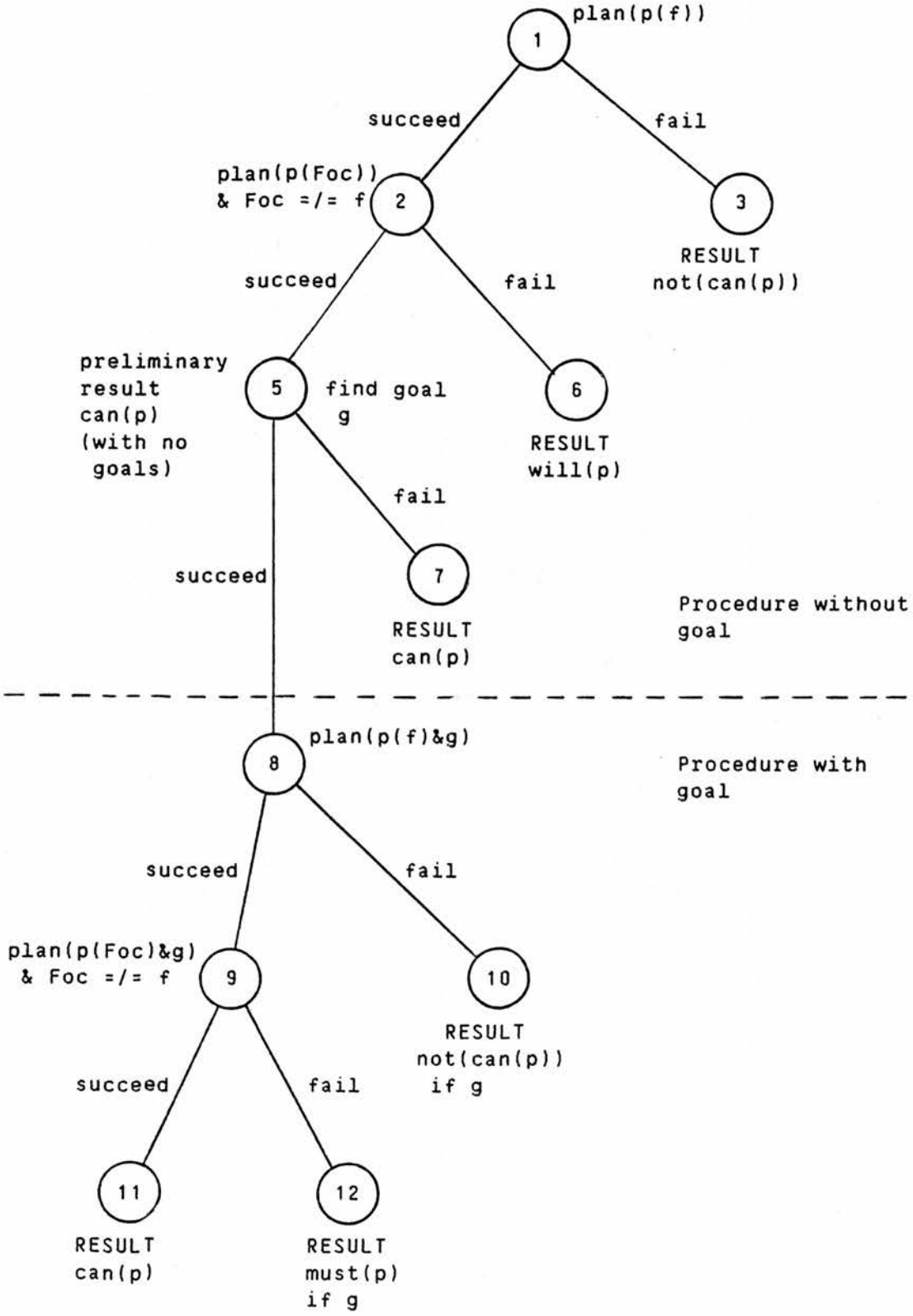


Figure 4 : Extended decision tree for dynamic modality

possibility. This is given in Figure 4. It will be noticed that the operations do not proceed in the same order as in the network given in Figure 3. As was pointed out there, the order of the steps is not crucial.

4.1.4. The *can* of ability

The difference between the *can* of feasibility and the *can* of ability is that the former depends on circumstances, on conditions being satisfied, whereas the ability to do something is a fact about an entity, a property of an entity.

To answer questions like

(34) Can you drive?
(35) Can you swim?

only the condition

knowhow(drive, Agent)

or knowhow(swim, Agent)

needs to be checked. *Can* in this sense is equivalent to "know how". ('knowhow' is here used in the sense "have the knowledge and skill to perform the action".)[3]

But *can* can also ask or be interpreted to be about "fitness". Thus (34) could be about either (or both)

(i) knowhow(drive, Agent)

(ii) fit(drive, Agent).

It could be interpreted as (ii) in a context in which the

questioner had some reason to think that the respondent perhaps was unable to drive (e.g. if he had hurt his arm).

Considering a question about possibility in general (not only ability), in a question like

(36) Can you drive me to the airport on Saturday?[4]

'drive' can be thought of as an operator (in the sense that operators are used in a planner like STRIPS; see Fikes & Nilsson 1971). The operator 'drive' has the following parameters:

drive(X, (Y), From, To, Time)

(X drives (Y) from somewhere to somewhere at a time)

The first argument, the agent, is compulsory. The second argument, Y (the passenger or object) is optional. But if any argument other than X is present, then the last three arguments have to be filled in: if they are not given in the question, they have to be supplied from the context. In (36) the From parameter has to be supplied and could be for instance "here" or "home".

The operator 'drive(X, (Y), From, To, Time)' would have preconditions like the following:

able(X, drive)

at(From, X, Time)

at(From, Y, Time)

at(From, Car, Time)

etc.

In other words, the driver, the passenger and the car have to be at place From at Time and the driver has to be able to drive. The precondition of ability has to be satisfied even when the question is about possibility in general.

The question ´Can you drive?´ may of course also refer to a specific time and place and have for example the meaning "Can you drive from here onwards?" if the missing parameters can be filled in from the context. But in a context in which the only parameters are the agent and the action, the only conditions that the modal procedure can check are

```
knowhow(drive, Agent)
fit(drive, Agent)
```

These are preconditions of the operator ´drive´.

Some activity verbs like DRIVE, SWIM, PLAY (an instrument, a game), when considered as operators in a plan, always have ´knowhow(Activity, Agent)´ as one of their preconditions. It would seem that many dynamic verbs also have, at least implicitly, a condition that stipulates that the agent be fit enough, i.e. physically capable of performing the action. The latter condition is more general than the former: the ´knowhow´ condition does not always apply although the ´fitness´ condition does apply. For example, verbs like COME, GO, LEAVE do not have it:

(37) Can you come?

can only refer to possibility that depends on circumstances

or goals, or on physical ability ("Are you well enough to come?"), but it cannot be interpreted as "Do you know how to come?".

Some verbs may have the 'knowhow' condition in some contexts; in other contexts it may be taken for granted that this condition is satisfied. For example, with WALK, as in

(38) Can he walk?

the 'knowhow' condition is not relevant and the above question would be interpreted as referring to physical ability unless 'he' refers to a child of the age at which children learn to walk.

All the essential preconditions, including the 'knowhow' and 'fitness' conditions, have to be checked even if the point of view of the question is the goal(s) of the interlocutor(s). The question

(39) A: Can you drive me to the station in the morning?

may be asking whether B is free to do the requested action, but B may give a negative answer on account of an unsatisfied precondition, e.g.

B: No, I can't drive.
B: No, because my car has just broken down.

Expressions that refer to ability (in the sense of "knowing how" or "having the skill" to do something) differ from all other modal expressions in that there is only the

dichotomy "able"/"not able"; there is no third value, no corresponding necessity (cf. Palmer 1979: 23). This is explained by the fact that ability is a property of the agent; the agent either has the property or does not have it. This fact is not affected by goals that might be present in a given situation.

There are two ways of dealing with ability can in our procedural framework. The first is to let the modal procedure handle it in essentially the same way it handles other dynamic questions containing can, using the planner to find a plan.

Let us compare how the procedure would deal with the questions

(40) Can you drive me to the airport on Saturday morning?

and (41) Can you drive?

With question (40), the first step is for the respondent to see if he has a relevant goal (i.e. plans for Saturday morning); the next step is to try to solve the proposition and the goal as a conjunct. The physical preconditions as well as the ability of the agent (the respondent) to drive are checked in the process of looking for plan for the proposition (drive(b,a,...)) because they are preconditions of the operator 'drive'. (For completeness, the permission constraint would also have to be checked; cf. below.) If no goals are found, only the conditions for p need to be checked.

If 'drive' has only one argument, as in (41), and no other arguments are found in the context either, the only applicable preconditions are 'knowhow(drive,b)' and 'fit(drive,b)': the two components of "inherent ability". These are arrived at as it were by default; other conditions, those relating to time and place etc., can be assumed to be irrelevant, since these parameters are not given. This leaves the ability sense of <u>can</u>, without the (circumstantial) feasibility sense.

There is still the problem that the 'point' of (41) may be either the 'knowhow' condition or the fitness condition or both, i.e. it can mean "Do you know how to drive?" or "Are you capable of driving?". This is, however, a general problem: the 'point of view' of a <u>can</u> question may be goals, but it may also be any of the preconditions for the proposed action.

It may seem unsatisfactory that even if we have the pure inherent ability sense, this system would try to find whether there is a plan for the action, as the plan then consists of nothing more than the fact about x that "x knows how to drive". The defense of this way of handling <u>can</u> is that in many cases in which the hearer understands the speaker's intention, he will go through the whole plan needed to realize that intention, including goals and permission, anyway. If A asks B 'Can you drive?' in a context where B realises that A's intention is to get him to do some driving, B might answer

(42) Yes, but I haven't got a British driving licence.
("I know how to but the law does not permit me to")

Another argument for including ability can within the same procedure as other dynamic uses is the fact that can is very often indeterminate between the different senses (Coates 1980). (Even though the context usually guides the interpretation, it does not always solve the indeterminacy.) Having all the senses in the same network explains the indeterminacy better than regarding can as polysemantic. The alternative for treating the can of inherent ability is to have a separate entry for this sense of can in the dictionary. In other words, can would have the distinct meanings of ability on the one hand and a more general possibility or feasibility sense on the other hand; the latter may however also include the ability sense, depending on what kind of action can is attached to.

4.1.5. Requests

Direct can-questions are frequently used with the illocutionary force of a request:

(43) Can you pass the salt?

(44) Can you drive me to the station tomorrow morning?

They have been favourite examples of indirect speech acts (see Gordon and Lakoff 1971, Searle 1975, Lyons 1977: 785, Morgan 1978).

There is agreement in studies on speech acts that utterances like (43) and (44) have two illocutionary forces: the force of a question and the force of a request, both being intended simultaneously by the speaker (Searle 1975: 59-60, Lyons 1977: 785). According to Leech (1983: 97) a can-question may function both as a question and as a conditional request: the speaker wants the hearer to do the action if the preparatory condition of the hearer being able to do the action is satisfied.

While accepting that utterances like (43) and (44), when intended as requests, do have these two forces, it is still the case that the hearer, when responding to these, has to decide whether to interpret the utterance as a question (= a request for information) or as a request (for action) as well as a question.

The hearer's reaction is different depending on which interpretation he makes (and the speaker obviously expects a different reaction).

The previous sections have dealt with how to respond to an information-seeking question.

Responding to a request consists of either performing the action or agreeing to perform the action in the future (depending on how near or distant a future the request refers to):

(43') A: Can you pass the salt?
B: Here you are.

(44') A: Can you drive me to the station tomorrow morning?
B: Yes.

Notice that B's answer in (44') constitutes an agreement to perform the requested action (in addition to providing the information that the action is possible); after this conversation A can expect B to be there in the morning with his car.

If the hearer reacts differently depending on whether the utterance is a request or a question, the hearer has to determine what illocutionary force the utterance has before responding. (It should perhaps be pointed out that this does not imply that utterances like (43) are ambiguous semantically. There is no ambiguity in the propositional content (except as far as can is held to be ambiguous), but the utterance may have different pragmatic forces.)

The main problem then that requests in the form of questions present in a procedural framework is the problem of how the hearer determines whether the utterance has the illocutionary force of a request or that of a question.

There may be some formal clues.

Searle (1975: 69) mentions that the intonation of a request often differs from the intonation of the same sentence when it is uttered with its literal illocutionary force.

Palmer (1979) lists some formal differences between information-seeking questions and questions used as requests. The most obvious one is that requests may have please added to them. Any and compounds of any- do not occur in requests: some is used. In requests, the subject is

2nd or 3rd person, but not 1st person (Palmer 1979: 168-169).

This section has so far juxtaposed questions asking about the feasibility of an action against requests asking the hearer to perform an action. The issue becomes more complicated when it is considered that can-questions are also used to ask for permission, and permission-seeking questions can sometimes also be used indirectly as requests:

(45) Can I have the salt (please)?

(46) May I have the salt (please)?

In utterances like these the subject is 'I' but it has the role of patient, not agent.

Permission-seeking utterances may contain please even when they are not intended as requests:

(47) May I please leave the room?

Can may also be used in an epistemic sense in questions:

(48) Can he be so foolish?

If can is epistemic, the utterance cannot be a request. Features that force an epistemic interpretation include a stative or progressive main verb. If these features are present, the utterance cannot be an instance of root modality, and cannot be a request.

In addition to clues like these which may be viewed as formal although their explanation lies in semantics or

pragmatics, there are other, different types of clues, which depend on the principles of conversation.

An important factor in determining the illocutionary force of an utterance are the speaker's goals.

In

(49) Can you reach the salt?

the speaker's goal is either to get the hearer to give him information about the hearer's ability to reach the salt or to get the hearer to pass the salt to the speaker. The hearer has to determine which goal the speaker is most likely to have in the context of the utterance. If the question is uttered at a meal time at the table, and the topic of the conversation has been something other than the hearer's abilities, it is more likely that the speaker's goal is to get the salt. If the hearer has hurt himself and the conversation is about to what extent he can move his limbs, the goal may be to get the information.

Yet another consideration in determining the illocutionary force is the interlocutors' beliefs of what the other participant knows. If Mary says to John

(50) Mary: Can you play an eightsome reel?

and John knows that Mary knows that John can play one, John will interpret the utterance as a request.

The considerations about the speaker's goal(s) and about the interlocutors' beliefs about what the other participant knows, can be explained in terms of the maxims that form

Grice's co-operative principle (Grice 1975).

Taking the interlocutor's goal(s) into account when responding is how the speaker makes his contribution relevant to the hearer.

The second point has to do with the maxim of quantity ("Make your contribution as informative as is required, but not more informative than is required"). If the speaker of (50) knows that p and the hearer knows that the speaker knows that p, the question by the speaker, asking whether p, cannot be interpreted literally (as an information-seeking question), but must have another illocutionary force.

A procedural model would have to have some heuristics, using factors such as the ones mentioned above, to determine the illocutionary force of a question, to decide whether it is a request for information, for action, or for permission.

4.2. A note on BE ABLE TO

Although the meaning of BE ABLE TO is often given as "having enough strength, power, means (to do a thing)", that it is in fact equivalent to CAN in the ability sense (Chambers Twentieth Century Dictionary; Quirk et al 1972: 97, 264), it has recently been pointed out that BE ABLE TO does in fact have a wider usage than that.

Coates (1980: 206) gives examples of BE ABLE TO in all the senses that CAN may have: ability, permission, and (circumstantial) possibility.

What then are the differences between CAN and BE ABLE TO? The following is a summary of the differences given by Palmer (1979) and Coates (1980).

In addition to the differences in syntax and form (see e.g. Palmer 1979) they are:

Firstly, BE ABLE TO occurs infrequently compared to CAN (Coates 1980: 213).

Secondly, when used in an assertion in the past tense, BE ABLE TO is factive, unlike could (Palmer 1979: 80). It may be factive also when used in the present tense; Coates (1980: 211) gives the following example:

> As Hardy develops as a writer it is interesting to observe the growing maturation of this device of pictorial illusion, which in his hands becomes a unique skill. In the later novels he is able to employ it in ways that go far beyond a purely descriptive intention.

can would be inappropriate in this example.

Thirdly, BE ABLE TO is not used in senses that depend on implication, as in requests, suggestions, etc. (Palmer 1979: 76). Compare

(51) Can you open the window?

(52) Are you able to open the window?

(53) We can send you a map.

(54) We are able to send you a map.

Fourthly, because BE ABLE TO can combine with WILL and SHALL, it may be used if the speaker wants to make it clear that the reference is to the future.

Palmer (1979) does not come to a clear conclusion about whether BE ABLE TO is sybject-oriented. It would seem that it does not have to be subject-oriented in the sense of 'inherent ability': not being able to do something may be caused by circumstances; for example

(55) If we move the van back, he'll be able to drive the car straight into the garage.

However, it does seem to be more restricted than can in that, for example in

(56) A: Will the painters be able to start on the 29th?
B: Yes.

although their 'ability to start' may be related to circumstances and goals affecting the painters, it seems impossible to interpret this question as relating to the goals of either A or B (such an interpretation is possible with can).

4.3. Must

4.3.1. The senses of must

Most descriptions of must recognise the two senses of "obligation" and "confident inference on the basis of facts known to the speaker", in other words a deontic and an epistemic sense (Chafe 1970: 179; Halliday (1970), Coates

1980: 75). Quirk et al. (1972: 101-2) give the senses "obligation" and "logical necessity", but their terms 'logical possibility and necessity' cannot refer to strictly logical modality: all their relevant examples of must are examples of confident inference rather than logical deduction.

The epistemic sense of "confident inference" must be distinguished from logical necessity. In scientific texts must is frequently used to express logical necessity, e.g.

> (57) If a prime p is a factor of the product ab, then p must be a factor of either a or b. (Courant & Robbins, What is Mathematics, 1963: 24)

but in non-scientific, everyday contexts this sense is not as common as the sense of subjective inference from facts known to or believed by the speaker (Lyons 1977: 805; Coates 1980: 89). Karttunen (1972) discusses how logical modalities differ from epistemic modality: the main observation with respect to necessity is that in natural language an assertion containing must is a weaker assertion than a factual assertion without must (compare 'John must have left' with 'John has left'), whereas in modal logic, 'necessarily p' is stronger than 'p' (Karttunen 1972: 11-12).

In monosemantic accounts of the modals, must is defined in terms of some circumstance or aspect of the world which requires the action or situation of the main predication (Ehrman 1966: 67; Perkins 1980: 134). Ehrman's examples can be grouped into deontic, dynamic and epistemic, since the 'aspects of the state of the world'

that she lists (Ehrman 1966: 67,68) include rules and regulations (example 58), ways to achieve an end (59) and the speaker's view of the probable consequences of the relevant factors (60). Also the logical use is subsumed under this definition: the predication is a logical conclusion 'required' by the premises.

(58) The officer had told him that both lists must be checked.

(59) ... a number of critical meteorological parameters must be met for an aerosol to exhibit optimum effect.

(60) The cars must have had their gas pedals pushed down to the floor boards.

(Ehrman's examples nos. 261, 265, 274)

Palmer (1979) also discusses the dynamic use of must.

In the following, the dynamic sense of must will be discussed first.

4.3.2. Procedure for dynamic must

While a question containing can often, as in our context, serves as a preliminary to arranging something, a question containing dynamic must or HAVE TO, such as

(61) Must the painters start on the 29th?

(62) Do the painters have to start on the 29th?

would not normally occur without some preceding discussion of the feasibility or necessity of the proposed action (cf. Palmer 1979: 96).

The differences in meaning between MUST, HAVE TO and HAVE GOT TO will be discussed in section 4.4. Here we will not be concerned with the question of which one of these, the modal must or either one of the quasi-modals, would be used or would be more common in questions like (61) and (62). The procedure deals with must; apart from syntactic considerations, the procedure for HAVE TO and HAVE GOT TO in this context would have to cover much of the same ground as that for must (cf. Leech 1971: 67: HAVE TO "cannot be semantically separated from MAY, MUST, and CAN").

The answers that the scheduling system can give to the question

(61) Must the painters start on the 29th?

are:

(63) a. No, they do not have to.

b. Yes, they must if the goal is to be met.

c. No, they cannot start on that day if the goal is to be met.

d. No, it is impossible.

e. Yes, they will start on that day.

In order to answer the question the system must check that it is feasible to start the job on the 29th and that it is not feasible to start at any other time, given the goal of finishing the house by time x. These two steps are achieved by calling the planner first with plan(p(f) & g) and

then with plan(p(Foc) & g) & Foc=/=f. If both succeed, the result is that p is feasible but not necessary, and the appropriate answer is (63a).

If plan(p(f) & g) succeeds but plan(p(Foc) & g) & Foc=/=f fails, p is necessary with respect to the goal, but the procedure has to check whether p is necessary on its own as well. It calls plan(p(Foc)) & Foc=/=f; if this succeeds, the result is "must(p) if g", and the answer is (63b). If it fails, the result is "will(p)", and the answer is (63e).

If the first call, plan(p(f) & g), fails, the procedure has to see whether p is feasible, and it calls plan(p(f)). If this succeeds, the result is "not(feasible(p)) if g", and the answer is (63c). If it fails, the result is "not(feasible(p))", and the answer is (63d).

This is exactly the same procedure as that employed for direct questions with <u>can</u>. The results it gives are repeated here for convenience:

(R1) can(p)

(R2) must(p) if g

(R3) not(can(p)) if g

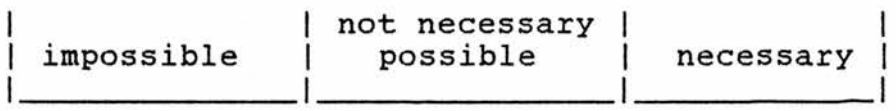
(R4) not(can(p))

(R5) will(p)

Feasibility as the meaning of <u>can</u> was defined as "possible but not necessary". In the procedure 'not have to' is defined identically to <u>can</u>: 'not have to' means "possible (i.e. not impossible) and not necessary".

In the same way that possibility was defined in the

narrow sense as the meaning of can, not-necessary is defined as excluding impossibility for the meaning of 'not have to'. Thus dynamic 'possible' and dynamic 'not necessary' are interpreted as covering the same area on the scale of modality:



The system uses one and the same procedure for producing the results that have been discussed for both can and must questions. Thus, for

(64) Can the painters come on the 29th?

and (65) Must the painters come on the 29th?

the process for finding the degree of modality for the proposition is exactly the same. The actual natural language answers to the question may of course differ depending on what form (what modal) the question has. The difference is clearest when the result is (R1): can(p), in which case the answer to (64) is

Yes, they can.

and possible answers to (65) are

No, they don't have to.
No, they needn't come then.
No, they don't need to.

Since the actual natural language answer depends not only on the result of the modal procedure but also on the

form of the question, the answer is formulated after the degree of modality has been established, on the basis of both the degree of modality and the form of the question.

The responses that the system gives to can and must questions are repeated here:

Can the painters start on the 29th?

a. Yes, they can.
b. Yes, in fact they must if the goal is to be met.
c. No, they cannot if the goal is to be met.
d. No, it is impossible.
e. Yes, they will come on that day.

Must the painters start on the 29th?

a. No, they don´t have to.
b. Yes, they must if the goal is to be met.
c. No, they cannot start on that day if the goal is to be met.
d. No, it is impossible.
e. Yes, they will start on that day.

As was explained in connection with can, the logical equivalent of ´necessary(p)´, ´not possible that not p´, is not testable within our framework, since the planner cannot deal with plan(not(p)). Instead, the procedure tests whether there is a state in which p is not true by altering the focus within p and then calling the planner. (Cf. discussion in section 4.1.2)

What support is there for the claim that must is always goal- oriented, not only in the deontic but also in the dynamic sense?

Those linguists who analyse root must univocally in terms of "x requires p", claim that there is a causal link between the action p and some condition x that requires it

(in the deontic sense, x is the deontic source) (Perkins 1980: 65; Ehrman 1966: 67; cf also Halliday: 1970, and Antinucci & Parisi 1971: 28-29 for causation in connection with must).

The use of the words like 'requires' or 'demands' indicates that the link between the cause and the action is something which is perceived as to some extent flexible: the 'cause' does not inevitably lead to p but only demands p. Perkins' (1980) definition "K(C entails X)" implies that he does not share this view.

The claim here is that if there is no choice about whether the action or process will take place, a speaker would not necessarily use must, or any other necessity modal. He could use some factual expression with the sense that p is the case or will be the case.

The if-clauses in the following examples are different in that in (66) the if-clause expresses a factual condition, and in (67) the if-clause expresses a condition dependent on the subject's intention:

(66) If you press the red button, the central heating will come on.

(67) You must come by seven if you want to see the film.

Both of these are of course normal 'if p then q' sequences, (67) being something like 'if want(agent,p) then nec(q)'. An interesting fact about these examples is that in (66) the if-clause expresses a condition for achieving the event of the main clause, whereas in (67) the proposition

under 'must' in the main clause expresses a condition for the event expressed under 'want' in the if-clause.

4.4. A note on HAVE TO and HAVE GOT TO

There is a lot of discussion in the literature as to the distribution of MUST compared to that of HAVE TO or HAVE GOT TO with respect to whether they differ in meaning, and if so, how.

Palmer (1979: 93) sees no difference in meaning between HAVE and HAVE GOT TO and consequently treats these as one verb with two variants. According to him, MUST is either deontic in a subject-oriented sense, or it is 'neutral', which seems to be intended to refer to dynamic modality. HAVE (GOT) TO on the other hand is either neutral or external (the necessity comes from an external source, not the subject).

The study by Brown and Miller (1975) on the use of the modal verbs by Scots indicates that the differences between how MUST and HAVE TO are interpreted are less clear-cut than is suggested by the usual categorisation of HAVE TO as expressing objective and MUST as subjective deontic modality.

Coates (1980), after pointing out that there are differences in meaning and distribution between HAVE TO and HAVE GOT TO (see Coates 1980: 108), claims that HAVE TO is always used to convey objective deontic modality whereas both MUST and HAVE GOT TO can be either

subjective or objective (Coates 1980: 113). She does not discuss the differences between these verbs when they are used in a dynamic sense.

All of these verbs occur in questions (Palmer 1979: 96):

(68) Must you go?

(69) Have you got to go?

(70) Do you have to go?

4.5. Wh-questions

This study has so far dealt with modalised yes-no questions and how to answer them. The next task is to see how the framework could handle wh-questions.

Direct questions ask about the truth or falsity of a proposition; modalised direct questions about the feasibility, possibility or necessity of the proposition. Wh-questions ask for a value for a variable. If we have a dynamically modalised wh-question it asks for a value with which the proposition is feasible or necessary.

For the questions

(71) When can the painters start?

(72) When must the painters start?

the parser returns a simple semantic representation in the form

question(when, can(starttime(painters, T)))

question(when, must(starttime(painters, T)))

(the first parameter, here 'when', indicates the type of question, i.e. yes-no or wh-question).

To answer (71) the procedure must find a value for T such that when T is instantiated to that value, the proposition is feasible (the system finds a plan for the proposed action p).

What kind of answer might (71) get, assuming again that the question is related to the goal of finishing the house at time X?

If it is found that the job of painting is not on the critical path, (73a) and (73b) are possible answers:

(73) a. They can start any time between t1 and t2.

b. They can start for instance on t1.

Answer (73a) is probably the most helpful in this particular context; in other contexts (73b) might occur. This would be in contexts where it is impossible or very cumbersome to give a list of all the possible values for T.

It would seem that in computing an answer for a modalised wh-question one has to take into account whether the set of possible values is finite and small, finite but large, or infinite. In the latter two cases it either may not be worth computing all possible values or it is impossible; instead, the first value that is found could be offered in the answer. (Or it may be easier to list the values with which p is not possible, stating that all other values make p possible; e.g. 'They can start any time except Wednesday'.)

If the job is on the critical path, we will assume again that it is more natural to answer

(73c) They must start on the 29th (if the house is to be finished by day 32).

rather than

They can start on the 29th.

The result that is the basis of answer (73c) is achieved using the same procedure as for the equivalent answer (must(p) if g) for direct questions: it is ascertained that there is only one value for T which is compatible with the goal.

With wh-questions, as with direct questions, the answer

(73d) They will come on day t1.

is given if the starting time has been fixed by agreement.

Giving a negative answer to a wh-question is more complex than giving a negative answer to a yes-no question. Direct questions ask whether something is true or false and expect either 'Yes' or 'No' as the answer. A wh-question, e.g.

(74) When can we start building the swimming pool?

presupposes that the swimming pool can be started at some time (cf. Lyons 1977: 758). A negative result means that the presupposition is wrong, and this has to be expressed in the answer. If a swimming pool is not mentioned in the

contract, the answer may be something like

(75) There is not going to be a swimming pool in this house.

As with direct questions, with wh-questions there is the distinction between impossibility because of physical (inflexible) circumstances and impossibility because of incompatibility with a goal: p may be impossible with any value, or impossible with any value if the goal is to be achieved.

This means that the responses as regards the degree of modality are the same as to direct questions, and the procedure proceeds in a similar way.

The obvious differences are the following. There is no problem in finding the focus: it is the wh-element. The task is to find a value or values for the unknown element, which is an uninstantiated variable at the beginning of the procedure, and the degree of modality, and give both the value for the focus and the degree of modality in the response.

Figure 5 given below represents the steps needed to answer wh-questions which contain <u>can</u> and <u>must</u>. If it is compared to the tree for direct questions with <u>can</u> and <u>must</u> in Figure 3, it will be seen to be similar except for two things. The first is that the focus variable Foc is not instantiated at the beginning: at node 3 the procedure tries to find a value f for focus such that plan(p(f) & g) succeeds. If it succeeds in this, it proceeds to see if

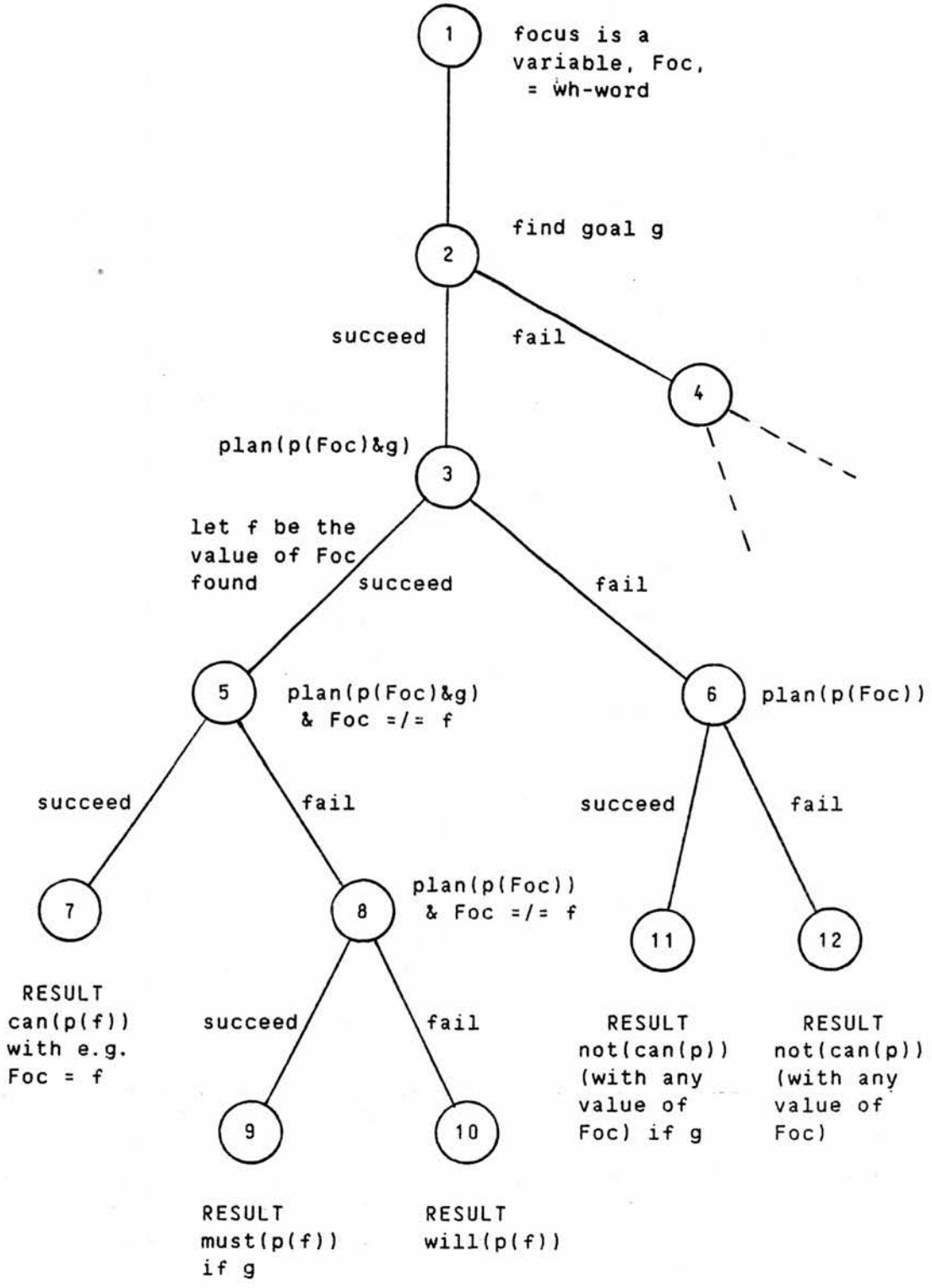


Figure 5 : Decision tree for answering wh-questions containing dynamic can or must

there is another value f1 such that f1=/=f, at which plan(p(f1)&g) also succeeds, and so on (cf. discussion of direct questions and Figure 3 in section 4.1.2.).

Secondly, in the results, the value of focus has to be inserted into p. The answers can then be formulated. The possible results and (some) equivalent answers are listed below. As was pointed out above, negative answers may take various forms because they contain a denial of a presupposition.

(R1) can(p) with e.g. Foc=f

(R2) must(p) with Foc=f if g

(R3) not(can(p)) with any value of Foc if g

(R4) not(can(p)) with any value of Foc

(R5) will(p) with Foc=f

(A1) They can start for instance on day 20.

(A2) They must start on day 20 if the goal is to be kept.

(A3) It is impossible to do that job if the goal is to be kept.

(A4) It is impossible to do that job.

(A5) They will start on day 20.

Negative wh-questions present a problem. Negative direct questions can be processed by dropping the negation, finding out whether p is true or false, and responding 'Yes' or 'No' respectively. For example,

(76) a. Has this bill been paid?

b. Has this bill not been paid?

(76a) and (76b) will both get the answer 'yes' if the bill has been paid, and 'no' if the bill has not been paid. The answer to (76a) depends on the same set as the answer to (76b). But the answers to

(77) a. Which bills have been paid?

b. Which bills have not been paid?

will belong to complementary sets.

This is true also of modalised wh-questions. For

(78) When can the painters come?

one has to find the set of times such that there is a strategy for 'starttime(painters, Time)', and for

(79) When can the painters not come?

a set of times such that there is a strategy for 'not(starttime(painters, Time))'. This is impossible in the present framework, as the planner cannot be asked for a plan for a negated event. However, the answer to (79) could be achieved by listing those instantiations of Time at which the planner fails to find a plan for 'starttime(painters, Time)'.

5. DEONTIC MODALITY

The purpose of this section is to look at how the procedure for dynamic modality, described in the previous section, could be extended to cover deontic modality as well. The claim is made here that with some modification the system can handle all root modality, both dynamic and deontic. It must be pointed out, however, that the procedure has not been tested in the form that would handle deontic modality.

In this section we are concerned with direct questions containing modals and answers to these questions.

Deontic modality differs from dynamic modality in that there is a 'deontic source', an authority from which the permission, obligation, or prohibition derives. In interpreting questions like

(1) Can I borrow your car?

(2) Can I take this book out (of the library)?

one has to detect from the context and the content of the question whether it is a situation in which there is a relevant authority, whether the possibility depends on some authority which is, or needs to be, consulted. 'Authority' here may mean a person who is entitled to have control in that situation, or rules, regulations, laws or a moral or an ethical code that are relevant in the situation. In example (1) the permission granting authority is the addressee;

example (2) refers to the authority of the regulations of the library.

Examples of situations in which one can determine from the extralinguistic context that authority and therefore permission sense may be involved, are situations where the social status of the interlocutors is not the same, or situations in which one interlocutor's ownership or being in his territory gives him the right to control some actions. Sometimes inferences like these can be made easily from the question itself, as in (1) from 'your car' and 'borrow' (the question 'Can I borrow ...?' would probably always be regarded as a question about permission). Other times they have to be deduced from the context only, as for example, if Bill says to John in John's house

(3) Can I make a phone call?

or presumably sometimes the interlocutors know or believe that the case is such: in example (2)

(2) Can I take this book out?

the context contains an authority in the sense that library regulations determine for example that some books are allowed out of the library and others are not, and the speaker of (2) shows that he believes that such regulations may exist.

In contrast to examples (1-3),

(4) Can I get a train to Inverness?

is less likely to be intended or interpreted as a question about permission because normally everybody is allowed to travel by train and no authority applies to the situation.

The concept of frames (see Minsky 1975) is perhaps useful here; there are frames in which there is an authority that is regarded as having control over some things in that frame. In the same way, with dynamic modality, one has to assume frames that make it possible for the system to choose the goals that are relevant in each case; the system has to pick only the goals that are relevant for the question at hand, rather than consider a random selection of goals.

Confronted with a question containing <u>can</u> the procedure should first see whether the situation is such that there might be an authority who controls the proposed action.

Apart from the general points mentioned above, how a deduction process like this works is not known, and it is not possible at this stage to speculate what kind of algorithm would be needed for this. The following discussion is based on the assumption that the system has some mechanism of pinpointing the authority if there is one.

There are the two obvious cases: either the person the question is addressed to is the authority (subjective deontic modality), or the authority is something external to the interlocutors: a third person, or an institution, laws, etc. (objective deontic modality).

To answer questions of subjective deontic modality, such as

(1) Can I borrow your car?

(5) May I borrow your car?

(6) Must you go?

(7) Could Mr Smith see you tomorrow?

(where the respondent is the authority), the respondent has to look at his own goals in order to decide what the modal status of the proposed action is: whether it is permissible, required, or to be prohibited. In checking the compatibility of the proposed action with his goals, the same procedure can be used as for dynamic <u>can</u> or <u>must</u> (if the question asks for permission or about an obligation).

If the modality is objective, i.e. the authority is some external source, the answer depends simply on whether the speaker happens to know what the decision of the deontic source is: either the speaker knows what the answer of the deontic source to the question is and is able to report it, or he does not know it and cannot answer the question except by acknowledging that he does not know.

In a question-answer situation it is the respondent who has to decide, irrespective of the meaning intended by the questioner, whether there is an authority and who or what it is (whether an answer will be an instance of subjective or objective deontic modality or neither). For example,

answers to (1) might be

(8) a. It is not my car, it is my brother's, you'll have to ask him.
b. It is actually not my car,it is my brother's, but he never lends it to anybody.

If the system finds that it is the system's permission, etc. that is being asked about, it will then look at its own goals to see whether they are in conflict with the proposition. The respondent's goals may include the goal of asserting his own authority.

The network presented in section 4.1.3 for dynamic can and must can be modified to accommodate deontic can and must by adding nodes for searching for an authority, and for deciding whether the deontic source is the speaker or a third party. An (informal) modified network is given in Figure 6.

It might be argued that if in the case of subjective deontic modality the speaker simply assesses his own goals in relation to the proposition before granting or refusing the permission, one could access these goals directly, rather than going through the concept of 'deontic source'. So the difference between subjective deontic and dynamic modality would be in whose goals are taken into account: subjective deontic modality would look (mainly) at the speaker's goals, whereas with dynamic modality the goals of anybody who is involved in the action may need to be looked at (the goals of the hearer, the speaker, the agent or mutual goals). Perhaps also objective deontic sources

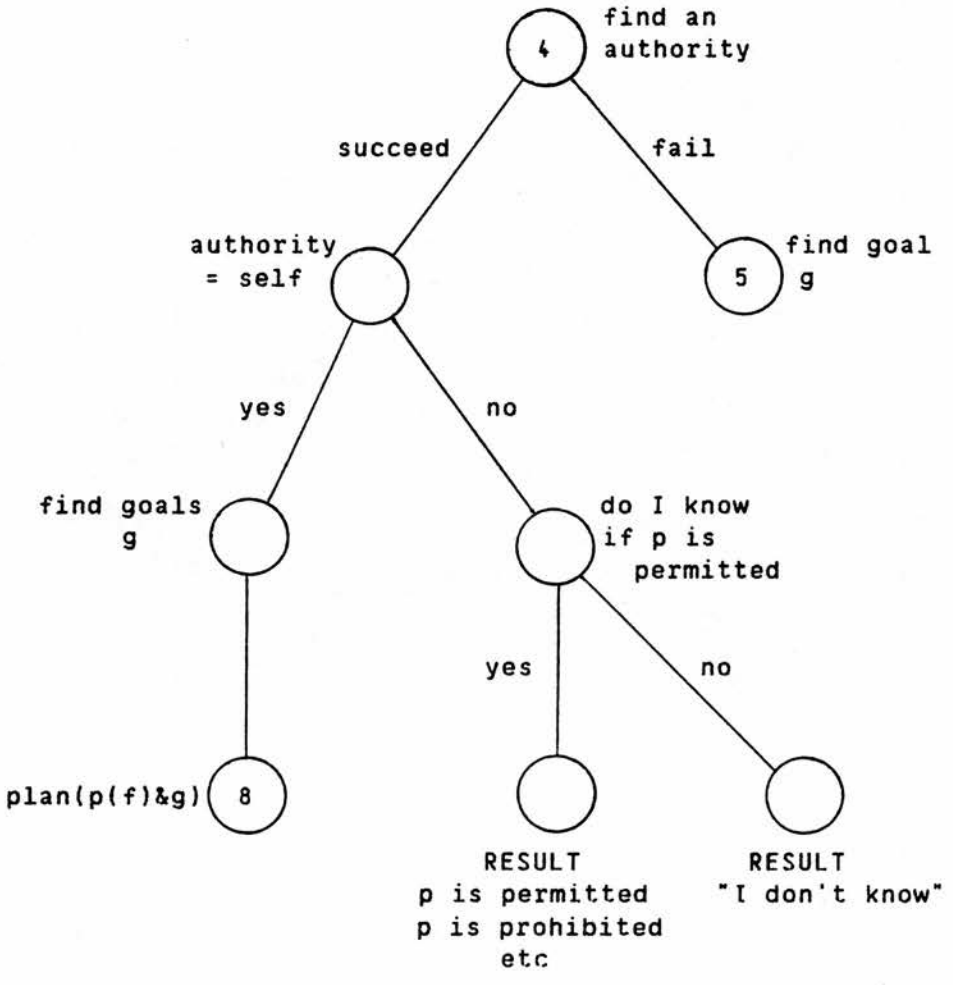


Figure 6 : Decision tree for deontic modality (this tree can be combined with the one given in Figure 4. Node 4 of this figure is inserted between nodes 2 and 5 of Figure 4.)

could be represented in the same format as goals. However, this approach would fail to account for direct appeals to authority, as in conversations like the following, which are perhaps unusual but not impossible:

(9) A: Can I borrow your car/take this book out/ take the afternoon off?
B: No.

(10) A: Why not?
B: Because I say so/because that is what the rules say/because the boss says so.

It can again be argued that as soon as we have a question with <u>can</u> even if it is intended in the deontic sense, like (1), the planner is called and it checks whether there is a plan for the proposition which takes into account all factors (physical preconditions, goals, permission by deontic source), not only the permission of the deontic source. Example (1)

(1) A: Can I borrow your car?

might get the answer

(11) B: I don´t mind you borrowing my car, but it doesn´t work.

and in fact a mere ´Yes´ would be an odd answer if B knew that it did not work.

6. EPISTEMIC MODALITY AND ITS RELATION TO ROOT MODALITY

This section will deal mainly with subjective epistemic modality. This is meant unless otherwise specified.

The discussion is concerned with the following points: Firstly, some well-known differences between epistemic and root modality are listed, with comments on them and on some of the modal verbs that occur in the epistemic sense. Secondly, there is a discussion of the feasibility of procedural treatment of epistemic modality, and to what extent the procedures for root and epistemic modality are parallel to each other.

Root and epistemic modality differ from each other in the following respects, most of which are well-documented.

Root modality has to do with the feasibility or necessity of events or actions (dynamic modality) or the permission or obligation to perform an action (deontic modality). Epistemic modality has to do with propositions: it expresses the speaker's assessment of the truth of a proposition (Palmer 1979: 3-4, Lyons 1977: 797).

In sentences containing root modality the event (the main predication) is always in the future (or, more exactly, future in relation to the point of reference). With epistemic modality, no such restriction applies: the event or state referred to in the main predication may be past, present, or future.

With root modality, there are restrictions on the nature of the process referred to by the main predication. The verb has to be a voluntary, agentive verb. The modality is linked to actions or events; typically, stative verbs and the progressive aspect do not occur with root modality. This is clearly the case with deontic modality, which is about permission etc. for an agent to do something.

There are examples which contain superficially stative verbs like

(1) You must know this by Monday.

but in these the predication cannot be interpreted as semantically stative but has acquired a meaning that implies a transition from one state to another (cf. Steedman 1977: 229-230). Similarly,

(2) You must be careful.

can be interpreted as an activity: "You must behave carefully"[1].

As regards dynamic modality, <u>can</u> requires an agentive predication; when it occurs with inanimate subjects, e.g.

(3) This record player can play for four hours without attention.

an implicit agent has to be assumed, i.e. "if somebody chooses to make it do that". Another explanation for examples like this is that people sometimes attribute human properties to machines: machines may be talked about

as if they were agents.[2]

With must, however, there are examples of the following type, which cannot be explained in terms of acquisition of another, event-type meaning:

(4) Assistant Quantity Surveyor wanted urgently... Must have experience in rehabilitation and restoration works (The Scotsman 14.8.1984)

(5) You must be very fit to be accepted on this mountaineering course.

These could be paraphrased as meaning something like "to achieve goal X, the subject must have property Y". These examples can only be interpreted in the root sense if the subject is indefinite. Thus (4) says "For this job, we will only consider applicants who have experience in rehabilitation and restoration". (5) can only be interpeted in the root sense if 'you' is the impersonal you (= one). If 'you' refers to the interlocutor, the sentence does not express a requirement but an inference, and is epistemic.

If these observations are correct, then dynamic modality may have a stative (as well as a dynamic) complement: a transition to the goal state has the precondition that the subject have property Y.

No restrictions as to the nature of the main predication apply to epistemic modality. The main predication may refer to a state, an action in progress, or to an event in the past, present or future.

Because of the restrictions on the types of predication that may occur with root modality, ambiguity (or

indeterminacy) between root and epistemic modality can only occur when the reference is to a non-progressive action in the future (cf. Lyons 1977: 826):

(6) He must come tomorrow.

But Palmer (1979: 44) points out that the use of sentences like (6) in the epistemic sense is rare because of the ambiguity and the possibility of misinterpretation as a mand, and if an epistemic sense is intended, there is an alternative, unambiguous form available:

(7) He must be coming tomorrow.

However, utterances with <u>may</u> and a non-progressive action can be used in either epistemic or root sense:

(8) He may come tomorrow.

(cf. discussion below).

Some scholars have claimed that epistemic modality differs from root modality also in that epistemic modality contains an assessment and an expression of the probability of the action. Wertheimer (1972: 125) attributes the difference between <u>can</u> and <u>may</u> to this. Halliday (1970: 349) says that (epistemic) "modality is the speaker's assessment of probability and predictability". According to Lyons (1977: 800) at least objective epistemic modality may be a qualification in terms of probability, although he rejects the idea that in everyday use this expression of probability has anything to do with probability as a

scientific, calculable concept.

Intuitively, it does seem that at least with epistemic must and can't, as in

(9) a. It must be raining.

b. It can't be raining.

the speaker somehow expresses his assessment of the likelihood of whether the proposition is true or not (based on some argument that he has in mind). However, these sentences do not mean the same as

(10) a. I think it is likely/probable that it is raining.

b. I think it is unlikely/improbable that it is raining.

Rather, they mean "Something makes me think/conclude that it is raining".

As for may, it is even less appropriate to paraphrase it with any reference to probability. The following is the interpretation of may given by Cresswell (1979: 294): "Nothing in my knowledge rules out the possibility that [it is raining]".

Modal verbs that are used in the epistemic sense include may, might, could, must, should, ought to, and the semi-modals HAVE TO and HAVE GOT TO. In non-assertive contexts can also occurs. May does not occur in questions in an epistemic sense. In negated sentences we get cannot, could not (and cliticised forms) if the modality is negated (i.e. wide scope: it is impossible that p is true), and may

not, might not, if the main predication is negated (narrow scope: it is possible that not p).

The differences between the uses and senses of can and may are especially interesting and have been discussed by some linguists.

Coates (1980) observes, on the basis of statistical analysis of corpuses, that in everyday language, the uses of can and may hardly overlap at all: can is primarily used for root possibility, may is primarily used for epistemic possibility; both are used for permission, but not as frequently as for possibility (see Coates 1980: 176).

Coates (1980) and Leech and Coates (1980) claim that although the different senses of can and may overlap (the deontic sense), and although the root senses of can may be seen as related to each other, yet may is clearly ambiguous: the root and the epistemic senses are clearly different (cf. discussion on the monosemy/polysemy issue in chapter 7).

R. Lakoff (1972) explains the difference between can and may in terms of the actual world versus possible worlds: may refers to possible worlds, can to the actual world that the speaker is in. However, her analysis is limited to sentences of the type

(11) Football players may be sex maniacs.

(12) Football players can be sex maniacs.

in which can occurs in a quantifier sense.

Perkins (1980) takes a strong monosemantic view in

proposing that can and may have an identical core meaning, "K(C does not preclude X)", and that they differ only in that they are used with different sets of contraints on the possible values of K, C and X. The following table, adapted from Perkins (1980: 61, 65-6), gives the types of values the variables may have, what the modality in question is and the ground that can and may cover:

	dynamic	deontic	epistemic
K:	natural laws	/ laws of the society	/ laws of reason
C:	empirical circumstance	/ deontic source	/ evidence
X:	event	/ event	/ proposition

can

may

Perkins' wording of the definition of can has been criticised in section 4.1.1. Apart from that, the level of this analysis is so general that not much can be said against it; but it is very difficult to see how to apply this formula either in interpreting a modalised utterance or in producing one.

Another difference between epistemic and root modality is seen in cases of necessity, in the nature of the 'cause' or 'reason' for the necessity. In root modality the necessity depends on a goal of one or both of the interlocutors (or the agent); the goal is obviously in the future:

(13) You must go now (if you want to see the film).

In epistemic modality, as in

(14) They must have left.

interpreted as "Something that I know causes me to infer that p is the case", the reason or cause of the inference is not restricted in time.

Epistemic modality is rare in questions (Palmer 1979: 56). One would normally use a direct (non-modalised) yes-no question in asking whether a proposition is true or false (cf. R. Lakoff 1972: 242). Compare

(15) Are they on holiday?

(16) a. Can they be on holiday?
b. Could they be on holiday?
c. Might they be on holiday?

(15) is the normal question, unless the questioner has some reason to believe that the respondent does not know (for certain) whether the proposition is true or not. The modalised question asks for the respondent's assessment of the truth of p.

(15) may elicit an epistemically modalised response if the respondent does not know the answer.

In question-answer dialogues one would thus expect epistemically modalised statements to be produced in responses to non-modalised questions (as well as, of course, to modalised questions), unlike with root modality, where the question that prompts a modalised response is normally modalised.

If the respondent does not know whether p is true (whether they are on holiday), the answer is either 'I don't know', or a (subjectively) qualified statement; the latter

could be a modalised statement such as 'They may/must/can't be'.

In order to see how epistemic modality could be approached procedurally, let us consider in what respects such a procedure would be similar to and how it would differ from the procedure proposed in the previous chapters for root modality.

Arguments in favour of treating epistemic modality in a procedure similar to the one already proposed include the following considerations:

(1) There is the obvious and clear parallelism in the degrees of modality. Epistemic possibility can be defined by excluding both impossibility and necessity, as was done with root possibility.

(2) Examples of dynamic modality can be turned into examples of epistemic modality: the goal that causes the necessity in dynamic modality becomes the reason or explanation that justifies the inference in epistemic modality:

(17) a. If John wants to see the film, he must come by seven.

b. If John saw the film, he must have come by seven.

c. If John wanted to see the film, he must have come by seven.

(18) a. If you want to get this job, you must have experience in restoration.

b. If he got the job, he must have experience in restoration.

In examples (17), coming by seven is a precondition for seeing the film; in (17a) the goal of seeing the film creates the necessity of satisfying the precondition; in (17b) and (17c) it causes the inference that John must have come by seven.

In these examples, epistemic and dynamic modality are exactly parallel: dynamic modality can be turned into epistemic modality by assuming that the goal has been achieved (is true in the present state of the world).

But this argument only applies to cases in which root modality is possible, that is where a transition from one state to another is involved. There is no root equivalent of examples like

(9) It must be raining.

(3) There is the argument (cf. Wertheimer 1972, Perkins 1980) that epistemic and root senses can be defined in the same way if we allow the factors that determine the modality to be of different types. Root modality is concerned with preconditions for actions; epistemic modality would (in addition) look at general rules about the world, relevant facts that the speaker knows about the entities/issues in question (e.g. their properties), and so on, and make inferences on the basis of factors like these.

Arguments <u>against</u> treating epistemic modality in the same way as root modality include:

(1) When epistemic modality occurs with stative or progressive verbs, one cannot analyse the sense of the

modal in terms of plans and goals, as with root modality:

(19) a. This must be John´s car.
b. This can´t be John´s car.
c. This may be John´s car.

(20) a. John must be fishing.
b. John may be fishing.
c. John can´t be fishing.

Statives and progressives describe what is true in the present state of the world: no change of state is involved. Plans and goals involve transitions from one state to another.

(2) There is a difference in the status of the proposition (= the complement of the modal verb):

Root modality describes whether it is feasible, necessary, permitted, etc., to get from one state to another (whether there is a plan/a strategy for the action). The status of the proposition is open in the sense that the action has not taken place and the modality is not about whether it will take place or not, but about whether it is feasible (if an agent chooses to carry it out). The modality itself is part of the ideational meaning of the utterance, to use Halliday´s (1970: 343) term, and the whole statement is a confident statement which the speaker believes to be true (cf. also Lyons 1977: 843).

In (subjective) epistemic modality, on the other hand, the status of the proposition is not open in the sense that the proposition is actually true or false (at least in cases in which the reference is to the past or the present), but the speaker does not know which is the case.

The fact that a modal verb is used may itself be taken as a signal that the speaker is not sure but is only giving his comment. (However, the logical (objective) use of must and cannot simply indicates that what the speaker says is based on an argument, that it is a deduction, and the speaker may be confident about the truth of what he is saying; e.g. 'Today is Tuesday, so tomorrow must be Wednesday').

This difference in the status of the modality and of the proposition explains why one can say

(21) John can come tomorrow but I know he won't.

but it does not seem possible to interpret

(22) John may come tomorrow, but I know he won't.

in an epistemic sense.

The only instances in which, in epistemic modality, the status of the proposition is open in the actual world occur when epistemic modality is connected to a future (non-progressive) action, as in

(23) John may come tomorrow.

(We take the progressive 'John may be coming tomorrow' to mean "John is intending to come tomorrow" where the intention is in the present time.)

(23) may be interpreted in two ways, as equivalent to "John may be coming tomorrow" (cf. Steedman 1977), where the truth of p is not open (but the speaker does not know

whether p is the case), or in the sense "It is possible that John comes tomorrow", i.e. he may decide to come tomorrow (nobody knows at the moment whether p will be the case). Other examples which are open in the sense that the event may happen or not are

(24) It may rain.

(25) He may win (the game).

(3) In determining root modality, when it is linked to a goal or goals, we need to consider only those factors that are relevant for the goal(s), i.e. the preconditions for that goal. Thus in

(26) Can the painters come on the 29th?

if the understood goal is finishing the house by time X, the respondent needs only to consider the critical path network in determining the modality. If the point of the question is whether the painters are free then, the respondent needs only consider the painters' diary and can ignore other considerations.

Epistemic modality is not restricted to the consideration of a limited number of factors in that way. To answer

(27) Might the painters come on the 29th?

one has to consider anything that might have a bearing on their coming.

In view of all the points that have been made so far, an attempt to find a procedural treatment for epistemic modality would encounter a number of problems, which can only be presented here as open questions.

(1) It seems that in answering a question like

(28) Are the painters coming tomorrow?

if the respondent does not know whether their coming has been arranged or not (which is what the question is interpreted to be about here), the first point to decide is whether to answer 'I don't know' or to express a qualified statement as an opinion. Why do people utter statements containing their assessment of the truth of something? The reasons may be social (cf. Halliday 1970, where epistemic modality is considered as part of the interpersonal function of language). Perhaps it has also to do with how much the speaker thinks he knows about the matter in question. This problem is one of the things that make it difficult to envisage epistemic modality in a language understanding system. How does one motivate the system to volunteer an opinion?

(2) Although the procedure for determining the modality would in general terms have to proceed in the same way as for root modality, there are differences that make it impossible to combine the procedures. The similarity appears in the way that the degree of modality is defined: in both cases, it is first determined whether p is possible; if so, whether p is necessary. If there is nothing that

makes p necessary, the result is (in epistemic modality) may. Thus may, like can, may be defined as "not impossible and not necessary".

However, the tree for root modality which looks for plans (and whether they are consistent with goals) cannot handle the cases mentioned above where no plans are involved, e.g. stative predicates. The procedure would have to make inferences or deductions on the basis of other factors (general principles, knowledge about individuals, etc.), and it would require some mechanism to search for these factors.

It is possible, however, at least in cases in which the predication denotes an action, that the tree for establishing the degree of root modality is used as part of the epistemic procedure: if something is dynamically or deontically impossible, possible or necessary, this fact may be used in the inference process for determining whether it is epistemically impossible, possible or necessary.

(3) Another problem is the number of factors involved. As was pointed out above, in determining root modality, when the 'point of view' of the question is a particular goal, only those factors that relate to that goal need to be considered. With epistemic modality, e.g responding to (27) or (28), the respondent answers on the basis of what he knows, but also knowing or believing that he does not know all the possibly relevant facts. But he has to try to consider the question from all points of view, rather than one point of view.

It would seem on the basis of the foregoing that it would be necessary to have a separate, if in many ways parallel procedure for determining epistemic modality. This procedure might have access to and use (when appropriate) the root modality procedure as part of its inference making process.

Assuming that there is an ambiguity in some modal verbs between epistemic and root sense, what are the clues that are used in deciding whether to interpret a modalised utterance in a root or epistemic sense?

There are some very straightforward clues (already discussed): If the predication is stative, progressive, non-agentive, or non-future, the modality is typically epistemic (cf. Coates 1980)[3]. If the predication does not have any of these features, as in

(29) John may come tomorrow.

(30) John may not come tomorrow.

(31) You must read novels.

the clues have to come form the context in which the utterance is made (or the utterance remains ambiguous).

Consider also the difference between (31), which may be deontic, dynamic, or epistemic (if the activity is habitual), and

(32) You must read this novel.

where the fact that <u>this</u> <u>novel</u> is definite causes the interpretation of a (single) future action to be strongly

favoured and makes epistemic interpretation unlikely. Definiteness and indefiniteness may thus also play a part.

7. SOME GENERAL CONSIDERATIONS

7.1. Monosemy versus polysemy and indeterminacy

One of the questions on which there has been a lot of debate in connection with modals is whether they are monosemantic or polysemantic: whether each modal has a basic meaning, of which its different 'senses' are variations, or whether the senses are discrete, not explainable in terms of a 'basic' or 'core' meaning.

Exponents of the monosemantic view include Ehrman (1966), Wertheimer (1972), and Perkins (1980), and of the polysemantic view Leech (1971) and Palmer (1979).

Many studies on mood and modality have either not taken a strong view or have taken an intermediate view: Leech and Coates (1980) consider root senses to be closely related to each other but conclude that there is a discrete demarcation between root and epistemic meaning.

Recently Palmer (1979), Leech and Coates (1980), and Coates (1980) have discussed the related problem of indeterminacy. With many modals, indeterminacy between the different senses is common in actual usage. It is therefore necessary to find a description of the modals that can handle it.

Indeterminacy is more of a problem for the polysemantic than for the monosemantic hypothesis. If each modal is assumed to have a basic meaning, the different senses are

seen as variants of this meaning, rather than as discrete meanings; indeterminacy between the senses is allowed for. If, however, the senses of a modal are regarded as discrete meanings, indeterminacy between the meanings is more difficult to explain (cf. Leech and Coates 1980: 79-80).

Here the question of polysemy versus monosemy will be taken up first and then the problem of indeterminacy.

What does it mean for an element to be polysemantic or monosemantic? If polysemy means that the hearer has to choose between two or more meanings, as for instance in

You must read novels.

between epistemic and root must (inference vs. obligation or dynamic necessity), then elements like must and may certainly are polysemous. But often the hearer also has to choose between different root senses in order to make an appropriate response, as with

Can the painters start on the 29th?

(cf. earlier discussion).

On the other hand, one can claim that modals are monosemantic if their meanings are defined in a very general way, i.e. if there is a common denominator which is present in all occurrences of the modal. Thus, for example, must can be defined as "something requires p" (cf. Perkins 1980). This level of generality has the consequence that can and may will be defined in an identical way ("nothing prevents p") (as is done in Wertheimer 1972 and Perkins

1980). These generalisations may be correct and useful in that they describe what is common in all the uses of a given modal (or two modals), but they do not provide the means for interpreting a modal utterance, because they necessarily remain at a high level (in a procedural sense) and ignore the complications and divergences at the lower levels.

Looking at the issue from the procedural viewpoint, rather than to ask whether modals are mono- or polysemantic, it might be more reasonable to ask at what level in the procedure the differences between the senses occur, i.e. to what extent a modal is monosemantic or polysemantic.

The points made in the section on epistemic modality about the differences between root and epistemic modality led to the tentative conclusion that these are discrete and would probably have to be treated by different, although similar, procedures.

We could say that the difference between the epistemic and the root senses is at a higher level in the procedural tree or network than the differences between the different root senses, which are lower down the tree.

This agrees with the view of Leech and Coates (1980) that there is a demarcation between root and epistemic modality, but that different root senses (and different epistemic senses) are closely related to each other. However, in the procedural approach, the polysemy versus monosemy issue does not have to be seen as a burning

question but can be seen as a matter of viewpoint and degree.

We will next look at a description of indeterminacy and then at how the procedural system envisaged here could handle it.

Leech and Coates (1980) divide indeterminacy into three types: gradience, ambiguity and merger.

Gradience occurs when there are two senses, or two categories of meaning (e.g. can "permission" and "possibility") with cases which are intermediate between the two categories. The intermediate cases may be interpreted as being closer to one or the other of the meanings, i.e. there is a continuum from one meaning to another.

A modal is ambiguous if the addressee has to choose either one meaning or another; contextual clues generally resolve the ambiguity.

A merger of two meanings occurs if the two interpretations are mutually compatible in the given context, so that the modal may be interpreted as having both meanings (see Leech and Coates 1980).

They point out that neither indeterminacy in general nor these three types of it are restricted to modals.

Coates (1980) distinguishes the same three types of indeterminacy, although there are some differences in the way she defines them, as in her study the meanings of modals are described in terms of fuzzy set theory.

In both studies, the examples given of gradience are continua between two root senses (or, in one instance,

between two epistemic senses), but there are no examples where gradience would cross the line between root and epistemic modality. The examples of ambiguity and merger, on the other hand, are cases in which one of the two senses involved is root and the other epistemic. In a merger the difference between the epistemic and the root sense is neutralised.

I will only discuss the kind of indeterminacy that occurs between different root senses, i.e. what Leech and Coates call gradience. In view of the fact that the discussion of epistemic modality presented earlier is itself very tentative it seems unwise, on the basis of this groundwork, to extend the discussion to cases of merger, even if this is an important question.

In order to see how the proposed procedural system would handle gradience, the examples given by Leech and Coates (1980) will be reproduced below and examined.

The modal <u>can</u>, they claim, has two gradients, one from the sense of "permission" to the sense of "possibility", and the other from "ability" to "possibility". As examples of the former they give:

(1)	You can't do that - I forbid it.	permission
(2)	You can't do that - it's against the rules.	
(3)	You can't do that - it would be breaking the law.	
(4)	You can't do that - everyone would think you are mad.	
(5)	You can't do that - it wouldn't be reasonable.	
(6)	You can't do that - it wouldn't be right.	
(7)	You can't do that - it is contrary to the law of gravity.	possibility

The procedures outlined in chapters 4 and 5 would deal with these as follows: In (1) the speaker is the authority; the prohibition is based on the speaker's goals. In (2) and (3) the authority is not the speaker but an external authority (rules, the law). In (7) the action is impossible because it is incompatible with a law of nature; in other words, the physical preconditions do not exist or are not satisfied.

Examples (4), (5) and (6) have to do with a code of behaviour, ethics or morals. These are the most difficult to fit into a simple network like the one given in Figures 4 and 6, and the problems they present are philosophical rather than linguistic. Is a moral or an ethical code to be regarded as something imposed by an external authority, or imposed by the speaker himself (in which case it could be seen as a goal, i.e. the goal of acting in accordance with the code)?

None of these examples exhibit a gradient in the sense that there would be a <u>degree</u> of permission and a <u>degree</u> of possibility present in the same token.

However, it <u>is</u> possible to envisage examples in which you do have both the permission and the possibility sense present in one utterance, for example, the one discussed earlier (section 4.1.3.):

(8) A: Can you drive me to the airport tomorrow morning?

B1: No, I can't, because...
B2: Yes, I can.

which may question all aspects of feasibility (are you allowed to, are you free, is your car working, etc), or the question may be vague and not intended in any such specific sense; and the respondent may take any or all of these aspects into account when answering the question.

The network given in Figures 4 and 6 proceeds in the following way (but as we have stressed, this is only one possible order of the steps that are needed): it checks the physical preconditions: if p is possible with respect to these, it checks whether we have a context where there is an authority that may have control over the possibility, etc. of p; if there is no authority, it checks whether there are any relevant goals, and whether p is possible, etc., with respect to these. But one might want to check the goals also in a case in which there is an authority, if the authority permits p. To do this, we need access from the results of the deontic network to the part of the procedure that tries to find a goal and then see whether p is possible with respect to the goal.

It might be better to regard the three parts of the network (the part without goals, the part that deals with deontic modality, and the part that deals with goal-oriented dynamic modality) as separate procedures, with the option that one or more of them could be called, depending on the point of view of the question and whether there is any indeterminacy.

This procedural approach would be able to deal with indeterminacy if indeterminacy is understood in the sense

that one or more senses may be present in a modalised utterance.

However, one cannot interpret utterances with can, for example (8:A) or (8:B1), as having some degree of the permission sense and a degree of the possibility sense, i.e. as occupying some place on a gradient from one sense to the other. Although one can have more than one sense present in one utterance, each of these senses is either there completely or not at all. This is how the (extended) procedural system would work. For example, the system finds that p is possible or impossible as regards physical preconditions; if it finds a goal, p is either possible, impossible or necessary with respect to the goal. The results of the different components are not quantified in terms of their respective importance.

(It may be the case, however, that with modals like should and ought to one does get a gradient from one sense to another, since these modals themselves seem to express some kind of modified rather than absolute necessity.)

The remarks made above apply also to the indeterminacy between the "ability" and "possibility" senses of can, and between the senses of root must.

7.2. Semantics and pragmatics

Leech (1983) discusses the roles of semantics and pragmatics in the study of language. He argues for a view where these are seen as two separate domains and as

operating in different ways from each other, but the domains are complementary. Semantics belongs to the formal, abstract system of language, pragmatics is the study of the meaning of utterances in use. Leech lists a number of significant characteristics which distinguish pragmatics from semantics and which derive from this essential difference of semantics having to do with form and pragmatics with function (Leech 1983).

The procedural account of can and must presented in chapter 4 does not make a clear difference between semantics and pragmatics. The meaning of the modals is not distinguished from their use in a context; the two are interwoven in the same procedure.

One could say that the procedure defines 'can(p)' as either "p is feasible" or "p is feasible but not necessary" or "p is permitted", and 'must(p)' as either "p is necessary" or "p is obligated", and that these are the semantic representations of can and must.[1]

However, in computing these results the procedure makes use of pragmatic knowledge and principles. The most important of these is taking into account the goals of the interlocutors, which in the context of modality are necessary in determining from what 'point of view' the feasibility or necessity is considered, and in excluding or establishing necessity, since we have argued that dynamic necessity is always linked to goals. An essential feature in the procedure for determining the degree of modality is the use of the focus of the question, which can also be

seen as something which is governed by pragmatic factors: what is in focus is linked to what the speaker's conversational goals are (cf. ch. 3).

But in the procedure the semantics and the pragmatics interact in a way that would make it impossible, for example, to extract the two aspects into separate parts or procedures, although one can obviously describe some elements of the procedure as semantic and others as pragmatic.

This seems to be incompatible with some of the postulates presented by Leech (1983) about the differences between semantics and pragmatics: the first postulate says that the "semantic representation (or logical form) of a sentence is distinct from its pragmatic interpretation" (Leech 1983: 5). In the procedural account, the sense of the modal (within the semantic representation of the sentence) is not defined as distinct from the pragmatics of the context; it is only defined after the pragmatic considerations have played their part in the definition of the meaning of the modal. However, if the semantic representation of <u>can</u> is given simply as "feasible" or "possible" and that of <u>must</u> as "necessary", then there is no incompatibility. But in defining what "feasible" or "possible" or "necessary" mean the semantics has to be combined with pragmatics.

The interaction of semantics and pragmatics is perhaps not incompatible with all the postulates given by Leech (1983), e.g. with the points that semantics is rule-governed,

and its explanations primarily formal, whereas pragmatics is principle-controlled and its explanations primarily functional (Leech 1983: 5). The steps in the modal procedure which have to do directly with the degree of modality (whether here is a plan for p or not, whether p is necessary) could be seen as semantic rules, whereas finding the ulterior and conversational goals (including finding the focus of the sentence) is done primarily on the basis of pragmatic principles.

In conclusion, in the procedure developed here for determining the degree of modality when answering questions, the semantic side cannot proceed without the pragmatic side; the semantic rules cannot be applied without pragmatic information.

8. CONCLUSION

This thesis has explored how some aspects of modality could be treated in a procedural framework.

Procedures were defined for the dynamic senses of the modals can and must. Although the domain within which this was done was restricted, the procedural definitions for the modals are general and could be used in other domains.

The framework of a language understanding system brought out some problems which are general and would probably be encountered in any such system but turned out to be central to the processing of modals, and these had to be addressed before a procedural definition of the modals was possible. The problems that were of most concern here are the concept of focus in questions on the one hand, and the higher level (non-linguistic) goal(s) of the interlocutors, on the other.

The focus of a question was used in the sense of what the speaker's attention is focused upon, and was related to the conversational goal of the speaker in uttering the question.

It was argued that questions containing dynamic modality are often intended and interpreted in the sense of whether the proposed action is possible or necessary from a particular 'point of view'. The point of view is in fact often some (non-linguistic) goal of one or both of the interlocutors.

The procedure for dynamic modals therefore has to look for a possible goal to which the modalised question may be related. It was argued that dynamic necessity is always linked to a goal or goals, and that possibility often is but not always.

Recognition of the focus of a question and of the possible higher level goal of the addressee is obviously important in making an utterance relevant to the hearer.

These conclusions accord with the results of recent research in both linguistics, where there has been a lot of work on pragmatic concepts like relevance, and in artificial intelligence, where there has been interesting work in the development of plan inference mechanisms.

The procedure for dynamic modals was presented in detail in chapter four. Determining whether an action is feasible (possible) is done by asking the general purpose planner, which is accessible from the system, to find a plan for the proposed action. The procedure also checks whether the question is related to a goal, and whether the plan for the proposed action (if there is one) is compatible with achieving the goal.

In order to exclude, or establish, necessity of the action, the procedure has to check whether an alternative plan is possible and compatible with the goal. An alternative plan is found by changing the value of the focus of the question.

The procedure can give the any of the following results: The action is possible and compatible with the goal

("can(p)"); the action is possible but incompatible with the goal ("not(can(p)) if g"); the action is impossible ("not(can(p))"); the action is necessary if the goal is to be achieved ("must(p) if g"); and the action is predicted to happen ("will(p)"). For answering a question, the result is then formulated into an appropriate response.

The same procedure can be used for determining the degree of modality in answering questions containing can as well as those containing must, if there is a separate component for formulating the answer on the basis of the result (the degree of modality) given by the modality procedure.

Both yes-no and wh-questions were discussed. The procedures that determine the degree of modality for these are similar. Apart from the obvious difference that wh-questions ask for a value for the focus, there are also some differences between the two types as regards negation.

It was proposed that deontic modality could be handled procedurally in a similar network as dynamic modality, or as part of one enlarged network which would cover both dynamic and deontic modals. The procedure for deontic modality has to establish first whether the situation or context of utterance is one where there is an authority who has the right to control some actions. If the speaker is the authority, the relevant goals are his own goals. The procedure can then proceed as with dynamic modality, relating the modality to these goals. How it is established

whether there is an authority in the situation is an open question.

Epistemic modality and especially its relation to root modality was also discussed, although at a more general level. The tentative conclusion was that although there is a clear parallelism between these modalities, the differences between them suggest that it is better to regard them as distinct categories. An example of the close connection between the two modalities is that there are examples of root necessity which are exactly parallel to epistemic examples: in the first, the necessity of p is caused by a goal; in the second, the knowledge that the goal has been achieved is the reason for the inference that p must be the case. Both are based on the fact that the action is a precondition of the goal.

But the differences seem to outweigh the similarities. For example, the point made above does not apply to stative and progressive predications, but only to dynamic ones. Root modality may be concerned with one point of view only (related to one goal), whereas in epistemic modality all factors that might have a bearing on the truth of the proposition have to be considered.

The chapter on epistemic modality also dealt with some of the problems that would be encountered in the attempt to define epistemic modality procedurally.

In the light of the proposals made in this study, the question of whether modals are monosemantic or polysemantic was not seen as an either-or question but as

a question about the level of analysis on which the differences between the senses of a modal begin to appear.

In the proposed system semantics and pragmatics are combined in the same procedure. Although it would be possible to point to parts of the procedure and call that semantics, the semantic parts cannot function without pragmatic information about the conversational and other goals of the interlocutors.

Since this thesis concentrated on a few modals in their dynamic senses, there is obviously a great deal of work to be done if this approach is to be extended to cover other modals, as well as the uses of modals with reference to past time.

Several interesting questions were treated only tentatively: one of the most important ones is the distinction between root and epistemic modality and how people tell which one is intended when they are interpreting a modalised utterance.

Apart from these large areas, more work is needed also on some phenomena that were highlighted by the procedural approach of this study. It has not been possible to explore here how the focus and the goals are found. Of the possible clues for these, some are likely to be linguistic and some non-linguistic. One could say that inferences that are based on non-linguistic clues need not concern the linguist; yet anyone who wants to include modals in a language understanding system will have to confront these problems. Although some work on how to find the focus

and how to infer the interlocutor's goals has been done in both linguistics and artificial intelligence, the results so far are only a beginning and more research needs to be done.

APPENDIX

This appendix contains a listing of the program, comments on the parser, and examples of input and output of (i) the parser and (ii) of the whole program.

First, the routines for the parser and for answering modalised questions and an example database are listed. To run the program, the planner (see Warren 1974) is needed, and a goal has to be asserted.

LISTING OF THE ROUTINES

```
/* Top Level Procedures  */

talk:- readandparse(SR,Fma,Fla),
       pickfocus(SR,Fma,Fla,Foc),
       answer(SR,Foc).

readandparse(SR,Fma,Fla):- read_in(X), removelast(X,X1),
                           sentence(SR,Fma,Fla,X1,[]).

/* 'read_in' and 'removelast' convert the input sentence
    into a Prolog list */


pickfocus(qu(Wh,Qm),Fma,Fla,Wh):- whword(Wh).
pickfocus(SR,Fma,Fla,Fma):- nonvar(Fma),!.
pickfocus(SR,Fma,Fla,Fla).

whword(when).

answer(qu(yesno,can(P)),Foc):- answercan(P,Foc,can).
answer(qu(yesno,must(P)),Foc):- answercan(P,Foc,must).
answer(qu(when,can(P)),Foc):- answerwhencan(P,Foc).
answer(qu(when,must(P)),Foc):- answerwhencan(P,Foc).

removelast([X|L],[X|L1]):- not(L=[]), removelast(L,L1).
removelast(X,[]).
```

```
/* Parser starts here.  For 'aux' read 'modal auxiliary'.
   Variables ending in -m are for semantic representations.
   'Fma' and 'Fla' are focus variables. */


?- lib(dcgs).      /* This calls the Prolog grammar rule
                      notation routine */



/* Rule 1: */

sentence(qu(Type,Qm), Fma,Fla) --> question(Type,Qm, Fma,Fla).


/* Rule 2: */

sentence(decl(Declm), Fma, Fla) --> decl(fin, Declm, L, Fma, Fla).


/* Rule 3: positive yes-no questions */

question(yesno, M, Fma, Fla) --> aux(pos, Scope, Auxm),
                                 decl(base, Sm, [yesno], Fma, Fla),
                                 {M=..[Auxm,Sm]}.


/* Rules 4-7 are for wh-questions */
/* Rule 4:  wh + can + not, wide scope */

question(Wh, no(M), Fma, Fla) --> whword(Wh), aux(pos,scope1,Auxm),
                                  np(Nm, Fma), [not],
                                  vp(base,Nm,Sm,[Wh],Fla),
                                  {M=..[Auxm,Sm]}.

/* Rule 5:  wh + can't, wide scope */

question(Wh, no(M), Fma, Fla) --> whword(Wh), aux(neg,scope1,Auxm),
                                  decl(base,Sm,[Wh], Fma,Fla),
                                  {M=..[Auxm,Sm]}.

/* Rule 6:  wh + mustn't, narrow sope */

question(Wh, M, Fma, Fla) --> whword(Wh), aux(neg,scope2,Auxm),
                              decl(base,Sm,[Wh], Fma,Fla),
                              {M=..[Auxm,no(Sm)]}.

/* Rule 7: wh + can or must (+ not), narrow scope if negative */

question(Wh, M, Fma, Fla) --> whword(Wh), aux(pos,Scope,Auxm),
                              decl(base,Sm,[Wh], Fma, Fla),
                              {M=..[Auxm,Sm]}.
```

```
/* Rules 8-13 are for negative yes-no questions */
/* Rule 8:  aux + not, positive expectation */

question(agree, M, Fma,Fla) --> aux(pos, Scope, Auxm), np(Nm, Fma),
                                [not], vp(base, Nm, Sm, L, Fla),
                                {M=..[Auxm,Sm]}.

/* Rule 9:  can + not, neg. exp., wide scope */

question(agree, no(M), Fma, Fla) --> aux(pos, scope1, Auxm),
                                     np(Nm, Fma), [not],
                                     vp(base, Nm, Sm, L, Fla),
                                     {M=..[Auxm,Sm]}.

/* Rule 10:  must + not, neg. exp., narrow scope */

question(agree, M, Fma, Fla) --> aux(pos, scope2, Auxm), np(Nm,Fma),
                                 [not], vp(base, Nm,Sm, L, Fla),
                                 {M=..[Auxm,no(Sm)]}.

/* Rule 11:  aux + n't, pos. exp. */

question(agree, M, Fma, Fla) --> aux(neg, Scope, Auxm),
                                 decl(base, Sm, L, Fma, Fla),
                                 {M=..[Auxm,Sm]}.

/* Rule 12:  can + n't, neg. exp., wide scope */

question(agree, no(M), Fma, Fla) --> aux(neg, scope1, Auxm),
                                     decl(base, Sm, L, Fma, Fla),
                                     {M=..[Auxm,Sm]}.

/* Rule 13:  must + n't, neg. exp., narrow scope */

question(agree, M, Fma, Fla) --> aux(neg, scope2, Auxm),
                                 decl(base,Sm, L, Fma, Fla),
                                 {M=..[Auxm,no(Sm)]}.


/* Rule 14: */

decl(Type, Declm, L, Fma, Fla) --> np(Nm,Fma),
                                   vp(Type, Nm, Declm, L, Fla).


/* Rules 15-20  rewrite the verb phrase */
/* Rule 15:  can + not */

vp(fin, Nm, no(M), L, Fla)--> aux(pos, scope1, Auxm),
                              [not], vp(base, Nm, Sm, L, Fla),
                              {M=..[Auxm,Sm]}.
```

```
/* Rule 16:  can't */

vp(fin, Nm, no(M), L, Fla) --> aux(neg, scope1, Auxm),
                               vp(base, Nm, Sm,L, Fla),
                               {M=..[Auxm,Sm]}.

/* Rule 17:  aux + vp ; if vp is negative, narrow scope */

vp(fin,Nm, M, L, Fla) --> aux(pos, Scope, Auxm),
                          vp(base, Nm, Sm, L, Fla),
                          {M=..[Auxm,Sm]}.

/* Rule 18:  mustn't */

vp(fin, Nm, M, L, Fla) --> aux(neg, scope2, Auxm),
                           vp(base, Nm, Sm,L, Fla),
                           {M=..[Auxm,no(Sm)]}.

/* Rule 19:  the condition in curly brackets ensures  that all
             negative yes-no questions are parsed as agreement
             questions */

vp(base, Nm, no(Sm), L, Fla) --> [not], vp(base,Nm, Sm, L, Fla),
                                 {L\==[yesno]}.

/* Rule 20: */

vp(Type, Nm, Vpm, L, Fla) --> verb(Type, Vm), prep_p(Advm, L, Fla),
                              {Vpm=..[Vm,Nm,Advm]}.



/* Rule 21: */

np(Nm,Fma) --> det(Detm), noun(Nm,Fma).


/* Rule 22: */

prep_p(Num, L, Num) --> prep(Prepm), noun(Time, Fma), numeral(Num).

/* Rule 23: */

prep_p(when, L, Fla) --> [], {nonvar(L), L=[when]}.



det(the) -->[the].

noun(plumbers,Fma) --> consumeword(plumbers, Fma).
noun(electricians,Fma) --> consumeword(electricians, Fma).
noun(plasterers, Fma) --> consumeword(plasterers, Fma).
noun(painters, Fma) --> consumeword(painters, Fma).
```

```
aux(pos, scope1, can) --> [can].
aux(neg, scope1, can) --> [can],['''],[t].
aux(neg, scope1,can) --> [cannot].
aux(pos, scope2, must) --> [must].
aux(neg, scope2, must)--> [mustn],['''],[t].


verb(Type, starttime) --> [come].
verb(Type, starttime) --> [start].


whword(when) --> [when].


prep(on) --> [on].
noun(day, Fma) --> [day].

numeral(X) --> [X], {integer(X)}.


consumeword(W, W) --> ['I'], [W], !.
consumeword(W,Fma) --> [W].

/* End of parser */



/* Procedures for answering yes-no questions
   containing 'can' or 'must' */


answercan(P, Foc,Mod):- addgoals(P,G,PG), anscan1(P,PG,Foc,Mod).

anscan1(P,PG,Foc,Mod):- plans(PG,start), !, anscan2(P,PG,Foc,Mod).
anscan1(P,PG,Foc,Mod):- anscan3(P,PG,Foc,Mod).

anscan2(P,PG,Foc,Mod):-
    alter(Foc,X,PG,PG1),
    plans(PG1,start),
    X\== Foc,
    write_either(Mod,'Yes', 'No, they dont have to'), nl.

anscan2(P,PG,Foc,Mod):- alter(Foc,X,P,P1), anscan4(P1,Foc,X,Mod).

anscan4(P1,Foc,X,Mod):-
    plans(P1,start),
    X \== Foc,
    write_either(Mod,
        'Yes, indeed they must if the goal is to be met',
        'Yes, they must if the goal is to be met'),
    nl, !.

anscan4(P1,Foc,X,Mod):-
    write('Yes, they will start on that day'), nl.
```

```
anscan3(P,PG,Foc,Mod):-
    plans(P,start),
    write('No, they cant if the goal is to be met'),
    nl, !.

anscan3(P,PG,Foc,Mod):- write('No,it is impossible'), nl.

write_either(can,X,Y):- write(X).
write_either(must,X,Y):- write(Y).


/* Procedures for answering when-questions
   containing  'can' or 'must' */


answerwhencan(P,Foc):- addgoals(P,G,PG), answhencan1(P,PG,Foc).

answhencan1(P,PG,Foc):- alter(Foc,T,PG,PG1),
                        alter(Foc,T1,PG,PG2),
                        alter(Foc,T1,P,P1),
                        plans(PG1,start), !,
                        answhencan2(P1,PG2,T,T1).
answhencan1(P,PG,Foc):- answhencan3(P,PG,Foc).

answhencan2(P1,PG1,T,T1):-
    plans(PG1,start), T1 \== T,
    write('They can start for instance on day '),
    write(T), write('.'), nl, !.
answhencan2(P1,PG1,T,T1):-  answhencan4(P1,T,T1).

answhencan4(P2,T,T1):-
    plans(P2, start), T1 \== T,
    write('They must start on day '),
    write(T), write(' if the goal is to be met.'), nl, !.

answhencan4(P2,T,T1):- write('They will start on day '),
                       write(T), write('.'), nl.

answhencan3(P,PG,Foc):-
    alter(Foc,T,P,P1),
    plans(P1,start),
    write(
      'It is impossible to do that job if the goal is to be met.'),
    nl, !.

answhencan3(P,PG,Foc):- write('It is impossible to do that job'),
                        nl.


/* Utilities for 'answercan' and 'answerwhencan' */


alter(P,Q,SI,SO):- SI=..[G|A],
                   alter1(P,Q,G,GO),
                   alterlist(P,Q,A,AO),
                   reconstruct(SO,GO,AO).
```

```
alter1(P,Q,P,Q):- !.
alter1(P,Q,G,G).

alterlist(P,Q,[],[]).
alterlist(P,Q,[HI|TI],[HO|TO]):- alter(P,Q,HI,HO),
                                 alterlist(P,Q,TI,TO).

reconstruct(X,X,[]):- !.
reconstruct(SO,GO,AO):- SO=..[GO|AO].


addgoals(P,G,PG):- findset(X, goal(X), G), attach(P,G,PG).

attach(PH&PT, G, PH&Result):- !, attach(PT,G,Result).
attach(P,empty,P):- !.
attach(P,G,P&G).

findset(X,Y,G):- setid(ID), ID1 is ID+1,
                 retract(setid(ID)),
                 assert(setid(ID1)),
                 putallindb(ID1,X,Y),
                 getallfromdb(ID1,G).

putallindb(ID1,X,Y):- Y, not(storeforset(ID1,X)),
                       assert(storeforset(ID1,X)),
                       fail.

putallindb(ID1,X,Y).

getallfromdb(ID1,X):- retract(storeforset(ID1,H)), !,
                      getallfromdb(ID1,T),
                      merge(H,T,X).

getallfromdb(ID1,empty).

merge(H,empty,H):- !.
merge(H,T,H&T).
setid(0).



/* Example database */


add(fixedstart(J),fixstart(J,T)).
add(starttime(J,T),fixstart(J,T)).

del(unfixedstart(J),fixstart(J,T)).

can(fixstart(J,T),
      unfixedstart(J)&predecessor(J,J1)&fixedstart(J1)
      &starttime(J1,T1)&duration(J1,T2)&(X is T1+T2)
      &biggerthan(X,T,3)).
```

```
biggerthan(X,T,0):-!,fail.
biggerthan(X,X,N).
biggerthan(X,T,N):- NM1 is N-1, XP1 is X+1, biggerthan(XP1,T,NM1).

predecessor(begin,none).
predecessor(plumbers,begin).
predecessor(electricians,plumbers).
predecessor(plasterers,electricians).
predecessor(painters,plasterers).
predecessor(finish,painters).

duration(begin,0).
duration(plumbers,2).
duration(electricians,3).
duration(plasterers,5).
duration(painters,4).
duration(finish,0).

given(start,fixedstart(begin)).
given(start,starttime(begin,1)).
given(start,unfixedstart(plumbers)).
given(start,unfixedstart(electricians)).
given(start,unfixedstart(plasterers)).
given(start,unfixedstart(painters)).
given(start,unfixedstart(finish)).
```

Some comments on the parser

A general outline of the parser was given in section 2.2.2. A summary of the main points is given below together with some more detailed comments on specific rules.

The parser was written using the method described in Pereira and Warren (1980: 252-253). It takes the input sentence and produces a semantic representation (SR) as its output. Each syntactic constituent is assigned a SR as it is being parsed and the SR of the sentence is built up from the SRs of the constituents as they are passed up the parse tree.

Many of the rules are based on rules given by Gazdar, Pullum and Sag (1982).

Aspects of grammar that have no direct bearing on the rules for modal auxiliaries are treated as simply as possible. For example, noun phrases are treated as unanalysed wholes. Input sentences are first rewritten either as declarative or as interrrogative (questions). Questions are of three types: yes-no questions, negative yes-no questions, and wh-questions.

Negative yes-no questions are distinguished from positive yes-no questions: the latter are neutral as to the answer they expect (positive or negative), whereas the former are usually biased towards either a positive or a negative answer. For example,

Shouldn´t I tell him (that the roof leaks)?

can be interpreted either (1) as having the assumption "I should tell him" and asking for agreement (i.e. "Isn´t it the case that I should tell him?"), or (2) as having the assumption "I should not tell him" and asking for agreement (i.e. "Is it the case that I should not tell him?"). Notice that the answers are different depending on which interpretation is chosen; for (1) the answers could be

Yes, you should.	(agreeing)
No, you don´t have to.	(disagreeing)

but for (2) they would be

No, you shouldn´t. (agreeing)

Yes, you can (if you wish). (disagreeing)

(See Bolinger 1957: 99-110 for a detailed discussion of conduciveness in negative questions.) Negative yes-no questions will here for convenience be called agreement questions.

As yes-no questions in this parser are always positive, only one rule is needed to rewrite them (rule 3).

For wh-questions and agreement questions more rules are needed because of variations in meaning caused by different scopes of negation with different modals, and because questions with negative forms of modal auxiliaries (e.g. can´t, mustn´t) have to be dealt with by separate rules from questions with negation by not (see below). In the case of agreement questions more rules are needed also because of the different possible bias (positive or negative) of the question.

Similarly, several rules are needed for rewriting verb phrases in order to handle negation and scope correctly.

The grammar employs features which can be attached to non-terminals and to terminals. For example, verb phrases and verbs have either the feature ´finite´ or the feature ´base´, depending on whether the verb phrase contains (or the verb is) a finite or a base form of the verb.

Modal auxiliaries are defined by means of features as either positive or negative and as having either wide or

narrow scope. The need to define each modal as either positive or negative arises because modals with so-called clitic negatives are treated as lexical items and have entries in the dictionary as independent units (this view of modals with the so-called clitic negation is taken for instance by Dowty 1979: 348 and Gazdar, Pullum & Sag 1981; for arguments for its defense, see also Zwicky and Pullum 1982). The scope definition is needed for the obvious reason that with some modals negation has wide scope, with others narrow scope; cf.

> The painters can´t come tomorrow.
> (= It is not possible that the painters come tomorrow)
>
> The painters mustn´t come tomorrow.
> (= It is necessary that the painters do not come tomorrow).

Modal auxiliaries thus appear in the dictionary in the following form:

```
aux(pos, scope2, must) --> [must].

aux(neg, scope2, must) --> [mustn´t].
```

The third argument of ´aux´ is the semantic representation of the word. The fact that mustn´t is negative cannot be represented in the SR because mustn´t has narrow scope, i.e. it is not the modality which is negated in a sentence containing mustn´t but the proposition under must. Instead the feature ´neg´ is passed up the tree with the semantic representation "must" and made use of when the SR of the whole sentence is being built.

Some information about the parse tree for the sentence is passed down the tree in the parsing process and this information is used as a condition for the application of some rules. In this fragment this is necessary in two cases.

Firstly, the rule for (positive) yes-no questions contains the feature 'yesno' (the list L is instantiated as [yesno]); this feature is passed down the tree. The rule that rewrites a verb phrase as a negated verb phrase,

```
vp(base,..., L,...) --> [not], vp(base,..., L,...),
                        {L\==[yesno]}.
(Rule 19)
```

(given here in a simplified form) has the condition that it cannot be applied if L=[yesno], that is if it is part of a tree that has the feature 'yesno'. This is to ensure that all yes-no questions with not are parsed as 'agreement' questions, not as 'yesno' questions.

Secondly, wh-questions pass down the wh-word with which the question begins, so that it can be inserted into the right slot in the SR once the parser finds a zero realization of the right syntactic constituent.

For instance, in the question

When can the painters come?

the wh-word is found to be when and the context variable L is instantiated as [when]. This argument is passed down through rules 14 and 20 to the rules which expand prepositional phrases (rules 22-23). The prepositional phrase

(here a time adverbial) can be realized as zero if the context variable is instantiated to a wh-word (in this case [when]).

Example input and output of the parser

The following is a list of examples of the types of sentences that the parser can handle. It is capable of giving a set of all the possible interpretations in cases where the sentence is ambiguous. (This feature is not exploited by the system as a whole, however, because the planner cannot handle all the types of sentences that the parser can.) The input to the parser is preceded by the symbol |: and the output is printed on the lines following the input. Indication of the focus is not included in these examples.

```
|: The painters can start on day 5.
decl(can(starttime(painters,5)))

|: The painters must start on day 5.
decl(must(starttime(painters,5)))

|: The painters can't start on day 5.
decl(no(can(starttime(painters,5))))

|: The painters can not start on day 5.
decl(no(can(starttime(painters,5))))
decl(can(no(starttime(painters,5))))

|: The painters mustn't come on day 5.
decl(must(no(starttime(painters,5))))

|: The painters must not come on day 5.
decl(must(no(starttime(painters,5))))

|: Can the painters start on day 5?
qu(yesno,can(starttime(painters,5)))
```

```
|: Can't the painters start on day 5?
qu(agree,can(starttime(painters,5)))
qu(agree,no(can(starttime(painters,5))))

|: Can the painters not come on day 5?
qu(agree,can(starttime(painters,5)))
qu(agree,no(can(starttime(painters,5))))

|: Must the painters come on day 5?
qu(yesno,must(starttime(painters,5)))

|: Mustn't the painters come on day 5?
qu(agree,must(starttime(painters,5)))
qu(agree,must(no(starttime(painters,5))))

|: Must the painters not start on day 5?
qu(agree,must(starttime(painters,5)))
qu(agree,must(no(starttime(painters,5))))

|: When can the painters come?
qu(when,can(starttime(painters,when)))

|: When can't the painters come?
qu(when,no(can(starttime(painters,when))))

|: When can the painters not come?
qu(when,no(can(starttime(painters,when))))
qu(when,can(no(starttime(painters,when))))

|: When must the painters start?
qu(when,must(starttime(painters,when)))

|: When mustn't the painters start?
qu(when,must(no(starttime(painters,when))))

|: When must the painters not start?
qu(when,must(no(starttime(painters,when))))

|: The painters can not not start on day 5.
decl(no(can(no(starttime(painters,5)))))
decl(can(no(no(starttime(painters,5)))))

|: The painters must not not start on day 5.
decl(must(no(no(starttime(painters,5)))))

|: The painters can't not come on day 5.
decl(no(can(no(starttime(painters,5)))))

|: Can't the painters not come on day 5?
qu(agree,can(no(starttime(painters,5))))
qu(agree,no(can(no(starttime(painters,5)))))

|: Mustn't the painters not come on day 5?
qu(agree,must(no(starttime(painters,5))))
qu(agree,must(no(no(starttime(painters,5)))))
```

Examples of input and output of the query program

(A goal has been asserted)

```
|: Can the plumbers come on day 1?
Yes, indeed they must if the goal is to be met.

|: Can the plumbers come on day 2?
No, they cant if the goal is to be met.

|: When can the plumbers start?
They must start on day 1 if the goal is to be met.

|: Must the electricians start on day 3?
Yes, they must if the goal is to be met.
```

(After the goal has been changed to a later date:)

```
|: Can the plumbers start on day 1?
Yes.

|: Must the plumbers start on day 1?
No, they dont have to.

|: When can the electricians start?
They can start for instance on day 3.

|: Can the electricians start on day 4?
Yes.
```

NOTES

Notes to Chapter 1

[1] For the delimitation and definition of modal verbs from the grammatical point of view, see Huddleston (1976), Palmer (1979: section 9.4), Coates (1980: section 2.1), Huddleston (1980). For comprehensive accounts of the details of grammar and of the use of modals in English, based on corpus analysis, see Ehrman (1966), Palmer (1979), Coates (1980). Brown and Miller (1975) and Brown and Millar (1980) analyse the use of modals in some dialects of Scottish English.

[2] Lyons' treatment of epistemic and deontic modality has some aspects in common with Boyd and Thorne (1969). They also include illocutionary force in the analysis of the meanings of modals, assigning subjective epistemic possibility and necessity to the component that expresses illocutionary force, and deontic modality to the propositional component.

Notes to Chapter 2

[1] This example has been adapted from one given in Wiest and Levy (1977), A Management Guide to PERT/CPM. The example given by them has already been simplified and does not contain all the information that a contractor would possibly require. I have simplified it further for the purposes of this study.

[2] In a real situation, the system would have to be able to adjust the network according to how the situation changes, what is agreed, etc. For example, if the plumbers can start any time between days 7 and 9, and it then agreed that they start on day 9, the possible starting times for the rest of the network will have to recalculated.

[3] See Fikes and Nilsson (1971). There is a good description also in Bundy (1983).

[4] The problem of how to deal with statements from the hearer's point of view is not tackled here. See Winograd (1972: 142) for a discussion of the considerable problems encountered in dealing with statements.

Notes to Chapter 3

[1] In their terminology, the nucleus of a tone unit is "the peak of greatest prominence within the tone unit" (Quirk et al. 1972:1044).

[2] Sidner's proposal is controversial, but it is not possible to discuss it in detail here.

Notes to Chapter 4

[1] There are at least two other uses of <u>can</u> which are often distinguished from these three senses: (i) It can express existential modality (Palmer 1979: 152-153), as in

Welshmen can be tall.

Rocks can fall on this path.

In this sense, <u>can</u> is regarded as equivalent to the quantifier <u>some</u> or <u>sometimes</u>.

Can has the advantage over some in sentences like the above that it does not have to refer to a particular time; with some, the verb has to be tensed; for example, 'Some Welshmen are tall' has a narrower interpretation than 'Welshmen can be tall'.

Can is the only one of the modals that has this existential sense. This usage is different from other senses also in that the negation does not have a quantifier sense:

Welshmen cannot be tall. =/= No Welshmen are tall.

The link between the existential sense of can and the possibility sense is the position they occupy in the tripartitions that they belong to: the first sense is in the intermediate position of the tripartition of quantification (e.g. all - some - none); the second sense the intermediate member in the modal tripartition (necessity - possibility - impossibility) (Jespersen 1924: 324-5; cf. also von Wright 1951; Palmer 1979: 152-3).

There will be no attempt to tackle the existential sense of can in this study.

(ii) CAN is used with the so-called private verbs like SEE, HEAR, REMEMBER, etc. It can be argued that this usage is non-modal, since in a sentence like

I can see the moon.

'I can see' is equal to 'I see' (Vendler 1967: 105). (The sentence can of course also have a modal interpretation.) See Vendler (1967: 105) for a discussion of the difference between this usage and the ability sense of CAN. The usage with the private verbs is perhaps an idiomatic feature of English; in many other languages (e.g. German, Swedish, Finnish) no modal is used in equivalent sentences.

Palmer (1979) mentions other, derived uses of CAN.

[2] To be consistent, one would also have to give this answer as an explanation of 'No, they cannot' if the reason for the impossibility was that the starting time had already been arranged, e.g. 'No, they cannot; they will come on the 28th'. Cf. below)

[3] It is interesting that many languages use separate verbs for the sense of possibility, on the one hand, and for the sense "know how, have the skill" to do something, on the other hand; for example, French pouvoir and savoir , Finnish voida and osata.

[4] We will ignore for the time being the goals that might have to be considered (in this case those of the respondent) and the fact that this utterance is a request. These points do not alter the fact that the feasibility of the proposed action has to be ascertained.

Notes to Chapter 6

[1] The fact that a predicate like be careful can also occur in the progressive, e.g.

She is being very careful not to upset her mother.

also shows that it may have a dynamic interpretation.

[2] When can occurs in the existential sense, it may have an inanimate subject:

Rocks can fall on this path.

There are also cases with inanimate subjects which are more difficult to explain. Ehrman (1966: 13) gives the example

A dark bathroom can be pretty scary ...

and Leech (1971: 220)

The monsoon can be dangerous.

Although it is possible to interpret these in the existential sense, neither Ehrman nor Leech do that. Ehrman interprets her example as having an implied if-clause: if circumstances are right, a dark bathroom is scary (Ehrman 1966:13). But cf. Austin (1961) for arguments against analysing can as expressing conditionality.

[3] Coates (1980) gives statistical correlations of features like these with different modalities. She also describes a program which is designed to determine whether must is epistemic or root in a given utterance (Coates 1980: 100ff.).

Notes to chapter 7

[1] In Figures 1-4 as they stand at the moment the results of the procedures are given in terms of the modal verbs themselves. They should be given as representations like the above to capture both the detail of the sense in question, and to show the possible overlap of the senses of different modals.

BIBLIOGRAPHY

Allen, J.F. & Perrault, C.R. (1980). 'Analysing intention in utterances'. Artificial Intelligence 15. 143-178.

Antinucci, F. & Parisi, D. (1971). 'On English modal verbs'. In Papers from the 7th Regional Meeting of Chicago Linguistic Society. Chicago.

Austin, J.L. (1961). 'Ifs and cans'. In Urmson, J.O. & Warnock, G.J. (eds.) Philosophical Papers (1979). Oxford: Clarendon Press.

Bolinger, D.L. (1957). Interrogative structures of American English (The direct question). Alabama: University of Alabama Press.

Boyd, J. & Thorne, J.P. (1969). 'The semantics of modal verbs'. Journal of Linguistics 5. 57-74.

Brown, K. & Millar, M. (1980). 'Auxiliary verbs in Edinburgh speech'. Transactions of the Philological Society. Oxford: Blackwell.

Brown, K. & Miller, J. (1975). 'Modal verbs in Scottish English'. Work in Progress 8, Department of Linguistics, University of Edinburgh.

Bundy, A. (1983). AI 1 Problem solving notes. Occasional Paper No. 30, Department of Artificial Intelligence, University of Edinburgh.

Burton-Roberts, N. (1984). 'Modality and implicature'. Linguistics and Philosophy 7. 181-206.

Chafe, W.L. (1970). Meaning and the structure of language. Chicago & London: University of Chicago Press.

Chafe, W.L. (1976). 'Givenness, contrastiveness, definiteness, subjects, topics, and point of view'. In Li, C.N. (ed.) Subject and topic. New York: Academic Press.

Coates, J. (1980). The semantics of the modal auxiliaries: a corpus-based analysis with special reference to spoken English. Ph.D. Dissertation. University of Lancaster.

Cresswell, M.J. (1979). Review of Lyons (1977). Linguistics and Philosophy 3. 289-295.

Dakin, J. (1970). 'Explanations'. Journal of Linguistics 6. 199-214.

Dowty, D. (1979). Word meaning in Montague grammar. Dordrect: Reidel.

Ehrman, M.E. (1966). The meanings of the modals in present-day American English. The Hague: Mouton.

Fikes, R.E. & Nilsson, N.J. (1971). 'STRIPS: A new approach to the application of theorem proving to problem solving'. Artificial Intelligence 2. 189-208.

Gazdar, G., Pullum, G.K. & Sag, I.A. (1982). "Auxiliaries and related phenomena in a restrictive theory of grammar'. Language 58. 591-638.

Gordon, D. & Lakoff, G. (1971). 'Conversational postulates'. In Papers from the 7th Regional Meeting of Chicago Linguistic Society. Chicago.

Grice, H.P. (1975). 'Logic and conversation'. In Cole, P. & Morgan, J.L. (eds.) Syntax and Semantics 3: Speech Acts. New York & London: Academic Press.

Grosz, B.J. (1978). The representation and use of focus in dialogue understanding. Ph.D. Dissertation. University of California, Berkeley, California.

Grosz, B.J. (1979). Focusing and description in natural language dialogues. SRI International, Technical Note 185. Menlo Park, California.

Hakulinen, A. & Karlsson, F. (1979). Nykysuomen lauseoppia. (Suomalaisen kirjallisuuden seuran toimituksia 350). Jyväskylä, Finland: Gummerus.

Halliday, M.A.K. (1970). 'Functional diversity in language as seen from a consideration of modality and mood in English'. Foundations of Language 6. 322-361.

Hare, R.M. (1970). 'Meaning and speech acts'. Philosophical Review 79. (Repr. in Hare, R.M. (1971) Practical Inferences, London & Basingstoke: Macmillan.)

Hintikka, J. (1960). 'Aristotle's different possibilities'. Inquiry 3. 18-28.

Huddleston, R.D. (1976). 'Some theoretical issues in the description of the English verb'. Lingua 40. 331-383.

Huddleston, R.D. (1980). 'Criteria for auxiliaries and modals'. In Greenbaum, S., Leech, G. & Svartvik, J. (eds.) Studies in English linguistics for Randolph Quirk. London & New York: Longman.

Isard, S.D. (1974). 'What would you have done if ...?'. Theoretical Linguistics 1. 233-55.

Isard, S.D. & Longuet-Higgins, C. (1973). 'Modal tic-tac-toe'. In Bogdan, R.J. & Niiniluoto, I. (eds.) Logic, language and probability. Dordrect: Reidel.

Jespersen, O. (1924). The philosophy of grammar. London: Allen & Unwin.

Kaplan, S.J. (1979). Cooperative responses from a portable natural language data base query system. Ph.D. Dissertation. University of Pennsylvania, Philadelphia.

Karttunen, L. (1972). 'Possible and must'. In Kimball, J. (ed.) Syntax and Semantics 1. New York & London: Seminar Press.

Kratzer, A. (1977). 'What 'must' and 'can' must and can mean'. Linguistics and Philosophy 1. 337-355.

Lakoff, R. (1972). 'The pragmatics of modality'. In Papers from the 8th Regional Meeting of the Chicago Linguistic Society. Chicago.

Leech, G.N. (1971). Meaning and the English verb. London: Longman.

Leech, G.N. (1983). Principles of pragmatics. London & New York: Longman.

Leech, G.N. & Coates, J. (1980). 'Semantic indeterminacy and the modals'. In Greenbaum, S., Leech, G.N. & Svartvik, J. (eds.) Studies in English Linguistics for Randolph Quirk. London & New York: Longman.

Lyons, J. (1977). Semantics. Cambridge: Cambridge University Press.

Minsky, M. (1975). 'A framework for representing knowledge'. In Winston, P.H. (ed.) The psychology of computer vision. New York: McGraw-Hill.

Morgan, J.L. (1978). 'Two types of convention in indirect speech acts'. In Cole, P. (ed.) Syntax and Semantics 9: Pragmatics. London & New York: Academic Press.

Palmer, F.R. (1979). Modality and the English modals. London & New York: Longman.

Pereira, F.C.N. & Warren, D.H.D. (1980). 'Definite clause grammars for language analysis - A survey of formalism and a comparison with augmented transition networks'. Artificial Intelligence 13. 231-278.

Perkins, M.R. (1980). The expression of modality in English. Ph.D. Dissertation. The Polytechnic of Wales.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1972). A grammar of contemporary English. London: Longman.

Reichenbach, H. (1947). Elements of symbolic logic. London & New York: Macmillan.

Ross, J.R. (1969). 'Auxiliaries as main verbs'. In Todd, W. (ed.) Studies in philosophical linguistics. Evanston, Ill: Great Expectations Press.

Searle, J.R. (1969). Speech acts. London & New York: Cambridge University Press.

Searle, J.R. (1975). 'Indirect speech acts'. In Cole, P. & Morgan, J.L. (eds.) Syntax and Semantics 3: Speech acts. New York & London: Academic Press.

Sidner, C.L. (1979). Towards a computational theory of definite anaphora comprehension in English discourse. Ph.D. Dissertation. Artificial Intelligence Laboratory, M.I.T.

Steedman, M.J. (1977). 'Verbs, time and modality'. Cognitive Science 1. 216-234.

Vendler, Z. (1967). Linguistics in philosophy. Ithaca, N.Y: Cornell University Press.

von Wright, G.H. (1951). An essay in modal logic. Amsterdam: North-Holland.

von Wright, G.H. (1968). An essay in deontic logic and the general theory of action (Acta Philosophica Fennica 21). Amsterdam: North-Holland.

Warren, D.H.D. (1974). WARPLAN: A system for generating plans. Department of Artificial Intelligence Memo No. 76, University of Edinburgh.

Wertheimer, R. (1972). The significance of sense: meaning, modality and morality. Ithaca, N.Y.: Cornell University Press.

Wiest, J.D. & Levy, F.K. (1977). A management guide to PERT/CPM with GERT/PDM/DCPM and other networks. Englewood Cliffs, New Jersey: Prentice-Hall.

Winograd, T. (1972). Understanding natural language. Edinburgh: Edinburgh University Press.

Zwicky, A.M. & Pullum, G.K. (1982). Cliticization versus inflection: English n't. Bloomington, Indiana: Indiana University Linguistics Club.