(De)coding epistemic modality in English and Swedish

ANNA WÄRNSBY

1 Introduction

When hearers are confronted with modal expressions in their native language, especially expressions that contain a modal verb, they are normally able to interpret these expressions immediately as being either epistemic or non-epistemic[1]. What it is that enables hearers to interpret modal expressions correctly? In order to account for the fact that (epistemic) modal meaning is readily recognized as such we should consider a variety of factors that are of importance for the interpretation; including the syntactic environment, the semantic contribution of the different parts of the utterance, and the pragmatic context of this utterance.

Coates (1983), amongst others, has suggested that expressions of epistemic modality can be differentiated from non-epistemic expressions in terms of certain contexts and environments. Using a corpus-based approach, Coates (1983:245) has argued that epistemic meaning is "typically associated with the following syntactic [sic] features":

(i) HAVE+EN construction
(ii) Progressive aspect
(iii) Existential subject
(iv) Stative verb
(v) Inanimate subject

She admits, however, that these associations may vary in strength. Since the aim of her study was "to interpret the data, not to impose some neat, preconceived system on it" (Coates 1983:247), the findings were not used further to form a model. On the other hand, some of the findings seem to have been incorporated in her later work in the model proposed in Coates (1995).

With the findings presented in Coates (1983) as the point of departure, the features described above were investigated with respect to their importance to interpretation of modality in utterances in the English-Swedish Parallel Corpus. The investigation is focused primarily on the systematic relationship between the environment in which modals appear in the utterances and epistemic interpretation. In this paper, the English modals *may* and *must*, and the Swedish modals *kan* and *måste* were investigated in terms of the environments in which they appear in the English-Swedish Parallel Corpus. To achieve uniformity in the data used in this paper, these two pairs of modals were chosen not only because they are often translational equivalents when used epistemically in the two languages, but also because these modals cover epistemic possibility meaning and differ only in the degree of speaker commitment to the truth of the proposition expressed in the utterance. The features discussed are used further to formulate a number of constraints that are relevant to the interpretation of

[1] This picture is of course simplified, since many of the occurrences of modals are ambiguous between the two interpretations (see Leech and Coates (1979) for the discussion of this phenomenon). In the corpus investigation on which this paper is based I exluded ambiguous occurrences of modals.

epistemic modality. Moreover, it is argued that these features are related to the notion of transitivity as described in Hopper and Thompson (1980).

For the purpose of the present investigation the situational context of modal utterances is not taken into account, nor are any generalisations concerning the meaning of the modals made in this study. Instead, it is argued that since modality in general, and epistemic modality in particular, seem to be universal categories in languages, the features discussed in this paper may prove to be valid when we are concerned with the analysis of modality in languages other than English, for example, Swedish.

2 Material and method

Data was collected from The English-Swedish Parallel Corpus (ESPC) (http://129.240.19.4/index-s.html). This corpus consists of four parts; original texts in English and Swedish and translations of these original texts into Swedish and English respectively. Since this study is not concerned with translation strategies, the modals *may*, *must*, *kan* and *måste* were searched for in the originals only. Searches were conducted in both the Fiction and the Non-fiction parts of the corpus. The instances of these modals in the corpus were analysed in terms of the epistemic/non-epistemic distinction. The retrieved examples were further analyzed in terms of the environment in which the modals appeared in the utterances. The features initially taken into consideration were:

(i) the type of the subject, i.e. whether introductory or not;
(ii) the form of the subject, i.e. whether expressed by a definite or an indefinite NP or other;
(iii) the type of the predicate, whether it is a state or an event/dynamic verb; and
(iv) whether the verb was aspectually or otherwise modified

These features correspond largely to the features mentioned in Coates (1983:245) as being associated with epistemic interpretation. The results of this analysis are presented in Table 1 below.

	MUST		MÅSTE		MAY		KAN	
	E	NE	E	NE	E	NE	E	NE
1. NP+AUX+state e.g. John must love Mary.	52	10	30	5	127	0	52	0
2. Intr.S+AUX+state e.g. There must sit a cat on the mat.	23	0	22	1	18	0	19	0
3. Ind.NP+AUX+state e.g. A car may be blue.	0	2	0	0	0	0	2	0
4. NCl+AUX+state e.g. What John said is interesting.	0	0	0	0	6	0	0	0
5. CondCl+AUX+state e.g. Whether or not we go for a walk may depend on the weather.	0	0	0	0	2	0	0	0
6. NP+AUX+PERF+state e.g. John must have been ill.	40	0	13	0	19	0	6	0
7. Intr.S+AUX+PERF+state e.g. There must have been a party here.	8	0	10	0	19	0	4	0
8. NP+AUX+PERF+event e.g. John must have taken the dog out.	27	0	32	0	21	0	2	0
9. Ind.NP+AUX+PERF+event e.g. An axe must have destroyed the desk.	3	0	4	0	2	0	0	0
10. NP+AUX+PROG+event e.g. John may be going out with Mary.	2	0	0	0	4	0	0	0
11. NP+AUX+event e.g. John must do it.	0	68	7	65	73	30	21	46
12. Ind.NP+AUX+event e.g. A soldier must do his duty.	0	1	2	0	11	0	1	1
13. Intr.S+AUX+event e.g. Det kan sluta med bråk. It MOD end-INF with fight 'A fight may finally break out'	0	0	0	0	0	0	3	0
14. NP+AUX+FUT+event e.g. Lasse kan komma att spela fiol. 'Lasse may FUT play the violin'	0	0	0	0	0	0	2	0
TOTAL:	155	84	120	55	302	30	112	70

Table 1. The environments in which modals occur in the corpus

In Table 1 the epistemic examples forming a patterned distribution are white, whereas the non-epistemic examples forming a patterned distribution are shaded light-grey. All exceptions from these alleged patterns are shaded dark-grey.

3 Discussion

In the following sections features which distinguish between epistemic and non-epistemic modality are discussed in connection to the data found in the ESPC. In Section 3.1 the syntactic/semantic environments in which epistemic (and non-epistemic) utterances appear in

the corpus are presented. In Section 3.2 constraints on the interpretation of epistemic modality which involve the above mentioned features are formulated. In Section 3.3 these features are discussed in connection to the notion of Transitivity proposed by Hopper and Thompson (1980).

3.1 Distinguishing between epistemic and non-epistemic utterances

It is possible to conclude from the data displayed in Table 1 that epistemic utterances found in the corpus differ systematically from non-epistemic ones in terms of the environments specified by the above-mentioned features. Some of the environments in which the modals appear may be grouped together on the basis of positive correlation between these environments and the interpretation of the utterance along the epistemic/non-epistemic dimension. In the following subsections these correlations are discussed.

3.1.1 State constructions in English and Swedish

Epistemic interpretations predominantly appear in all the different variations of state constructions, while non-epistemic interpretations are infrequent in these environments as illustrated in Table 2[2].

	MUST		MÅSTE		MAY		KAN	
	E	NE	E	NE	E	NE	E	NE
1. NP+AUX+state e.g. John must love Mary.	52	10	30	5	127	0	52	0
2. Intr.S+AUX+state e.g. There must sit a cat on the mat.	23	0	22	1	18	0	19	0
3. Ind.NP+AUX+state e.g. A car may be blue.	0	2	0	0	0	0	2	0
4. NCl+AUX+state e.g. What John said is interesting.	0	0	0	0	6	0	0	0
5. CondCl+AUX+state e.g. Whether or not we go for a walk may depend on the weather.	0	0	0	0	2	0	0	0
6. NP+AUX+PERF+state e.g. John must have been ill.	40	0	13	0	19	0	6	0
7. Intr.S+AUX+PERF+state e.g. There must have been a party here.	8	0	10	0	19	0	4	0
TOTAL:	123	12	75	6	191	0	83	0

Table 2. State constructions in English and Swedish

Some of the types of state constructions that were found in the data are illustrated by the examples below.

(1) "I **may** have something for you that you'll like even more," said Harcourt-Smith. (FF1)[3]

[2] Relevant parts of Table 1 are repeated here and in subsequent sections.

[3] Here and in the subsequent examples the abbreviations in parenthesis is the notation used in ESPC to indicate the source text.

(2) "That **must** be a dangerous job", Macon said politely. (AT1)

(3) Och kvinnan ombord **måste** vara skotskan som han
And woman-DEF aboard MOD be-INF Scotswoman-DEF COMP he
hade funnit på någon av öarna och vars liv
have-PAST find-PART on some of island-PL-DEF and whose life
han hade riskerat. (BL1)
he have-PAST risk-PART
'And the woman aboard must be the Scotswoman that he had found on one of the islands and whose life he had risked.'

(4) Det **kan** också vara chocken att upptäcka att hon lever. (BL1)
It MOD also be-INF shock-DEF to realize-INF COMP she live-PRES
'It can also be the shock of discovering that she is alive.'

In (1) the state verb *have* is present. In (2) we find copula *be*. (3) and (4) are examples of utterances containing the Swedish copula *vara* 'be'.

A strong tendency is displayed in the data for the epistemic interpretation to coincide with state constructions in the two languages. There is, however, a number of utterances which are interpreted as non-epistemic, in spite of the fact that they are state constructions. In all these (few) exceptions, the modal part of the utterance specifies either explicitly or implicitly a condition in which the proposition contained in an utterance is/will be true. Thus, these seeming exceptions are nonetheless systematic; their "exceptional" status is due to the fact that only four features (i.e. type and form of subject, type of verb, and aspectual modification) were initially included in the analysis. Sentence (5), for example, is a case of a non-epistemic state construction with an explicitly stated condition.

(5) To climb the tower you **must** be in a group, be aged over 11 and
have a letter of permission from your MP or embassy. (SUG1)

Sentence (6), on the other hand, exemplifies a non-epistemic state construction with an implicitly stated condition (the speaker in (6) is a prisoner and has to be brave in order to retain her dignity).

(6) I **must** be brave, I **must** maintain my own high standards. (ST1)

The non-epistemic interpretation in (5) and (6) seems to arise due to the presence of a condition in these sentences. If more features are included in the analysis, it is possible to show that the relation between the environment and the interpretation is systematic also in the so-called exceptional cases.

3.1.2 Aspect constructions in English and Swedish

The aspect constructions, i.e. English utterances modified by the progressive or the perfective aspect, and Swedish utterances modified by the perfective aspect, that were found in the data only have epistemic interpretations as can be seen in Table 3 below.

	MUST		MÅSTE		MAY		KAN	
	E	NE	E	NE	E	NE	E	NE
6. NP+AUX+PERF+state e.g. John must have been ill.	40	0	13	0	19	0	6	0
7. Intr.S+AUX+PERF+state e.g. There must have been a party here.	8	0	10	0	19	0	4	0
8. NP+AUX+PERF+event e.g. John must have taken the dog out.	27	0	32	0	21	0	2	0
9. Ind.NP+AUX+PERF+event e.g. An axe must have destroyed the desk.	3	0	4	0	2	0	0	0
10. NP+AUX+PROG+event e.g. John may be going out with Mary.	2	0	0	0	4	0	0	0
TOTAL:	80	0	59	0	65	0	12	0

Table 3. Aspect constructions in English and Swedish

The following utterances exemplify some of the aspect constructions found in the data.

(7) We **may** all of us be being punished now for sins we are about to commit. (FW1)
(8) He even said to her that they **must** have read the same book. (RR1)
(9) Flickan **kan** inte ha varit en vacker syn. (HM1)
Girl-DEF MOD not have-INF be-PART a beautiful sight
'The girl must not have been a pretty sight.'
(10) Han tänkte hastigt att hon **måste** ha burit den
He think-PAST fast COMP she MOD have-INF carry-PART it
med sig hela tiden. (HM1)
with REFL whole time-DEF
'He had a fleeting thought that she must have been carrying it with her the whole time.'

Sentences (7) and (9) contain state predicates (the English copula *be* in (7), and the Swedish copula *vara* in (9)) and are modified by progressive and perfective aspect respectively. (8) and (10), on the other hand, both contain events or dynamic verbs (*read* in (8), and *bära* 'carry' in (10)), while being modified by the perfective aspect.

The absence of the non-epistemic aspect constructions in the data can be explained by the nature of the non-epistemic modality represented by the modals under investigation (PERMISSION and OBLIGATION), i.e. it is performative (cf. the description of deontic modality in Palmer 1990:69), and thus is immediate to the time of the utterance. Furthermore, the action described by the main verb in such utterances is not actualized (if at all) until the listener chooses to act in accordance with or defiance of the permission, prohibition or command given by the speaker, which makes it quite impossible to express these notions in an aspectually modified utterance. For example,

(11) You **must** go to the party. Command
(12) You **must** have gone to the party. *Command
(13) You **must** be going to the party. *Command

While the above may serve as an explanation as to why modals which have both deontic and epistemic interpretations are not normally interpreted as deontic in an aspectually modified utterance, a complication may arise in connection to the Swedish *kan*, which has an additional interpretation, namely ABILITY[4], as in (14). However, aspectually modified utterances such as (15) do not allow the ABILITY interpretation either.

(14) Han **kan** spela piano. Ability
He MOD play-INF piano
'He can play the piano.'

(15) a. Han **kan** ha spelat piano. *Ability
He MOD have-INF play-PART piano
'He may have played piano.'

b. *Han håller på att **kunna** spela piano. *Ability
He PROG-PRES MOD-INF play-INF piano
'He is being able to play the piano.'

The impossibility for an aspectually modified utterance containing *kan* to be interpreted as expressing ABILITY can be explained as follows. In a subject-oriented ABILITY reading the subject is described as having either innate ability or ability which has been learned. In both cases such ability is a long-term quality of the subject (cf. Teleman et al 1999: 302). Palmer (2001:8) claims that "[…] dynamic modality refer[s] to events that are not actualized, events that have not taken place, but are merely potential". Thus, if one has the ability to play the piano, it is not necessarily the case that this person is actually going to play the piano. Similarly, if one had the ability to play the piano in the past, it is not necessarily the case that the person in question exercised his/her ability at any time in the past. Perfective aspect, on the other hand, indicates that the event described by the main predicate is not only actualized, but also completed. It is exactly because one cannot combine the notion of potentiality with the notion of completeness that the ability reading is impossible in (15a). (15b), modified by the (lexical) progressive aspect, is quite ungrammatical in Swedish. It is possible to conclude that, among other things, the progressive aspect, indicating that the event described by the main predicate is on-going, i.e. realized, 'clashes' with the notion of potentiality in ABILITY.

3.1.3 Event constructions in English and Swedish

The predominant occurrence of event constructions with non-epistemic meaning can be explained since this is how permission, obligation, command, etc. are expressed in modal utterances in both languages, i.e. all of these will normally contain a dynamic verb. The distribution of event constructions in English and Swedish is illustrated in Table 4.

[4] The modal normally used to express PERMISSION in Swedish is *få*. *Kan*, however, can also be used as a modal of PERMISSION, as in *Du **kan** ta kakan* 'You may take the cake' (cf. Teleman et al 1999: 298-300).

	MUST		MÅSTE		MAY		KAN	
	E	NE	E	NE	E	NE	E	NE
11. NP+AUX+event e.g. John must do it.	0	68	7	65	73	30	21	46
12. Ind.NP+AUX+event e.g. A soldier must do his duty.	0	1	2	0	11	0	1	1
13. Intr.S+AUX+event e.g. Det kan sluta med bråk. It MOD end-INF with fight 'A fight may finally break out'	0	0	0	0	0	0	3	0
14. NP+AUX+FUT+event e.g. Lasse kan komma att spela fiol. 'Lasse may *FUT* play the violin'	0	0	0	0	0	0	2	0
TOTAL:	0	69	9	66	84	30	27	47

Table 4. Event construction in English and Swedish

There are, however, quite a number of examples where such constructions are interpreted as epistemic. These cases are uniform, in that the interpretation is triggered by the presence of either an epistemic adverbial or an inanimate subject. Consider examples (16) and (17).

(16) Den saken var något som Torsten nog **måste**
This thing-DEF be-PAST something COMP Torsten probably MOD
finna ut på egen hand. (LG1)
find-INF out on own hand
'That was something Torsten would obviously have to find out for himself.'

(17) Insects are not confined to the flowers, however, and a careful search of the foliage **may** reveal the amazingly colourful rhododendron leafhopper as well as oak bush crickets and speckled bush cricket. (SUG1)

In (16) the epistemic interpretation stems from the presence of the epistemic adverbial *nog* 'probably' (without the adverbial the utterance is ambiguous between the two interpretations). In (17) the presence of an inanimate subject *a careful search of the foliage* influences the interpretation in a similar fashion. Again, the 'exceptional' status of these and similar examples depends on the limited number of the features initially chosen for the analysis.

3.1.4 Kommer att V construction in Swedish

The special case of the Swedish '*kommer att V* construction' is presented in example (18):

(18) Det lugnar mig under denna tid då jag anar att
It calm-PRES me under this time when I suspect-PRES that
mycket oönskat **kan** komma att inträffa. (AP1)
much unwanted MOD come to happen
'It calms me during this time when I suspect that many undesirable things may happen.'

This kind of construction is possible in Swedish due to the nature of *komma att* 'come to'. It can be used both as a finite modal or as a modal infinitive in combination with another modal, as in example (18) above. It may be argued that *komma att* has not fully developed

into a modal yet[5], and has retained much of its original lexical meaning and syntactic behaviour, which in turn allows it to enter into constructions such as the one in (18). The infinitive *komma att* is used only when the event time is not defined by the context of an utterance. Furthermore, it is frequently used as a complement to a discourse-oriented verb (e.g. epistemic *kan*) (see Teleman et al 1999: 252).
The future marker *komma att* can be seen as more subjective in Swedish than the variant future marker *ska*. Consider the following Swedish sentences.

(19) Han ska åka till Stockholm på söndag.
He FUT go-INF to Stockholm on Sunday
'He is to go to Stockholm on Sunday.'
(20) Han kommer att åka till Stockholm på söndag.
He *kommer att* go-INF to Stockholm on Sunday
'He will go to Stockholm on Sunday.'

According to Teleman et al *ska* is a temporal auxiliary verb which indicates that the event described by the main predicate follows the time of the utterance. Furthermore, *ska* implies that this future event is planned, either by the subject, or by somebody else who has the authority to impose his/her will on the subject (Teleman et al 1999: 246). Thus, not only *ska* is seen as more objective, but a parallel may also be drawn between the future and the deontic interpretation of *ska*. *Komma att*, on the other hand, does not imply that the event described by the main predicate is planned or intended, but indicates a more or less competent prognosis on the part of the speaker, and that the speaker relies on (external) factors to make such a prognosis (Teleman et al 1999: 244). Thus, *komma att* may be seen as more subjective, and, thus, related to epistemic modality. The subjectivity of *komma att* allows it to combine with another subjective element (epistemic *kan* in (18), whereas such a combination is not well-formed with the objective *ska* (cf. *…då jag anar att mycket oönskat kan *skola/ska inträffa*). The fact that occurrences of *kan komma att V* are never non-epistemic supports this claim.

3.2 Formulating some constraints on the interpretation of epistemic modality

To summarize the above discussion, we see that the following features are relevant to differentiating between epistemic and non-epistemic modal utterances:

(i) Presence of an epistemic modal adverbial
(ii) Introductory subject
(iii) Aspectual modification
(iv) State/dynamic verb
(v) Explicit/implicit condition
(vi) Animacy of the subject
(vii) Specificity of the subject
(viii) Future

[5] The recent development of *komma att* involves the loss of the infinitive marker, so that constructions such as *komma göra* 'will do' instead of *komma att göra* are more and more frequent in both formal and informal Swedish discourse (see Teleman et al 1999: 244). When this development is completed and *komma att* achieves full modal status, it may be predicted that constructions as in (18) will disappear (cf. **may will do, *kan skall göra*).

Further, on the basis of the data analysed it is possible to make a tentative suggestion as to how these features may be ranked with respect to their importance for the interpretation of an epistemic utterance. This list is not necessarily exhaustive in a cross-linguistic perspective however.
In Section 3.1.3 it was mentioned that modals in combination with dynamic verbs without aspectual modification exhibit a strong tendency to appear together with another epistemic element (e.g. modal adverbials or modal particles).

(21) If an export manager sent out by his company cannot communicate on the business and social level with the foreign customer, and has no respect for the different ways of conducting business and behaving socially, the customer **may** well choose another partner. (JPM1)

The effect of an epistemic adverbial on the interpretation of an utterance containing a modal verb is discussed, *inter alia*, by Quirk et al (1985:583-586), Hoye (1997:149-152), and Wärnsby (1999), who all argue that the presence of an epistemic modal adverbial may neutralize, disambiguate or reinforce the meaning of the modal in an utterance. The presence of an epistemic adverbial in an utterance (both in harmonious and disharmonious combinations) takes precedence over all other features, since these adverbials are sentential, and therefore take scope over the utterance as a whole. Thus, it is possible to formulate the first constraint:

A. Utterances containing an epistemic adverbial or epistemic particle are epistemic.

The fact that no aspectual modification was found in utterances with a non-epistemic interpretation supports the initial assumption that aspect is associated with epistemic modality. This association is grounded in the nature of non-epistemic modality itself. Consider, for example, the following aspectually modified sentences which contain different types of predicates; in (22) there is a state verb, whereas in (23) there is an event verb.

(22) Eller som min föregångare - vem tusan han nu **kan** ha varit
Or as my predecessor - who thousand he now MOD have-PRES be-PART
– antingen innan de blev alltför fulla eller alltför skumögda för
– either before they become-PAST far-too drunk or far-too dim-sighted to
att föra det hela till slut, och utkastade frampå eftermiddagen. (LG1)
to bring it whole to end, and thrown-out later-on afternoon-DEF
'Or as my predecessor – who the hell he now may have been – either before they got too drunk or too dim sighted to bring the whole thing to an end, and got thrown out late in the afternoon.' Epistemic

(23) But they didn't go that way, they **must** have gone by the by-road, and then there wasn't much of a chance. (RR1) Epistemic

Both sentences, however, are interpreted as epistemic. This suggests that this feature is more important for the interpretation than the state/event distinction. On the other hand, it is possible to find sentences which are interpreted as non-epistemic, despite being aspectually modified, as demonstrated in (24).

(24) You **must** have finished your homework by tomorrow. (Papafragou 2000:102)

Thus, aspectual modification alone cannot be seen as criterial for epistemic interpretations. On the other hand, the difference between (24) and the epistemic examples found in my data lies in the time reference for the proposition. Thus, (22) and (23) have past time reference, whereas (24) has future reference. Consequently, it is possible to formulate the second constraint:

B. Utterances with perfective aspect and past time reference are epistemic.

With respect to utterances modified by the progressive aspect, all the English examples found in the corpus are epistemic. However, the fact that the progressive aspect is so strongly associated with epistemic interpretations should not be taken to prove the initial hypothesis that aspect is associated with epistemic modality. It is possible to find examples when an utterance modified by the progressive aspect can be interpreted as non-epistemic, as in (25).

(25) *We **must** be leaving soon.* (Papafragou 2000:102)

Papafragou claims that this example is non-epistemic. However, out of context, it is not possible to decide in favour of one or the other interpretation. That is, if appropriate contexts for this sentence are created, it can be interpreted as either epistemic or non-epistemic. (26) is then interpreted as epistemic, whereas (27) is non-epistemic.

(26) I see that my husband has started packing, so we **must** be leaving soon. Epistemic
(27) We **must** be leaving soon, if we are to catch the 9 o'clock train. Non-epistemic

Both (26) and (27) have future reference. This indicates the impossibility of distinguishing between epistemic and non-epistemic utterances modified by the progressive aspect on the basis of the time reference alone. In (26) for example, the speaker expresses a confident judgement on the basis of the evidence at hand. (27), on the other hand, is a conditional statement, where the conditions are specified for the proposition. (27) contains an explicit condition, but the condition does not have to be expressed explicitly. It can also be left implicit and be inferred from the (immediate) context of the situation. If such a condition is neither explicitly expressed nor implicitly present, the sentence in (25) is to be interpreted as epistemic. It is now possible to formulate a third constraint:

C. Utterances with the progressive aspect in which conditions are not specified are epistemic.

In cases where the utterance is not aspectually modified, the distinction between state and event verbs is crucial. Consider the following examples.

(28) You **may** not know about this one: it's a modern sin. (FW1) Epistemic
(29) Han reste sig från bordet: Ursäkta mig – jag **måste** ta ut några
He rise-PAST REFL from table-DEF: Excuse me – I MOD take-INF out few
bullar ur ugnen. (APR1)
buns out-of oven-DEF
'He rose from the table: Excuse me – I have to take some buns out of the oven.'
Non-epistemic

Sentence (28) contains a state verb, and is interpreted as epistemic, whereas sentence (29) contains an event verb and is interpreted as non-epistemic. However, as demonstrated in

Table 1, utterances containing dynamic verbs can have epistemic interpretations, and utterances containing state verbs can be interpreted as non-epistemic.

Consequently, it is possible to argue with regard to the utterances with state verbs that the presence of a state verb is not enough on its own to determine that such an utterance is to be interpreted as epistemic. An additional feature has to be taken into consideration, i.e. whether such an utterance also contains an explicit or implicit condition. Consider again example (6) repeated here as (30), and the following example (31):

(30) I **must** be brave, I **must** maintain my own high standards. (ST1)
Implicit condition
Non-epistemic

(31) Så är det hemma hos oss, man **måste** vara uppmärksam
So be-PRES it at-home with us, one MOD be-INF watchful
på det, annars **kan** förfarliga saker inträffa. (AP1)
on it, otherwise MOD terrible things happen-INF
'This is how it is at home, one has to be observant of this, otherwise terrible things may happen.'
Explicit condition
Non-epistemic

The presence of a (specified) condition in an utterance containing a state verb is interpreted as non-epistemic, as demonstrated in the examples above.[6] Thus, it is possible to formulate a fourth constraint on the interpretation of epistemic modality:

D. Utterances with state verbs in which conditions are not specified are epistemic.

Furthermore, a correlation between introductory subjects and epistemic interpretations was found in the data, as exemplified in sentences (32-33) below:

(32) Neuropsychologically, there is little or nothing you can do; but in the realm of the Individual, there **may** be much you can do. (OS1) Epistemic

(33) Det **måste** finnas medicin mot sådant, något upplösande kanske? (PCJ1)
It MOD exist-INF medicine against such, something dissolving maybe?
'There must be some medicaments against things like these, something dissolving maybe?'
Epistemic

This correlation may be explained by appealing to the fact that sentences containing introductory subjects are existential in nature and always involve some copula or state predicate. A contributing factor is that formal subjects in existential sentences are non-specific. The presence of a copula or state predicate in these constructions is seen as crucial; therefore utterances containing introductory subjects are suggested to be handled by the above-mentioned constraint.

When analysing utterances without aspectual modification containing a dynamic verb, it is important not only to examine whether an epistemic adverbial/particle is present in an utterance (cf. the discussion above), but also to check whether the subject in an utterance is inanimate. Sentence (34) below contains an animate subject in combination with an event verb, and is interpreted as non-epistemic. Sentence (35), on the other hand, contains an inanimate subject in combination with an event verb, and is interpreted as epistemic.

[6] The fact that these utterances have future implication is not considered of crucial importance here, since it is possible to find epistemic utterances with future implication as well (e.g. *John may arrive tomorrow*).

(34) Glöm inte att vi **måste** gå till fots till Eolsberg. (APR1)
Forget NEG COMP we MOD go-INF to feet to Eolsberg
'Don't forget that we have to walk to Eolsberg.' Non-epistemic

(35) Although the noise **may** detract from some people's enjoyment, it certainly seems to do little to upset the birds, who carry on feeding regardless. (SUG1) Epistemic

Thus, all things being equal, the combination of an inanimate subject and a dynamic verb not modified by aspect results in an epistemic interpretation of this utterance, as shown in examples (34-35). Given this observation, I formulate the fifth constraint:

E. Utterances containing dynamic verbs and inanimate subjects are epistemic.

With regard to the feature listed as (vii) above, 'specificity of the subject', conflicting evidence was found in the data. Firstly, both specific and non-specific subjects appear in the corpus in combination with state verbs, which indicates that the state/event distinction is a higher-order distinction when compared to the specificity of the subject.

(36) An animal, or a man, **may** get on very well without "abstract attitude" but will speedily perish if deprived of judgment. (OS1) Non-specific subject
Epistemic

(37) He kept his distance from me because he thought he **must** smell of her arm and shoulder pressed against his. (NG1) Specific subject
Epistemic

Secondly, it is possible to conclude that this patterning is also valid for the modals *may* and *kan* in combination with an event verb in the data. In (38), for example, a non-specific subject combines with a dynamic verb, and the utterance is interpreted as epistemic.

(38) There is always the feeling that, at any moment, something **may** happen that will lead to a new discovery. (DM1) Epistemic

The only example which seems not to fit the pattern is (39) below, which is interpreted as non-epistemic contrary to the hypothesis.

(39) A Contracting Party **may** at any time raise a matter of concern at the level of the EEA Joint Committee or the EEA Council according to the modalities laid down in Articles 92 (2) and 89 (2), respectively. (AEEA1)
?Non-specific subject
Non-epistemic

On the other hand, although *a Contracting Party* has the form of an indefinite NP, it can be argued that in legal texts, where this example was found, a Contracting Party is an entity, specified in detail in the text, and can, thus, be considered to constitute a specific subject in (39). This line of discussion, however, is not continued here, since the focus of this paper is primarily on the systematic relationship between the environment in which the modals appear in the data and the epistemic interpretation.

Another non-epistemic example is displayed in (40).

(40) Yet, deprived of books, a man **must** fall back on thought, and out of thought, if he can learn to order it, will come the urge to record, and in extremity, if he has no means of recording, recitation, the ordering of memory which leads to metre, to commemoration. (DW1)

Condition
Non-specific subject
Non-epistemic

In (40) the part of the utterance containing the modal verb specifies a condition which is to be fulfilled in order for the proposition to be true. It may be argued then, in analogy with the discussion accompanying sentences (34-35), that feature (vii), specificity, is secondary to other features in the cases above.

On the other hand, even though no examples of epistemic *must* or *måste* were found in the corpus in combination with a non-specific subject and a dynamic verb, it does not mean that such examples are ungrammatical. Consider the following sentences.

(41)	Somebody **may** do it, even if you won't.	Epistemic
(42)	Somebody **must** do it, even if you won't.	?Epistemic
(43)	Somebody has to do it, even if you won't.	Non-epistemic

Sentence (41) is quite unproblematic, since *may* works well in this environment. It is, on the other hand, difficult to get an epistemic interpretation of sentence (42), while (43) is unproblematically non-epistemic. Here, for the first time, the modals clearly exhibit a difference in their behaviour. A tentative conclusion may however be reached, i.e. that feature (vii) applies to the modals expressing possibility, and not to the modals expressing certainty in this analysis, although a careful investigation into this problem is needed. At the present stage no constraints involving this feature are formulated.

Let us return to the question raised at the beginning of this paper – what is it that enables hearers to correctly interpret modal expressions in their native language In view of the discussion in this paper, the possible answer to this question may be that when facing the task of interpreting an utterance as epistemic or non-epistemic the hearer activates a "scanning device" of some kind, and probes the utterance in question with respect to the features discussed above.

As was suggested at the beginning of this section the features may be arranged according to the degree of their importance for the interpretation of epistemic modality. Thus, the order in which the constraints are formulated reflects the fact that each subsequent constraint involves a deeper analysis of an utterance, i.e. more and more features are involved in the analysis. This indicates that the constraints apply in a successive manner, which in turn determines the nature of the above-mentioned scanning device.

Such a device may work as follows. With a modal utterance as input the device scans for the presence of the modal adverbial or particle in the utterance. If such an adverbial or particle is present, the utterance is interpreted as epistemic. If, however, no epistemic adverbial or particle is found, the device probes further whether the utterance is modified aspectually. If this is the case then the device scans for the type of aspectual modification, i.e. whether the utterance is modified by the perfective or the progressive aspect. If the utterance is modified by the perfective aspect, the device examines whether the time reference is past or not. If the time reference is past the utterance is interpreted as epistemic, if not, then the device needs to probe the utterance further. This is represented schematically in Figure 1 below.

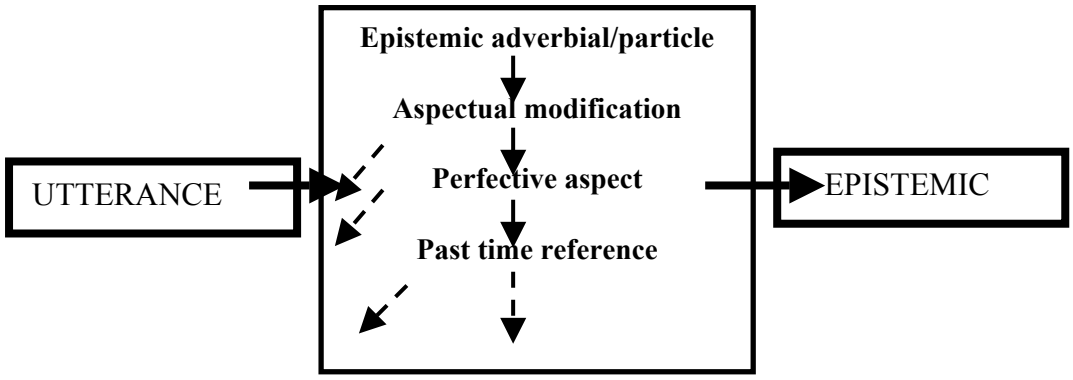

Figure 1. "Scanning device"

An objection to this hypothesis can be raised, however, – since these features are not only complex but also in most cases unrelated, it may be too demanding for the language user to go through such a "checking" procedure. An indirect support for the existence of the suggested scanning device can, on the other hand, be found in Coates (1988), who observed that the system of modal meanings of 8-year-olds is only rudimentary, and that even by the age of twelve the child's modal system is not yet isomorphic with that of the adult. The fact that modality is acquired at a later stage of language development suggests that it requires considerable cognitive abilities on the part of the speaker or the listener.

3.3 Modality and Transitivity

In the previous section it was mentioned that the features discussed seem to be unrelated. This may however prove to be untrue, if we consider what Hopper and Thompson (1980) call Transitivity.

Hopper and Thompson define Transitivity as a composite notion, which is only partly concerned with the presence of an object of the verb. Traditionally, the notion of Transitivity is understood in terms of "an activity […] 'carried-over' or 'transferred' from an agent to a patient', and thus involves 'an action which is typically EFFECTIVE in some way" (Hopper and Thompson 1980:251). This view is modified by the authors in such a way that they identify several component parts of Transitivity, each of which implies a scale of Low to High Transitivity for sentences in a language. Each of these components is said to "involve […] a different facet of the effectiveness or intensity with which the action is transferred from one participant to another" (Hopper and Thompson 1980:252). These components are presented in Figure 2 below.

	HIGH	*LOW*
A. PARTICIPANTS	2 or more participants, A and O	1 participant
B. KINESIS	action	non-action
C. ASPECT	telic	atelic
D. PUNCTUALITY	punctional	non-punctual
E. VOLITIONALITY	volitional	non-volitional
F. AFFIRMATION	affirmative	negative
G. MODE	realis	irrealis
H. AGENCY	A high in potency	A low in potency
I. AFFECTEDNESS OF O	O totally affected	O not affected
J. INDIVIDUATION OF O	O highly individuated, i.e. O is:	O non-individuated, i.e. O is:
	(i) proper	(i) common
	(ii) human, animate	(ii) inanimate
	(iii) concrete	(iii) abstract
	(iv) singular	(iv) plural
	(v) count	(v) mass
	(vi) referential, definite	(vi) non-refrential

Figure2. The components of Transitivity (adopted from Hopper and Thompson 1980)

The following summary is based on Hopper and Thompson's account of the above components (Hopper and Thompson 1980:252-255).

Component A refers to the number of participants in the clause. Since an action can only be said to have been 'transferred' if there are at least two participants, clauses containing an A(gent) and an O(bject) are generally considered more transitive than those that contain only an A[7].

Component B, Kinesis, reflects the fact that actions can be transferred from A to O, whereas states cannot. Thus, in the *The boy kicked the ball* the ball is likely to undergo some transition, e.g. from being immobile to being set in motion. In *The boy hates ice-cream*, on the other hand, no change is imposed on the ice-cream.

By making use of component C, Aspect, Hopper and Thompson are able to view an action in terms of telicity[8]. Thus, telic actions which are high on the Transitivity scale, are marked by the past tense or the perfective aspect, whereas the present tense and the progressive aspect mark atelic actions, which are low in transitivity. In *John ate the cake* the transfer of the action is completed. In *John is eating the cake* the transfer is not completed. Consequently, this latter utterance is seen as being lower in Transitivity than the former.

The inclusion of component D, Punctuality, reflects the fact that "actions carried out with no obvious transitional phase between inception and completion have a more marked effect

[7] Note, however, that since the notion of transitivity is regarded as componential and scalar, in some cases a clause with no O can be considered to be higher in transitivity than a clause containing both an A and an O, as in (i) and (ii) respectively:

(i) Susan left. (+action, +telic, +punctual, +volitional)

(ii) Jerry likes beer. (+ two participants, -volitional)

For more details see Hopper and Thompson (1980:254).

[8] This account of aspect and telicity refers to the external or grammatical phenomenon, and should not be confused with the internal or lexical telicity, which is a component of verbal *aktionsart*. *Aktionsart* of the verb can be said to be roughly reflected by component D, Punctuality, in Hopper and Thompson's model (cf. Hopper and Thompson 1980:271).

on their patients than actions which are inherently on-going" (Hopper and Thompson 1980:252). Thus, clauses containing punctual verbs (e.g. *break*) are considered to be more transitive than clauses incorporating non-punctual verbs (e.g. *play*).

Component E, Volitionality, illustrates the fact that A can act purposefully. *I wrote your name* (volitional), is therefore considered to be higher in transitivity than *I forgot your name* (non-volitional) (Hopper and Thompson 1980:252).

Component F, Affirmation, indicates whether a clause is affirmative or negative, where the former is considered to be higher in transitivity than the latter. The reason for this is, of course, that in an affirmative statement it may be reported that an action was transferred onto O, whereas a negative statement would be used a report that no such transfer took place.

The next component, Mode, indicates the mood of the clause, whether *realis* or *irrealis*. Actions described by clauses in irrealis mood are considered to be less effective than actions described by the clauses in realis mood. The latter clauses are then higher in Transitivity than the former.

Component H, Agency, shows that A's high in agency/potency[9] are able to transfer the action more effectively than those A's which are low in agency/potency. Thus, the normal interpretation of *George startled me* is that of a perceptible event with perceptible consequences; but that of *The picture startled me* could be completely a matter of an internal state" (Hopper and Thompson 1980:252).

Component I, Affectedness of O, deals with the degree that O is affected by the action, whereas component J, Individuation of O, deals with the degree to which O is distinct from A, and with the degree to which O is distinct from its own background. In *John peeled potatoes for two hours*, for example, O is less individuated (since *potatoes* is common, plural and inanimate), and not totally affected (since there may be some potatoes left to peel). In *John peeled the potatoes in two hours*, on the other hand, O is more individuated (since *the potatoes* is concrete, count and definite), and it is also totally affected (since there is an implicature that all of the relevant potatoes were peeled).

Hopper and Thompson provide extensive typological evidence that Transitivity, as a compositional notion, is a central relationship in human language. The question asked by the authors is why the morpho-semantic correlations that they found to be relevant to their notion of Transitivity are so regular cross-linguistically? From here we move into the area of linguistic universals. Hopper and Thompson suggest that linguistic universals "originate in a general pragmatic function, and that the universal is not explained until this function has been isolated and related to the universal" (Hopper and Thompson 1980:280). It appears that the function Transitivity has (in discourse) is to distinguish between the backgrounded and foregrounded information[10] in such a way that clauses high in Transitivity tend to be foregrounded, whereas clauses low in Transitivity tend to provide background information.

The foregrounded/backgrounded distinction fundamental to the organisation of discourse is also said to be "a universal – having its origins in central communicative and perhaps psychological functions" (Hopper and Thompson 1980:283). In languages where

---

[9] The notion of agents high in potency is compatible with the properties of the Proto-Agent discussed by Dowty (Dowty 1991 in Goldberg 1995:116). Also consider the so called Agency Hierarchy (Siverstein 1976 in Hopper and Thompson 1980:273):

1st Person > 2nd Person > 3rd Person > Proper name > Human > Animate > Inanimate

This hierarchy is arranged so that As located in the left end of the hierarchy are considered to be higher in agency/potency than those located in the right end of the hierarchy.

[10] Hopper and Thompson define the notions of background and foreground as follows. "That part of discourse which does not immediately and crucially contribute to the speaker's [communicative] goal, but merely assists, amplifies or comments on it, is referred to as BACKGROUND. By contrast, the material which supplies the main points of the discourse is known as FOREGROUND." (Hopper and Thompson 1980:280)

foregrounded/backgrounded information is not marked by a single morphosyntactic feature, e.g. English and Swedish, foregrounding is not marked in absolute terms. Instead, it is indicated by encoding the foregrounded clause in such a manner that it will be high in Transitivity, i.e. such a clause will typically have more high Transitivity features as defined above, than a clause which is backgrounded. Hopper and Thompson concentrated their investigation on narrative texts, since "[n]arrative is a cultural universal, and hence readily accessible in a variety of languages. Other studies, however, conducted on different discourse genres, e.g. conversation and procedural discourse, have come to similar results (see Hopper and Thompson 1980:282).

A striking similarity may be observed between the features constituting Transitivity in Hopper and Thompson's account, and the features differentiating epistemic and non-epistemic modality discussed in the present paper. Thus, the distinction between action and non-action in the Kinesis component of Transitivity reflects the distinction between state and dynamic/event verbs used in this paper. (Grammaticalized) telicity in the Aspect component is related to what has been called 'aspectual modification' in previous sections. Punctuality, as well as Affectedness and Individuation of O are some of the features of *Aktionsart* as defined in Section 2 above. This account of aspect and telicity refers to the external or grammatical phenomenon, and should not be confused with the internal or lexical telicity, which is a component of verbal *aktionsart*. *Aktionsart* of the verb can be said to be roughly reflected by component D, Punctuality, in Hopper and Thompson's model (cf. Hopper and Thompson 1980:271). Volitionality and Agency relate to the distinction made previously between animate and inanimate subjects. The significance of this similarity is two-fold.

Firstly, the features discussed in this paper were shown to be prominent in terms of the unrelated notion of Transitivity and the discourse distinction between foregrounding and backgrounding. Since features constituting Hopper and Thompson's notion of Transitivity, as well as Transitivity itself, were shown to be universal in languages, it is possible to argue that speakers' awareness of these features is high. Furthermore, since the primary function of Transitivity, according to Hopper and Thompson, is to mark the distinction between foregrounded and backgrounded information in discourse, such a distinction is "a universal – having its origins in central communicative and perhaps psychological functions" (Hopper and Thompson 1980:280). Entertaining the idea that it is universal, it is possible to conclude that the features constituting Transitivity are cognitively salient. This can be seen as additional support for the hypothesis stated in the previous section, that speakers employ some kind of a device, scanning utterances for these features, in order to arrive at the interpretation of modality expressed in these utterances. Since these features are important for other language phenomena, unrelated to modality, the existence of such a scanning device seems less implausible. Such a device may already exist to take care of other notions than modality, and its assistance in the interpretation of modality may be a logical spin-off from its primary function.

Secondly, it is possible to analyse my data in terms of Transitivity possibly maintaining the distinction between epistemic and non-epistemic modalities as far as the degree of Transitivity and the discourse function of the utterances with different interpretations is concerned. Thus, 50 occurrences of each modal (*may, must, kan* and *måste*) in epistemic utterances and 50 occurrences of the modals in non-epistemic utterances in main clauses[11] were assessed in terms of the degree of Transitivity of each of these clauses. The average degrees of Transitivity found in these clauses are displayed in Table 5 below.

[11] The choice of main clauses as subjects of this investigation was dictated by the fact that only main clauses have the potential to be foregrounded.

	Degree of Transitivity in epistemic utterances	Degree of Transitivity in non-epistemic utterances
may	1.36	4.53
kan	0.87	3.77
must	2.42	3.54
måste	1.74	3.25
Average	1.6	3.8

Table 5. The degree of Transitivity of epistemic and non-epistemic clauses

Not surprisingly, both epistemic and non-epistemic utterances are relatively low in Transitivity[12], since both these modalities express irrealis[13]. Both epistemic and non-epistemic utterances can be said to comment on (epistemic) or (potentially) modify (deontic in particular) the discourse, thus providing some background information. This is consistent with our understanding of epistemic modality as propositional modality, expressing such notions as speakers' attitudes towards and the degree of commitment to the proposition, and of non-epistemic modalities as event modalities, referring to the events that are merely potential (see Palmer 1990, 2001).

Furthermore, there is a strong statistical tendency for epistemic utterances to have a lower degree of Transitivity than non-epistemic utterances. In many corpus- and cognitively-oriented linguistic theories, frequency is claimed to be a factor to which language speakers are sensitive. Leech and Coates, for example, maintained the notion of quantitative prototype, i.e. 'a particular stereotype which is quantitatively and hence psychologically predominant' (Leech and Coates 1979:88). Accordingly, it is possible to assume that speakers of English and Swedish are sensitive to the difference in the degree of Transitivity between epistemic and non-epistemic modalities, i.e. the speakers are able to pick up the statistical quantitative difference in the degree of Transitivity between the two modalities. In this way, modally modified utterances displaying a low degree of Transitivity will be first interpreted as epistemic on a probabilistic basis. Thus, the degree of Transitivity in an utterance may be seen as yet another feature differentiating between epistemic and non-epistemic modalities.

4 Conclusions

In this study, epistemic utterances were shown to differ systematically from the non-epistemic ones with respect to environments defined in terms of features such as the type and form of subject, aspectual modification, and the type of verb in both English and Swedish. Additional features, such as the animacy of the subject, the presence of an epistemic adverbial or particle, the presence of an implicit or explicit condition in (the immediate context of) an utterance, and time reference, were considered to be of importance for the interpretation of modality. A number of constraints on the interpretation of epistemic modality involving these features were formulated. A scanning device of some sort was suggested that may account for the ease of interpretation of modally modified utterances by native speakers. An argument was presented for the salience of the features that were found

[12] Consider, for example, the results of Hopper and Thompson's investigation, where they found that foregrounded clauses averaged 8.0 points in their degree of Transitivity, whereas backgrounded clauses averaged 4.1 points in their degree of Transitivity (Hopper and Thompson 1980:284).

[13] The basic distinction between the realis and the irrealis is that "[t]he realis portrays situations as actualized, as having occurred or actually occurring, knowable through direct perception", whereas "[t]he irrealis portrays situations as purely within the realm of thought, knowable only through imagination" ((Mithun 1999:173 in Palmer 2001:1).

to distinguish between the two modalities. This argument was supported by the fact that these features are components of the notion of Transitivity (Hopper and Thompson 1980), which has a crucial discourse function. Furthermore, there exists a difference between utterances with epistemic and non-epistemic interpretation respectively in the degree of Transitivity.

References

Coates, J. (1995): "The expression of root an epistemic modality in English", *The verb in contemporary English, theory and description*, eds. Bas Aarts and C. F. Meyer. Cambridge University Press. 145-156.

Coates, J. (1988): "The acquisition of the meanings of modality in children aged eight and twelve", *Journal of Child Language,* June 1988, 15:2, 425-434.

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

Goldberg, A. (1995): A Construction Grammar approach to argument structure. The University of Chicago Press, Chicago and London.

Hopper, P. & S. Thompson (1980): "Transitivity in grammar and discourse", *Language,* Vol.56, Num 2, June 1980, 251-299.

Hoye, L. (1997): *Adverbs and modality in English.* English language series. London and New York: Longman.

Leech, G. & J. Coates (1979): "Semantic indeterminacy and the modals", Studies in English linguistics, ed. S. Greenbaum, G. Leech & J. Svartvik. London and New York: Longman.79-90.

Palmer, F. R. (2001): Mood and modality. 2nd edition. Cambridge University Press.

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

Papafragou, A. (2000): Modality: issues in the semantics-pragmatics interface. Elsvier: Amsterdam-Lausanne-NewYork-Oxford-Shannon-Singapore-Tokyo.

Quirk, R., Greenbaum, S., Leech, G. and J. Svartvik (1985): A Comprehensive Grammar of the English Language. London: Longmans.

Teleman, U., Hellberg, S., & E. Andersson (1999): Svenska Akademiens grammatik. Norstedts Ordbok.

Wärnsby, A. (1999): "The interpretation of *can* in combination with epistemic sentence adverbials in spoken English". Unpublished ms, Department of English, Lund University

Appendix: abbreviations

A - agent
AUX – modal auxiliary
COMP – complementizer
CondCl – conditional clause
DEF – definite
E – epistemic
FUT – future
Ind.NP – indefinite noun phrase
INF – infinitive
Intr.S – introductory subject
MOD – modal
NCl – noun clause

NE – non-epistemic
NEG – negation
NP – definite noun phrase or proper noun
O - object
PART – participle (past or present)
PASS – passive
PAST – past tense
PERF – perfective aspect
PRES – present tense
PROG – progressive aspect
REFL – reflexive pronoun