Mark Newson

*English Modal Auxiliaries**

0   Introduction

The nine English words usually referred to as modal auxiliaries present some of the most difficult aspects of the language. From one point of view, their syntax and morphologically are relatively simple, but semantically they are very complex indeed. All of the modals have more than one meaning and there is a wide range of meanings that modals are used to express. Furthermore there is no simple correspondence between the meanings that individual modals have and the types of meanings that modals as a class express: we cannot state that modals of type X express meanings of type Y. Two modals can be virtually synonymous in some contexts and express very different meanings in others.

For many syntactic theories, the semantic complications that modals present are ignorable and their syntactic and morphological simplicity have made them rather uninteresting. The approach to syntax adopted in this paper however, makes the semantic complexities of these elements central by assuming that the choice of vocabulary items used to give realisation to a grammatical expression is part of the grammatical process itself and this is directly related to the semantic features from which grammatical expressions are built. The question of which modal is to be used in any particular expression is therefore a matter of the distribution of these vocabulary items and hence this aspect of their syntax is just as complex as their semantics.

This paper attempts to account for why particular modals are used to express particular modal meanings. Not only is this a complex issue in itself, but there are many other interacting factors which add further complexity. One such complicating factor which will not be dealt with in this paper is negation. Indeed, the issue of negative modals requires a separate paper, being just as varied and complex as the positive ones which the present paper treats[1].

---

[1]   Indeed, there are complexities added to complexities once we start to consider the interaction between modality and negation. For example, the positive *can* is not used to express 'possibility', but the negative *can't* is used to express impossibility. Thus the negative form has a greater range of possible meanings in this case. On the other hand, the negative *mustn't* is used to express only negative obligation whereas the positive *must* can express necessity as well as obligation. Here the positive modal has the wider range of meanings.

Another interacting factor is time reference and this we will give only a partial account of, again delaying a fuller treatment to another paper which deals with other auxiliaries besides the modals (Newson, forthcoming)[2].

In the following section I will introduce the main aspects of the syntactic framework used in this paper and thereafter we shall concern ourselves with the analysis of the English modals. This will necessarily involve the development of a semantic analysis in tandem with that of the syntactic system responsible for the production and 'spelling out' of grammatical expressions.

1   Alignment and Late Vocabulary Insertion

The analysis is based within the Alignment Syntax framework (Newson 2004, Newson and Maunula 2006), which is a limited version of an Optimality Theoretic grammar. It is assumed that the evaluation component of the grammar consists of just faithfulness and alignment constraints. Alignment constraints align input elements with respect to other input elements[3] in two ways. One kind of alignment concerns linear order, placing target elements before or after their hosts. The other kind of alignment concerns adjacency, placing target elements as near as possible to their hosts.

For the purposes of the present paper, however, not much will be made of this distinction as the phenomena we will be dealing with are relatively uniform, exclusively concerning precedence relationships. Moreover, in all cases, order is more important than adjacency and hence although elements prefer adjacent positions to their hosts, they will never 'switch sides' in order to achieve a position closer to the host when forced to be further from it on their preferred side. The ranking of the constraints therefore follows the pattern given below:

(1)   precedence > succession > adjacency

This simply means that it is more important for an element to precede its host than to succeed it and it is more important to be on the right side of the host

---

2   A reviewer of this paper also points out that stress is another complicating factor that is again not treated in this paper.

3   As first suggested by Gáspar (2005), an input element may not only be linearly ordered with respect to other individual input elements, but also to sets of input elements, called *domains*. These sets are defined in terms of input properties, e.g. the set of input elements dependent on a predicate is called the predicate domain of each member of the set. Alignments concerning two individual input elements may be seen as a restricted kind of domain alignment, where the domain set consists of a single element. In the present paper, none of the alignments will concern domains bigger than a single element.

than to be near it.

I will also be developing the framework, along lines suggested by Maunula (2006), to include a notion of 'late vocabulary insertion'. This idea is similar to that developed in Distributed Morphology (Halle and Marantz 1993)[4]. Essentially the claim is that the input consists of independent abstract semantic features rather than such features pre-bundled into lexical items. The task of the grammar is to impose a linear order onto these features, via alignment constraints. Once the grammatical ordering has been established, the features are spelled out through replacement by vocabulary items. These are phonological forms associated with sequences of semantic features. Replacement of output features by vocabulary items is done in terms of a notion of 'best match', which we will define more precisely below. Essentially, the vocabulary item(s) whose associated features match the best with a sub-sequence of features of the output will be used.

There are some general conditions that govern the process of vocabulary insertion. First, the more features replaced by a single vocabulary item the better. We can call this the principle of Minimal Vocabulary Access. Thus, suppose we have the vocabulary items below:

(2) $A \rightarrow \{[x], [y]\}$
     $B \rightarrow \{[x]\}$
     $C \rightarrow \{[y]\}$

Further suppose the features '… [x][y] …' turn up as a sub-sequence of an output. In this case 'A' will be selected to replace the features rather than 'BC'. If there were no vocabulary item *A*, then *B* and *C* would have to be used to individually replace the relevant features. Thus the choices a language makes concerning its vocabulary will have wide ranging effects on how it realises grammatical expressions.

A second aspect of the vocabulary insertion process concerns the notion of 'best match'. Importantly best match does not necessarily mean 'exact match', as it may turn out that there is no vocabulary item whose associated features match exactly with those of an output sub-sequence: the available vocabulary

---

[4] One difference between the present approach and Distributed Morphology is that the Subset Principle, often adopted in DM, which states that the set of associated features of a vocabulary item must be identical to or a subset of the set of the features to be replaced for the vocabulary item to be selected, will not be accepted here. Indeed, in most of the cases we will review, vocabulary items are associated with features which are not part of the output. These features do not contribute to the interpretation of the expression as it is a basic assumption of the framework that meaning is read off the input, not the output, or its realisation after vocabulary insertion.

items and the features they are associated with are language specific and mostly accidental. In this case there will be a search for the closest vocabulary item to match with the features.

One issue that must be made more precise is what is meant by a match. In principle we might look at feature matches in a number of ways. For example, we might consider a vocabulary item to match with a sub-sequence of features if it is associated with no feature that clashes with any feature of the sub-sequence. In this case, the vocabulary item may not be associated with some feature (or indeed any) of the output sub-sequence and still be considered a match. In other words, a vocabulary item not associated with any feature would be considered to match with all output sub-sequences under this view. A more restrictive view on matching would be to require that for every feature *f* of the output sub-sequence the vocabulary item must be associated with *f* in order to be considered a match. On this view, a vocabulary item not associated with any feature would not match with any sub-sequence of an output. To clarify, consider the following example. Suppose three vocabulary items as follows:

(3) $A \rightarrow \{[+f]\}$
       $B \rightarrow \{[-f]\}$
       $C \rightarrow \{[+x]\}$

Here both A and B are defined for feature [f], but with opposite polarity, while C is undefined for [f]. Suppose the output contains [+f], then under the strict requirement A would be chosen to replace it and B and C would not. Under the less restrictive condition, A and C would be optional replacements and B would be excluded. I will argue in this paper that strict matching is required to account for the replacement patterns found with the English modals.

Finally, we will see that there is some evidence that not all matching requirements are equal. As we said previously, matching is not a rigid requirement as there is often no vocabulary item associated with the exact set of features in a given sub-sequence of an output. As part of the process determining the best matching vocabulary item, one may be favoured over other possibilities because it matches with a certain feature rather than another. We will see that this plays a major role in determining which vocabulary modal is selected to realise sequences of modality related features. I will handle this in a standard optimality theoretic way, using a set of matching conditions which are ranked with respect to each other. Again, a brief example will serve to make the idea clear. Suppose the following vocabulary items:

(4) $A \rightarrow \{[x], [\ y]\}$
  $B \rightarrow \{[y], [\ z]\}$

Suppose further that it is more important to match with the type of feature represented by [x] than it is to match with one represented by [z]. We propose two matching conditions: Match x and Match z with the following ranking:

(5)   Match x > Match z

Now suppose we have an output sequence … [x][z] … Assuming no other vocabulary item which is associated with these features, the competition is between *A* and *B*. The following table demonstrates the result:

(6) … [x][z] …	Match x	Match z
☞ … A …		*
… B …	*!	

In this case *A* is the selected vocabulary replacement for the output sequence as it matches with [x], while *B* does not. The fact that *B* matches with [z] and *A* does not is irrelevant as this is a lower ranked requirement. Of course, had there been a vocabulary item *C* associated with both [x] and [z], this would be selected over *A*.

Having introduced the basic grammatical mechanisms we will be utilising, we will move on to consider the data we will be concerned with in the following analysis.

2   The Classification of Modals

The main problem facing any treatment of English modals is to account for the circumstances in which one modal may be used and another not. This is no simple matter as they are clearly not in a one to one relationship with any identifiable semantic feature or even a set of features. To give some idea of the extent of the problem, consider the use of *might* and *could*. In some contexts these modals seem to be interchangeable with no detectable interpretation difference:

(7) a   we might be being followed
  b   we could be being followed

It is sometimes claimed that there is a difference between such examples in

that the one containing *might* expresses something less certain than the other, though I suspect this is wishful thinking driven by the desire to find some difference on which to hang a neat description. Personally, I do not detect this difference, and other more careful descriptions of the English modals, such as Palmer (1987), have reached the same conclusion. However, this is not to say that *might* and *could* are always interchangeable:

(8)   a   in those days, I could speak Spanish fluently
      b   ? in those days, I might speak Spanish fluently

Clearly these two sentences do not mean the same and if (8b) has a sensible meaning at all, it is not the one associated with (8a). It seems then that we have to associate these two modals with different meanings. But if this is so, how can they be used to mean the same thing in cases such as (7)?

This is not just an isolated quirk of these two modals. It seems that similar observations can be made about all English modals: for every modal there is at least one other that can be used in a similar sense to it in some circumstances but the two must be interpreted differently in other circumstances.

It is this aspect that makes a 'late vocabulary insertion' analysis an attractive one. If the modals were simply in a one to one correspondence with a set of semantic features, it would be easy to associate them with those features in a lexical entry which simply carry them to the interpretative levels. But such a simple approach cannot work, as clearly modals are not in a one to one correspondence with semantic features. The idea of late vocabulary insertion allows competition between various vocabulary items and hence, given different circumstances, different ones may emerge as victorious.

Of course, this is not the solution to the problem in its entirety. We still need to establish what features modal vocabulary items are associated with and what the selectional process is that can account for their complex and overlapping distributions. In the rest of this section we will look at the modal system to try to establish the set of features we need to base the analysis on.

2.1 Degree and type of modality

There is a very common analysis of modals which utilises two axes of modal differentiation. On the one hand notions of *possibility* and *necessity* are used to distinguish between modals, as in the following examples:

(9)   a   a triangle must have three sides
      b   a triangle may have one 90° angle

Palmer (1987) refers to this axis as the degree of modality. The other axis refers to a distinction that we can call the type of modality, referring to notions such as *epistemic* (the do with modality contingent on knowledge) and *deontic* (to do with notions such as obligation and permission):

(10) a   that must be the right answer
      b   you must appear in court when summoned

In (10a) the modal is used in its epistemic sense, referring to a conclusion that follows from what is known. In (10b) the modal is used in its deontic sense, placing an obligation on some individual.

Although this two axis system is very common, different linguists have proposed a number of variations concerning how many distinctions can be made along each axis. Barbiers (2005) cites a popular classification of modals into two types: epistemic and root – citing Hofmann (1976) – where root modality covers all non epistemic meanings. Palmer (1987) claims this to be "too drastic" (p. 103) and cites behavioural differences between modals in the 'root' class, with respect to tense and negation, which lead him to conclude that this class is not homogenous. Instead he opts for a three point classification: epistemic, deontic and, what he terms, dynamic modalities, the latter referring to uses of *can* and *will* to indicate 'ability' and 'volition', as in the following:

(11) a   James can play the tuba
      b   the newsagent will usually give you change for a ten pound note

Lyons (1977) adds alethic modality, referring to logical as opposed to knowledge based possibility and necessity. Compare:

(12) a   if you didn't take the diamonds, the butler must be the thief
      b   if you are a bachelor, you must be unmarried

I am not convinced that this distinction really adds very much to our understanding of the English modal system and as far as I can tell, exactly the same modals are used to express alethic modality as epistemic, hence the system itself does not appear to distinguish between the two. Moreover, the difference between the two seems to have more to do with the content of the proposition than with the modality: what makes (12b) 'logical' is the relationship between the meaning of *bachelor* and *unmarried* rather than what the modal *must* contributes.

In the present paper, I will adopt Palmer's three point system, though I

dislike his term *dynamic* for the modality of ability and volition. Although this term has Greek origins, as do *epistemic* and *deontic*, its meaning in English has drifted from the original, 'power', to something specifically connected to movement, which has little to do with ability and volition. Moreover, the term is already used in the classification of verbs making its use in the description of the modal system unnecessarily confusing. Instead, breaking with tradition, I will opt for the Latin based *potential* to name the third type of modality.

Turning to the degree axis, again there are various proposals. Most commonly the two degrees of possibility and necessity are used, though Palmer (1987) adds a third unnamed degree referring to the following examples:

(13) a John must be in his office
      b John will be in his office
      c John may be in his office

I can only assume that Palmer senses some cline from *must*, through *will* to *may* here, though I am not sure what notion sits between necessity and possibility: Palmer's reluctance to name this level of degree presumably indicates his uncertainty on this issue.

Indeed, I am not altogether happy about the terms necessity and possibility to identify the points on the degree axis. These may be suitable for epistemic modality, but they do not fit so well for deontic and potential modalities. Clearly there is a distinction within these types of modalities, which bears some relation to the distinction made with epistemic modality: permission and obligation with deontic and ability and volition with potential modality. But these distinctions have little to do with possibility and necessity.

One thing that all modalities have in common concerning degree is that one degree seems to name a stronger condition than the other. For example, necessity is a stronger condition than possibility. However, there are also differences in this respect to be seen across the different modals. For example, while epistemic *must* may imply epistemic *may*, deontic *must* suggests deontic *can* but doesn't imply it:

(14) a he must be the thief → he may be the thief
      b he must pay the fine ↛ he can pay the fine

Potential *will* does in general imply potential *can*, though clearly willingness does not imply ability, which demonstrates that these concepts do no exactly describe this modality:

(15)   he will lend you the money → he can lend you the money

However, even here the relationship is not the same as the logical one as the following demonstrates:

(16)   a   he will lend you the money, if he can
       b   ? bachelors must be unmarried, if they may

    Because of these differences, one might be tempted to assume that there is a different notion of degree relevant for each modal type. But one crucial observation suggests that, however it is to be defined, there is just one notion of degree applicable for all modal types. It is quite common to find different modals used to express a number of modality types. For example, *must* and *may* may be epistemic or deontic and *could* can be used in all three ways:

(17)   a   he could be out of town
       b   you could have gone out, if you had finished you homework
       c   when I was a child I could speak Spanish

However, it is a curious observation that modal usage never crosses in terms of degree: *must*, *shall* and *will*, in all their uses, are 'strong' and *can* and *may* never are. Thus whatever modal type a modal is used to express, it maintains its degree and therefore degree must be identical across different modal types.
    It still remains to be decided how many levels of degree there are. As pointed out, Palmer's suggested three levels is problematic as it is unclear what the third level amounts to. Moreover, I am doubtful that the distinction that Palmer refers to with the examples in (13) is truly one of degree. Comparing the examples below, it seems to me that the distinctions between them are of a different nature:

(18)   a   this must be the right answer
       b   this may be the right answer

(19)   a   this must be the right answer
       b   this will be the right answer

In (18) we have the standard distinction between 'necessity' and 'possibility'. Let us refer to this distinction as 'high' and 'non-high', to anticipate the binary feature analysis I will eventually adopt. The distinction in (19) is different. I will argue that both these modals are high and that the difference between them has to do with the fact that *must* is associated in its epistemic use with

conclusion while *will* is associated with epistemic prediction. Essentially, (19a) means: given what we know, we conclude that this is the right answer, while (19b) means: given what we know, we predict that this is the right answer. The difference has to do with the amount of knowledge available. For conclusion, we are in possession of all the relevant knowledge needed whereas prediction is based on a lesser degree of knowledge. I think this is the basis of Palmer's intuition that *will* falls between *must* and *may*, it shares the same degree as *must*, but it expresses something which is slightly less sure.

The distinction between conclusive and predictive uses of modals falls into what I will refer to as minor features. Typically such features are relevant to a subset of modals and only in some of their types. Thus, *must* is conclusive only in its epistemic use and not in its deontic use and *can* is not used either conclusively or predictively.

Having settled on a binary analysis of degree, it should be noted that within the two degree levels there is another distinction to be found. This is clearly exemplified by the following:

(20)   a   he may know the answer
       b   he might know the answer

Though both of the modals involved in (20) are non-high, they are distinguished in that *might* expresses something more tentative or perhaps more dependent on conditions. A similar distinction can be detected with other pairs of modals in their various uses. Consider:

(21)   a   that must be the answer   that should be the answer
       b   I shall be there at 3   I should be there at 3
       c   that will be the answer   that would be the answer (if …)
       d   you must buy a ticket   you should buy a ticket
       g   he will lend you the money   he would lend you the money (if …)
       d   he may be out of town   he might be out of town
       e   you can go the party   you could go to the party (if …)
       f   I can play the tuba   I could play the tuba (if …)

I will refer to the use of modals exemplified by the second column as the lowered form (again anticipating a binary distinction between lowered and non-lowered degrees). Note that in the majority of cases, the lowered degree is expressed by what is traditionally considered the past tense form, with the exception of (21a), where *should* expresses the lowered degree of *must*. In fact, it seems that this is the real relationship between these modals in the present system rather than the tense distinction it originated in. The tense

distinction remains with the potential modals, though, and we will discuss tense at a later point.

The conclusion of this discussion is that there are four degree levels, with a major split between the high and non-high degrees. These are then split again into lowered and non-lowered degrees. The system can be described with the use of two binary features [±high], [±lowered]. Combining this with the three modal types we get the following system:

(22)

	epistemic	deontic	potential
[+high, -lowered]	must will shall	must shall	will
[+high, + lowered]	should would	should	would
[-high, -lowered]	may	may can	can
[-high, + lowered]	might could	might could	could

The claim made earlier that the modals readily cross over from one type to another, but never cross from one degree to another is very apparent in this table. Clearly this is something in need of explanation.

2.2 Temporality and modality

Above we mentioned the fact that what are traditionally referred to as 'past tense' modals are in fact more accurately described as lowered modals. The interaction between time reference and modality is an interesting topic in itself. Here I will briefly mention some rather spurious connections between the two before turning to the central issue to be treated in this paper.

It is a common view that the modal *will* expresses future tense in English. While this claim can be challenged on the grounds that this is an inaccurate use of the term 'tense', I would go further and claim that 'futurity' is not a semantic feature of this modal. For one thing, *will* is not always associated with a future meaning:

(23) a   that will be the postman
      b   you will have heard of him in the news

In (23a) the situation referred to is clearly a present one and in (23b) it is past.

In this last example it is not even the case that the meaning involves future from a past point. In these examples it is the predictive aspect of the meaning of this modal that is utilised and I suspect that it is this aspect of meaning that makes this modal compatible with a future interpretation: prediction is often associated with the future for the simple reason that we can have no other kind of knowledge about it. Predictions can be made about the past, as (23b) demonstrates, but this is not the only knowledge we can have of the past so the relationship between prediction and past is not so strong.

In support of these claims it should be noted that *will* is not the only modal which is compatible with future time reference. Indeed all modals with a predictive aspect to their meaning can:

(24)   a   it may/might rain tomorrow
       b   you shall/should have it tomorrow
       c   I can/could send off the package first thing in the morning

The first thing to note from these examples is that although it is obvious from them that these modals can be used in such a way, it is not usual to find the claim that these are 'future' modals. It is unclear therefore why *will* should be singled out in this respect. Second, if it is accepted that these modals can be used in situations with future time reference without being specifically marked for future, then the same possibility must be open for *will*.

Next we turn to apparent past tense uses of certain modals. As we know one set of modals originates from the past tense forms of the other set, though as we have seen, tense is not the primary distinction between them in the present system. However, it has been claimed that a certain aspect of past tense remains with these modals in that they are used to express the 'present' modals in clauses subordinate to a clause with a past tense verb:

(25)   a   he may leave   I said he might leave
       b   I shall be on time   I said I should be on time
       c   I can play the tuba   I said I could play the tuba

I suppose the argument is that there is some sort of agreement in the tenses of the main and subordinate clauses in the second column and hence given that the matrix is past, we should assume that the subordinate clause is past.

There are a number of reasons to reject this claim, however. The first is that these modals simply do not behave like past tense verbs in this situation. A past tense verb of a clausal complement of another past tense verb maintains its own past time reference and is not just in the past tense form because of an agreement requirement:

(26) a   she leaves for Paris tomorrow
      b   * she left for Paris tomorrow
      c   * I said she left for Paris tomorrow

Present tense verbs are compatible with future adverbials such as *tomorrow* but past tense verbs are not, as shown in (26a) and (b). As demonstrated by (26c), the past tense verb maintains its incompatibility with such adverbs even when embedded in a past tense clause. With modals, however, the situation is entirely different:

(27) a   she may go to Paris tomorrow
      b   I said she might go to Paris tomorrow

The fact that the modal maintains its compatibility with the adverb even when apparently 'agreeing' with the past tense verb of the matrix in (27b) demonstrates that whatever is going on here has very little to do with past tense.
     Another argument for the same thing can be made on the basis of the following data:

(28) a   I shall be there at 3
      b   I should be there at 3
      c   I said I should be there at 3

(28b) has a possible deontic interpretation excluded for (28a). Note however that (28c) is ambiguous between the two interpretations. This leads us to the conclusion that, assuming that *should* in (28c) agrees with the past tense form of the governing verb, *should* is that past tense form of both *shall* and *should*. This can only be made sense of if we assume that there are three different modals *should*: one which is the past tense of *shall*, one which is not a past tense modal at all and one which is the past tense version of this modal. This seems very unlikely.
     The way out of these problems is, I think, to drop the assumption that the relationship between the matrix and embedded clauses is an agreement one. If we assume instead that there is simply a form of the modal that appears in clauses subordinate to past tense clauses, though this form itself is not 'past tense', then the problems disappear. Let us call the relevant form the *shifted form*, which in general corresponds to the lowered form (with the exception of *must* which is its own shifted form). If we now assume that only shifted forms may appear in clauses embedded in past tense clauses we state the relevant

condition without the assumption that the form the modal takes has anything to do with past tense.

We now turn to consider a rather more important interaction between modals and tense. It is sometimes claimed that modals are in complementary distribution with finite verbal inflection. However, there are a number of observations that show this to be inaccurate as far as tense is concerned. First the potential modals demonstrate a clear tense distinction:

(29)   a   I can speak Spanish
        b   in those days, I could speak Spanish

(30)   a   I will pay anything for a ticket
        b   when I had the money, I would pay anything for a ticket

It is clear that in these examples *could* is used to refer to past ability and *would* for past volition, making these past tense versions of *can* and *will*. For other modals this distinction is not apparent: *might* never means 'past *may*', for example. Yet even in these cases, a past meaning can be present, only indicated by the perfective auxiliary:

(31)   a   he may/might go to Paris   he may/might have gone to Paris
        b   he must leave   he must have left
        c   I shall/should arrive at 3   I shall/should have arrived by 3

I will not deal with the issue of how or why past is realised by the perfective auxiliary in these cases, that is the subject of another paper, however, I will simply note that whatever else we might have to say, the past feature appears in a position following the modal. This feature is in complementary distribution with that which appears on the potential modals, as pointed out by Palmer (1987):

(32)   a   * he can have played the tuba
        b   * he will have lent me the money

(32b) is grammatical, but only with an epistemic reading. It cannot mean that he was willing to lend me the money at sometime in the past.

These observations indicate that the semantic features of time reference, let us refer to them as tense features, are not in complementary distribution with modality features. But how tense features are realised is dependent on the modality type: with potential modality, the tense feature is realised by the modal itself and with other modalities it is realised in an independent and

following position.

It is important to note that the distribution of the tense feature is determined by modality type and not the actual modal auxiliary used. This can be clearly seen by the fact that (32b) has a grammatical interpretation when *will* is interpreted epistemically. Thus, [+past] is not incompatible with the modal *will*, it just distributes differently depending on which modality this modal represents.

There is, however, an incompatibility between past and one kind deontic modality, for fairly obvious reasons. Non-high deontic modality grants permission, a speech act not usefully combined with past time reference:

(33)　　? Cinderella may have gone to the ball

That this is purely a meaning restriction is supported by the fact that past is not incompatible with high deontic modality, where obligation for some past condition makes perfect sense:

(34)　　you must have (already) bought a ticket to get into the concert

In conclusion, there is obviously an interaction between tense and modality features and this has an effect on modal choice. We will return to these issues in the last section of this paper.

2.3 Minor features

Finally in this section I will discuss some of the minor distinctions between the English modals.

In the classification system we have so far outlined it is often the case that a given set of features can be realised by one of a number of modals (see table (22)). For example, consider high epistemic modality. Under one condition or another this can be realised by *must*, *shall*, *should*, *will* or *would* (that is, all of the modals marked for high degree):

(35)　a　he must be there by now
　　　b　I shall/should be in my office at 3
　　　c　that will/would be the postman

Clearly, however, the modals in these sentences do not express the same thing. It is rare that languages have two equal ways of expressing exactly the same meanings and when there are two similar realisations for some underlying meaning it is often that they are separated by overtone or stylistic difference.

Let us start with the distinction between *must* and *will*. We have already seen that *will* is predictive in our discussion of 'future tense'. *Must*, by contrast, is associated with conclusions rather than predictions:

(36) a   the butler must be the thief
      b   the butler will be the thief

The above sentences might be uttered in the discussion of a crime novel. In (36a) the speaker offers an opinion which concludes on the basis of available evidence provided in the novel so far about the identity of the thief. (36b) offers a prediction, also based on previous knowledge, but probably of a more general kind (i.e. that in this sort of novel the butler is always the thief). Hence (36b) might be uttered before the speaker has read any part of the book.

The distinction between conclusive and predictive modals is restricted to epistemic modality, for obvious reasons. Conclusions and predictions are based on prior knowledge which is exactly what epistemic modality is concerned with. Deontic and potential modalities are incompatible with this distinction: there can be no predictive or conclusive obligation or volition.

For the epistemic modals however, the distinction is fairly pervasive. *Shall*, for example, is clearly predictive and as we shall see, is more like *will* than *must*. Interestingly, both *would* and *should* are used conclusively, and so appear to be lowered versions of *must*, but only *would* is used predictively as well, as the lowered version of *will*:

(37)   a   the butler should be the thief, if what I've been told is correct
       b   the butler would be the thief, if the book were written properly

(38)   he would say that, wouldn't he?

In (37) both modals are used conclusively: based on previous knowledge, it is concluded that the butler is the thief, contingent on certain conditions[5]. The modal use in (38) is interesting and can be best understood if we compare it to the following:

(39)   he will say those kind of things

---

[5]   As far as I can tell, the difference between the two modals is that *would* is used counterfactually here. The implication is that the butler isn't the thief precisely because the book was not properly written. With *should* the implication is that the conclusion is a valid one, providing the conditions are met.

In this example *will* is used predictively: based on what is known about the character of the subject, it is predictable that he says those kind of things. In (38) the modal is used in a similar way, though the lowering has an effect on the interpretation. In this case there is a kind of 'after the event' prediction. The sentence means something like: given that what is known of the character of the subject and given the set of circumstances, it was predictable that he said what he said. I cannot think of a similar use of *should* and hence it appears that this modal is not used predictively.

Consider now the distinction between *shall* and *will*. Obviously these differ in that only *will* is used potentially and only *shall* is used deontically. But both are used epistemically. In this case, for most speakers, *shall* is replaceable by *will* and indeed many avoid this use of *shall*. Tradition has it that *shall* is a first person form and that *will* is used for all other persons. However, it is unlikely that this is an agreement phenomenon, as, for one thing, English modals do not otherwise display agreement features. Moreover nowhere else in the language is there a distinction made between first person and all others for agreement purposes. The most common agreement distinction is third person singular versus the rest and with the verb *be* in the past tense we find first and third person singular versus all the rest. In the present tense, *be* distinguishes first person singular, but not against all the rest, as third person singular is also distinct and so there is a three-way distinction made. However, for no verb is the first person distinguished from all the rest.

To understand what is going on with *shall* used in its epistemic sense, we should consider it in its deontic sense too. In this case it contrasts with *must*:

(40)   a   you must go to the ball
       b   you shall go to the ball

Both of these sentences impose obligations, but the target of the obligation is different. With *must* the obligation for bringing about the situation described by the proposition sits squarely with the subject of the sentence, but with *shall* the obligation lies with the speaker: (40b) is a promise given by the speaker to the hearer that the speaker will bring about the situation described by the proposition. We might say, therefore, that *shall* is 'speaker oriented' in its deontic sense.

I believe that it is the speaker orientation of *shall* that influences its use in first person environments in epistemic contexts. This may also account for the use of shall in proclamative predictions:

(41)   he shall be king of all England!

Such proclamations are promissory and hence might also be taken to be speaker oriented. However, this is probably not an aspect of the grammar, as evidenced by the fact that for most speakers this restriction is a preference rather than a grammatical condition and many speakers find the usage uncomfortable and therefore avoid it and use *will* instead.

We should also consider the deontic usage of *should*. This modal is not speaker oriented and so it is not used as a lowered form of deontic *shall*. Instead it seems to be the lowered form of *must*. As Palmer (1987) points out, one of the differences between deontic *must* and *should* is that although both are used to express obligation (i.e. high deontic modality), *must* tends to state an inviolable obligation, whereas *should* imposes a somewhat less strong obligation:

(42)   a   ? he must sign the contract, but he won't
        b   he should sign the contract, but he won't

This fits well with the idea that *should* is a lowered version of *must*.

Finally we should mention the difference between deontic *may* and *can*. Both of these confer permission in pretty much the same way:

(43)   a   you may leave now
        b   you can leave now

Again the difference seems to be a stylistic one, *may* being restricted to formal registers. I will not attempt to answer the question of how this is to be handled in the grammatical system, if at all, in this paper. There are several possibilities, however. We might consider formality as a kind of minor feature which restricts the use of *may* to those situations marked as formal. Alternatively we can view different registers themselves to be different systems and as such formal and informal expressions do not compete with each other. A third possibility is that registers do not differ grammatically, but have access to separate vocabularies and hence there are different choices made in selecting vocabulary items to realise grammatical expressions.

3   The Grammatical System

Because of the separation of the process of vocabulary insertion from the system which determines the grammatical organisation of input elements into sequences, the analysis of the use of modals falls into two parts: one dealing with the alignments which fix the features of modality into a given order and place them within the expression with respect to other features and one which

determines the principles governing the choice of vocabulary item to be inserted into the expression. We will take these issues in this order.

3.1 Feature alignment

Clearly there are a number of features associated with modality. Following from the above discussion I will assume that there are two main types and possibly a number of minor features. The main features determine the type and degree of modality. The observations we have made suggest the following features:

(44) type: [epistemic]
                [deontic]
                [potential]
       degree: [±high]
                [±lowered]

For the cases we will be dealing with, only one of the type features will be involved[6]. This will be combined with both of the degree features to make up the whole modality meaning.

I will take the type feature to be the modal head. This means that the degree features will be aligned with respect to this and also the general distribution of modals will be determined by the alignment requirements of the type features.

The alignment of the degree features with respect to the type feature is not a complicated nor particularly interesting aspect of the system. As far as I can tell these features are never separated and therefore realised by independent elements. As they are always realised by a single vocabulary item, it is also difficult to demonstrate what order they appear in. I will assume that their alignment is determined by the general head-dependent conditions of the language, which for English is head first[7]. If we also assume that the [±high]

---

[6] It is probable that more than one type feature can be selected to modify a single predicate as dialects which allow the use of double modals testify. In standard English double modalities are possible with the second one being realised as verbs and adjectives with modal meaning, such as *might have to* or *must be able to*, etc. There are restrictions on this however and double modal dialects do not allow the combination of all modals nor any order of the modals that are able to combine. I will not treat this issue in this paper.

[7] This also addresses the issue of learnability. Clearly the language learner will not be presented with evidence concerning the order of these features. But assuming their head-dependence relationships to be universal, overt evidence from the language will be present to enable to relevant constraint ranking to be learned.

feature is a head with respect to the [±lowered] feature, which seems reasonable, it follows that this feature will be aligned after the head degree feature. Hence the general organisation of the modal features will be:

(45)   [type] [±high] [±lowered]

This pattern, as mentioned, seems never to be interrupted. However there are other elements which are subject to the general head-dependent alignment constraint, lexical heads and their complements for example, where the general pattern may be disrupted under certain circumstances. Objects appear at the front of the expression if they are interrogatively marked or marked for topic status, etc. The reason why such processes never affect the modality features is, I suspect, due to their relative semantic simplicity. A wh-argument is semantically complex, carrying both properties of its argument and interrogative status. It can therefore be subject to contradicting requirements on all the aspects of its meaning. But there are no contradictory requirements on modality features and thus they will only be required to occupy the position relative to their head.

Next we turn to the issue of the placement of the modality features within the string of elements that constitute a proposition. For the majority of cases, the modal precedes the main predicate[8]. However the requirement that it be adjacent to the main verb is fairly weak as there are many things that can intercede between them: adverbials, every other auxiliary verb, the tense feature and negation, etc[9]. This indicates that while the precedence constraint aligning the modal features, specifically the type feature, to the left of the predicate is higher ranked than the constraint that would place it to the right, its adjacency constraint is relatively low ranked: lower than the adjacency constraints for the non-modal auxiliaries, for example. We can thus discern the following rankings:

(46)   P(mtype, pred) > F(mtype, pred) > … > A(pass, pred) >
         A(prog, pred) > A(perf, pred) > A(mtype, pred)

The constraints referred to here are:

---

[8]   The main exception to this concerns VP fronting phenomena. We will not treat this in this paper, though it is clear how such a treatment would proceed in an OT framework.
[9]   Inversion phenomena also indicate that there may be a fronting condition playing a role in the distribution of finite elements.

(47)   P(mtype, pred)   the modality type feature precedes the predicate
       F(mtype, pred)   the modality type feature follows the predicate
       A(mtype, pred)   the modality type feature is adjacent to the predicate
       A(pass, pred)   the passive is adjacent to the predicate
       A(prog, pred)   the progressive is adjacent to the predicate
       A(perf, pred)   the perfective is adjacent to the predicate

Note that because the adjacency between the modality type feature and the predicate is relatively weak, this will not affect the alignment requirements holding between the modality features themselves, which are never disrupted. From this we can conclude that the head-dependent alignment constraints are higher ranked than A(mtype, pred), as demonstrated below:

(48)

	F(dep, head)	A(dep, head)	A(mtype, pred)
☞ [type] [±high] [±lowered] V			**
[type] V [±high] [±lowered]		*!	
[type] [±high] V [±lowered]		*!	*

The optimal order has the modality features contiguous, at the expense of the adjacency requirement between the type feature and the predicate. The suboptimal candidates better satisfy this adjacency requirement, but at the expense of violating the head-dependent alignment constraints, which are stronger requirements.

In conclusion, then, the modality features will always cluster together in an expression and hence are able to be replaced by a single vocabulary item. We will now turn to the process of vocabulary insertion itself.

3.2 Vocabulary insertion

What is to be explained in this section is why particular modal vocabulary items are selected to replace any given cluster of modality features. This explanation relies on two aspects of the system. First there is the analysis of individual modal vocabulary items in terms of what feature sets they are associated with. Second there are the principles of the selection process which use the associated features of vocabulary items to decide which is the best one to use. But these are obviously so interwoven that we cannot analyse one without the other. Therefore the analysis of these must proceed in tandem.

The simplest cases are those in which there is only one modal used to represent a certain set of features. All being equal, we can assume that this modal is associated with exactly these features. For example, only *may* is used

for non-high, non-lowered epistemic modality and hence we might assume it to be associated with the features {[epistemic] [-high] [-lowered]}. Similarly, only *can* is used for non-high, non-lowered potential modality and so we assume its associated features are {[potential] [-high] [-lowered]}. Let us therefore assume the following feature associations:

(49)   *may*   →   {[epistemic] [-high] [-lowered]}
       can   →   {[potential] [-high] [-lowered]}

However, both *may* and *can* are used to express non-high non-lowered deontic modality, {[deontic] [-high] [-lowered]}, and therefore there must be no modal associated with exactly this combination of modality features, otherwise this would be chosen instead of *may* or *can*. Obviously *can* and *may* are the closest to this set of features and as they both differ from it in just one feature, the type feature, we must either assume that every other modal differs from this set by more than one feature, or that differing on the type feature is less harmful than differing on the degree features. This last conclusion is supported by the fact that no modal is selected to realise modalities of different degree, though it is quite common for one modal to realise different types of modality at the same degree level.

This situation can be captured under the assumption that the selection of vocabulary items is determined by a number of ranked matching conditions. In order to be selected as the realisation of a set of features, vocabulary items are evaluated on whether their associated features match with those to be realised. The evaluation considers the relevant features one at a time, eliminating vocabulary items that do not have a matching feature. Obviously the features considered first will play more of a role in determining the selection of the vocabulary items than those considered later. In the case we are considering, the degree features are given priority over type features.

Let us take a few examples to demonstrate how this works. Suppose the target features are: {[epistemic] [-high] [-lowered]}. For the sake of the argument, let us assume a modal *X* which is associated with the same features as *may* except that it is [+high] rather than [-high]:

(50)   *X*   →   [epistemic] [+high] [-lowered]

It turns out there is no such vocabulary item in English, but the point is, even if there were it would not be selected in this case, as demonstrated in the following table:

(51)

[epistemic] [-high] [-lowered]	Match degree[10]	Match type
☞ may		
can		*!
X	*!	

Obviously, as *may* has exactly matching associated features, this vocabulary item will be selected. The same would be true of *can* for the target features {[potential] [-high] [-lowered]}.

The interesting case is of course deontic modality. In the following table we evaluate the modals for the target features {[deontic] [-high] [-lowered]} and this time assume the modal *X* to be marked as deontic, though having non-matching degree features:

(52) $X \rightarrow$ {[deontic], [+high], [-lowered]}

Recall that there is no vocabulary item associated with exactly this set of features:

(53)

[deontic] [-high] [-lowered]	Match degree	Match type
☞ may		*
☞ can		*
X	*!	

In this case, both *may* and *can* demonstrate the same level of matching and hence both can be selected as the realisation of these target features.

We have pointed out that these modals are not equivalent, even in their deontic use, as *may* appears more formal than *can*. If different registers amount to the use of different grammatical systems, then this distinction should not be dealt with here. If different registers make use of the same grammatical system but access different vocabulary items, this might be handled by having vocabulary items marked for register and having a high ranked matching condition specific to register. Thus, suppose we have the same target features as for the previous case, though this time marked for formal register, the following would result:

---

[10]   I assume here that the degree features are evaluated together under one matching condition. This is not a necessary assumption and indeed there is reason to believe that the degree features may be evaluated separately, with [±lowered] being a lesser consideration. This has to do with the behaviour of the negative *mustn't* which we will not review in this paper, but see Newson (forthcoming) for an analysis. For reasons of simplicity at this point, we will not adopt this assumption here.

(54)

[deontic] [-high] [–lowered]	Match register	Match degree	Match type
☞ may			*
can	*!		*
X		*!	

Questions arise, however, concerning how register is marked in a linear way, given that it is a feature that potentially affects the whole expression rather than just individual elements within it. This suggests that vocabulary items of different registers are entirely separate and do not enter into competition with each other.

Let us now turn our attention to the lowered equivalents of *may* and *can*, i.e. *might* and *could*. It would seem reasonable to take these to be associated with the same set of features as previously discussed, except for being associated with [+lowered]. However one observation suggests that this is not entirely correct. While it is true that like their non-lowered equivalents, *might* and *could* can be used epistemically and potentially respectively and that both can be used deontically, the fact that *could* can also be used epistemically prevents us from adopting the straightforward analysis. If *might* were marked for [epistemic] and *could* for [potential], the latter would never be able to be used in epistemic contexts as the former would always block it. However, if both *might* and *could* were marked for [epistemic] and there were no [potential] modal at this level, we would predict that, contrary to fact, both modals would be used in potential contexts. The fact that both modals can be used in epistemic and deontic contexts indicates that no modal is marked for these modality types at this degree level. In which case, the question that needs answering is what features is *might* associated with? As *could* is the only modal to be used to express potentiality at the appropriate degree, it seems that this modal is marked for this type. The only way to get both *might* and *could* to be equally used in the other contexts would be to mark *might* as neither. And hence, this modal seems not to be associated with any type feature at all:

(55)   *might*   →   {[-high], [+lowered]}
       could   →   {[potential], [-high], [+lowered]}

Let us consider a few examples. The most straightforward case is that of non-high, lowered potential modality:

(56)

[potential] [-high] [+lowered]	Match degree	Match type
☞ could		
might		*!
can	*!	

As previously, the modal which is associated with the same features as the target set will be the only one selected.

For deontic and epistemic modalities, things are different:

(57)

[deontic] [-high] [+lowered]	Match degree	Match type
☞ could		*
☞ might		*
can	*!	
may	*!	

(58)

[epistemic] [-high] [+lowered]	Match degree	Match type
☞ could		*
☞ might		*
may	*!	

In both cases only *could* and *might* are marked for the relevant degree and so these are the only two that could be used in these contexts. However neither of them match for the type feature in either case: *could* because it is marked for [potential] and not [epistemic] and *might* because it is associated with no type feature at all. We see here the importance of the decision concerning what counts as a violation of the matching conditions. If it were better to fail to be associated with a feature than to be associated with a conflicting feature, there would be no way to prevent *might* being chosen over *could* in these cases. Only if non-association with a feature and association with a conflicting feature are taken to be equal violations can we account for the distribution of these modals.

Now we turn to the high modals, which are to some extent more complex than the non-high ones. The potential modals at this degree level are straightforward with only *will* used for high and non-lowered and *would* for high and lowered. As previously, this suggests that these modals are associated with exactly these features so there can be no other modal used to realise them:

(59)   *will*   →   {[potential], [+high], [-lowered]}
       would →   {[potential], [+high], [+lowered]}

The other modalities are more complex. For high deontic both *must* and *shall* can be used for non-lowered cases and *should* is used for the lowered version. For high epistemic modality *must*, *shall* and *will* are used in non-lowered cases while *should* and *would* are used in lowered cases. This means that it is deontic modality which is the more restricted in terms of its selected realisations and indeed any high modal can be used to express high epistemic modality. Given our previous arguments, this would suggest that no modal associated with a high feature is associated with the epistemic feature, otherwise this would always be the best realisation of this modality. Therefore we conclude that *must*, *shall* and *should* are all associated with the deontic type feature. This explains why *will*, unlike its non-high equivalent *can*, is never used to realise deontic modality. We therefore assume the following feature associations in the vocabulary:

(60)   *must*   →   {[deontic], [+high], [-lowered]}
     shall   →   {[deontic], [+high], [-lowered]}
     should   →   {[deontic], [+high], [+lowered]}

We will not demonstrate how these feature associations lead to the correct modal selection for the realisation of combinations of modality features as this should be fairly obvious by now.

The distinction between *must* and *shall*, however, merits some discussion. We have said that the distinction is twofold: *shall* is speaker oriented which results in a deontic reading functioning similar to a promise and an epistemic reading restricted to first person subjects and *must* is conclusive rather than predictive in its epistemic use. Neither of these are to do with register and so they are not to be dealt with in the same way as the distinction between deontic *may* and *can*. We can assume that the conclusive/predictive restriction is a minor feature of modality, which is restricted to epistemic contexts. This appears to be a semantic restriction as the distinction is not compatible with the meaning of deontic and potential modalities. If this is so, then we may assume that the presence or absence of these features is an input condition and not determined by constraint ranking. In the presence of a conclusive feature *must* will be selected as opposed to *will*. In the presence of a predictive feature, *must* will not be selected.

The behaviour of *should* and *would* is interesting in this respect. Recall that epistemically, *would* can either be used predictively or conclusively whereas *should* can only be used conclusively. That both can be used conclusively follows from the assumption that there are no modals associated with the features [+high] and [epistemic]. That *should* cannot be used

predictively indicates that it is not associated with a predictive feature, but *would* is. It follows that *should* is neither associated with a predictive nor a conclusive feature. We therefore conclude that these modals have the following associated features:

(61) *should* → {[deontic], [+high], [+lowered]}
       would → {[potential], [+high], [+lowered], [predictive]}

To demonstrate the usage of *would* and *should* in predictive and conclusive contexts, consider the following tables:

(62)

[epistemic] [+high] [+lowered] [conclusive]	Match degree	Match type	Match minor
☞ would		*	*
☞ should		*	*

(63)

[epistemic] [+high] [+lowered] [predictive]	Match degree	Match type	Match minor
☞ would		*	
should		*	*!

In the first table, neither *would* nor *should* are associated with the epistemic type feature nor the minor conclusive feature. Hence both are equivalent and may be used to replace the given features. In the second table, though neither are epistemic, *would* is associated with the predictive feature and hence *should* is not selected to realise these features.

Speaker orientation is a complex issue about which it is difficult to come to definite conclusions. On the one hand, it seems that the speaker orientation of *shall* in its deontic sense is semantic, bearing all the hallmarks of a minor feature, being restricted to a certain modal in a particular usage. However its effects in the epistemic use of *shall* are not so clear cut. Moreover it is not at all clear that the two are to be taken as similar phenomena: deontic speaker orientation changes the speech act performed from granting permission to making a promise, but in epistemic contexts it merely predisposes towards the use of first person subjects. In its deontic use there is no substitute vocabulary item which is able to express this particular modality, but with epistemic modality *shall* can be replaced by *will* and indeed there is a tendency to do so. In both of its uses, it seems to me, that there is some discomfort with its use

and I suspect that it is the least used modal of all[11]. For these reasons, I think it is likely that the speaker orientation of *shall* in epistemic contexts is not grammatically or semantically based and may even be extralinguistic, influenced by prescription and possibly used in a 'set' or 'fossilised' way. I will not speculate about this further in the present paper.

4 Potential Modality and Tense

The final topic we will attend to in this paper is the relationship between modality and tense features. As was discussed previously, there is a difference in this relationship depending on whether the modality is potential or one of the others. For the potential modals, the choice of modal will obviously depend on the presence of a past tense feature (for *could* and *would*) or a non past tense feature (*can* and *will*). This will be straightforward under the assumption that each modal is associated with the relevant feature. This time, however, the more complex issue concerns the alignment of the tense and modality features.

It is perhaps a traditional view that the tense distinctions to be seen in combination with different modalities is more of a semantic issue. For example Palmer (1987) describes the distinction in terms of whether it is the (potential) modality that is tensed or whether it is the proposition: from this point of view, the proposition is marked for tense by the presence of the auxiliary *have*[12]. Thus the following express a 'past ability' to perform some action (64a) and the possibility of some 'past action' (64b):

(64)   a   I could play the tuba
       b   I may have played the tuba

However, this may not be the best way to view the situation as it fails to account for the complementary distribution of the two kinds of tense marking. Even if we assume a semantic restriction that only potential modality can be

---

[11]   As a very rough indication of this, a Google search for *will* found approximately 4,690,000,000 hits while *shall* found only 274,000,000. Of course, one has to bear in mind that *will* has other uses than its modal and this would account for some of this difference. Without a more sophisticated search engine than Google is able to provide however, it is impossible to get more accurate figures.

[12]   Palmer clearly uses the term tense to cover not only the form of verbs but also the form of larger sequences of words. For example he defines the 'proposition' as "what is expressed by all that follows [the modal], including the main verb" p. 98-99. From this perspective, while a verb may be marked for tense by a morpheme, the proposition is marked for tense by the addition of an auxiliary verb.

associated with tense features, it is not clear why we could not have a past tense modality and a past tense proposition in such a situation. It seems that there can be just one tense feature per clause and this can be either realised by the (potential) modal or by an auxiliary following the modal.

Another approach would claim that there is just one tense feature in (64) and that its interaction with the modality is to be seen in terms of its scope: either the modality or the tense feature has wide scope. Specifically the scope relationships seem to be as follows:

(65)   epistemic, deontic > tense > potential

Support for this scope hierarchy comes from dialects which allow the use of 'double modals', such as certain Scots dialects (Brown 1991). Cormack and Smith (2002), on the basis of Brown's data, claim that only *can* and *could* appear in the second modal position:

(66)   a   He should can go tomorrow
            'he ought to be able to go tomorrow'
       b   He would could do it if he tried
            'he would be able to do it if he tried'

Although Cormack and Smith refer to all non-epistemic modalities as deontic, it is clear from the examples they cite that the second modal is used in its potential sense. Thus this provides evidence that potential modality lies in the scope of the other modality types when they are able to appear together in one propositional unit.

The problem faced by this view is how the syntactic differences can be made to follow from the semantic differences. The only point of contact between syntax and semantics within the present framework is the input, which consists of abstract semantic features. It is difficult to see how something like scope can be represented in terms of a feature and therefore how it can be made a legitimate input element. But if scope is a pure semantic property, there is nothing in the input which would mean that the tense feature should be reaslised differently in the grammatical expression, i.e. as part of the modal or as a separately pronounced perfective auxiliary.

The answer, I think, lies not with the syntactic system at all, but with the mechanism of vocabulary insertion. I will assume that the tense feature is positioned with respect to the verb and the modal in exactly the same place no matter what modal type is present. The fact that when the tense feature is realised as a separate element from the modal it follows the modal and precedes the verb (as in (64b)) suggests that this is its position in general:

(67)   modal tense Verb

This is easily achieved via the following constraint and ranking:

(68) P(tense, pred)   the tense feature precedes the predicate
      F(tense, pred)   the tense feature follows the predicate
      A(tense, pred)   the tense feature is adjacent to the predicate

(69)   P(tense, pred) > F(tense, pred) > A(tense, pred) > A(mtype, pred)

The ranking of the first three constraints is the typical pattern, with precedence the highest ranked condition. Therefore tense will be positioned before the predicate. The ranking of the tense adjacency constraint above that of the modality type ensures that the tense feature will be closer to the predicate than the modal, producing the order in (67). Again, assuming the high ranking of the head-dependent constraints, this will ensure that the modality features will cluster together in front of the tense feature.

As vocabulary items, English has potential modals which are also associated with tense features, a fact which is probably influenced by the semantic relations between them. But its other modals are not associated with tense features. Hence replacement of the modality and tense features by one element is only possible with potential modality and with the other modalities the tense feature must be realised by a separate vocabulary item.

As previously stated, it lies beyond the scope of this paper to account for the realisation of the tense feature as a perfective auxiliary when it accompanies non-potential modalities. Here I will concentrate on the realisation of tense on the potential modals. First, we must amend the vocabulary entries associated with the relevant modals:

(70) *can*   →   {[potential], [-high], [-lowered], [-past]}
      could   →   {[potential], [-high], [+lowered], [+past]}
      will   →   {[potential], [+high], [-lowered], [-past]}
      would   →   {[potential], [+high], [+lowered], [+past]}

Suppose the alignment system produces the following subsequence of input features:

(71)   … [potential] [-high] [+past] …

In principle, there are two ways this could be realised. The modality features

could be spelled out by the most appropriate modal, in this case *could*, and the past tense feature could be spelled out independently, presumably by *have*. Alternatively all three features could be spelled out by a modal, again *could*. It is clear that the latter is preferable given the principle of Minimal Vocabulary Access. Obviously with other modalities the second option is not available and hence the tense feature has to be spelled out separately.

I observe that there appears to be some connection between the [+lowered] feature and the [+past] feature, which I am unfortunately unable to fathom. This shows itself in two ways. First, modals associated with [+past] are also associated with [+lowered] and second, the two features are in complementary distribution[13]. Thus there is no tense distinction with the lowered potential modalities. For this reason both *could* and *would* are ambiguous when used potentially, expressing either the lowered form of *can* and *will* or the past tense form of these modals:

(72) a   I could sing falsetto
          "I used to be able to sing falsetto"
          "I would be able to sing falsetto (if …)"
      b   I would lend him money
          "I used to be willing to lend him money"
          "I would be willing to lend him money (if …)"

I have no idea why these features are connected in this way, though it obviously suggests an even more abstract analysis in which they are collapsed into one. What this abstract analysis is to be based on and how it can be restricted to just the potential modalities are mysteries.

5   Conclusion

This paper has attempted to demonstrate how it might be possible to account for some of the most puzzling aspects of the English language. Clearly, only the surface has been scratched and probably more questions have been raised than answered. Possibly this is inevitable given the subject matter and much more research will be necessary before less tentative conclusions can be reached. I feel however that the approach presented here makes sufficient progress to warrant interest. It should also be pointed out that it forms only a

---

[13] This complementary distribution, however, breaks down when the two features are realised separately. Thus with other modalities, both [+lowered] and [+past] can appear together:
    i)   he might have been in his office

part of an ongoing research programme which aims to provide an analysis of English auxiliaries in general. A number of the questions raised in this paper but not attended to fall within the realms of this wider research and it is hoped that they can be handled in a consistent way to that presented here.

References

Barbiers, Sjef. 2005. The syntax of modal auxiliaries. In:. Martin Everaert and Henk van Riemsdijk (eds.) The Blackwell companion to syntax. Oxford: Blackwell. 1-22.

Brown, Keith. 1991. Double Modals in Hawick Scots. In: Peter Trudgill and Jack K. Chambers (eds.) Dialects of English. London: Longman. 74-103.

Cormack, Annabel and Neil Smith. 2002. Modals and Negation in English. In: Sjef Barbiers, Frits Beukema, and Wim van der Wurff (eds.) Modality and its Interaction with the Verbal System. Amsterdam/Philadelphia: John Benjamins. 133-164.

Gáspár Miklós. 2005. Coordination in Optimality Theory. PhD. dissertation. Budapest: ELTE.

Halle, Morris and Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In: Kenneth Hale and Samuel Jay Keyser (eds.) The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger. Cambridge, MA: MIT Press.111-176.

Hofmann, T. R. 1976. Past tense replacement and the modal system. In J. D. McCawley (ed.) Syntax and Semantics 7: Notes from the Linguistic underground. New York: Academic Press. 85-100.

Lyons, John. 1977. Semantics. Volumes I, II. Cambridge, England: Cambridge University Press.

Maunula, Vili Oskari. 2006. Case marking, agreement and word order. MA dissertation. Budapest: ELTE

Newson, Mark. (forthcoming) Negation and modality. Ms. Budapest: ELTE.

Newson, Mark 2004. Deforestation in Syntax. In László Varga (ed.) The Even Yearbook 6 Budapest: Eötvös Loránd University. 135-147.

Newson, Mark and Vili Maunula. 2006. Word order in Finish: whose side is the focus on? In László Varga (ed.) The Even Yearbook 7. http://seas3.elte.hu/delg/publications/even/2006.html#nm.

Palmer, Frank. R. 1987. The English Verb (second edition). London: Longman.

Mark Newson
Department of English Linguistics
ELTE
newson@ludens.elte.hu