STOCKHOLM UNIVERSITY

Department of English

"It is certain that it can be argued a million times over" – expressions of epistemic modality in L1 and L2 writing

Tina Ericsson

Degree Project

Linguistics

FW 2008

Supervisor: Alan McMillion

Abstract

This corpus-based study analyzes different types of epistemic markers used in argumentative essays by University students. More specifically it compares Swedish L2 writers and English L1 writers. The scope of the analysis covers epistemic modal verbs, lexical verbs and adverbs. A number of markers are counted to see which expressions are preferred by L1 and L2 writers respectively and if the frequency rates differ between the two groups. Further, it discusses whether the non-native writers use epistemic markers appropriate to an academic register, and an attempt is made to see whether the L1 and L2 writers show similar patterns of 'committing' to and 'distancing' themselves to their arguments. The results reveal a few notable differences between the Swedish and English writers. A tendency is seen among the L2 writers to 'overuse' certain expressions, particularly in the category of lexical verbs. Compared to the native writers, the L2 writers display higher frequency rates when it comes to markers that are most commonly found in spoken conversation. Further the L2 writers seem to display more 'writer visibility' than the L1 writers do, which could perhaps be due to differences in writing culture. The findings also suggest that Swedish L2 writers, even on a relatively advanced level, may have difficulties in mastering modal expressions in English.

Table of contents

1 Introduction

1.1 Background

Understanding the aspects of argumentative writing is a skill needed and expected in the academic world. In this type of texts, the writer has to be able to make claims and then discuss these claims. Doing that involves putting across attitudes and opinions to some degree. Expressions in language that are attitudinal, or non-factual, fall under the concept of *modality*. Modality in a broad sense can be described as an aspect of language through which we can express possibilities, or "the area of meaning that lies between yes and no" (Halliday 1985:335).

Epistemic modality, more specifically, has to do with judgment of the truth in propositions. It is defined by Nuyts (2000:xv) as "a speaker's evaluation of the likelihood of a state of affairs, as expressed in language". Another definition is given by Vold (2006:65), who suggests that "epistemic modality markers [are] linguistic expressions that qualify the truth value of a propositional content". Linguistically they can be expressed in a number of ways. In English, these include among others modal auxiliary verbs (e.g. *might*, *will*), lexical verbs (e.g. *think*, *seem*), and adverbs (e.g. *definitely*, *perhaps*). They all serve to modify propositions and to either strengthen or weaken the truth value of what is being said. Moreover, certain types of pronouns (e.g. *everybody*, *all*, *nothing*) may also serve as epistemic markers. Longer phrases, e.g. *if you know what I mean* (Hinkel 2005), can also be used as such expressions. Regardless the type of expression, they are all used by writers/speakers as a means of "commenting on and committing or distancing themselves from their propositions" (Hyland 1998:2).

Modality is a relatively complex aspect of language, as "writer commitment can be expressed in an enormous variety of ways and these expressions can convey a wide range of meanings" (Hyland and Milton 1997:184). It has further been claimed that native speakers of a language have an early acquisition of these modal meanings (Papafragou 1998:375). In view of that, and because of the fact that different languages have different ways of expressing modal meanings, second language learners of English are said to often have problems with such expressions in language production (Hyland and Milton 1997:184).

In brief, the topic of this essay is to investigate epistemic modality markers in second language (L2) versus first language (L1) written texts. In particular, it compares Swedish and English students. The scope of the analysis is limited to certain types of epistemic markers, namely modal verbs, lexical verbs and adverbs. It will become clear that even though Swedish

and English are structurally quite similar languages, there are still differences in the preferences and frequencies of epistemic markers among the L2 writers compared to the L1 writers.

1.2 Previous research

Modality is an area within linguistics that has been given quite a lot of attention. Many researchers have focused on modal auxiliary verbs, as opposed to other modality markers. It has been suggested that one reason for this is "the fact that the verb is the most central part of the sentence" (Palmer, quoted in Hoye 1998:3). In English, moreover, modal verbs represent the most central means used to express modal meanings (Aijmer 2002:72).

Several researchers have dealt with the polysemous nature of the modal verbs, i.e. their various semantic meanings. One author is Coates, who has presented a detailed outline of the English modal auxiliaries (1983). In her corpus-based study, she discusses the numerous meanings of the modal verbs, which include for instance possibility, probability, ability, permission and obligation (for a more detailed discussion of these meanings, see Coates, 1983:25). Researchers seem to agree that "meanings do not reside in the modals themselves, but are assigned to utterances which contain them" (Hyland 1998:106). In other words, many of the modal expressions may or may not be epistemic depending on the context they appear in. Much research within the area has therefore focused on how to interpret and distinguish these different meanings.

Epistemic modality has been studied in connection to second language writing. Hyland and Milton (1997) investigate expressions of certainty and doubt in L1 and L2 students' argumentative writing. The skill in this type of writing is to "present claims that are neither overstated nor understated in relation to evidence or reasonable assumption" (Hyland and Milton 1997:186). To make an overstatement is often referred to as "boosting" or "intensifying" while understating a claim can be denoted "hedging" (cf. Crompton 1997; Hinkel 2005; Hyland 1998; Hyland and Milton 1997). In Hyland and Milton's study, English L1 writers are compared to Chinese L2 writers of English. The results show that the native speakers differ notably in some respects from the non-native speakers in their use of epistemic markers. Hyland and Milton find that the L1 writers have a greater variation of markers in their texts than the L2 writers. They also observe that the L2 writers, through their choice of epistemic expressions, make stronger claims than L1 writers (1997:193). In other words, the non-native writers appear to boost their arguments more than the native writers do, by preferring stronger modal expressions. Strong modals refer to those that convey a high degree

of certainty (e.g. *will*, *indeed*) compared to those conveying uncertainty (e.g. *might*, *perhaps*). In their study, Hyland and Milton present "discrete epistemic categories, locating particular patterns and lexical items on a scale extending from maximum to minimum certainty" (1997:192). As the researchers point out, however, the semantic meanings of modal expressions are sometimes ambiguous, and these categories may overlap.

Another study is done by Hinkel (2005), who analyzes the use of hedges and intensifiers ('boosting' devices) in academic essays of L1 and L2 writers. Similar to Hyland and Milton, Hinkel finds that the non-native writers are more limited in their use of epistemic markers (particularly hedging devices) than the native writers, and that they to a greater extent use expressions common in conversational discourse rather than an appropriate academic register (2005:47).

In a Swedish-English perspective, corpus-based research on epistemic modality has been carried out by Aijmer. She analyzes advanced L2 writers' use of the expression *I think* (2001), as well as modal auxiliaries and epistemic adverbs (2002). These studies point to an overuse (a significantly higher frequency in use) among Swedish L2 writers when it comes to certain types of modal expressions. Among other things, the overuse of the expression *I think* makes the writing style of the Swedish L2 writers more personal and direct than that of the native writers, meaning that the writer becomes 'visible' in the text, sometimes to a degree not appropriate in academic writing.

2 Aim and scope

The aim of this essay is to investigate how epistemic modality is expressed in argumentative texts written by University students. More specifically, it looks at the frequency of epistemic markers in texts produced by Swedish students writing in English as their L2 compared to English students writing in their L1. As noted above, the epistemic markers included in this study are modal verbs, lexical verbs and adverbs. They are counted as separate items, but analyzed in their context (as part of a phrase, clause or sentence). This essay will neither make an assessment of grammatical correctness nor evaluate whether the markers are used with an appropriate degree of certainty/uncertainty. Further, modality markers often occur in combinations, such as *could perhaps* or *will certainly*. The limited scope of the essay does not allow any discussion of how such combinations work.

In many studies comparing native and non-native English writers, the L2 writers are native speakers of languages that are quite distinct from English, e.g. Chinese, Japanese and Arabic. From this perspective, it is interesting to compare two languages that are structurally similar, such as English and Swedish. As mentioned above, some research has been done in a Swedish-English perspective, analyzing particular expressions in-depth as well as modality in a broader sense.

The focus of this study is to look at a relatively large number of epistemic markers, to get an overview of the different kinds of expressions used by L1 and L2 writers of English. Additionally, an attempt is made to compare the results both to other existing studies on Swedish L2 usage of modality (cf. Aijmer 2001; 2002) and to studies of East-Asian and Arabic L2 usage of these language features (cf. Hyland and Milton 1997; Hinkel 2005).

Since it has been established that L2 writers often have difficulties with modal expressions, studies of this kind are important because they may have implications for L2 learning and teaching. From this standpoint, I set out to answer the following questions:

1. Do the L1 and L2 writer use the same selection of epistemic markers, and do these markers occur with the same frequency?

2. Do the L1 and L2 writers choose epistemic modal verbs, lexical verbs and adverbs with the same variation?

3. Are there differences in the extent to which L1 and L2 writers choose epistemic markers associated with either conversational or academic style?

4. Are there differences in the extent to which L1 and L2 writers either commit to or distance themselves from their claims?

3 Material

The material used in this study consists of two corpora, one with essays written by native speakers of English and the other with essays by non-native speakers of English. The essays were produced by first semester students at the English Department at Stockholm University and were originally part of a research project carried out there. They are compilations of short argumentative essays, all on the same topic, i.e. to discuss the question "Is it true that only rich countries can afford to worry about the environment?". The non-native speakers wrote their essays on two occasions in 2002 and the native speakers wrote theirs on one occasion in 2005. For this study, only the first of the two sets of L2 essays has been used for analysis.

In the L1 corpus there are 30 essays, 8882 words in total, while the L2 corpus adds up to 20888 words, comprising 107 essays. The mean length of the L1 essays is 296 words; the L2 texts are somewhat shorter, with a mean length of 195 words. All the essays were hand-written by the students and thereafter transcribed into computer files. In the material a few spelling errors were detected and counted, but there could be additional ones not discovered.

It should be noted that, in comparison to the material used in the other studies mentioned throughout the essay, the corpora used for analysis in this study are relatively small. Also, there are limitations as far as text type is concerned. Argumentative texts belong to a quite narrow genre, and the conclusions drawn in this study only refer to this type of writing. Thus the results should be read with some caution, since they can merely point to tendencies, not provide exact answers.

Moreover, I have only limited information regarding the students' age, academic and social background, and writing experience, and these are aspects that may influence the data. The essay writers are probably not a completely homogenous group in this respect, and it is important to keep this in mind when reading the results. On the other hand, the groups are comparable in view of the fact that all students were given the same instructions on how to write the essay, and they all had 30 minutes to complete the task.

4 Method

Before the actual search for epistemic markers in the material could be conducted, several steps were taken. To begin with, the L1 and L2 texts were arranged in two separate text files. This preparation included removing elements that were not produced by the subjects themselves, such as comments added by those responsible for word-processing the texts. Also, a couple of L2 essays occurred twice in the raw material, and these detected doubles were removed.

4.1 Selecting epistemic markers

It was then decided which epistemic markers to include in the analysis. On this matter, a first decision was made to include only single-word items and not longer phrases. The exceptions are the adverbial expressions *of course* and *in fact* which are in fact adverb phrases, though in this study treated as single units (cf. Hoye 1997:205).

Furthermore, in selecting markers, I have leaned on the work of Hyland and Milton (1997). Their study of epistemic expressions includes modal verbs, lexical verbs, adverbs, adjectives and nouns, and within these grammatical classes they present a list of the most frequently used epistemic items in academic writing (1997:205). This list was used as a basis in selecting markers in this study. A pilot calculation on my material showed, however, that epistemic nouns and adjectives were not as frequent as the other classes, in line with the findings of Hyland and Milton. Thus, to limit the scope of the analysis, only modal verbs, lexical verbs and adverbs were included.

Finally, Hyland and Milton's list was modified slightly in regards to individual markers. Two items not listed by the researchers, *really* and *maybe*, were found frequently in the L2 corpus and therefore added to the record. Contrastively, some markers listed by Hyland and Milton did not occur in either of my corpora. Needless to say, only markers that occurred at least once in either corpus were recorded for further analysis.

4.2 Recording epistemic markers

As observed by Vold (2006:65), many of the markers in question are polysemous, i.e. they have different meanings attached to them. Because of this, it was necessary to distinguish the epistemic instances from the non-epistemic ones. In many cases this turned out to be quite problematic. Therefore, drawing on various theories and ideas, a set of criteria for classifying the markers was eventually developed (see 4.3.4. below).

The computer program WordSmith Tools 4 (Scott 2004) was then used to count instances of epistemic expressions in the two corpora. More specifically, a concordance function enabled me to search for every item separately and then see all occurrences in context (sentential or larger if needed). Instances of each marker in the L1 and L2 corpora were compared side by side, one marker at a time, to ensure that they were assessed in the same way, applying the criteria. Inflected verb forms (e.g. *knows*, *argued*), contractions and negations of modal verbs (*couldn't*, *won't*), as well as a number of detected misspellings were also included in the count.

4.3 Development of criteria for epistemic classification

Many different theories and points of view exist in as regards interpreting epistemic modality markers. Below follows a short discussion of a few of the concerns regarding epistemic classification. It demonstrates some of the issues taken into account when developing the criteria which were thereafter used to separate epistemic instances from non-epistemic ones.

4.3.1 Some aspects concerning modal verbs

Hyland (1998:105) points out that modal verbs can have either epistemic meaning or root meaning, and that root meanings of modal verbs involve "the will, ability, permission or obligation to perform some action or bring about some state of affairs" (c.f. Barbier 2002:11 and Nuyts 2001:25). Among the root meanings, the one most commonly referred to is deontic meaning. Hoye points out that deontic modality typically has to do with an action or event, rather than a claim about things being a certain way, and that the "point of issue is not whether something is true but whether something is going to be done by others, or by the speaker" (Hoye 1997:46; c.f. Hyland 1998:44). The problem of classifying modal verbs as either epistemic or deontic is that "sentences are often ambiguous between the two readings" (Barbier 2002:11). A sentence of that type is presented below.

(1) You *should* be there by 4 o'clock.

The sentence above could either be interpreted as meaning 'it is likely that you will be there by 4 o'clock' or it could be interpreted as 'you are obliged to be there by 4 o'clock'. Distinguishing between examples such as these in written text is not an altogether straightforward task, since it requires knowledge of the writer's intention, which is not always reflected clearly in the text.

When conducting the count of epistemic markers in this study, many ambiguous instances were encountered. One concern was the auxiliary verb *will*, which may have epistemic meaning, or simply denote future tense. As Hyland (1998:116) states, "it is extremely difficult to distinguish an epistemic from a future interpretation where will refers to a future action, as reference to the future inevitably involves some uncertainty or doubt". In ambiguous cases, I applied a couple of strategies used by Vold, which involved either substituting the modal in question with a different one, or adding a new epistemic expression to the sentence (2006:72). The idea was to see if it was possible to find 'overlapping' expressions that gave the sentence epistemic meaning without altering the basic meaning of it. To illustrate this, two contrasting examples from the corpora are given below. In example (2) we can for instance add the epistemic adverb *perhaps* or change the modal auxiliary to *may*, and the result appears to be a more hedged version of the sentence. However, conducting the same procedure in example (3) seems to alter the basic meaning of the sentence, from denoting intention to expressing possibility/probability. As far as these examples are concerned, (2) has been classified as epistemic while (3) has not.

(2) This *will* bring them the same problems with garbage that we are struggeling with in the western world today. (L2)

(3) This short essay *will* analyse the question in both contexts. (L1)

4.3.2 Some aspects concerning lexical verbs

Lexical verbs such as *think*, *believe*, *appear*, *tend* can be used to express epistemic modality. As in the case of modal verbs, however, lexical verbs may have meanings other than epistemic attached to them. For instance, *appear* conveys epistemic meaning in (4) but not in (5) below.

(4) It *appears* that all students passed the exam.

(5) Suddenly he *appeared* on stage.

Another concern in epistemic classification of some lexical verbs is what Crompton refers to as "the issue of responsibility for utterance" (1997:283). He suggests that verbs such as *claim*, *believe* and *suggest* are epistemic only when writers "use them to report their own proposition" (1997:283). Crompton gives the following examples, and asserts that the first sentence counts as epistemic, while the second does not.

(6) *I suggest* that pigs fly.

(7) *Smith suggests* that pigs fly.

This view is not shared by all researchers. Hyland and Milton bring up similar examples, but contrary to Crompton, they claim that the writer may use the opinion of someone else as a "means of disguising the epistemic source" (Hyland and Milton 1997:283). In this study, Hyland and Milton's view was adopted, partly to make my results as comparable as possible to theirs. Consequently, in a construction such as the one below, *knows* was recorded as epistemic:

(8) And it's neither the knowledge of how "the environment" works, because almost *everybody knows* that we have to take care of where we live. (L2)

4.3.3 Some aspects concerning adverbs

Concerning the adverbs, one difficulty of epistemic classification lies in the various functions of the adverbs, depending on their placement in a clause. The adverbs may be so called sentence adverbs, modifying the whole sentence or clause, of they may be word or word-group modifiers, modifying only a specific part of the sentence (Hoye 1997:145). As Lyons states, sentence adverbs "are used by the speaker in order to express, parenthetically, his opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes" (1977:452). It would be hard to argue that the sentence adverbs in the examples given below, extracted from the corpora, do not have epistemic meaning:

(9) *Clearly*, wealthy countries have greater financial and research resources to use in environmental protection projects. (L1)

(10) *Of course* it is true that the poor countries have lots of other primary problems to think about, like hunger, diseases and poorness. (L2)

However, some of the sentence adverbs sometimes function as word modifiers, and vice versa, and thus it is often hard to distinguish when these markers are used to modify only a part of the sentence or the whole proposition (Quirk et al, 1972:440). One of the examples given by Quirk et al is:

(11) I can't *really* believe him.

In the example above, the adverb *really* can be seen as modifying the modal negation, the main verb, or even the whole clause. On this matter, there are researchers who argue that "epistemic interpretation only occurs with sentence adverbs" (see e.g. Drubig 2001:10), and others who assert that word or word-group modifiers can also convey epistemic meaning (see e.g. Hyland 1998:138).

For this study, adverbs of this second category were also recorded as epistemic, when they occurred in sentences such as (12) and (13) below:

(12) Many environmental initiatives *actually* reap monetary rewards in their own right. (L1)

(13) But in the sametime it's *really* important for the rich countries to help the poor ones. (L2)

4.3.4 Adopted criteria for epistemic classification

It is beyond the scope of this essay to argue in depth for a particular set of categories for epistemic markers. However, the following criteria were used consequently when classifying the markers:

1. For a marker to be classified as epistemic, it had to either strengthen or weaken the truth value of a proposition (c.f. Vold, 2006:65).

2. Markers were excluded if they unambiguously conveyed deontic or any other meaning other than epistemic.

3. In ambiguous cases, where an epistemic reading was possible, but not entirely clear-cut, the marker was still recorded as epistemic.

4. In the case of lexical verbs, markers were counted if they added to the proposition an attitude or opinion put across by the writer either explicitly, e.g. *I believe that...*, or implicitly, e.g. *some people argue....*

5. Adverbs were recorded as epistemic both in cases where they modify a whole sentence or clause, as well as only part of a sentence or clause, as long as criterion no. 1 was realized.

6. Modality markers occurring in questions were not recorded as epistemic. One could argue that for instance tag-questions and rhetorical questions may convey some epistemic meaning. However, with reference to Halliday (1985:86), who asserts that "in a statement the modality is an expression of the speaker's opinion … whereas in a question it is a request for the listener's opinion", questions of all kinds were excluded from the count.

The principles above were not chosen at random, as different theories and views were evaluated before deciding on this particular set. However, it is not certain that they agree completely with the principles employed by other researchers in the studies used for comparison. An effort was made to draw as much as possible on the views of Hyland and Milton (1997), Aijmer (2001; 2002) and Hinkel (2005), but these studies did not always give an account on how epistemic meaning was distinguished.

5 Results and discussion

All in all, 46 distinct markers are included in this analysis (for a complete list see Appendix).
Occurrences of these markers in the two corpora have been counted, and epistemic
occurrences distinguished from non-epistemic ones. The results of these counts are presented
in Table 1 below. Over 30 % of the markers in the L2 corpus and over 25 % of the
occurrences in the L1 corpus have not been classified as epistemic, and will therefore not be
accounted for any further in the analysis. A presentation of these numbers seems important,
however, since my interpretation of the expressions is not the only one possible. In other
words, the results would most likely turn out differently if a different set of criteria was used
for epistemic classification.

Table 1
Ratio of epistemic markers compared to total no. of markers

	L1	L2
Total no. of markers /1000 words	27.70	34.04
No. of epistemic markers /1000 words	20.49	23.08

Due to the fact that the corpora differ in size, it is not of interest to compare raw
numbers. Instead the number of occurrences per thousand words is given. The numbers in
Table 1 show that the epistemic markers counted are somewhat more frequent among the non-
native writers. About 20 words of 1000 in the L1 corpus are epistemic markers, compared to
23 words of 1000 in the L2 corpus.

Concerning the question of whether or not L2 writers are less varied in their use of
epistemic expressions, this appears to be the case, but only with a slight difference compared
to L1 writers. A general remark is that, of the 46 epistemic markers included in the study, the
L1 writers employ 39 of them, and the L2 writers make use of 38 of them. Table 2 below
presents the 10 most frequent epistemic markers ranked by their rate of recurrence in native
and non-native texts respectively. Again, occurrences per thousand words are given. If we
sum up the items ranked 1-5 in each group, it shows that these markers constitute just above
half of the total usage in both L1 and L2 texts. Thus so far the variation of epistemic
expressions seems relatively equal between the Swedish and English students. A look at the
frequency rates of the top 10 expressions, however, will reveal a difference in the L2
compared to L1 corpus. The items ranked 1-10 constitute about 72 % of the L2 writers' total
use of epistemic markers. The corresponding figure among the L1 writers is 66 %. This is

comparable to the results of Hyland and Milton's study (1997:189), although my study shows a smaller difference between L1 and L2 writers in this respect.

Table 2
Most frequent epistemic markers

L1 (native speakers)			L2 (non-native speakers)		
Rank	Item	#/1000	Rank	Item	#/1000
1	would	3,94	1	think	3,35
2	may	2,14	2	will	3,11
3	will	1,91	3	of course	1,96
4	believe	1,24	4	would	1,91
5	often	1,13	5	believe	1,44
6	seem	0,90	6	probably	1,15
7	perhaps	0,68	7	maybe	1,10
8	think	0,56	8	might	1,01
9	appear	0,56	9	really	0,96
10	could	0,45	10	may	0,62
1-10	*Top rank*	13,51	1-10	*Top rank*	16,61
11-46	*Other markers*	6,98	11-46	*Other markers*	6,51

A closer look at Table 2 reveals that there are differences regarding which specific markers are preferred by L1 and L2 writers. At the very top of the L1 list we find three modal verbs, *would*, *may* and *will*. It appears that the native writers prefer modal verbs as a means of expressing certainty or doubt. The L2 writers also make use of the modals *will* and *would*, but there are two additional markers that stand out at the very top of the list; the lexical verb *think* and the adverb *of course*, both of which rank low among the native writers.

Concerning the variation of markers used by each essay group, it is also interesting to note that there are differences in word class distribution, i.e. the inclination to choose modal verbs, lexical verbs or adverbs. We can see that, even though L2 writers make extensive use of the epistemic marker *think*, and to some extent also *believe*, these are the only two lexical verbs present on the top 10 list in Table 2. The L1 writers on the other hand display a more mixed palette of lexical verbs: *believe*, *seem*, *think* and *appear* are all among the top 10 most frequently used epistemic markers. Before discussing specific markers in more depth, however, it is useful to have a look at how the markers are spread across the three grammatical classes accounted for. This distribution is shown in Figure 1 below.

Figure 1

**Grammatical distribution of epistemic markers.
Chi-square test tells whether difference is statistically significant:
p-value <0.050 = significant.**
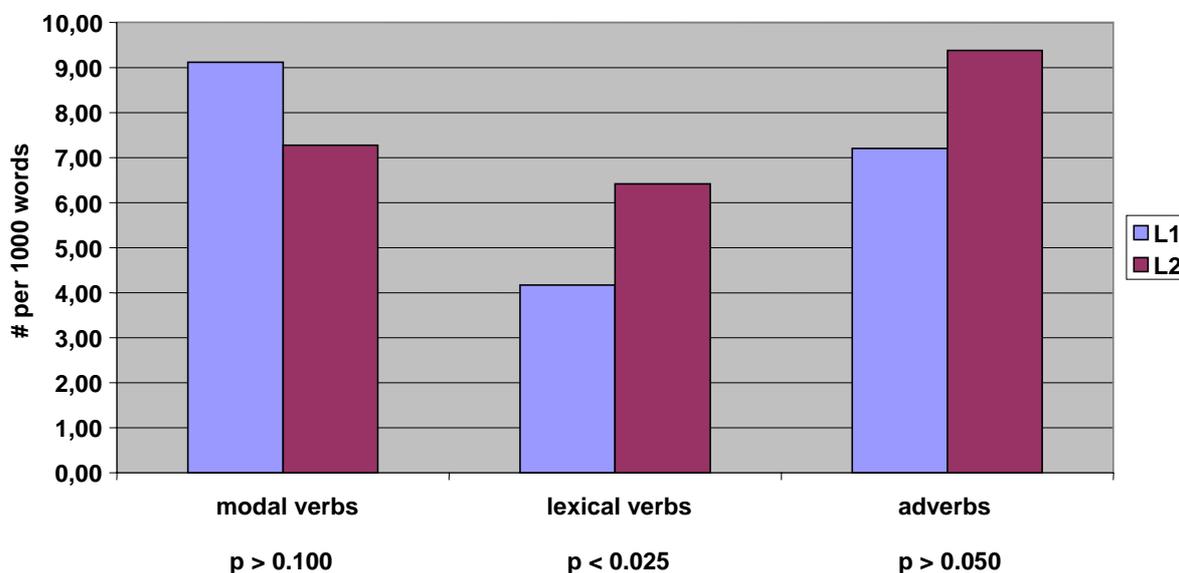
As Figure 1 shows, modal verbs are the most common means used by L1 writers to express epistemic modality. The L2 writers, on the other hand, seem to prefer epistemic adverbs. Concerning lexical verbs, this is the least frequent category among both native and non-native writers. At the same time, it is the category that shows the most notable difference between L1 and L2 writers.

In the case of modal verbs, *would* and *will* constitute the two most frequent epistemic markers occurring in the two corpora regarded together. *Would* is by far the most common marker in the L1 corpus, appearing about twice as often as *will*. In the L2 corpus *will* is the second most frequent marker, occurring 1.5 times as often as *would* does. The discrepancy between *will* and *would* in the L1 and L2 texts corresponds to some extent to the findings of Hyland and Milton (1997:188). First of all, the L2 writers in their study also employ the marker *would* with a lower frequency than the L1 writers do. Additionally, as with the Swedish L2 writers of this study, the Chinese writers also use epistemic *will* to a greater extent than the native writers. According to Hyland and Milton, the preference to use *will* over *would* indicates that the L2 writers make stronger claims, and that L1 writers consequently tend to be more cautious in their arguments (1997:188). Similarly, Aijmer's study reveals that the L2 writers prefer *will* over *would*, although she finds an overuse of both these modals

among L2 writers compared to the native writers. Since Aijmer's analysis includes epistemic as well as root/deontic uses of the modals (2002:64), however, a direct comparison to her results on this particular matter is difficult.

Another observation in connection with modal verbs is that the native writers make extensive use of *may* in comparison to *might*. Among the non-native writers, *might* is the preferred marker of these two, even though it ranks relatively low. In this case there seems to be a disagreement with Hyland and Milton's study, in which both L1 and L2 writers use *may* more frequently than *might* (1997:196). Interestingly, Aijmer also reveals an overuse of *might* among Swedish writers. One possible explanation given by Aijmer is that there could be differences in writing culture, and that L1 and L2 writers may have "different preconceptions about the degree of directness and certainty required in academic writing" (2002:63).

The category of lexical verbs shows the greatest discrepancy between L1 and L2 writers. Regarding individual markers, there are also differences in which lexical verbs are preferred by the two groups.

The most popular epistemic item among non-native writers is the lexical verb *think*, occurring six times more often in their essays compared to those of the native students. The excessive use of *think* among L2 writers is similar to the findings of Hyland and Milton. Their results demonstrate that epistemic *think* is four times more common among the Chinese students than among the English students (1997:189). Interestingly, Aijmer's study of L2 learners points to almost exactly the same variation between Swedish and English students (2001: 248). An observation should be made, however, regarding the results of my study. Apart from the fact that *think* is six times more frequent in my results as compared to four times in the studies of Aijmer and Hyland and Milton, it is notable that the frequency of *think* in my L1 corpus is twice as high as in the other two studies. In this case it is hard to say whether my results would be comparable with Aijmer's, whose L2 writers are also Swedish, or with Hyland and Milton's, whose L2 writers are closer to my subjects when it comes to educational level. In the case of Aijmer, the differences could at least partly be due to the fact that her L2 writers are more advanced than the L2 writers of this study. However it could also depend on other aspects, such as differences in the writing task, as well as length of the essays.

The marker *think* is almost exclusively found in the construction *I think (that)…*. The following sentence is a typical example extracted from one of the L2 essays:

(14) *I think that* everybody can afford to worry about the environment since we all have
to go on living on this planet.

As pointed out by Hyland (1998:127) and Aijmer (2002:71), *I think (that)* is commonly used in spoken conversation, though not as frequent in the more formal, academic register. Therefore it is reasonable to believe that "the overuse of *I think* is due to influence from spoken language" (Aijmer 2002:71), and that non-native writers may have more difficulties in knowing which expressions are appropriate to different registers.

Further, Aijmer states that, even though *I think* may be employed both as a hedging or boosting device, it is primarily "a marker of certainty and authority", used to formulate convincing arguments (see Aijmer 2001:256 for a more detailed discussion of the different functions of *I think*). This expression also makes the style of L2 texts more straightforward and 'personal', a technique which L2 writers may find easy (Aijmer 2001:256). In connection with this it is also suggested that there may be differences in the writing cultures of the native and non-native writers. It is possible that Swedish and other writing cultures allow "writer visibility" in the text to a greater extent than English does (Aijmer 2001:256; 2002:71). Additionally, Hyland and Milton make a similar comment on the fact that the Chinese L2 writers use such personal forms more often than the L1 writers of their study (1997:199).

Epistemic *believe* is used in similar constructions as *think*, i.e. *I believe (that)* followed by a proposition. Because of the similarities of these two epistemic markers, perhaps one could have expected a difference in L1 compared to L2 usage in this case as well. It is interesting to note though that this expression is used with almost the same frequency in both L1 and L2 texts.

In addition to *I think* and *I believe*, the L2 writers also make use of *I guess* and *I know* in a similar way. These expressions are absent in the L1 texts.

Concerning L1 usage of epistemic lexical verbs, the item *appear* stands out in regard to frequency ratio between L1 and L2 writers. Although not present at the very top of Table 2, this marker occurs much more frequently among native writers than among non-native writers. In fact, only one instance of epistemic *appear* is found in the L2 texts, compared to five in the L1 texts. If we consider the difference in size between the two corpora, this points to *appear* occurring almost 12 times as often among L1 writers. Another epistemic lexical verb preferred by the L1 writers is *seem*, occurring almost twice as often in their texts compared to L2 texts. However, since the occurrences are very few in both these cases it is only possible to talk about tendencies.

Thus it becomes clear that L1 and L2 writers differ in some respects in their use of epistemic lexical verbs. The differences are shown first and foremost by the L2 writers'

extensive use of the expression *I think*, which is claimed to be used most of all as a marker of certainty (Aijmer 2002:256). Secondly, items such as *appear* and *seem*, which are both seen as hedging devices (Hyland 1998:125), occur less frequently in the L2 texts compared to the L1 texts.

As illustrated in Figure 1 above, the total use of epistemic adverbs is more frequent among L2 writers though the difference is not remarkable. Table 2 shows that among the top 10 markers the non-native writers make use of adverbs almost three times as often as the native writers (5.17 vs. 1.81 per thousand words), and it also shows that among these top 10 markers the L2 writers use twice as many individual adverbs.

The most noticeable epistemic adverb among the L2 writers is the item *of course*, followed by *probably*, *maybe* and *really*. Together these four items constitute over half of the epistemic adverb expressions used in the L2 corpus. In contrast, these markers all rank low among the L1 writers, and comprise no more than 1/8 of their total use of epistemic adverbs. When using epistemic adverbs, the native writers instead seem to rely on expressions such as *often* and *perhaps*. Moreover, the native writers use a number of adverbs which, although not the most frequent ones, do not occur at all in the non-native corpus. Examples of such adverbs are *largely*, *presumably*, *indeed* and *clearly*. This would indicate that the L1 writers still demonstrate a wider range of epistemic adverbs than the L2 writes, even though they do not employ the adverb category as frequently.

In addition to being the third overall most frequent epistemic marker among the Swedish writers, *of course* is nearly nine times more common among them compared to the native writers. It is suggested that *of course* functions as a rhetorical device, "easing the receiver's approach to the argument" (Barbaresi quoted in Aijmer 2002:70). The following sentence is an example from the L2 corpus:

(15) The grand example is *of course* the US, where people are used to live the life of megaconsumption giving not a single thought to earth their walking on.

Regarding *maybe*, it should be noted that while it is one of the most frequently used expressions among L2 writers, not one instance of the item was found in the L1 corpus. A typical L2 example is:

(16) *Maybe* the solution to the lack of interest in the environment, could be to cooperate more.

In addition, *probably* is five times more common in the L2 texts, and *really* is twice as common in L2 texts.

The frequency in L2 usage of these adverbs is in line with the findings of Aijmer (2002). In her study, Aijmer shows that the expressions *probably*, *maybe* and *of course* are all overused by Swedish L2 writers compared to English native writers (2002:68). Similarly, in their study on L2 learners' use of adverbial connectors, Altenberg and Tapper demonstrate that *of course* is found more frequently among Swedish L2 writers than among English L1 writers (1998:86).

However, it seems that Swedish L2 writers of English are not the only ones overusing these epistemic adverbs. Gilquin and Paquot also show that the markers *of course*, *maybe*, *really* are much more frequent in L2 writing than in L1 writing (2007:3). Their study includes L2 writers of various mother tongues, among them Chinese, French, Russian and Swedish. They assert that the adverbs found frequently among L2 writers are more common in conversational discourse than in a formal register, and that the overuse may be due to a "lack of register-awareness" among the L2 writers. A similar argument is offered by Hinkel (2005:46), whose non-native writers (East-Asian and Arabic) overuse markers that are more common in and appropriate to conversational and informal registers, e.g. *really*.

Hyland and Milton's study includes the epistemic markers *probably* and *of course*, though no further details are presented on either L1 or L2 usage concerning the latter. However, an observation can be made concerning the marker *probably*. Hyland and Milton's results show that this individual marker is quite rare among L1 and L2 writers alike, and that both the English and the Chinese students use the expression with roughly the same frequency (1997:196). In fact, the Swedish L2 writers use the item *probably* almost three times more often than the Chinese L2 writers.

As a final remark, then, the distribution of epistemic adverbs reveals clear differences between the L2 and L1 writers in this study. Above all they show a tendency among the L2 writers to use expressions more common in spoken language than academic writing.

6 Conclusion

This essay has focused on an investigation of the use of epistemic markers in argumentative texts written by Swedish students (L2 writers) and English students (L1 writers). The results indicate several differences between the two groups. Firstly, it appears that the L2 writers are slightly more restricted in their choice of specific markers since they make use of fewer markers than do the L1 writers. This smaller set of markers is used with relatively high frequency, however. This can be seen most clearly in the case of epistemic lexical verbs, where the L2 students primarily rely on a single expression, viz. *I think*. Similarly, a small number of epistemic adverbs, *of course*, *really*, *probably* and *maybe*, are also used with relatively high frequency, perhaps "overused", by the L2 writers.

Regarding the distribution of markers across grammatical classes, it should be noted that only one category (lexical verbs) showed statistically significant differences. Nevertheless, it appears that the L2 writers use adverbs and lexical verbs as their principal means of expressing certainty and doubt, while the L1 writers seem to prefer modal auxiliary verbs. It also appears that the L2 writers' preferences concerning certain modal verbs differ from those of the L1 writers. The L2 writers' tend to use *will* more frequently than *would*, while the L1 writers prefer the latter. A similar difference concerns the uses of *may* and *might*, with the L2 writers preferring *might*.

Although any explanation for these differences in the employment of epistemic markers between L1 and L2 writers would be speculative at this point, some possibilities should be mentioned. Regarding the uses of modal verbs, we see that these devices are commonly used by the L1 writers to hedge arguments, but not by the L2 writers. A possible reason is that the L2 writers choose English epistemic markers that 'match' expressions of certainty and doubt used in Swedish, their mother tongue. In English, modal verbs are often assumed to be the primary means of expressing modal meanings. In Swedish, on the other hand, modality is "more often realized as an adverb or as an adverb plus a modal verb" (Aijmer 2002:72). Consequently, one might conclude that the Swedish writers have not mastered the modality potential of English modal verbs. In other words the L2 writers may not have mastered the vague, 'fuzzy', and often subtle meanings of the modal verbs. The very noticeable differences in the use of *may* and *might* indicate that L2 writers are not sufficiently aware of the semantic differences between the two auxiliaries.

Furthermore, it appears to be the case that the L2 writers, to a greater extent than the L1 writers, prefer a more personal style when arguing their case. Based on the high frequency of such items, it could be assumed that the Swedish writers use expressions such as *I think (that)* and *of course* in order to affect a straightforward and personal style of writing. This may depend on differences in writing culture and educational traditions. Such devices make the writer more 'visible' in the text and are apparently acceptable in Swedish essay writing. However, such visibility is usually avoided in English L1 argumentative writing (cf. Aijmer 2001:256). Interestingly, the tendency to enhance the visibility of the author has been found in studies of East-Asian and Arabic L2 writers as well (Hyland and Milton 1997; Hinkel 2005).

Another aspect of explanation is that of "register-awareness" (Gilquin and Paquot 2007:3). The results of the present study indicate that L2 writers use markers that are common in spoken language rather than those of standard L1 academic writing. The extensive use of such markers as *maybe*, *really* and *I think* by L2 writers may be indicative of a lack of awareness regarding what markers are appropriate for the text type they are producing. However, use of such spoken language markers does not seem to be characteristic of Swedish L2 learners exclusively, since this tendency is also found among English L2 from many different linguistic and cultural backgrounds (Hinkel 2005:46).

Both the present research and the previous research discussed above indicate that regardless of language and cultural background, L2 writers of English seem to share certain tendencies in the use of epistemic markers. In addition to an overuse of features common in conversation, the research suggests that there is also a tendency for the L2 writer to overuse writer-visibility markers, at least to an extent that is more extreme than what is considered appropriate for English academic writing.

This study has focused on written production and has not considered the receptive side of epistemic modality markers. Whether there may be a corresponding mismatch between L1 and L2 readers of epistemic modality markers in comprehension is an open question.

Future research on L2 use of English epistemic modality markers should include a survey of the extent to which epistemic modality is explicitly taught in primary and secondary school English classes. If students are left to learn English epistemic modality entirely implicitly, there may be considerable transfer effects, which may account for the data described above. The importance of clear, and often subtle, use of epistemic markers in academic English would seem to make this an area for further research.

References

Aijmer, Karin. 2001. "*I think* as a marker of discourse style in argumentative student writing".
In Aijmer, Karin (ed.). *A wealth of English: Studies in Honour of Göran Kjellmer.*
Göteborg: Acta Universitatis Gothoburgensis, 247-257.

Aijmer, Karin. 2002. "Modality in advanced Swedish learners' written interlanguage". In
Granger, Sylviane; Hung, Joseph; Petch-Tyson, Stephanie (eds.). *Computer Learner
Corpora, Second Language Acquisition and Foreign Language Teaching.* Amsterdam:
John Benjamins Publishing Company, 55-76.

Barbiers, Sjef. 2002. "Current issues in modality". In Barbiers, Sjef; Beukema, Frits; van der
Wurff, Wim (eds.). *Modality and its Interaction with the Verbal System.* Amsterdam:
John Benjamins Publishing Company, 11-27.

Coates, Jennifer. 1983. *The Semantics of the Modal Auxiliaries.* Beckenham: Croom Helm.

Crompton, Peter. 1997. "Hedging in academic writing: some theoretical problems". *English
for Specific Purposes* 16(4), 271-287.

Gilquin Gaëtanelle; Paquot Magali. 2007. "Spoken features in learner academic writing:
identification, explanation and solution". Available: http://ucrel.lancs.ac.uk/publications/
CL2007/paper/204_Paper.pdf [2008, Sept. 2]

Halliday, M.A.K. 1985. *An Introduction to Functional Grammar.* London: Edward Arnold
(Publishers) Ltd.

Hinkel, Eli. 2005. "Hedging, Inflating, and Persuading in L2 Academic Writing". *Applied
Language Learning* 15, 29-53.

Hoye, Leo. 1997. *Adverbs and Modality in English.* London: Longman.

Hyland, Ken; Milton, John. 1997. "Qualification and certainty in L1 and L2 students'
writing". *Journal of Second Language Writing* 6(2), 183-205.

Hyland, Ken. 1998. *Hedging in Scientific Research Articles.* Amsterdam: John Benjamins
Publishing Company.

Lyons, John. 1977. *Semantics.* Cambridge: Cambridge University Press.

Nuyts, Jan. 2001. *Epistemic Modality, Language, and Conceptualization: A cognitive-
Pragmatic Perspective.* Amsterdam: John Benjamins Publishing Company.

Papafragou, Anna. 1998. "The acquisition of modality: Implications for theories of semantic
representation". *Mind and Language* 13(3), 370-399.

Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; Svartvik, Jan. 1972. *A Grammar of
Contemporary English.* London: Longman.

Scott, Mike. 2004. *WordSmith Tools* version 4. Oxford: Oxford University Press.

Vold, Eva Thue. 2006. "Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study". *International Journal of Applied Linguistics* 16(1), 61-87

Appendix

Occurrences of items in L1 and L2 corpora
(Total and epistemic occurrences in raw and relative numbers)

Items	L1 corpus (8882 words)				L2 corpus (20888 words)					chi-square value[1]
	total raw number	total words/1000	epistemic raw number	epistemic words/1000	total raw number	total words/1000	epistemic raw number	epistemic words/1000		
Modal verbs										
could	15	1,69	4	0,45	41	1,96	8	0,38		n/a
may	20	2,25	19	2,14	14	0,67	13	0,62		**p < 0.001**
might	5	0,56	4	0,45	21	1,01	21	1,01		**p < 0.025**
should	29	3,27	2	0,23	75	3,59	5	0,24		n/a
will	25	2,81	17	1,91	92	4,40	65	3,11		p > 0.050
would	41	4,62	35	3,94	66	3,16	40	1,91		**p < 0.005**
Subtotal	135	15,20	81	9,12	309	14,79	152	7,28		p > 0.100
Lexical verbs										
appear	5	0,56	5	0,56	1	0,05	1	0,05		n/a
argue	5	0,56	3	0,34	5	0,24	5	0,24		n/a
assume	1	0,11	1	0,11	2	0,10	2	0,10		n/a
believe	12	1,35	11	1,24	32	1,53	30	1,44		p > 0.250
claim	0	0,00	0	0,00	1	0,05	1	0,05		n/a
guess	0	0,00	0	0,00	5	0,24	5	0,24		n/a
know	2	0,23	0	0,00	24	1,15	6	0,29		n/a
seem	8	0,90	8	0,90	10	0,48	10	0,48		p > 0.100
suppose	0	0,00	0	0,00	1	0,05	1	0,05		n/a
tend	4	0,45	4	0,45	4	0,19	4	0,19		n/a
think	6	0,68	5	0,56	97	4,64	69	3,30		**p < 0.001**
Subtotal	43	4,84	37	4,17	182	8,71	134	6,42		**p < 0.025**

[1] *Chi-square values p < 0.050 mark significant differences between epistemic occurrences in L1 and L2 corpora.*

Appendix (cont.)

Items	L1 corpus (8882 words)				L2 corpus (20888 words)				chi-square value
	total raw numner	total words/1000	epistemic raw number	epistemic words/1000	total raw number	total words/1000	epistemic raw number	epistemic words/1000	
Adverbs									
actually	5	0,56	4	0,45	12	0,57	6	0,29	n/a
almost	2	0,23	1	0,11	9	0,43	8	0,38	n/a
always	3	0,34	3	0,34	13	0,62	12	0,57	n/a
certainly	3	0,34	3	0,34	6	0,29	6	0,29	n/a
clearly	3	0,34	3	0,34	0	0,00	0	0,00	n/a
definitely	0	0,00	0	0,00	2	0,10	2	0,10	n/a
essentially	1	0,11	1	0,11	0	0,00	0	0,00	n/a
evidently	1	0,11	1	0,11	0	0,00	0	0,00	n/a
generally	1	0,11	1	0,11	0	0,00	0	0,00	n/a
in fact	1	0,11	1	0,11	1	0,05	1	0,05	n/a
indeed	2	0,23	2	0,23	0	0,00	0	0,00	n/a
largely	3	0,34	3	0,34	0	0,00	0	0,00	n/a
likely	0	0,00	0	0,00	1	0,05	1	0,05	n/a
maybe	0	0,00	0	0,00	24	1,15	23	1,10	**p < 0.005**
never	1	0,11	1	0,11	5	0,24	3	0,14	n/a
obviously	1	0,11	1	0,11	2	0,10	2	0,10	n/a
of course	2	0,23	2	0,23	41	1,96	41	1,96	**p < 0.001**
often	10	1,13	10	1,13	13	0,62	12	0,57	p > 0.100
perhaps	6	0,68	6	0,68	13	0,62	13	0,62	p > 0.250
possibly	1	0,11	1	0,11	2	0,10	1	0,05	n/a
presumably	2	0,23	2	0,23	0	0,00	0	0,00	n/a
probably	2	0,23	2	0,23	24	1,15	24	1,15	**p < 0.025**
quite	5	0,56	4	0,45	10	0,48	8	0,38	n/a
really	4	0,45	4	0,45	29	1,39	20	0,96	**p > 0.100**
relatively	2	0,23	2	0,23	1	0,05	1	0,05	n/a
sometimes	2	0,23	1	0,11	7	0,34	7	0,34	n/a
surely	3	0,34	3	0,34	3	0,14	3	0,14	n/a
undoubtedly	1	0,11	1	0,11	0	0,00	0	0,00	n/a
usually	1	0,11	1	0,11	2	0,10	2	0,10	n/a
Subtotal	68	7,66	64	7,21	220	10,53	196	9,38	p > 0.050
Total sum	246	27,70	182	20,49	711	34,04	482	23,08	