ON THE SYNTACTIC FORM OF EPISTEMIC MODALITY

Hans Bernhard Drubig

University of Tübingen

2001
0. Introduction

One of the goals of our project has been to show that typological data, when systematically taken into consideration, can be a decisive factor in the analysis of a particular language, such as English. Typological data can become indispensible when a specific analysis suggests itself but is underdetermined by language particular evidence. I have chosen some aspects of the grammar of epistemic modality in English to show that a typological perspective can yield interesting results and helps strengthen language particular analysis.

I will proceede as follows: After a brief consideration of the relationship between epistemic modals and evidentiality I will survey some important syntactic properties of epistemic modals in English, (relating to Polarity and Tense) and show how epistemic modals interact with focus and givenness marking. The results of this discussion will then be checked against the evidence of multiple modal (MM) constructions in some variations of English and the typology of evidential systems.

1. Epistemic Modality and Evidentiality

Some recent work on the semantics of epistemic modals (Westmoreland 1998) challenges commonly held views of the meaning on English verbs such as must and may or might. Westmoreland proposes an analysis that I find interesting both in relation to what we already know about the syntactic behavior of epistemic modals and from a typological point of view. Westmoreland argues that epistemic must and may/might are not versions of the modal logical necessity and possibility operators and that they are neither duals nor quantifiers over possibilities. More specifically:

1) Epistemic must (in contrast to root modals) is not a modal but must be analyzed as an evidential marker labelling the proposition in its scope as a deduction. It relates a proposition \( \phi \) to some other information that serves as evidence for \( \phi \).

2) The traditional view that may \( \phi \) or might \( \phi \) is true (or acceptable) just as long as \( \phi \) is consistent with the context does not suffice to capture its meaning; rather: an expression such as might \( \phi \) is used to mean that the context contains causal factors that make \( \phi \) plausible.

In general we may say: just as a question marker takes a proposition and derives a question, an epistemic modal takes a proposition and yields an evidentially labelled proposition. An epistemic modal, such as must or may, according to this view,
is part of the metalogical vocabulary, i.e. not equivalent to the square □ but more like the sign .∴ marking the deduction of an inference.

Let us look now at one of Westmoreland’s examples from Alice in Wonderland.

(1) So she began. "O Mouse, do you know the way out of this pool? ..." ... (Alice thought this must be the right way of speaking to a mouse; ...)

(2) Alice thought ["O Mouse"] must be the right way of speaking to a mouse...

It is important to see that MUST φ in (2) is not the proposition which functions as the complement of the propositional attitude verb think. The propositional complement of think is φ. In other words, MUST φ does not affect the context; rather, stating MUST φ makes φ available to the context. As Westmoreland convincingly argues, MUST φ in (2) has a semantic content which is a pair < φ, DEDUCTION>, where DEDUCTION is one of a set of labels that the language has for indicating the source of information. The same type of labelled proposition can also occur as the complement of other propositional verbs and ‘stance’ verbs (cf. Hegarty 1992), as shown in (3).

(3) I think/hope/doubt that Jane must be in the attic

These observations suggest a number of important conclusions. In contrast to root modals with a deontic or dynamic reading, the epistemic modals must and may/might are evidential markers which do not contribute to the informative proposition, but disclose the source of information. Insights into the evidential character of epistemic modals lead to a better understanding of their syntactic (in particular, scopal) properties and their place in the functional hierarchy of the clause. This allows us to evaluate the epistemic modal system of English in the wider context of the typology of evidential systems. In fact, it is not difficult to show that English epistemic modals correspond rather closely to certain evidential forms occurring in languages with more fully developed evidential systems such as Tuyuca, a Tucanoan language, discussed by Barnes (1984).

(4) The Tuyuca evidential system (Barnes 1984: 257):

   a. díiga apé - wi
      soccer  play  - VISUAL
      'He played soccer.' (I saw him play.)

   b. díiga apé - ti
      soccer  play  - NONVISUAL
      'He played soccer.' (I heard the game and him, but I didn't see it or him.)
c. diiga apé - yi
soccer play - APPARENT
'He played soccer.' (I have seen evidence that he played: his distinctive shoe print on the playing field. But I did not see him play.)

d. diiga apé - vigi
soccer play - SECONDHAND
'He played soccer.' (I obtained the information from someone else.)

e. diiga apé - hiyi
soccer play - ASSUMED
'He played soccer.' (It is reasonable that he did.)

As shown in (4), Tuyuca has no less than five distinct inflectional forms of the statement ‘He played soccer’, each one reflecting a different source of information. The inflectional paradigm can be ordered into a hierarchy reflecting the relative defeasability of the source of information. (4)a. is the strongest form of a statement, which can only be used when speaker had direct visual evidence, while b. is used when the evidence supporting the statement is nonvisual. With respect to (4)c. Barnes’ discussion suggests that this is the form used when no direct perceptual evidence is available. The APPARENT form states a necessary inference supported by indirect evidence. As the examples discussed by Barnes show, this is the evidential form which most closely corresponds to the use of the English epistemic modal must in sentences like (1) or (2). (4)e. is weaker than c. and expresses a ‘reasonable assumption’ which is compatible with the evidence but not strictly implied by it, an interpretation which suggests a certain similarity to the evidential modals may or might in English.

Finally, the SECOND HAND form in (4) d. appears to be very close to the ‘hearsay’ modals of Dutch or German, as shown in (5).

---

1 The exact meaning of should in the sense here relevant, which some authors (e.g. Coates 1983) classify as an epistemic modality whereas others (e.g. McDowell 1987) do not, is tangential to the concerns of this paper, but I would like to briefly indicate how it would fit into the overall system of inferential/evidential ‘modality’. Particularly relevant here is the discussion in Woisetschlaeger (1976). According to Woisetschlaeger, a cook may say (i) in a situation where some kind of perceptual (perhaps visual or olfactory) evidence is available to him, whereas (ii) seems to imply the absence of such evidence and may report an assumption exclusively based on the knowledge that all relevant steps of the operation guaranteeing the desired result have been correctly executed.

(i) The roast must be done now
(ii) The roast should be done now

Note that (ii) rather than (i) is felicitous with a continuation such as ‘unless a catastrophe has happened’ or ‘of course, you never know …’. (i) expresses an inference, (ii) a guarded assumption. Cf. Woisetschlaeger (1976) for further examples and an insightful discussion.
This shows that epistemic and hearsay modals correspond to certain slots in more elaborate evidential systems. In a sense, English and other European languages can be claimed to have developed the rudiments of an evidential system, which arose in a historical process starting off with a number of modal auxiliaries. Furthermore, if Westmoreland’s approach to the semantics of epistemic modality is on the right track, which I assume it is, then we may conclude that when epistemic MUST $\varphi$ occurs in the scope of a propositional attitude predicate, must is 'invisible' to the higher predicate and BELIEVE (MUST $\varphi$) is interpreted as if it were BELIEVE $\varphi$; MUST $\varphi$ cannot be interpreted as a propositional argument: epistemic MUST is inaccessible to propositional operators. In other words, an epistemic modal cannot occur and scopally interact with another epistemic modal (or a hearsay modal), and, furthermore, a circumstantial or deontic modal can be in the scope of an epistemic modal (or a hearsay modal), but not vice versa:

(6) John may have to be available for consultation

In (6), the periphrastic modal have to has a deontic reading. Beyond that, the extrapropositional status of an epistemic modal implies that it takes scope over all propositional operators. In order to show this, I will now briefly inspect the interaction of epistemic modals with a number of propositional operators, in particular negation, question operators, tense/aspect, and focus phenomena.

2. Syntactic Properties of Epistemic Modality

2.1 Negation

The scopal relationship between epistemic modality and negation is controversial and has given rise to considerable discussion. On the one hand, it is clear that certain modals cannot be in the scope of negation. In English, must has wide scope over the negative no matter whether its interpretation is epistemic or deontic, a seemingly idiosyncratic fact which, however, can also be observed in numerous other languages (de Haan 1997). Should in its sense closely related to must displays the same behavior. On the other hand, the scope-taking properties of may are known to vary with its interpretation. Thus, (7) a. has wide scope under its epistemic reading, but is in the scope of negation when the interpretation is deontic.

\footnote{Cf. note 1}
According to the standard view, represented by such authors as de Haan (1997), Palmer (1986), Papafragou (1998) and many others, this again must be considered a semantic idiosyncracy of some sort, which shows nothing about the scopal status of epistemic modality in general. According to this view, \textit{can} is the suppletive form of epistemic \textit{may}_ep in the scope of negation. Hence, epistemic modales do not display any unusual properties with respect to negation since they can be both external and internal to its scope. The fact that a suppletive form is necessary in the latter case must be considered an accidental fact. Furthermore, scopal order is unrelated to the order of elements in the functional hierarchy: negation follows the modal or is cliticized to it, independently of its semantic rank.

In the following, I would like to argue that this view, irrespective of its popularity both among semanticists and syntacticians, is essentially mistaken. While there can be no doubt that a sentence with \textit{can}, such as (7)b., uttered in the face of certain knowledge, does have a reading which can be characterized as ‘epistemic’ insofar as the notion of possibility is obviously not equivalent to ‘logical possibility’ but related to knowledge and certainty. As we will see below, however, modals like \textit{must} and \textit{may/might} are ‘epistemic’ in a very different sense: they disclose the nature of the evidence with which the speaker of (7)a. is ready to back up the truth claim of his statement if required. In contrast to the commonly recognized ‘epistemic’ reading of (7)b., the notion of epistemicity attached to sentences with \textit{must} or \textit{may/might} is unrelated to – and, in particular, cannot be reduced to – the modal operators $\Box$ or $\Diamond$. This does not mean that the semantic approach to natural language modality initiated by the work of Angelika Kratzer and others is necessarily on the wrong track, but it does mean that such approaches, no matter how successfully they handle the problems of deontic and other types of ‘root’ modality, cannot be extended to the epistemic domain, which is not susceptible to reductive attempts of this kind, inspite of repeated claims to the contrary. The account of epistemic modality proposed in Westmoreland (1998) provides us with a wider typological perspective which offers more explanatory approaches to otherwise idiosyncratic facts.

Let us now turn to some sentences which throw light on the interaction between negation and epistemic modality. (8) makes immediately clear that epistemic \textit{may} cannot be in the scope of negation: a negative quantifier in a position guaranteeing sentential scope allows only a nonepistemic (in this case: deontic) reading.

(8) Nobody may be at home
If we temporarily disregard the above-mentioned scopal idiosyncracy of must, the following examples suggest that any construction involving unambiguous wide scope negation filters out the epistemic reading in favour of the deontic one. In (9)a. the order of modal and preverb allows may to scope over negation, which induces an epistemic interpretation. In (9)b. and c., the scopal order is inverted, forcing a deontic reading. Similar to Negative Constituent Preposing, shown in (9)b., preposed constituents introduced by only also imply sentential scope negation readings, which trigger Subject Auxiliary Inversion, license the potential occurrence of negative polarity items anywhere in IP (not shown in (10)b.) and exclude the epistemic interpretation of modals, as we would expect:

(9)  
   a. John may never leave early  
   b. Never may John leave early  
   c. John never may leave early

(10)  
   a. John may visit only three people  
   b. Only three people may John visit

The complements of adversative verbs (doubt, deny, etc.) are another context from which epistemic readings are excluded (Progovac 1988: 152 ff.):

(11)   I doubt that John may leave early tonight.

Following Laka (1990) we may assume that adversative predicates select negative complementizers which import a wide scope negation reading to the clausal constituents they head, thereby providing the type of negative context incompatible with epistemic modality.

Under the hypothesis that syntactic structure is the formal basis of semantic composition, we are lead to the conclusion that epistemic and circumstantial modals must be base-generated in different positions, as schematically indicated in (12):

(12)   
   \[ M_{ep}P \]  
   \[ PolP \]  
   \[ M_{nonep}P \]

---

3 Cf. McDowell (1987) for similar observations.
4 Similar suggestions are made elsewhere in the literature, e.g., in Marrano (1998) for Spanish and English, or Shyu (1995: 17 ff.; 56 ff.) for Mandarin Chinese. A forerunner of this analysis is Proposed by McDowell (1987) according to whom root modals are base-generated in VP and epistemic modals in INFL.
The analysis in (12) is not easy to motivate as long as the evidence is exclusively taken from English, because English, unlike very many other languages (e.g. Romance languages, and also all other Germanic languages), allows only a single modal in each clause, a wellknown restriction which, is crosslinguistically highly uncommon. Besides that, English modals are auxiliaries subject to head movement, a much more frequent phenomenon which also contributes to the opaqueness of INFL-complex in English. At this point, comparative evidence is not unwelcome because it may make an analysis such as (12) appear less unmotivated. Malay is a language with multiple modal structures whose components have traditionally been analyzed as free forms or “indices” (cf. Gonda 1952: 37, among others) which we may perhaps interpret as free form instantiations of modal or mood projections occurring in base order at the surface, because they seem to be no auxiliaries in the familiar sense, i.e., light verbs which move to check (non-modal) features in particular positions. Malay is similar to ‘English-type languages’, however, in that its epistemic and root modals show essentially the same formal correspondences. Among the Malay modals we find the following.

\[
\begin{array}{ll}
\text{(13) Epistemic:} & \text{Root:} \\
boleh jadi \ 'may' & boleh \ 'may/can' \\
mesti \ 'must' & mesti \ 'must' \\
\end{array}
\]

In (14)a. the weak epistemic modal (boleh jadi) precedes tidak, primarily the marker of neutral, i.e. non-focussed sentence negation in Malay (Cf. Schindler 1998 for discussion). The inverse order in (14)b. is ungrammatical.

\[
\begin{array}{ll}
(14) & a. \ \text{Dia boleh jadi tidak suka saya} \\
 & \ \text{he may NEG like me} \\
 b. \ *\text{Tidak boleh jadi dia suka saya} \\
 & \ \text{NEG may he like me} \\
 & \ \text{‘He may not like me’}
\end{array}
\]

The strong epistemic modal (mesti) shows the same pattern.

\[
\begin{array}{ll}
(15) & a. \ \text{Dia mesti tidak belajar} \\
 & \ \text{he must NEG study} \\
 b. \ *\text{Tidak mesti dia belajar} \\
 & \ \text{NEG must he study} \\
 & \ \text{‘He must not study’} \\
 & \ \text{(I am certain that he does not study)}
\end{array}
\]

5 All Malay examples are taken from Idris (1980).
Nonepistemic modals generally require the opposite order:

(16) Dia tidak *boleh belajar
He NEG may study

a. ‘He may not study (is not permitted to study)’
b. ‘He cannot study (is not able to study)’

The evidence of modal adverbs supports the analysis in (12) insofar as the relationship between epistemic modality and negation is concerned. It has been known for a long time that modal or evidential adverbs like *probably, supposedly, evidently, obviously*, etc., which we may assume to be specifiers of $M_{ep}P$, as suggested in Cinque 1999: 85 f., cannot occur in the scope of negation (Jackendoff 1972).

(17) a. John probably never ran so fast
   b. *Never did John probably run so fast

Furthermore, as McDowell (1987: 278 f.) points out, ‘dislocated’ modal adverbs in right-peripheral position lack pitch accents and are preceded by a pause, which forces the closure of the clausal constituent and enables the adverb to have scope over it.

(18) John never ran so fast, *probably / obviously

Certain other modal adverbs, in particular necessarly, do occur in the scope of negation, but lack epistemic interpretations.

The same scopal asymmetry also shows up in another well-known restriction: Modal/evidential adverbs neither have morphological nor constituent negation (Bellert 1977: 346):

---

6 Chinese is similar with respect to order. Whereas in English the interpretation of a modal has an unambiguous interpretation only when it follows a preverb such as *never* (cf. (9) above), the order of such elements seems to be more rigid in Chinese. Chinese has a future marker *hui* ‘will’, base generated in INFL according to Huang (1982), which is probably neither an affix nor an auxiliary, but rather a free form and therefore inert with respect to head movement. Just as the English future modal *will* or the future tense affix in Spanish and other languages, the marker *hui* also has an epistemic interpretation. The fact that epistemic *hui* occupies a higher base position than the future marker becomes clear if we observe that the two readings strictly correlate with the position before and after preverbs such as *dou*, which has the semantics of a distributor (Cf. Cheng 1991: 137 ff. for discussion).

(i) Tamen dou hui lai
They all FUT come
‘They will all come (in the future)’

(ii) Tamen hui dou lai
They EPMODall come
‘They will all come (It is possible that they will all come)’

7 Cf. Bing (1979) for an in-depth discussion of the prosodic properties of dislocated sentence adverbs.
It is interesting to note that modal/ evidential adjectives are free from this constraint when they occur in predicative (or attributive) function.

In keeping with what was said above, adverbs like *necessarily* show prefix negation (cf. (21)), as well as constituent negation.

The fact that the epistemic interpretation only occurs with sentence adverbs supports the conclusion that epistemic M must be linked to a higher level of the extended projection of V. Since modal adverbs in some languages, including French, can be adjoined to a level of clausal structure headed by the complementizer, as shown in (22), we may assume that the higher clausal level from which epistemic modality operates is in fact CP. (Examples from Rochette 1990: 388f.). When *que* is missing, complex inversion is possible, as Dubuisson and Goldsmith (1976) observe. This provides additional support for the association of epistemic modality with COMP.

This again suggests that epistemic modality is not part of the proposition but takes the entire proposition into its scope, as we suggested at the beginning.

### 2.2 Nonassertive Contexts

Jackendoff (1972: 100ff.) was perhaps one of the first to notice that epistemic modals do not occur in polarity questions (cf. also Coates 1983, Leech (1971: 62;72) and Hermerén 1978: 939) for similar observations, but cf. Brewer 1987: 67ff. for an opposing view). Compare (23) a. and b.
(23)  a.  \[ \text{John must leave early tonight (epistemic/deontic)} \]

\[
\begin{align*}
\text{must} \\
\text{may}
\end{align*}
\]

b.  \[
\begin{align*}
\text{Must John leave early tonight?} \\
\text{May}
\end{align*}
\]

\[ (*\text{epistemic/deontic}) \]

McDowell (1987) has shown that the same is true for WH-questions.

(24)  \[
\begin{align*}
\text{Why must John leave early tonight?} \\
\text{may}
\end{align*}
\]

\[ (*\text{epistemic/deontic}) \]

(25)  \[
\begin{align*}
\text{Who must leave early tonight?} \\
\text{may}
\end{align*}
\]

\[ (*\text{epistemic/deontic}) \]

The epistemic reading is also not available when the modal occurs in an indirect question.

(26)  \[
\begin{align*}
\text{I wonder why John must leave early tonight.} \\
\text{whether} \\
\text{if}
\end{align*}
\]

\[ (*\text{epistemic/deontic}) \]

Furthermore, (27) (cf. Progovac 1988) shows that epistemic modals are also excluded from the protasis of a conditional:

(27)  \[
\begin{align*}
\text{If John must leave early tonight, ...} \\
\text{may}
\end{align*}
\]

\[ (*\text{epistemic/deontic}) \]

As we would expect, all the above restrictions affect not only modal auxiliaries, but also the adverbial exponents of epistemic modality. Contexts in which a modal auxiliaries cannot have an epistemic reading also exclude all those sentence adverbs which are exclusively epistemic or evidential, such as probably, possibly, obviously, evidently etc. (cf. Jackendoff 1972). Since not only interrogatives but also conditionals are involved, the relevant restriction, seems to be that epistemic are generally absent from nonassertive sentences. Exclamatives, on the other hand, are not affected since both wh- and inverted exclamatives (as well as the other types) presuppose the truth of their propositional complements.\[8\] This suggests that the relevant category is nonassertive, rather than nondeclarative.\[9\]

(28)  Must he be rich!

8 This was pointed out to me by Katrin Axel, University of Tübingen.
9 The facts are somewhat less clear in the case of English tag-questions. McDowell (1987:267) claims that (i) has no epistemic reading, in contrast to Palmer 1986, who claims that such an interpretation is at least marginally possible.

(i) John must leave early, mustn’t he?

The crucial part is, of course, not the modal in the declarative host clause, but the one in the tag. Varying reactions to such sentences seem to depend on the extent to which speakers are inclined to interpret the tag as a more or less frozen idiomatic form which simply conveys a request for confirmation of a proposition held to be true rather than a full-blown polarity question with an elliptical predicate.
The point made above, however, has been contested by a number of authors. Brennan (1993:69) and more recently Papafragou (1998, 1998a) have pointed out that epistemic modals are at least marginally acceptable in deliberative, self-addressed questions.

(29) ⑦Must Alfred have cancer?

The same is true for the German hearsay modal sollen in (30).

(30) Soll sie eine gute Sängerin sein? (³hearsay/deontic)

Whatever the exact status of such examples may be, I do not believe that deliberative questions of this type are sufficient to refute the claim made here that epistemic modals and other evidentials (including, e.g., German or Dutch hearsay modals) are in principle excluded from sentence types with nonassertive force. It should be clear that such a sentence as (29), which is marginally at best, can never be a request for the truth value of a proposition like John must have cancer, since epistemic must, as argued in § 1. above, is not part of the propositional representation but necessarily external to the proposition it accompanies. In fact a deliberative question is not a question in the relevant sense at all. When the doctor in Papafragou’s scenario utters (29) while running through John’s medical files, he does not ask a question in a linguistically relevant sense but wonders whether the evidence is strong enough to draw a conclusion. The interrogative form of the sentence articulates an uncertainty but it does not indicate a question. On the contrary, the observation that (29) cannot be uttered as a felicitous question and, insofar as it is possible at all, calls for an indirect interpretation not mediated by linguistic structure alone underlines the fact that epistemic modals (and other evidentials) cannot occur in nonassertive contexts.

The fact that epistemic modals and other evidential markers appear in INFL but seem to operate from a position in COMP has motivated some researchers, most notably McDowell (1987), to propose that epistemic modals (and perhaps truth conditional operators in general) must be in INFL at PF but undergo obligatory raising from INFL to COMP at LF. Rochette (1988: 277) proposes a similar account (cf. also Roberts 1983). McDowell develops the most explicit proposal of its type. Her account actually rests on two assumptions. Besides LF movement of truth conditional operators she assumes the COMP-indexing algorithm argued for in Aoun, Hornstein and Sportiche (1981):

---

10 Cinque (1999:86) shows that some of Brennan’s examples have nonepistemic modals and are therefore irrelevant. This does not apply, however, to Papafragou’s evidence.

11 This is similar to the interpretation of questions in free indirect discourse, a use of linguistic form which is taken to be a direct representation of the stream of consciousness of a third party. Cf. Banfield (1982) for insightful discussion.
This assigns to a sentence like (23) a. the LF representation in (32).

\[(32) \quad [CP [COMP, must_i] John [INFL t_i [VP leave early tonight]]]\]

In (32), \textit{must} has moved to COMP and indexes it through the COMP-indexing algorithm, which allows \textit{must} to antecedent govern its trace in INFL. In (24), on the other hand, COMP is indexed by the WH-operator \textit{why} at PF. When the epistemic modal raises at LF, it is contraindexed with the WH-operator (and therefore with COMP) and cannot antecedent govern its trace in INFL, which accounts for the ill-formedness of (24).

\[(33) \quad [CP [COMP, why_i, must_j [IF John [INFL t_j [VP know the answer t_j]]]]]\]

In cases like (23)b. or (27) we may assume that an empty operator moves to COMP and triggers COMP-indexing. In (29) and (30), the effect of contra-indexation is somehow suspended due to some sort of interpretive procedure. Contra-indexation alone, however, does not account for the data presented above because WH-movement blocks the epistemic reading only if the sentence type it derives is incompatible with the sentence type requirements of epistemic (or evidential) modality, as (28) shows. McDowell does not consider this type of data. Furthermore, (28) shows that the epistemic (or evidential) modal can in fact surface in COMP at PF, a movement perhaps triggered by an empty element in the exclamative COMP position (as in other types of head movement to \textit{C°}). Some authors, in particular Ormazabal (1995) have argued that empty functional heads are always affixal and trigger head movement for support. In this case all we have to assume is that in non assertive sentences \textit{C°} carries a sentence type feature which clashes with an inherent feature of raised evidential modal. Since deontic and other nonepistemic modals do not clash with any sentence type feature in COMP, we may conclude that they do not carry an inherent feature of this kind.

An alternative account (considered, e.g., in Rochette 1988) would depart from the observation that COMP and INFL behave as discontinuous elements. Well-known cases of this kind are selectional constraints between COMP and INFL, such as the obligatory presence of the finite declarative complementizer in the English mandative subjunctive in (34) which has an empty modal (equivalent to \textit{should}) according to Roberts (1987) or the selection of a future-orientated or ‘unactualized’ INFL-element (Control \textit{to}) by the complementizer \textit{for} in Control-complements such as (35) (cf. Martin 1996).

\[(34) \quad \text{John insisted *(that) he SUBJ be reelected}\]
Rochette proposes that modals in Romance languages (including the epistemic variety) arise in VP and raise to INFL in order to make their semantic content accessible to the INFL/COMP discontinuity. Under this approach, we must assume that epistemic modals are lexically marked as positive plurality items of some sort, which must move to a position external to the scope of negation. Picallo (1985), on the other hand, observes that in Catalan epistemic modals are situated above negation, whereas nonepistemic modals are below it. Since in Catalan only epistemic modals have the properties that characterize English modals in general (i.e. they are excluded from all infinite positions and lack infinitive and gerund morphology) she assumes that both epistemic and nonepistemic models are raising predicates. Furthermore she proposes that Catalan epistemic modals are associated with a sentential operator in COMP and base-generated in INFL (above PolP), while all other modals originate in verbal positions. Something similar might be appropriate for German and Dutch. Of course we are still faced with the question of why certain modals, which obviously form a natural class, take scope over negation and why such modals participate in the INFL-COMP discontinuity. The answer must be sought in the semantic content of epistemic/evidential modals, which forces them to operate from an extrapositional position at least at LF. Under an approach combining base-generation in INFL with postsyntactic movement of COMP an explanatory account of epistemic modality can be envisaged: base-generation in a layer of INFL above PolP is necessary in order to avoid an intervention effect caused by negation. Nonepistemic modals are below PolP because they are part of the proposition. All this suggests the following conclusion: epistemic modals belong to a class of operators which can be characterized as split (or discontinuous) signs. The defining property of ‘split signs’ is their capacity to merge their PF-and LF-interpretable properties in distinct syntactic positions. Under an alternative representational account of the PF/LF-discontinuity of INFL and COMP, the PF- and LF-interpretation sites would be linked without movement. The essential point is the disassociation of the locus of phonetic realization and the place of semantic delivery. The LF location of epistemic modality fixes its scope and at the same time determines its incompatibility with operators with nonassertive force. Since the scopally

12 The idea that epistemic modals, which take epistemic/evidential adverbs as their specifiers (Cinque 1999), participate in the INFL-COMP continuity through the linking of two head positions might also turn out to be advantageous for an account of adverb placement, which must explain why adverbs of Jackendoff’s class III (comprising modals and evidentials; cf. Rochette 1990) can occur both in sentence peripheral and in ‘AUX-positions’.

13 I borrow the term ‘split sign’ from Cormack and Smith (1997), who use it in a somewhat different, but related sense.
relevant location of epistemic modality is in COMP, it has scope over the entire proposition (including the subject), although its PF properties are located within the sentence. A further consequence of the COMP-relatedness of epistemic modals is their relationship to quantifier scope. As already observed in Leech (1971: 73), quantifiers seem to be unable to take scope over such modals.

(36) Every clause must contain a finite verb.

The deontic reading of (36) is compatible with wide and narrow quantifier scope, whereas the epistemic modal always can only have wide scope. The placement of modal adverbs seems to be in agreement with this analysis of epistemic modality, which is further supported by the occurrence of certain specificity effects.15

2.3. Tense

Now that we have established that the INFL part of an epistemic modal has scope over sentence negation we need additional evidence to back our claim that it occupies a high position within the INFL range. An important aspect of the grammar of epistemic modals is their relationship to tense. Cinque (1999) postulates a universal functional hierarchy in which epistemic modality and tense are ordered as given in (37) (Cinque 1999: 72).

(37) Mod<sub>ed</sub> > T (Past)

Cinque’s evidence for this claim derives from a typological survey of auxiliary order and the ranking of suffixation in various languages. Some languages, including Guyanese Creole, are said to have free froms representing tenses. They appear to invariably follow other free forms representing epistemic modality in sentences where both occur.

14 A similar point was made earlier with respect to negation (cf. Drubig 1994). It is well known that ‘assertive’ or neutral negation is a predicate operator which does not take scope over the subject and cannot license negative polarity items in this position, unless a negative quantifier or an auxiliary affixed with ‘contracted negation’ moves to COMP. In contrast to this, contrastive negation, which associates with focus, takes scope over the entire proposition and may bind a focussed subject, inspite of the fact that it is not within its c-command domain at PF. In effect, Drubig (1994) proposes that contrastive negation is a ‘split sign’, similar to epistemic modality, which is pronounced in INFL but interpreted in COMP at LF.

15 The widely discussed obviative effect that can be observed in the subject position of subjunctive complements in Romance languages (cf. Kempchinsky 1986 for discussion and references) has also been claimed to provide evidence for the COMP-related nature of epistemic modality (Picallo 1985, Rochette 1988). The obviative phenomenon (or Subject Disjoint Reference Effect, following Kempchinsky 1986) is characterized by the obligatory disjoint reference of subjunctive subjects in relation to their matrix subjects.

(i) Je voudrais qu’il finisse son article
   ‘I would like that he finish (SUBJ) his article’

(ii) Il voudrait qu’il finisse son article
   ‘He would like that he finish (SUBJ) his article’

Picallo (1985: 260ff.) points out that in Catalan the effect disappears when the subjunctive complement contains an epistemic modal. According to Picallo this is due to the fact that epistemic modals are associated with a
Other languages have both tense and modality-marking (epistemic) suffixes. In keeping with the Mirror Principle, the tense suffix is closer to the lexical head, which is the expected order.

(39) Garo (Sino-Tibetan)
    Anti-ci re’an- aha- kon
    market-to go- PAST PROBABLE
    ‘He probably went to the market’

Cinque establishes this order on the basis of a considerable amount of comparative evidence. Since his framework is universalistic, the projected hierarchy is also claimed to be instantiated in languages which do not show the patterns represented by (38) and (39) and which may involve further complications, as is the case in English, where all modal auxiliaries occur in INFL. This invites the question of whether we can find language-specific evidence in English which confirms the ranking in (37). I will now make an attempt to answer this question in the affirmative. In order to do this I will inspect a somewhat neglected aspect of the English auxiliary complex, the so-called Past Tense Replacement (PTR) effect. The PTR effect is a phenomenon that has rarely been mentioned and never been adequately accounted for in the literature on English tenses. A first explicit discussion occurs in Hofmann (1976), where it is shown that the well known restriction on the use of the Present Perfect in English has a systematic exception. English, unlike Romance languages, continental Germanic and most Scandinavian languages, has a Present Perfect which cannot be used with reference to a specific time in the past, as a quasi-substitute of the past tense morpheme.

(40) a. He left last Tuesday
    b.  *He has left last Tuesday
    c. He has already left

In languages where PTR is generally available, the Perfect as PTR form is often restricted to spoken or colloquial varieties. Besides that, its acceptability often varies across dialects. Compare the following examples from German.

---

sentential operator in COMP, which creates an ‘opaque domain’ in which the subject may corefer freely. Cf. Rochette (1988) for further discussion.
(41)  a. Sie ging gestern nach hause
     'She went home yesterday'
     b. Sie ist gestern nach hause gegangen
     'She went home yesterday'
     c. Sie ist schon nach hause gegangen
     'She has already gone home'

While PTR is unacceptable in all finite contexts in Standard English, it is systematically available in the following nonfinite contexts first described by Hofmann (1976):

1.) **Epistemic modals:**
(42)  a. He may have left last Tuesday
     b. He may have already left

2.) **ECM infinitives (object-raising):**
(43)  a. I believe John to have left last Tuesday
     b. I believe John to have already left

3.) **Subject-raising infinitives:**
(44)  a. John appears to have left last Tuesday
     b. John appears to have already left

4.) **Gerunds:**
(45)  a. John's having left last Tuesday ...
     b. John's having already left ...

Within the theoretical framework on which Hofmann's (1976) discussion (which originally appeared in 1966) is based, a possible account of PTR could be based on the assumption that the a. examples in (42) - (45) have past tense morphemes in underlying form. A transformational rule then eliminates the past tense morpheme and derives the Perfect, as shown in (46).

(46) Hofmann (1976): **Past Tense Replacement Rule**
    \[ X \text{ ed (Perf)} \rightarrow 1 \text{ Perf } \text{Ø} 4 \]

The type of construction specific rule on which Hofmann's account relies is not available in a principle-based theoretical framework, but there can be no doubt that the proposal is observationally adequate. This becomes clear when we observe that there is independent evidence that nonfinite perfect constructions do indeed have different interpretations when they occur with adverbs that are otherwise restricted and can only occur either with the Past or the Perfect, as shown in (42) - (45). Evidence for the PTR effect can be found in the different
interpretations that occur in subordinate tenses. As is well known, a sentence like (47) is ambiguous and can be understood as reporting two different speech events: John may either have said 'I am tired' (the Sequence-of-Tense (SOT) reading), in which case the embedded past tense must be interpreted as the 'past-harmonized' version of an embedded Present, or he may have said 'I was tired', which corresponds to the so-called 'past-shifted' reading.

(47) John said last night that he was tired (I am tired/I was tired)

The SOT reading is absent when the verb in the complement of the speech act report is an event-denoting predicate, as in (48), since such predicates do not occur in the present tense, unless the asterisked example (48) has a generic interpretation.

(48) John said last night that he          *I leave the house two hours earlier/
                   left the house two hours earlier       I left the house two hours earlier

Since the Present Perfect is a 'retrospective present' rather than a past tense, it cannot induce an SOT effect.

(49) John has already said that he was tired *I am tired/I was tired

Evidence for the PTR effect can be seen in the following sentences (cf. Baker 1989: 456):

(50) John must have said last night that he was tired (I am tired/I was tired)
(51) John must have already said that he was tired (*I am tired/I was tired)

As already shown in (42) - (45) the epistemic modal in (50) and (51) creates a context in which the PTR effect occurs. While PTR in English is restricted and occurs only in the nonfinite constructions listed in (42) - (45), the other languages mentioned above also exhibit PTR effects in finite clauses.

(52) Hans muss\{ a. gestern \} gesagt haben, dass er krank war (Ich bin krank/Ich war krank)
       \(\text{b. jetzt gerade}\) (*Ich bin krank/Ich war krank)

(53) Hans hat \{ a. gestern \} gesagt, dass er krank war (Ich bin krank/Ich war krank)
       \(\text{b. jetzt gerade}\) (*Ich bin krank/Ich war krank)

We can observe the following pattern:

(54) Past Tense Replacement effects:
The Perfect construction may have a Past reading
a) in nonfinite contexts: English, Danish
b) in nonfinite and finite contexts: Continental Germanic, Romance
PTR effects support the conclusion that the complement of an epistemic modal is a tensed proposition. This raises the question under what conditions PTR effects can occur and how we can account for them. In the following I would like to sketch an approach which rests on four assumptions.

1. Tense and aspect are relational categories that can be accounted for within a post-Reichenbachian framework of the kind proposed by Stowell (1995, 1996) and more recently extended to aspectual categories by Demirdache and Uribe-Etxebarria (1998), which strictly correlates semantic composition with morphosyntactic form;

2. The English present tense neither dissociates event time and evaluation (or 'reference') time nor binds the event argument of a verb, i.e., it is semantically vacuous (cf. Enç 1996);

3. English futurate will is not a tense but a (nonepistemic) modal which, like other modals, involves quantification over possible worlds (in the case of will: those that are consistent with predictions) and - unlike epistemic modals - shifts evaluation time to the future.16

4. Auxiliaries are 'light verbs' which, like full verbs, are associated with a temporal argument, but, unlike full verbs, lack lexical content (including a theta-grid of participant roles); light verbs can also be 'expletive', i.e. semantically inert, in which case they also lack an event variable and have no effect on temporal relations.

1.-3. are now widely assumed in the literature. 4. was first suggested in Drubig (1998), but a similar proposal was independently made in Brugger (1997). Following Stowell (1995, 1996) I assume that 'tense' is articulated in terms of two distinct functional heads, one (T^0) expressing a temporal relation (AFTER in the case of past tense), the other (Z^0) representing the morphological tense features ϕ realized on the verb.

---

16 Enç (1996:354) suggests that root modals are sentence operators and defines their interpretation as follows:

(i) MODAL[S] is true at w;i iff in every world w' accessible to w there is an interval i' such that i < i' and S is true at w';i'

In (i), i is utterance time, which functions as the primary time of evaluation, whereas i' is a future time with respect to which the modal sentence is evaluated. Cf. also Barbiers (1995) for an extensive discussion of the 'polarity-transition property of root modals'. Modals with the property defined in (i) license the free choice quantifier any (Carlson 1981:10). Since free choice any is licensed if and only if its licenser is in the immediate scope of any, epistemic modals cannot be such licensors since they do not occur in the scope of quantifiers. In addition, of course, they also lack future orientation (alias 'the polarity transition property')
The external argument of $T^0$ in (55) is a variable (PRO), which is identified either by the context of utterance in the case of matrix clauses and in double access contexts or by the event time of the embedding verbs in the case of temporal subordination (SOT and Past Shift). The internal argument ZP (the equivalent of DP in the nominal domain) provides for an operator binding the event variable of the lexical verb to a referring expression denoting the event time interval. According to Demirdache and Uribe-Etxebarria (1998), the aspectual category (traditionally AspP) is essentially the same configuration embedded into (55), hence AspP is a TP embedded into another TP (cf. also Giorgi and Pianesi 1997).

Let us now address ourselves to the problem of PTR and its representation in a syntactic theory of tense. As we have seen, the PTR reading of a sentence like (56) implies that a past tense must be embedded under the projection headed by the epistemic modal. We may assume that this implies that TP is generated below $M_{ep}P$. But then how do we have to analyse the syntactic form of a sentence like (56), which allows both a perfect and a PTR reading?

(56) John must have slept

Let us first focus on the embedded perfect reading. An epistemic modal, as we have already seen, does not shift evaluation time away from utterance time, in contrast to deontic must and all other nonepistemic modals. As Enç (1996: 354) observes, it is reasonable to assume either that the tense node is empty in such a case, or that there is no TP at all, since the temporal operation in the case of the present is vacuous. In the absence of evidence to the contrary I will decide without argument in favor of the latter alternative and represent the perfect reading of (56) as in (57) a. ZP is neglected here.
In the absence of an effective TP, the configuration R, S in (57) is the default choice. As a result of the presence of a non-inert auxiliary verb and an embedded past tense TP, which represents the temporal semantics of the past participle affix, the representation compositionally derives the configuration in (57) b., the modal perfect reading of (56).

Now let us assume that *have* is an expletive auxiliary, similar in function to the copula. Such an auxiliary meets the selectional properties of the epistemic modal, which requires a verbal complement, but it fails to introduce an additional event argument which could act as a reference time argument for the TP below it. In this case, the external argument of the TP is identified by the context of utterance. The external argument of TP is then S, the primary reference time, hence the configuration determined by (58) a. is b., a representation of the past tense. We now have accounted for the PTR effects.
(58) a. $M_{ep}P$

John $M_{ep}'$

$M_{ep}$ VP

$V'$

have S PRO T' E ______ S

TV P

AFTER

slept

E _____ S

b. E _____ S

The approach to PTR exemplified in (58) analyzes the auxiliary as a semantically empty and referentially inert prop verb which serves the sole function of satisfying the categorial selection features of the epistemic modal. The relational meaning of the PTR construction is provided by the TP which is headed by the affix of the participial component. The crucial factor in the derivation of a PTR effect is the referential inertia of the expletive auxiliary, which in effect eliminates one of the two constitutive relations of the Present Perfect, thereby reducing it to a relation of anteriority (S ____R), which is equivalent to a simple past. Alternatively, we may say that 'PTR' is the only possible representation of tense in the complement of an epistemic modal, which, like other modals, selects [+V -N]. This suggests that epistemic modals have scope over tense. Since scopal order is reflected in the functional hierarchy, the ranking must be as follows:

(59) $M_{ep}P$

PolP

TP

The use of a 'Perfect' construction with an expletive auxiliary can now be seen as a last resort necessitated by the language-specific fact that the marker of epistemic modality in English developed from a (root) modal auxiliary selecting bare infinitives. Whenever the configuration in (59) arises, an expletive auxiliary must merge between $M_{ep}$ and TP (past
tense). In English, this device is restricted to epistemic modals and the other nonfinite constructions listed in (42) - (45), all of which select verbal stems. In English, PTR occurs only in conjunction with this selectional requirement. In the other languages referred to in (54), the PTR phenomenon has been extended to finite (i.e., present tense) forms thereby loosing its last resort flavor.

The fact that epistemic modals lack the future-shift of evaluation time characteristic of nonepistemic modals suggests a closer relationship between epistemic modality and the present tense. Whereas the event-time denotation (E) in a past tense sentence is prior to evaluation time (which is S in matrix clauses), the event-time denotation in a present tense sentence is evaluation time. Following Enç (1996) we may conclude that present tense, which does not dissociate event and evaluation time and in this sense contrasts both with the past tense and the futurate modal will, is not a tense at all (Cf. also Stowell 1996:284). In fact, there is more that supports such a conclusion. Enç observes that in English nonstative (event-denoting) predicates do not occur with an episodic (specific) reading in the simple present. Such predicates do occur with a generic/habitual reading or when accompanied by a quantificational or frequency adverbial. Stative predicates are unconstrained. This contrasts with the past tense, where no such restrictions can be observed.

(60) a. John walks habitually/every day/*right now
   b. John walked habitually/every day/yesterday

Following Enç we may assume that event-denoting predicates have an event argument, a variable which must be bound by a temporal operator. It is characteristic of tense that it not only expresses a relation ( anteriority/posteriority) between evaluation time and event time denotation, it also is associated with an operator which binds the event time argument of a nonstative predicate. Since the 'present tense' lacks these characteristics, it cannot be a tense. This conclusion gains added attractiveness when we look at the morphosyntactic evidence: verb forms with present time reference in English only mark agreement, forms with

---

17 Stative predicates lack an event argument, which would explain why they are insensitive to the restrictions shown in (60). Chierchia (1995), however, proposes that stative predicates have event variables that are lexically bound. This seems to allow the explanation of certain phenomena which remain unaccounted for under Kratzer's (1995) proposal. Event-denoting predicates must be in the syntactic domain of a temporal binder which we assume to be present in SpecZP (cf. (55) above). ZP is selected by TP. Alternatively, the event variable can be bound by a quantificational or frequency adverbial, as shown in (60), or its invisible counterpart GEN, a 'default' generic operator (cf. Chierchia (1995) for references).

18 Stowell's (1995, 1996) insight that traditional TP must be split into two component projections, one headed by the temporal ordinary predicate, the other hosting the temporal operator in its specifier (cf. (57) above), which was originally motivated by the need to account for the crosslinguistic variation of SOT, also explains the two-sided nature of tenses in Enç's argumentation: since TP exclusively selects ZP, the temporal ordering and binding properties of tense are either simultaneously present or simultaneously absent.
past time reference only mark (past) tense. This motivates the conclusion that the inflectional affix marking the English 'present tense' actually expresses subject agreement, which must be checked in AgrS$^0$, whereas the past tense affix corresponds to T$^0$. Once we have taken this step, the parallelism between epistemic modals and simple present tense forms becomes obvious: like the simple 'present tense', epistemic modals do not affect temporal ordering. As Steedman (1977) first pointed out, they show the same restrictions as the simple 'present': event-denoting predicates lack episodic readings. Aspectual constructions (Perfect and Progressive) pattern with stative predicates, as is well known, and as such are unconstrained both in the 'simple present' and under epistemic modals. The Perfect construction exhibits the only difference between the two: in English at least, the PTR effect occurs in the scope of epistemic modals, but not in the 'simple present'. Since this effect can be derived as shown above, we may draw the following conclusion: epistemic modality and AgrS$^0$ are alternative instantiations of a category which projects either AgrSP or M$_{ep}P$, as shown in (61).

\[
\begin{array}{c}
\text{AgrS/M$_{ep}P$} \\
\text{PolP} \\
\text{TP}
\end{array}
\]

The disjoint category in (61) could be circumscribed by the term 'truth modality'. Following Klein (1991/1992) I assume that finiteness (i.e., $\phi$-features) represents the assertoric force of a statement. Epistemic modals characterize 'qualified' assertions (or 'apodictic judgements' in Frege's parlance; cf. Westmoreland 1998), which are qualified in a sense: they indicate that the evidential source (DEDUCTION, or merely ASSUMPTION) does not warrant an 'unqualified' statement.

We now have accounted for PTR effects occuring in the context of epistemic modality. We have demonstrated that it can be derived from the following two assumptions:

1. Epistemic modals, which select verbal stems, may occur in tenseless matrix finite sentences (in the sense of Enç 1996: 354).

---

19 Cf. the discussion in Kayne (2000).
20 Cf. McDowell (1987) for more discussion and references.
21 Cf. Sola (1996) for an analysis in which M$_{ep}P$ and AgrSP are alternative choices of the same category.
22 This is the background of what is sometimes referred to as 'Karttunen's effect' (Horn 1972). Karttunen (1972) notes that $\Phi$ is taken to be stronger than MUST $\Phi$ inspite of the assumed implicational relationship. Cf. also Palmer (1986: 86ff.)
2. When an epistemic modal occurs in a tensed sentence, the TP hosting the (past) tense morpheme is embedded in the projection headed by the epistemic modal as shown in (61). In the latter case PTR is necessary to satisfy the selectional requirement of the epistemic modal.

Nonepistemic modal auxiliaries on the other hand, can occur only in tensed sentences. I will assume that such modals are marked [+tense], which could mean that they arise in T⁰ and (in English) obligatorily move through Pol⁰ to AgrS⁰. This would explain their surface position relative to the exponents of PolP, as well as the affixation of the 'contracted' form of negation to nonepistemic modals. Since in English all modals compete for the same PF position, they are mutually exclusive. Unlike their epistemic relatives, nonepistemic modals do not select against event-denoting predicates. Since such predicates have event variables that need to be bound, we must assume that a nonepistemic modal forms a syntactic domain in which the variable of an event-denoting predicate is saturated: whereas the (past) tense binds the temporal argument and fixes the time (in the past) at which the event denoted by the VP occurs, a nonepistemic modal, under the Enç's (1996) proposal, triggers a default mechanism which identifies the denotation of the temporal argument with evaluation time. This is the same mechanism we saw at work in the case of epistemic modals and agreement, their epistemically 'unmarked' counterpart, nonepistemic modals shift evaluation time to future intervals in the possible worlds over which they quantify. This is the semantic content for which the syntactic feature [+tense] is intended to indicate. Syntactically this implies that nonepistemic modals project a functional category MP which is an alternative to TP, just as M_epP is an alternative to AgrSP. Neglecting ZP, we can now represent the complete picture as follows:

(62) \[
\begin{array}{c}
M_{ep}/AgrSP \\
\quad PolP \\
\quad \quad M/TP
\end{array}
\]

Since English nonepistemic modals move to AgrS⁰, the simultaneous choice of M_{ep} and M in (62) leads to ungrammaticality.

---

23 The expression of multiple modality in English sentences requires periphrastic modals such as have to and others.

24 In the case of stative predicates, the event argument is either nonexistent or lexically bound.
We must now briefly turn to another type of construction, namely infinitives with the particle *to*, which at first sight appear to be unrelated to modal sentences. As it will turn out, the modal category and infinitival *to* are closely related and can be identified as instantiations of the schema in (62). We see the parallels when we look at the remaining constructions that may induce PTR effects (cf. (42) - (45)). Just as PTR effects in modal constructions seem to occur exclusively (or at least primarily) with epistemic modals, the infinitival constructions showing PTR also form a natural class: recent approaches reanalyse ECM constructions as Object Raising (OR), parallel and complementary to Subject Raising (SR). Hence, infinitival constructions associated with PTR can be jointly classified as Raising constructions, whereas PTR is entirely (or, perhaps, largely) absent from Control constructions. This is significant with respect to our concerns, because the Raising/Control contrast turns out to have a semantic correlate. Stowell (1982) observed that Control infinitives and Raising infinitives exhibit a difference in temporal interpretation that immediately reminds us of the epistemic/nonepistemic contrast: in (63), the event time of the infinitival predicate in the Control construction is 'unrealized' with respect to the event time of the matrix predicate, which lies in the past, whereas the OR construction (64) shows no such effect (cf. Ormazabal 1995 and especially Martin 1996 for extensive discussion of the temporal interpretation of infinitives).

(63) John promised to invite Mary
(64) Mary believed Bill to be at home

In (64) the event time denoted by the infinitival predicate coincides with the event time of the matrix predicate. As Martin (1996) shows, SR constructions are not as consistent as OR constructions and must be divided in two subclasses: sentences like (65), which traditionally are analyzed as SR constructions, show the shift to the future which is characteristic of Control constructions. Martin discusses a considerable amount of evidence which suggests that such infinitives must be analyzed as Control rather than Raising Constructions, which would explain their temporal properties. Sentences like (66), on the other hand, are true Raising constructions and pattern with OR constructions.

(65) The men are likely to leave
(66) John seems to be at home

25 I will set aside gerundive constructions, which do show PTR effects but are tangential to my topic.
As a rule, biclausal infinitival constructions\footnote{Unlike Raising and Control constructions, infinitival constructions formed with implicative predicates (such as \textit{try} or \textit{manage}) or aspectuals (like \textit{begin} or \textit{start}) do not compose distinct eventualities. Implicative and aspectual infinitival constructions denote phases or integral parts of complex events. This leaves no room for co-temporality or shifts to a future reference time. There are good reasons to assume that such constructions are not biclausal and hence cannot involve Control or Raising. They also do not pattern with the modal distinctions we are concerned with here (cf. Martin 1996: 55f.).} have finite paraphrases which make the difference in temporal difference visible: the finite paraphrases of Control constructions require a modal (usually \textit{will}/\textit{would}; in factive infinitives and indirect questions \textit{should}: cf. Martin 1996:54ff.), Raising constructions do not.

\begin{equation}
\begin{aligned}
(67) & \text{John promised that he would invite Mary} \\
(68) & \text{Mary believed that Bill was at home}
\end{aligned}
\end{equation}

The suggested parallel between infinitives and modal constructions goes even further: Like epistemic modals, Raising infinitives select stative predicates and allow embedded event-denoting predicates only with a generic or habitual reading, which we can account for in the same way as in the case of the epistemic modals and 'present tense' (alias subject agreement). Since Control-\textit{to} (\textit{to}_c) occurs with event-denoting predicates, we may assume that it saturates their temporal arguments in the same fashion as root modals. Martin (1996), similar to Mittwoch (1990:124), stipulates that every infinitival \textit{to} is equivocal to a modal auxiliary: \textit{to}_c is associated with the type of future-shift of evaluation time that also occurs with nonepistemic modals. In Control infinitives, the modal component is a feature of the lexical semantics of the governing verb. Something similar seems to apply to Raising constructions: According to Mittwoch (1990), Raising-\textit{to} (\textit{to}_R) “signal[s] only the nonentailment of the complement”. The governing verb, on the other side, either has an evidential meaning (\textit{seem}, \textit{appear}, etc.) or expresses a propositional attitude (\textit{believe}, \textit{expect}, etc.). In each case, the infinitival construction looks like the discontinuous counterpart of a modal auxiliary, which unites both components in a single head. Following Stowell (1982), Martin assumes that \textit{to}_c is [+tense] and \textit{to}_R [-tense], which is reminiscent of Picallo’s (1985) partitioning of modal auxiliaries. This corresponds to the schema in (62) above as follows: \textit{to}_R heads M_{op}P/AgrSP and, being [-tense], is temporally inert, like an epistemic modal;\footnote{Martin (1996: 62f.) presents an interesting pair of examples which he takes to be evidence against Enç’s (1996) claim that English has no finite present tense. (i) shows an important difference between \textit{to}_R and present tense/subject agreement: (i)a. is true iff Mary is pregnant at matrix event time, but (i)b. entails that she is still pregnant at the evaluative time of the matrix clause, which is utterance time. Thus, (i)a. is open to a double access reading, whereas b. forces it. This explains the contrast between (ii)a. and (ii)b., where the temporal adverbial entails a factual inconsistency.}

\begin{equation}
\begin{aligned}
(i) & \text{a. John proved Mary to be pregnant} \\
& \text{b. John proved that Mary is pregnant} \\
(ii) & \text{a. Two years ago, John proved Mary to be pregnant} \\
& \text{b. Two years ago, John proved that Mary is pregnant}
\end{aligned}
\end{equation}
evidential meaning of an epistemic modal (which is contributed by the governing verb), as well as \( \varphi \)-features; \( to_c \) heads \( M/TP \), and, being \( [+\text{tense}] \), manipulates evaluation time in the familiar way, but it lacks the circumstantial meaning of a root modal, which is contributed by the lexical content of the main verb; being a free form (not an auxiliary), it also lacks \( \varphi \)-features and therefore does not move to \( \text{Agr}S^0 \). Both \( to_R \) and \( to_c \) are not only syntactically, but also semantically dependent on their head governors, as we have seen.

Some of the predictions of this analysis seem to show that it is at least a move in the right direction: because \( to_c \) remains in situ, it is preceded by sentence negation. When negation follows \( to_c \), it can only be predicate negation (cf. Barrett 1998).

(69) a. I prefer not to go to church
b. I prefer to not go to church

In the case of \( to_R \), there are no strict correlations like this and negation both before and after \( to \) corresponds to sentence negation. An analysis according to (62), however, does not predict that the preferred placement of negation is in front of \( to_R \).

By way of summary, the main points made above can be represented as follows:

(70) Truth Modality vs. Control Modality

<table>
<thead>
<tr>
<th></th>
<th>( M_{\varphi} )</th>
<th>( M )</th>
<th>Raising Infinitive</th>
<th>Control Infinitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>future-shift</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>event-denoting predicates</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>PTR effect</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

According to Martin this contrast is in conflict with Enç’s (1996) account of the present tense, because the double access reading seems to be due to the presence of a present tense, and not just the absence of tense. Under the present approach to tense and truth modality, however, the double access effect is the combined effect of two factors, the absence of tense in keeping with Enç’s interpretation of the ‘present tense’ and the ‘sign of (unqualified) assertion’, which is rooted in the sentence’s finiteness features. Hence, the presence of the subject agreement in the b. examples indicates that the speaker guarantees the truth of the subordinate proposition at utterance time. The ‘empty tense’ marker \( to_R \), in Mittwoch’s (1990) sense, indicates the ‘nonentailment of the complement’, which suspends the double access effect in the a. examples. Interpreted in this way, Martin’s observation does not constitute a difficulty, but rather supports the approach defended here.

28 Barrett (1998:33ff.) argues at length against the idea that \( to \) is a \( T^0 \)-head, but fails to distinguish between \( to_R \) and \( to_c \), which is crucial, as Ormazabal (1995) and Martin (1996) have shown.

29 The central property of root modality is their capacity to displace evaluation time. The two other features in (70) can be derived from this property.
The syntactic structure to which (70) corresponds is (62). Due to head movement, the order of elements does not always reflect hierarchical relations. Again we can quote the evidence of languages which exhibit a greater amount of transparency. Compare the following sentences from Malay, which show the morphemes in the expected order.

(71) Dia telah tinggal di sana tahun lalu
He PERF live at there year last
(‘He lived/had lived/had been living there last year’)

(72) Dia mungkin telah menjual keretanya sedang akan
He may PERF sell car + POSS be selling
(‘He may have sold his car’)

(73) Dia telah boleh membaca
He PERF can read
(‘He was able to read.’)

In (71) we can see that the tense/aspect marker in Malay precedes the verbal stem. Since this marker is a free form, not an inflectional affix, it is not ‘lowered’ onto the verb. In (72) the marker of epistemic modality preceeds the tense/aspect marker: surface order reflects the scopal hierarchy. In (73) boleh expresses dynamic modality. It must follow the tense/aspect marker, which has scope over it.

3. Focus and Information Structure

3.1 Introduction

In the following section I inspect a number of phenomena related to information structure and try to show that the syntactic approach sketched in the preceeding sections can account for these phenomena when combined with an approach to information structure which I have presented elsewhere in more detail (cf. Drubig (1992, 1994, 1998a, forthcoming); Drubig and Schaffar (forthcoming)). Following Drubig (1994), Kiss (1998) and others I distinguish between presentational focus, which highlights new or context-implementing information,
and contrastive focus, which identifies a focussed item within a contextually given set by eliminating its alternatives.

My first topic will be presentational focus. In 3.2 I will discuss Verb Phrase Ellipsis (VPE) as a form of givenness marking and compare it with related constructions. Finally, there will be a few observations on contrastive focus and modality in § 3.3.

3.2 Presentational Focus and Givenness Marking

In § 2.3 I presented evidence supporting an inherent relationship between modals and infinitives: Control modals correspond to nonepistemic modals; Raising infinitives, if redefined in the way suggested by Martin (1996), correspond to epistemic modals. Since both groups are natural classes – the former is [+tense], the latter [-tense] – we expect their members to behave alike with respect to VPE and similar givenness constructions. The following discussion has profited considerably from the discussion of epistemic modals in McDowell (1987) and the recent treatment of VPE in López and Winkler (2000). McDowell (1987: 327ff.) presents a considerable range of data and comes to a conclusion which is similar to the position adopted here. As McDowell shows, VPE, as well as Pseudogapping, its ‘incomplete’ counterpart, exclude epistemic readings.

(74) John must wash his car every day and Peter must e too (*epistemic/deontic)
(75) John will often sit there and do nothing and Bill will e too (*epistemic/future)
(76) John may not obey his mother, but he must e his father (??epistemic/deontic)
(77) She certainly didn’t support the President but she may e the Governor
     (??epistemic/deontic)

As originally observed in Drubig (1998: 20), German lacks VPE, but has a proform es, which induces the same readings as English VPE (Cf. the more comprehensive discussion of German predicate anaphora in López and Winkler 2000).

(78) Hans \{ muss \} diese Aufgabe lösen und Petra \{ muss \} es auch
     \{ wird \} \{ wird \}

Another construction which shows the effect is VP Topicalization (VPT).

---

30 A more appropriate term would be ‘predicate ellipsis’, but I will continue to use the conventional label.
31 Interestingly, this does not seem to apply to the English propredicate so in sentences like (i), which do have an epistemic reading. This was pointed out to me by Anthony Kroch (University of Pennsylvania).

(i) Mary \{ must \} be happy, and so \{ may \} John

It should be noted that the occurrence of the epistemic modal in C° does not seem to cause a problem, which may be due to the fact that inversion after so does not effect the assertive force of the sentence.
(79) Peter said that Max must work for the KGB and work for the KGB Max must t
("epistemic/deontic")

When we now turn to infinitival constructions we find that the evidence of ellipsis and
topicalization confirms the parallels suggested in § 2.3 in a striking way. Ormazabal (1995)
and Martin (1996) convincingly demonstrate that the Control/Raising contrast is the decisive
factor in the distribution of VPE and VPT. Control infinitives again pattern with nonepistemic modals and occur both in VPE and VPT constructions, while Raising infinitives pattern with epistemic modals.

(80) a. John wasn’t sure he’d win the race, but he tried [PRO to e]
b. *Sam considered Sue to be clever, and Mike considered [Jane to e]

(81) a. No one believed John would fix the car, but [VP fix the car] John tried [PRO to t]
b. *No one expects John to know the truth but [VP know the answer] I believe [him to t]

The systematic absence of [-tense] categories from both VPE and VPT constructions calls for
an explanation. An obvious point of departure is the observation that both VPE and VPT are
forms of givenness marking: what is marked as ‘given’ is not a referring expression, but a
predicate which specifies a previously mentioned property (individual-level) or type of
event (stage-level/ episodic). In the first case, the givenness construction (VPE, VPT) affirms
the attribution of such a property to a given individual, in the second case the occurrence of an
event of a previously mentioned type. The second type, like all episodic sentences,
presupposes a stage topic (cf. Erteschick-Shir 1997), but it does not necessarily involve a
participant topic, as (82) shows.

(82) A: Was there a stranger in his room?
B: Yes, there was

No matter whether the givenness marking construction involves an individual-level or an
episodic predication, there must be an appropriate functional head in the extended projection of
V° which carries the focus features when the remaining portion of the projection is given. I

32 Rizzi (1990) already noted that VPT occurs with Control but not with Raising infinitives.
33 The correspondence between modals and infinitives is striking, but there is one context in which it breaks
down. Barrett (1998: 43) points out that in contrast to VPE, Pseudogapping is impossible with infinitival to –
both in Raising and in Control constructions:
(i) *John wanted to eat pizza and Mary to e spaghetti
Under the approach presented here this fact cannot be accounted for and I am forced to consider it an accidental
gap. This is unsatisfactory, because the restriction is too systematic to be attributed to the relative marginality
and idiosyncracy of the Pseudogapping construction.
34 Csúri (1996) calls this variety of anaphora ‘descriptive’. She distinguishes between ‘descriptive’ or ‘type
anaphora’ and ‘referential’ or ‘token anaphora’.
suggest that [-tense] categories are unable to assume this function, because they are extrapropositional, in contrast to [+tense] categories.

The assumption that [+tense] plays a role in the licensing of VPE (and presumably also VPT) appears natural, because [+tense] heads are involved in the binding of the event argument, which amounts to the internal closure of the presentational domain (Drubig 1992).

López and Winkler (2000) suggest that this role is played by polarity. Sentences like (83) look like strong evidence in favour of this assumption.

(83) The boys left, but the girls did not e

Although we now have a conflict between two competing conceptions of VPE licensing, there appears to be a way out of this dilemma. All the available evidence points to the conclusion that the licensing of VPE depends not only on heads involved in the saturation of the event argument (internal closure), but also on heads involved in the external closure or bounding of the domain, which are instances of Pol⁰. Presumably the licensing function is normally taken over by two heads in conjunction.

Now that we have acknowledged this possibility, we can focus on a type of VPE which shows that [-tense] heads can participate in VPE under certain circumstances. In (84), epistemic may occurs with VPE, but it is separated from the ellipsis site by negation, or negation plus an auxiliary.

(84) John may not be at home and Mary
   a. may not be e either
   b. may not e either

The essential condition seems to be that there is some element that intervenes between the epistemic modal and VPE. The presence of negation, as in (84)b., is only one way of accommodating VPE in an epistemic sentence. A 'stranded' auxiliary, as in (85) and (86), is

---

35 When VPE occurs in a finite clause, the functional head carrying the finiteness features is a default focus and must be prominent. López and Winkler (2000) present an interesting argument to the effect that the phonological consequences of the presentational default accent are preserved in VPE, even if a secondary focus structure is superimposed. Thus the second have in (i) has a reduced vowel, but cannot be nonmoraic, which López and Winkler interpret as an echo of the presentational default accent under the superimposed focus structure.

(i) The BOYS have LEFT and the GIRLS have \[ v \] e, too

When VPE occurs in Control infinitives, default accentuation occurs in a completely different form: to be is never accentuated and the default accent occurs on the governing predicate.

(ii) They asked Frank to do the job but he didn’t WANT to e

A possible account for this default phenomenon could depart from an observation in § 2.3: the governing predicate, not the infinitival head, is responsible for the future-oriented modality of to be in Control constructions.
sufficient. (86) shows that a stranded expletive auxiliary in a PTR construction serves the same purpose.

(85) John may have already left and Mary
   a. may have e too
   \[
   [\forall v]
   \]
   b. ?? may e too

(86) John must have left yesterday and Mary
   a. must have e too
   b. ?? must e too

(87) and (88) show that the intervention of a head has the same effect in elliptical Raising constructions.

(87) John seems to have left and Mary seems
   a. to have e too
   b. * to e too

(88) A: Do you think John is intelligent?
   B: I believe him
   a. to be e
   b. * to e

Finally, the same effect shows up in constructions with anaphoric pronominals in languages like German.

(89) Hans muss fleissig sein und seine Frau
   a. muss es auch sein epistemic/deontic
   b. muss es auch * epistemic/deontic

Sentences in which not is the sole intervener, such as (84)b., do not present a problem, because the polarity head closes the domain in which ellipsis occurs. The epistemic modal is external to PolP and therefore not involved in VPE. The auxiliary cases are more difficult to account for. Since neither $M_{ep}^0$ nor $to_R$ can license ellipsis, we are left with the inescapable conclusion that the licenser is the auxiliary below $M_{ep}$ or $to_R$, inspite of the fact that it is not finite. Since the phonetic form of the auxiliary have in (85)a. seems to show that the same principle of moraic conservation is here at work as in the more familiar finite type of VPE we may make the following assumption: the notorious close correlation between finiteness and the licensing of ellipsis is accidental. Any verbal head selecting TP (or AspP, i.e., an embedded TP) is [+tense]. Every such head, including the one carrying the finiteness features,
internally closes a potential focus domain and each 'closed VP' may be 'given'. The fact that the stranded auxiliary phenomenon is also available in sentences without $M_{ep}$ or $to_R$ shows that we are just dealing with alternative givenness options.

(90) John has been talking all the time and Mary
   a. has been e too
   b. has e too

3.3 Contrastive Focus and Association with Focus

In contrast to presentational focus and givenness marking, the relationship between epistemic modality and contrastive focus is relatively straightforward. A considerable number of authors (among them Kiefer 1986, Nuyts and Vonk 1999: 718ff. and Öhlschläger 1989) observe that epistemic modals resist contrastive focussing and are normally unstressed. Aijmer (1980) claims that under stress the epistemic reading of a modal disappears and the modal can only have its 'Grundbedeutung'. Westmoreland (1998: 111ff.) reports the results of a series of prosodic experiments which unanimously confirm Aijmer's observations. He presents cases like the following:

(91) It's clouding up, so it might RAIN today
(92) It's sunny and the sky is blue, but it MIGHT rain today

In (91) the modal statement moves in the direction in which the evidence points and the sense is epistemic. (92), however, highlights a mere possibility in the face of counter-evidence. The highlighting indicates the contrast shown in (93) and thereby eliminates the evidential character of the statement.

(93) It is POSSIBLE (but not NECESSARY)

The following examples show that only a nonepistemic modal can be associated with focus particles like *only*. The epistemic modal takes scope over the focus particle. In contrast to what we found in the case of negation, surface order is transparent with respect to scope, in this particular case.

(94) a. We only may live once
    b. We may only live once

36 Cf. López and Winkler (2000), as well as n. 36.
The conclusion we have to draw from these sentences is simple: epistemic modals are inaccessible to contrastive focussing. Narrow focus on a modal auxiliary is either metalinguistic or requires a nonepistemic reading. In other words, epistemic modality takes scope not only over presentational focus but also over the contrastive focus operator including various polarity elements, such as contrastive negation or focus particles, which license and associate with narrow focus.

3.4 Summary

The discussion of the modality and information structure in §§ 3.2. and 3.3 has shown that the syntactic analysis of epistemic and nonepistemic modality proposed in § 2 can account for the restricted cooccurrence of modality with focus and givenness reading. The position of epistemic modals within INFL in combination with the feature [-tense] predicts that epistemic modals cannot be adjacent to given or 'D-linked'/anaphorical' predicates, no matter whether they are deaccentuated in situ or topicalized, anaphorical pro-forms or elliptical. Epistemic modals do occur with such predicates, provided that an appropriate head intervenes and acts as the licenser. This is the only configuration in which an epistemic modal may carry the default accent of the presentational focus domain. Nonepistemic modals are [+tense] and license anaphorical predicates. The same feature specifications must also be assigned to Raising and Control infinitives, which pattern with epistemic and nonepistemic modals respectively and presumably occupy equivalent positions in INFL. Unlike modal auxiliaries, infinitival to is ineligible for default accentuation both in Raising and in Control constructions. Whereas the interaction of epistemic modals with presentational focus and givenness can be derived from their rank in INFL, the interaction with contrastive focus and association is determined by their COMP-related scope properties: epistemic modals are split into an overt exponent (the auxiliary) in INFL and a covert scope exponent in COMP, the latter of which gives them scope over the focus operator: in contrast to nonepistemic modals, which are restricted to INFL, epistemic modals relate to COMP and can neither be focussed nor associated with a focus-sensitive operator such as only or even.

37 Not surprisingly, the same is true for modal and evidential adverbs.
4. Comparative Evidence

4.1 Introduction

The hypothesis on the syntactic form of epistemic modality argued for in § 2. and 3. above was essentially derived from English evidence, with occasional looks across the fence. In the two following sections I confront the hypothesis with selected comparative evidence. In § 4.1 I test the hypothesis against the evidence in multiple modal (MM) constructions or multiple modals dialects of English. Such constructions have already been discussed in works dealing with the syntax of epistemic modality (cf. esp. McDowell 1987, Marrano 1998, Cinque 1999). I choose English MM dialects rather than MM languages such as Dutch, German or Scandinavian, because the English dialects show a number of interesting effects that cannot be observed in languages whose modals are verbs. There is solid evidence that the epistemic modal in a MM construction is not an auxiliary anymore but has developed into a nonverbal free form with interesting consequences.

4.2. The Evidence of Multiple Modal Constructions

MM constructions occur in a considerable number of varieties of English, including southern dialects in the U.S.A., Scottish dialects, African American English (AAE), creoles and Child English. Particularly striking is their occurrence in Child English because they also occur in areas where no MM dialects are spoken. (95) and (96) are quoted by Shepherd (1981: 107ff.).

(95) He must can do it
(96) I might will through out the meat

McDowell (1987: 269) was among the first to point out that the syntactic behavior of MM constructions in certain varieties of English mirrors the syntactic behavior of epistemic modals in the standard language. Under the analysis of epistemic modality proposed in this paper, MM dialects are different from Standard English and all other single modal dialects in one respect: both modal positions in (62) above (repeated as (97)) can be filled simultaneously.

(97) $M_{ep/AgrSP}$

\[
\begin{array}{c}
\text{PolP} \\
M/TP
\end{array}
\]
(97) predicts that MM constructions should have the following properties:

1. Only the first modal should bear an epistemic reading;
2. Negation should take scope over the second, but not the first modal;
3. A MM construction whose first modal is epistemic should be restricted to sentences with assertive force;
4. The second, but not the first modal should carry tense;
5. The second, not the first modal should license VPE and other constructions with anaphoric VPs.

When we inspect relevant data, we find that all these predictions are confirmed.

1. Only the first modal in a MM construction can be epistemic: This is the standard form in the MM dialects of the USA.

(98) You might could do it

2. Negation affects the second modal: In southern dialects in the USA, the clearest evidence for this is Neg-contraction, which can only occur with the second modal, as in (99).

(99) You might couldn't annoy Bill

Some authors report that the uncontracted form shows a certain amount of variation: it normally follows the second modal (100) but may occasionally precede it (101). Scopal order, however, does not seem to be affected by this variation.

(100) He might could not answer some questions, because he hadn't read all the material (Battistella 1991)

(101) I don't hear to well (...) I thought maybe I better put it [her hearing aid] on (or) I might not could understand you, ... (Battistella 1991)

3. MM constructions cannot be questioned (Labov, Cohen, Robins and Lewis 1968 on AAE; Boertien 1978 on southern dialects spoken primarily by white speakers in the USA; but cf. Coleman 1975 for a different view), unless the first modal is nonepistemic, which is often the case in Scottish MM dialects (Miller 1989):

(102) Will/*might he can give us a hand?

4. The second modal carries tense. Initial modals except should are epistemic and untensed (Boertien 1978: 26):

(103) It scared him that he may \{ could \} have been killed yesterday (Boertien 1978)
In (103) (Texan English) an SOT effect occurs on the second modal, which is equivalent to the same effect on a single modal in Standard English (104).

(104) It scared him that he could have been killed yesterday

(105) shows that the second modal carries the tense required by a past time adverbial

(105) In the past an ocean voyage might be very dangerous (Boertien 1986: 303)

Labov, Cohen, Robins and Lewis (1968) report a very interesting development in Northern AAE, where tense embedded under a single epistemic modal triggers *do support. (Cf. (106) and (107), both quoted from Labov, Cohen, Robins and Lewis (1968: 262):

(106) ... they must don't have too much in their wardrobe; right?
(107) [Conversation in an elevator:]
Father: Didn't you read the note?
Son: I read it
Father: Well, you must didn't read it too good!

Since the exponent of epistemic modality in sentences like (106)-(107) precedes what appears to be the finite auxiliary position, it is tempting to reconsider the proposal (originally made in Labov, Cohen, Robins and Lewis 1968) that *might in such sentences has been reanalyzed as an epistemic modal adverb (cf. also Labov 1972; 1973; and similarly Butters 1973 for southern American English varieties, as well as Miller 1989 for Scottish English; see also Battistella 1991, who tentatively proposes to extend an analysis along these lines to Standard English epistemic modals). In spite of its initial plausibility, the evidence speaks against this proposal: neither in Standard English nor in any of the MM varieties do epistemic modals show transportability properties characteristic of epistemic and evidential sentence adverbs. In § 2.2 I speculated that transportability effects might be related to the specifier status of sentence adverbs plus the COMP - related heads they modify. Unlike specifiers, however, the exponents of epistemic modality in the auxiliary complex of all English varieties rigidly occupy an initial position in INFL, which argues in favor of a head analysis (Cf. Pampell (1975: 115), Boertien (1979: 16) and di Paolo (1986: 10ff.) for arguments against an adverb analysis). Thus, epistemic modals in Standard English and MM varieties seem to pattern alike. In both varieties epistemic modals contrast with all other modals and seem to have undergone a significant historical change, in the course of which they lost their auxiliary...
status. Since there is no evidence of verb movement, I suggest that epistemic modals have been reanalyzed as free form instantiations (particles) of a category representing epistemic/evidential modality. The question of why Northern AAE lacks PTR effects and employs do support instead requires further investigation.

5. The second modal licenses VP Ellipsis:
As we saw in § 3.2, epistemic modals can cooccur with anaphoric VPs as long as there is an appropriate licenser. In MM constructions this is a [+tense] modal auxiliary.

\[(108) \text{[Can you get your civil rights without getting your head busted?] You might will e in the long run (Labov, Cohen, Robins and Lewis 1968: 262)}\]

\[(109) \text{You might could broad jump the Grand Canyon, and John might could e, too (Pampell 1975)}\]

The function of the second modal in VP anaphora constructions can also be observed in tag-constructions. Not surprisingly, tags are formed on the second of the two modals (Boertien 1978; Pampell 1975), which moves to C° and licenses VPE through its trace. This is possible because the tag duplicates its host with the exception of the epistemic modal, which counteracts a possible violation of the Head Movement Constraint. 38

\[(110) \text{You might could see Uranus if you e had a telescope, couldn't i you t, e?}\]

Summarising we may say that the evidence of MM constructions fully supports the analysis defended in § 2. and § 3.

4.3. The Evidence of Evidential Systems

4.3.1. Free Forms
In the preceding section I claimed that epistemic modals in the sense initially discussed in § 1. have moved away from the modal system and developed into evidential markers, comparable to those found in languages with more elaborate evidential systems. Many such languages employ free forms encoding evidential categories such as DEDUCTION, SECONDHAND and others. The restricted co-occurrence of such forms with markers indicating force, polarity

38 Unlike single modal sentences, MM constructions do not seem to be compatible with so-Preposing and auxiliary inversion.

(i) *You might could broad jump the Grand Canyon, and so might could John (Pampell: 1975)
I suggest that so-Preposing is excluded because auxiliary inversion has no well-formed option in such a configuration: neither can the two modals move conjointly into a single head position, as shown in (i), nor can the second modal move to C° alone, because this would lead to a violation of the Head Movement Constraint.
and tense aspect distinctions provides important typological evidence against which the predictions of the above analysis has to be tested.

Kazenin (1997) discusses both the evidential system and the focus marking system of Bagwalal, a SOV language of the Nakh-Daghestanian family spoken in East Caucasus. Bagwalal, according to Kazenin, has two evidential markers (EV), expressing SECONDHAND and DEDUCTION, which are free forms, but similar to auxiliaries, as Kazenin notes.

All the sentences in (111) – (113) show, the evidential categories have scope over the markers of polarity (cf. de Haan 1997:150ff. for crosslinguistic evidence on this scopal order) but also over tense distinctions:

(111) maHamad wa:wo weč’e Rala EV: SECONDHAND
Mohammed come- PERF NEG EV
‘Mohammed didn’t come (they say)’

(112) maHamad wa:-wo (ek_o’a) Rala EV: SECONDHAND
Mohammed come-PERF AFF EV
‘Mohammed has come (they say).’

(113) maHamad wa:-wo weč’e Ro EV: DEDUCTION
Mohammed come-PERF NEG EV
‘Mohammed certainly didn’t come’

Kazenin proposes the following analysis (in (114) ‘CL’ represents a morphological class marker):

(114)
Kazenin’s MoodP I interpret as essentially identical to the category M_{epP}/AgrSP in (97) in § 4.2.

Bagwalal evidentials interact in a very interesting way with the rather intricate focus marking system of this language. Bagwalal has a number of morphemes including the question marker \( (i)\text{št} \) which associate with narrow focus in the sense of Jacobs’ (1984) ‘relational focus theory’. \( (i)\text{št} \) may either be attached to the finite verb or right-adjoin to its associated narrow focus:

(115) maHamad - \( \text{išt} \) wa:-wo
Mohammed - Q come-PERF
‘Was it MOHAMMED that came?’

As we would expect, evidential markers (Rala: SECONDHAND; Ro: DEDUCTION) turn out to be mutually exclusive with the question marker \( (i)\text{št} \). Furthermore, they also associate with focus, as Kazenin shows. In this function, they occur in the same positions as the question marker: they either follow the finite verb or right-adjoin to the narrow foci they bind.

(116) a. maHammad wa:-wo ek\text{₀,}'a \text{Rala}
AFF EV
‘Mohammed has come (they say)’

b. maHammad \text{Rala} wa:-wo ek\text{₀,}'a
EV AFF
‘It was MOHAMMED that came (they say)’

c. maHammad ek\text{₀,}'a \text{Rala} wa:-wo
AFF EV

d. *maHammad ek\text{₀,}'a wa:-wo \text{Rala}
AFF EV

(117) a. maHammed wa:-wo weč’e \text{Ro}
NEG EV
‘Mohammed certainly didn’t come’

b. maHammed \text{Ro} wa:-wo weč’e
EV NEG
‘It certainly was MOHAMMED that didn’t come’

c. maHammed weč’e \text{Ro} wa:-wo
NEG EV
This evidence leads Kazenin to an interesting conclusion: In Bagwalal, *(i)št* (QUESTION) and *Rala/Ro* (SECONDHAND/DEDUCTION) are alternative focus binders. Kazenin’s analysis of Bagwalal independently suggests two fundamental assumptions on which my analysis of epistemic modality rests: 1. the scope superiority of epistemic modality over polarity as well as tense, and 2. the mutual exclusiveness of interrogative force and epistemic modality (cf. § 2.2). Furthermore, the evidence underlines the relatedness between evidentiality and epistemic modality, which I stipulated at the beginning (§ 1.).

4.3.2 Bound forms

The evidence of languages with inflectional evidentials, which has been intensively discussed in the relevant literature and which I will only briefly touch upon here, additionally confirms the conclusions of § 4.3.1, but it also has some additional implications.

On the one hand, what is known about inflectional evidentials in general unambiguously shows that in a language that also has inflectional sentence negation the evidential marker must have scope over negation. (117) and (118) are representative of this type of language:

(118) Turkish: (de Haan 1997:151)

*Ev - i - yok - mus*

house-POSS-NEG-EV

'He has no house (they say)'

(119) Tuyuca (Barnes 1994:331):

a. *Bué - ruku - ri - wū*

study - constantly - NEG - EV

'I did not study constantly (i.e., I studied, but not constantly)'

b. *Bué - ri - ruku - wū*

study - NEG - constantly - EV

'I constantly did not study (i.e., I was constant in not studying)'

On the other hand, it becomes clear that the evidential affix not only has scope over the polarity affix, these two affixes actually belong to two different layers which must be distinguished within the affixal structure of highly inflected languages. Jacobsen (1986) in his discussion of evidentiality and negation in Nootka and other languages shows that the inflectional affixes in this language form two different layers. The internal layer, to which

39 Alternatively, Bagwalal could perhaps be analyzed as a language which has ‘inflected relational focus markers’ that can be adjacent to focus and at the same time accommodate (and ‘locate’) the finiteness and force feature of the sentence in which they occur. Comrie (1995) discusses unusually interesting evidence from Malayalam which suggests such an interpretation for the focus particles of this language. The focus marker in Malayalam is formally identical to the copula. Cf. Kazenin (forthc.), Heine and Reh (1984), Harris and Campbell (1995) and Schaffar (2000) for discussion.
negation belongs, he calls the ‘relevant’ layer, following Bybee (1985); the external layer which hosts the evidential marker, is called ‘incremental’. A ‘relevant’ inflectional category is defined as follows: “A category is relevant to the verb to the extent that the meaning of the category directly affects the lexical content of the verb system.” (Bybee 1985:15) In other words, Bybee’s category ‘relevant’ defines a layer of inflectional elements which enter into semantic composition with the verbal stem. The external category called ‘incremental’ by Jacobson relates the semantic content determined by the internal layer to the context. The elements that belong to the ‘relevant’ layer must therefore be closer to the stem, as shown in (120).

I suggest that the distinction in (120) is reminiscent of the INFL/COMP layering of clausal syntax and related to the propositional/extrapositional dichotomy; i.e. the INFL/COMP structure of clausal syntax. If this is correct, it suggests a far-reaching conclusion: Affixation not only mirrors the syntactic hierarchy of functional heads, as stipulated in the Mirror Principle it also echoes the layering of propositional and extrapropositional markers.

The significance of this differentiation is not only relevant for inflectional morphology, it also affects the division of work between the inflectional and the derivational domain of morphology. In fact, there is another parallel between evidentials and epistemic modals: Evidential morphemes are always inflectional, never derivational (Bybee 1985). The same can be said about epistemic modality, which does not seem to be represented at the derivational level. Circumstantial (e.g. deontic modal) morphemes, on the other hand, may participate in lexical derivations. One such morpheme in English is –able, which attaches to verbs and derives adjectives. The semantic effect of this composition is equivalent to the syntactic combination of the modal can with a verb: The modal element derives a stage-level property into an individual level property (DiSciullo 1996: 9).

(121) a. analyz-able
    b. can analyze
The modal *can* cannot have an epistemic interpretation, but the suffix *–able*, its counterpart in derivational morphology, is also restricted to a root reading. This means that the scopal ranking of polarity and nonepistemic modality discussed in § 2.1 can also be attested on the morphological level.

5. Conclusion and Final Discussion

In § 2. and 3. I tried to show that English epistemic modals (such as *must, may/might*) are morphosyntactically and semantically distinct from root modals. Furthermore, I argued that epistemic modals must be analyzed as evidential markers. As such they are part of the extrapropositional layer of clause structure and take scope over all propositional operators, in particular negation and tense. Like evidential markers in languages with evidential systems, English epistemic modals can only occur in assertive contexts. Additional support comes from another side, namely the interaction between modality and focus/givenness marking, an important aspect of the phenomenon at hand which so far has been largely ignored in the relevant literature. Here, too, the behaviour of epistemic modals is determined by their extrapropositional status: in contrast to root modals, which can be narrow foci, associate with focus-sensitive operators and license deaccentuation as well as predicate ellipsis and fronting, epistemic modals turn out to be external to focus-background structure. They may cooccur with anaphoric predicates only if accompanied by a licensing head of the appropriate sort (negation or auxiliary), which then acts as the licensor. Exactly the same distinction was found to be relevant in infinitival constructions: Control infinitives pattern with root modals and license predicate anaphora, whereas Raising infinitives pattern with epistemic modals and can support an anaphoric predicate only when accompanied by a 'vicarious' licensor. Even more significant are the interpretive parallels between infinitives and modal auxiliaries: root modals and Control infinitives shift evaluation time and bind temporal arguments; i.e. they have the same selectional properties as tenses (in particular, past tense); epistemic modals and Raising infinitives, on the other hand, have no effect on temporal ordering and do not bind temporal arguments; i.e. they have the same selectional properties as the English 'present tense' (alias subject agreement). Under the approach to presentational focus structure presented in my earlier work on focus (cf. the references in § 3.1), this difference in behavior explains the selective cooccurrence between modals/infinitives and predicate anaphora.

40 Cf. Papafragou (1998:36ff.) for a convincing demonstration that this is true.
According to this approach a presentational focus domain is licensed by the combination of two factors: the domain internally closed through the binding of the temporal argument and externally bounded by a 'boundary marker', an appropriate operator in INFL (i.e., the polarity element); an anaphoric ('given') predicate is marked by deaccentuation of the predicate and default accent on a licensing head. Root modals can assume both functions, while epistemic modals can only bear default accents. In infinitives, the default accent can only occur on the governing predicate in the superordinate clause. The fronting of topicalized predicates must accordingly be seen as a separate effect (checking of the feature [+topic]; cf. Drubig fortc.), contingent upon givenness marking.

The above analysis of the grammar of modality not only draws a sharp syntactic boundary between epistemic and root modals, it also provides additional support for those semantic approaches which argue that epistemic modals are semantically close to the evidential markers of languages with evidential systems. Such an approach, if correct, would undermine any attempt to account for the meaning of all modals in terms of modal operators (\(\Box, \Diamond\)) and their duals. Under the evidential approach, as Westmoreland (1998) argues, epistemic modals are closer in meaning to the metalogical vocabulary (e.g., \(\therefore\)) than to modal operators. Furthermore, since must and may are no duals, the relations between epistemic judgements like MUST \(\phi\), MAY \(\phi\) etc. cannot be cast into such schemata as the traditional 'square of oppositions'.

---

41 This also applies to a related schema, the so-called 'duality square', which Löbner (1986) uses to predict the order of possible lexicalizations in the domain of generalized quantifiers, including various modal expressions. While Löbner's schema seems to make the right predictions for root modals and similar terms, it is inapplicable to exponents of epistemic modality in the sense defined in Westmoreland (1998), which underlies the present approach.
While the schema in (122) is arguably appropriate as a representation of root modalities (if we replace \textit{may} by \textit{can}), an interpretation in terms of epistemic modals is impossible, since \textit{may} in (122) cannot be identified with ‘\(\Diamond\)’, the dual of ‘\(\Box\)’. As we repeatedly pointed out, an epistemic possibility sentence such as MAY \(\varphi\) does not just mean that a certain possibility is open. It has a stronger meaning and indicates that the speaker has some grounds to claim that \(\varphi\) is true, no matter how weak they actually are. A statement like (123) a. tells us more than b.

(123) a. They may be in the desk drawer  
    b. They can be in the desk drawer

While (123) a. is indeed helpful when uttered as a response to a question such as \textit{Do you have any idea where my keys are?}, (123) b., in contrast is not. While \textit{can} is a good candidate for the lower left corner of (122) under a nonepistemic interpretation, \textit{may} cannot play this role in its epistemic counterpart. Since \textit{must}_{ep} and \textit{may}_{ep} are not duals, they cannot be combined into a simple square and, moreover, since neither one of them has an external negation, neither \textit{must}_{ep} nor \textit{may}_{ep} alone can form a square of this kind. As matter of fact, under the syntactic approach outlined above, such possibilities are systematically excluded: since duality is defined by the combination of external and internal negation, exponents of epistemic modality cannot be interpreted as duals, because they do not occur in the scope of negation. In this sense the syntactic form of epistemic modals reflects their semantic interpretations.  

\footnote{While I claim that epistemic modals in the evidential sense are semantically distinct from their nonepistemic counterparts, I do not want to deny that certain modal expressions, such as \textit{possible} and others, which are no}
6. References


Evidentials, are often interpreted against the background of 'what is known' and have 'epistemic interpretation' in this particular sense. As Hacking (1967) and following him Horn (1972) and many other students of natural language modality have argued at length, some expressions of this type, (e.g., possible) are nearly always used in this sense. The approach in the present paper, however, is exclusively concerned with those modals which have an evidential reading and distinct syntactic properties.


Gonda, Jan (1952) “Indonesian Linguistics and General Linguistics II.” *Lingua* 3, 17-52.


Hegarty, Michael V. (1992) *Adjunct Extraction and Chain Configurations*. Ph.D. Diss., MIT.


Kempchinsky, Paula (1986) *Romance Subjunctive Clauses and Logical Form*. Ph.D. Diss. UCLA.


