Epistemic modality markers in research articles: a cross-linguistic and cross-disciplinary study

Eva Thue Vold *University of Bergen*

Research on the use of hedging strategies in research articles has received increasing attention during the last few decades, but few have compared the use of hedges across languages and disciplines. This article explores the use of epistemic modality markers – an important and frequently used type of hedges – in research articles written in three different languages (English, French and Norwegian) and belonging to two different disciplines (linguistics and medicine). Gender differences are also examined. The material is compiled within the electronic corpus of the KIAP project (Cultural Identity in Academic Prose). Statistical analyses of the data (Kruskar-Wallis and Mann-Whitney tests) show that Norwegian- and English-speaking researchers use significantly more of these hedges than their French-speaking colleagues. Disciplinary affiliation and gender seem to have little influence on the proportion of epistemic modality markers in a text, but there are interesting differences between disciplines as regards the type of markers used. These findings may have implications for the teaching of academic writing as well as for cross-cultural understanding between academics.

Key words: epistemic modality, hedging, academic discourse, cross-linguistic and cross-disciplinary variation

Ces dernières décennies, la recherche sur l'emploi d'atténuateurs dans des articles de recherche a reçu un intérêt croissant. Cependant, il existe peu d'études comparant l'emploi d'atténuateurs à travers les langues et les disciplines. La présente étude examinera un type fréquent d'atténuateurs, à savoir les modalisateurs épistémiques, et leur emploi dans des articles de recherche rédigés en trois langues (l'anglais, le français et le norvégien) et appartenant à deux disciplines (la linguistique et la médecine). Les différences de sexe seront examinées aussi. Les matériaux sont tirés du corpus électronique du projet KIAP (Identité culturelle dans le discours scientifique). Les analyses statistiques (les tests de Kruskar-Wallis et de Mann-Whitney) révèlent que les auteurs anglais et norvégiens utilisent plus de modalisateurs épistémiques que leurs collègues français. L'appartenance disciplinaire et le sexe de l'auteur ne semblent pas influencer substantiellement la fréquence de ces marqueurs, mais on observe certaines différences intéressantes entre les disciplines quant aux types de marqueurs utilisés. Les résultats obtenus peuvent avoir des implications pour l'enseignement de la rédaction de textes scientifiques aussi

bien que pour la compréhension interculturelle entre chercheurs de cultures diverses.

Mots clés: modalité épistémique, atténuation, discours académique, variation interlangagière et interdisciplinaire

Introduction

Following an increasing focus in recent years on academic writing, numerous academic writing classes have been established in many different countries. Unfortunately, the teaching practices of such courses often seem to be based on traditional normative principles rather than on solid empirical evidence. It is important that the content of such classes be based on empirical results from analyses of actual language use rather than on traditional normative principles that in many cases need to be reconsidered.[1]

Academic writing becomes especially challenging when the text is to be written in a foreign language. English has become the lingua franca of academic discourse, and novices as well as established researchers must be able to express themselves in that language if they want to be fully accepted members of the international academic community. This has become more and more important over recent years as the pressure to produce work in English and publish internationally has increased. Due to this position of English as a lingua franca in academic settings, numerous Academic English courses for researchers and students at all levels are being offered in countries all over Europe. These classes, too, need to base their teaching practices on empirical research. For this purpose, analyses of contemporary usage of academic English are essential.

This article aims at contributing to the stock of empirical evidence that can be used for such pedagogical purposes. The feature focused on here will be hedging, more specifically the use of epistemic modality markers, which constitute a specific and frequent type of hedge. The article deals with the use of epistemic modality markers not only in English but also in French and Norwegian. The choice of a language-contrastive approach is motivated by the view that not only English academic language with its specific features should be described but also its similarities and differences with other languages, so that students and researchers from non-English backgrounds can easily compare and contrast the academic language of their own mother tongue with English academic language.

Hedging, or the mitigation of claims, is often seen as a rhetorical device used to convince and influence the reader. It is an argumentative strategy considered to be crucial to the writer of scientific texts (cf. Myers' 1989: 13 assumption that all statements conveying new knowledge are hedged, because they have not yet gained acceptance in the scientific community). This is related to the view that scientific discourse is not only informative and content-oriented but also aims at convincing the reader (see Markkanen

and Schröder 1997: 9). The traditional view of the research article as an objective, neutral, informative and factual-oriented genre has been questioned by numerous studies (see e.g. Bazerman 1988; Swales 1990; Hyland 1998, 2000; Fløttum, Dahl and Kinn in preparation; Vassileva 1998).

Mastering English academic discourse is therefore not restricted to mastering English vocabulary, syntax, morphology etc. and the genre schemata of the discipline in question. It is also crucial to master the rhetorical strategies and genre practices specific to English academic discourse. As Vassileva (1997: 203) points out: "It is . . . essential for any scientist anxious to become or remain a member of the international academic community to master the rhetorical organization of English academic discourse." As cultural differences in argumentation strategies and rhetorical means are embodied in language use, it is essential to have some knowledge of these differences while writing in a foreign language. Awareness of cultural differences within academic discourse, such as the differences in the use of hedges, is important for researchers who want to express themselves and read academic texts in languages other than their own.

In addition to the language aspect, disciplinary differences in the use of epistemic modality markers will be examined here. Much of the research investigating variation across disciplines has focused on differences in text structure, referring to formats such as the IMRAD model in the hard sciences and Swales' (1990) CARS model for introductions. But as Fløttum (2005a: 30) points out, "there is more to the genre of the research article . . . than a more or less straightforward structure to be identified and imitated by students". Although genre schemata are important, it is equally important to focus on the rhetorical organization that takes place within the text. As hedging is an important element in the rhetorical organization of a text, the study of hedges across disciplines can tell us something about the argumentative strategies used in different disciplines. It is important to be aware of disciplinary differences, because results from one discipline cannot automatically be transferred to other disciplines. Every discipline has its own terminology and also, it seems, its own preferred rhetorical strategies. The content of an academic writing course should therefore be adjusted to the appropriate research field.

Quite a considerable amount of research on hedges and their communicative functions in English academic discourse has been carried out during the last few decades (see for example Vihla 2000; Varttala 1999, 2001; Salager-Meyer 1994; Salager-Meyer et al. 2003; Hyland 1994, 1998, 2000; Skelton 1997; Meyer 1997). For many other languages, on the other hand, there is clearly a need for more research. Generally, very little research has been done on hedging in languages other than English (except perhaps for German, see Markkanen and Schröder 1997: 252–3 for references), and few have compared hedging strategies across languages. As regards variation between disciplines, Varttala (2001) has compared the hedging strategies of three different disciplines and found that the frequency of hedges may vary

considerably from one discipline to another. Markkanen and Schröder (1997: 10) suggest, on the other hand, that disciplinary variation in the use of hedges is not as great as has often been assumed. According to Varttala (2001: 41–2), very little empirical research exists comparing hedging in different disciplines, and we therefore still know very little about disciplinary variation.

In order to contribute to filling these gaps, this article aims at comparing the use of a specific type of hedges, viz. epistemic modality markers, in English, French and Norwegian research articles taken from the disciplines of medicine and linguistics. The overall aim is to explore to what extent there are differences between these two disciplines and the three languages when it comes to the use of epistemic modality markers in research articles. The present study is carried out within the framework of the Norwegian KIAP project (*Kulturell Identitet i Akademisk Prosa* = 'Cultural Identity in Academic Prose') which, through analyses of linguistic devices in texts, seeks to answer the question of whether cultural identities can be identified in academic prose, and, if so, whether these identities are national or discipline-specific in nature. Being an important rhetorical tool, epistemic modality markers constitute one of the linguistic devices that may contribute to an answer to this question. The material discussed in this article is a subcorpus of the electronic KIAP corpus (see more about the KIAP project on www. kiap.aksis.uib.no).

Epistemic modality markers

Despite the growing body of research concerning hedging strategies in academic discourse, the research field suffers from a lack of agreement on important terminological questions. No adequate and precise definition of the term 'hedge' has been agreed upon, and there has been considerable debate on how to understand the term. (cf. *English for Specific Purposes* 16.4 and 17.3). A commonly used definition is Hyland's (1998: 5), which states that hedges are "the means by which writers can present a proposition as an opinion rather than a fact".[2] This fairly broad definition, which has been criticised for being too vague (Crompton 1997), allows the inclusion of a series of different types of markers, ranging from markers of uncertainty and pointers to methodological limitations to what Salager-Meyer (1994: 154) calls "emotionally-charged intensifiers", i.e. expressions like *extremely interesting* and *surprisingly*. Whether one opts for a broad or a restricted definition of the term 'hedge', the group of markers selected for this paper – i.e. epistemic modality markers expressing uncertainty – remains a dominant and basic type of hedge (see Salager-Meyer 1994; Varttala 1999; Vihla 2000).

Inspired by Lyon's definition of epistemic modality, "[a]ny utterance in which the speaker explicitly qualifies his commitment to the truth of the

proposition expressed by the sentence he utters . . . is an epistemically modal, or modalized utterance" (Lyons 1977: 797), I will define epistemic modality markers as *linguistic expressions that qualify the truth value of a propositional content* (for example *perhaps*, *probably*). Epistemic modality markers thus mark to what extent one can rely on the information which is being conveyed by the proposition. Take, for example, the proposition 'smoking causes lung cancer'. The truth value of this proposition is marked in different ways in (a–c):

a) It is possible that smoking causes lung cancer.
b) Smoking probably causes lung cancer.
c) We know that smoking causes lung cancer.

The proposition is marked as a possibility in (a), a probability in (b), and a certainty in (c). The domain of epistemic modality ranges in principle from total uncertainty to absolute certainty (Le Querler 1996: 71). In this article, however, I will concentrate on expressions of *uncertainty* only. Markers that state or confirm the truth of a proposition (like example c does) are not hedges and will not be considered here.

The criteria I used to identify epistemic modality markers in the texts were the following:

i) The marker had to explicitly qualify the truth value of a certain propositional content.
ii) The marker also had to be a lexical or a grammatical unit (thus I am not talking about entire phrases or paragraphs that are used to tone down the findings).

See below for more information about the method used to distinguish epistemic and non-epistemic senses.

The criterion of explicit qualification is important: verbs such as *propose* and *claim*, which have been included in certain studies of epistemic modality markers (e.g. in Varttala 1999: 185), have not been included in this one. With these markers, there might be an implicit qualification of the truth value of the propositional content as questionable, but their main function is that of a reporting verb, introducing the propositional content without really qualifying its truth value. It should also be emphasized that epistemic modality is extra-propositional, i.e. an epistemic modality marker always modifies a propositional content. The lexical verb *assume*, which in many other contexts would be classified as an epistemic modality marker, is not classified as such in example (1) because there is no propositional content to be modified:

1) . . . explicit content is 'much more inferential and hence worthy of pragmatic investigation' than is **assumed** by pragmatists in the Gricean tradition (engling05).[3]

In order to illustrate what kind of markers have been included, let us look at some valid examples, taken from the KIAP corpus:

2) As has been discussed in the literature on relative clauses, gerundive relatives **seem to** have a reduced clause structure when compared to full relatives (see Williams 1975, Stowell 1982). (engling01)

Here we have an identifiable lexical unit, the semi-auxiliary *seem*, which modifies a propositional content, namely 'gerundive relatives have a reduced clause structure when compared to full relatives'. In the next example, we have two epistemic modality markers, the first one being the lexical verb *suggest* and the second a grammatical unit, the modal verb *might*:

3) Our findings in non-smoking mothers **suggest** that the positive associations previously reported among first births **might** simply reflect inadequate adjustment for confounding variables. (engmed01)

The proposition that is being modified is 'positive associations previously reported among first births reflect inadequate adjustment for confounding variables'.

It should be noted that markers like *might* and *suggest* are polysemous and do not always function as epistemic modality markers. I will return to the question of polysemous markers in the section concerning the classification of the occurrences.

Research questions and hypotheses

One of the purposes of this article is to investigate which factors influence the use of hedges or, more precisely, the use of the selected epistemic modality markers in research articles. It is reasonable to believe that factors such as the object of study, type of study, research field, language, the author's age or position in the scientific community etc. influence the author's use of such hedges. It is of course impossible to check for all possible factors in one single study. In this article, I examine three factors that might influence the author's use of epistemic modality markers: discipline, language and gender. The first two factors are the variables studied in the KIAP project and the ones that the KIAP corpus selection is based on.

Firstly, we shall have a look at the *discipline* factor. Does disciplinary affiliation influence the frequency of epistemic modality markers used? It could be assumed that one would find more hedges in the linguistic articles than in the medical ones, because this latter discipline is thought to be closer to the positivistic ideal, and thus the claims made would be more factual-oriented (cf. Markkanen and Schröder 1997: 10). However, medicine is a

discipline with a traditionally high frequency of hedges (Varttala 1999: 177). Secondly, we shall have a look at the *language* variable. Does the frequency of epistemic modality markers vary across languages? If so, what are the explanations for this? Thirdly, we shall examine the possible factor of *gender*. Are there differences between male and female authors when it comes to the use of epistemic modality markers? Some studies show that women use more hedges than men, making room for suggestions like "women are less sure of themselves as scientists" or "women are more polite and more attentive to the reader", while other studies show no such differences (see Markkanen and Schröder 1997: 8–9). Finally, it would be interesting to investigate whether the frequency of hedges correlates with individual measurable factors such as age and position in the scientific community. However, I have not checked for these here. In future research this could be a fruitful course to pursue.

Material and method

Material

In order to answer the questions posed above, I studied the frequency of epistemic modality markers in a selection of 120 research articles written in English, French and Norwegian and belonging to the disciplines of medicine and linguistics. The material is a subcorpus of the KIAP corpus, which is an electronic data base consisting of 450 research articles from three different disciplines and written in the three languages. In building up the KIAP corpus, articles written by native speakers were selected to the extent that it was possible to determine this. However, it was not always possible to establish the nationality of the author(s).[4] For English, both American English and British English are represented, but potential differences between the two will not be considered here.

The research articles constituting the KIAP corpus are all taken from well-recognized, high-quality refereed journals. For this study, the articles come from the sources listed in Table 1. All the articles were published during the last decade, most of them during the five most recent years. The research articles are evenly distributed over the six subgroups – English linguistics (engling), English medicine (engmed), French linguistics (frling), French medicine (frmed), Norwegian linguistics (noling) and Norwegian medicine (nomed). The material includes the body of the articles, i.e. the complete running text of the articles, excluding abstracts, notes, bibliographies, quotations, linguistic examples, tables and figures. Since the number of words is very unevenly distributed over the six subgroups (consider e.g. the length of the English linguistic articles), attention should be paid to the relative frequency of markers rather than to the number of occurrences.

Table 1. Sources used for this study

Language	Discipline	Words	Sources	No. of articles
French	Linguistics	68,727	*Travaux de Linguistique*	18
			Marges Linguistiques	2
	Medicine	61,180	*Annales de Médecine Interne*	15
			Maladies chroniques au Canada	5
Norwegian	Linguistics	90,579	*Norsk lingvistisk tidsskrift*	20
	Medicine	43,234	*Tidsskrift for Den Norske Lægeforening*	20
English	Linguistics	170,981	*Journal of Linguistics*	10
			English for Specific Purposes	5
			Language	4
			Linguistic Inquiry	1
	Medicine	59,410	*Journal of the American Medical Assoc.*	10
			British Medical Journal	5
			International Journal of Cancer	5
Totals		494,111		120

Method

The selection of markers was primarily based on frequency in an exploratory corpus consisting of 30 articles. In order to ensure (to the extent that it is possible) the representativeness of the exploratory corpus, it reflected in its composition the larger corpus' distribution over different journals, disciplines and languages. For the articles constituting the exploratory corpus, all epistemic modality markers were written down and counted. As mentioned above, in order to be considered an epistemic modality marker, the marker had to qualify explicitly the truth value of a particular propositional content and be a lexical or grammatical unit. The most frequent epistemic modality markers were submitted to a quantitative analysis of the corpus as a whole. These markers are listed in Table 2.[5]

In addition, some markers (French *peut*, *peut-être*, *indiquer* and *possible*, English *probably*, *could* and *possible*) were selected even though they were not particularly frequent in the exploratory corpus. These markers were selected because they or their equivalents in the other languages are often considered as prototypes of epistemic modality markers, and/or because they seemed particularly interesting to study from a contrastive perspective. English *probably*, for instance, was quite rare in the English exploratory corpus (2 occurrences), while the Norwegian equivalent *sannsynligvis* was the third most frequent marker in the Norwegian exploratory corpus. Epistemic *indiquer*

Table 2. The most frequent epistemic modality markers in the exploratory corpus by number of occurrences

French		Norwegian		English	
sembler	22	*kan*	25	*may*	38
paraître	8	*synes*	17	*assume*	31
pourrait	7	*sannsynligvis*	11	*suggest*	25
probable	5	*se ut*	10	*appear*	22
probablement	5	*mulig*	9	*might*	20
sans doute	5	*anta*	8	*seem*	16
suggérer	4	*kanskje*	7	*perhaps*	7
		trolig	7	*indicate*	6
		tyde på	7		
		antagelse	6		
		muligens	5		

was absent in the French exploratory corpus, whereas the near-equivalent in Norwegian, *tyde på*, was relatively frequent in the Norwegian exploratory corpus. And the epistemic use of French *pouvoir* in the present tense was surprisingly rare compared to the epistemic uses of English *may* and Norwegian *kan*. I wanted to check if tendencies like these would persist when studying the whole corpus, and therefore these additional markers were included. However, as already mentioned, most of the markers were chosen as a result of demonstrated frequency. Except for the few that were selected because of the relatively high frequency of their (near)-equivalents in the other languages, the selection of markers is not based on equivalence, and the selection of markers in one language is therefore independent of the selection in the other two languages. In sum, 11 markers from each of the three languages were chosen to be used in a quantitative analysis of the corpus as a whole.

Selecting markers on the basis of what is actually found in the articles seems to be an adequate place to start when trying to say something about the use of a semantic-pragmatic category like epistemic modality. In previous research, it has been common to select *a priori* some markers and then submit these to a quantitative analysis (see e.g. Vihla 2000). Such studies say something about the use of the selected markers, but there might of course be other important markers that have been overlooked. In other studies, for example in Salager-Meyer (1994) and Varttala (1999), the choice of markers was to some extent based on what are "commonly considered" (Salager-Meyer 1994: 154) or "typically interpreted" (Varttala 1999: 182–3) as hedges or epistemic modality markers, which is a criterion quite low in precision (see also Crompton 1997: 277). In the present study, precise criteria have been formulated in order to classify the markers more accurately (see the following section and the one on epistemic modality markers).

Classifying the occurrences[6]

Several of the expressions that I studied were polysemous, the epistemically modal sense being only one of the different meanings the marker in question could convey. The modal verb *may* is an obvious example. Among this marker's different uses are: permission (as in *you may go*), epistemic possibility – which is the function I am interested in (as in s*he may be at home now*), and pure possibility, stating a fact (as in *this marker may indicate permission, epistemic possibility and pure possibility*). Step two of the analysis was therefore to classify the occurrences according to their meaning in a particular context. The occurrences were divided into two groups: those occurrences that were classified as epistemic modality markers expressing uncertainty, and other meanings. Assigning a particular occurrence to one of these groups is not always a straightforward task, as the different meanings of a polysemous marker are more or less related to each other, and there is not always a clear-cut distinction between two related meanings. *Indicate* (and French *indiquer*), for example, may mean 'suggest' (French: *suggérer*), but also simply 'show' (French: *montrer*). In the former sense, *indicate* questions the truth value of the propositional content, while in the latter sense it is a marker of certainty rather than uncertainty. In most cases where *indicate* is used in the latter sense, it would probably not be considered an epistemic modality marker at all but rather a "neutral" marker. Example (4) is an obvious case of the first sense, although *indicate* is used here on a subordinate level in the text and does not really function as a hedge:

4) This analysis does not assume, as Rieber's does, that the contrastive import of *but* is expressed in a distinct proposition whose truth is **indicated** or suggested rather than asserted. (engling05)

Example (5) is an obvious case of the second meaning:

5) Fig. 6 **indicates** the forms of oral communication most often used by graduate business employees in team work. (engling13)

Most often, the occurrences are not so univocal, as examples (6) and (7) illustrate:

6) The survey **indicated** that employers are looking for graduates who have the capacity for ". . . independent and critical thinking" (Koutsoukis, 2000, p. 11). (engling13)
7) Again, the results **indicate** that oral communication skills are important for all of these aspects. (engling13)

From decontexualized sentences like (6) and (7), it is impossible to conclude whether *indicate* should be read as meaning 'show' or 'suggest'. The

distinction is not always relevant; the two meanings may co-occur in one single form and are not necessarily mutually exclusive (see also Johns 2001). *Indicate* (and *indiquer*) was classified as an uncertainty marker when this seemed the most plausible reading, e.g. when it was coupled with other epistemic modality markers expressing uncertainty, such as *may* (*the results may indicate*) or when the context otherwise suggested that there was uncertainty attached to the truth value of the propositional content. While *indicate* and *indiquer* at the outset were seen as near-equivalents of the Norwegian *tyde på*, a close reading of the occurrences in context made me realize that *tyde på* is nearer in meaning to *suggest* and *suggérer* than to *indicate* and *indiquer*.[7]

Another polysemous marker is *assume*, which may mean 'suppose' (example 8) but also simply 'adopt', 'take as a point of departure' (example 9). In this latter example, there is no modification of a propositional content either (cf. criterion (i) mentioned in the section on epistemic modality markers), so this occurrence is easily classified as non-epistemically modal. There are also other uses of *assume*, for example in *assume responsibility*, but these are more easily distinguished from the epistemically modal sense.

8) We therefore **assume** that these default preferences for the null complements in (19a), (20a) and (21a) have been established as part of the conventional meanings of the relevant verbs. (engling09)
9) I **assume** Hornstein's (1990) Reichenbachian approach to tense (engling01)

The different meanings of the modal verb *may* and the difficulty of distinguishing them have been discussed many times (e.g. Perkins 1983; Palmer 1986). Let me just point out that it might be hard to decide whether *may* expresses a fact (pure possibility) or an eventuality (hypothetic/epistemic possibility). Example (10) can be read in both ways ('can be heard' or 'is perhaps heard'):

10) In contrast, TCU-initial uses mark the introduction of touched-off material that **may** be heard as potentially topically relevant because of having been thus triggered. (engling20)

It should also be noted that with *may* there is often indeterminacy between different meanings: several meanings co-occur, and each meaning contributes to the understanding of the utterance. Which meaning is perceived as dominant (if there is one) may vary from reader to reader.

In order to make the classification of occurrences as systematic as possible, the criteria for epistemic modality markers needed some refinement. The question *Does the marker question the truth value of a proposition?* is not always easy to answer, and I therefore used several tests to help determine whether or not an occurrence should be classified as an epistemically modal. Following

Vihla (2000: 212), I used a substitution test, meaning that in cases of doubt I would try to replace the polysemic marker, for instance *may*, with an intrinsically epistemic marker, e.g. *perhaps*. If such a substitution was possible with no great change of meaning, the marker was classified as epistemically modal. Other helpful tests were trying to add a phrase like *but I'm not sure* after the marker/proposition and, following Crompton (1997: 282), trying to formulate a less hedged version of the proposition. If adding an 'uncertainty phrase' felt natural, the occurrence was classified as epistemically modal:

8') We therefore **assume** that these default preferences for the null complements in (19a), (20a) and (21a) have been established as part of the conventional meanings of the relevant verbs, *but we are not sure.*

If this was not possible, the occurrence would be non-epistemic:

9') *I assume Hornstein's approach, but I'm not sure.

Likewise, if a less hedged (or non-hedged) version of the proposition existed, i.e. if it could be reformulated in a more 'certain' way by changing the marker in question, the occurrence was classified as an epistemic modality marker expressing uncertainty. For example, *we know that* . . . is a non-hedged version of *we assume that.* . . .

Syntactic criteria might also be helpful, and these have been used to a certain extent in combination with semantic criteria of the type listed above. For example, constructions such as *assume(ing) that* and *be assumed to* go together with the epistemically modal sense, while *assume* + NOUN (for example *assume a theory*) do not. As for *appear*, which also has several meanings, the constructions *appear to be*, *it appears that* are related to epistemically modal senses, whereas *appear* + PREPOSITION (e.g. *in*) is related to other, non-epistemic senses. For markers such as *indicate* and *may*, syntactic criteria are less relevant.

It should be noted that all occurrences expressing an epistemically modal sense, regardless of who is responsible for the modalization, are included in Tables 3, 4 and 5 below. This means that the numbers include examples like (11) where the person responsible for the modalization is not the author but somebody else,[8] as well as hypothetical constructions like (12):

11) Hence, a fundamental problem with Kreitzer's account (as well as with Lakoff's), is that it **assumes** that the rich understanding we obtain about spatial scenes is derived entirely from what is coded by formal linguistic expression. (engling17)
12) Given general knowledge of shooting arrows and targets, most speakers would **assume** that whoever shot the arrow intended to hit the target but aimed too high. (engling17)

Occurrences like these have the strange property of being epistemically modal though not hedges. A more detailed analysis of the function of each occurrence could give an even clearer understanding of how and to what extent authors of research articles hedge their papers. It is, however, very unlikely that the exclusion of such occurrences would influence the main results in any considerable manner.

Results

Let us now move on to the results. Table 3 sums up the results for the French linguistics and medical texts, Table 4 for Norwegian, and Table 5 for English. The first column lists the epistemic modality markers in descending order of frequency.[9] The next column, *epistemic occurrences*, is divided

Table 3. Epistemic modality markers in the French research articles

Discipline	Marker		Epistemic		All	
			f/1000	no.	f/1000	no.
Linguistics	*sembler*	'seem'	0.89	61	0.90	62
	paraître	'appear'	0.32	22	0.39	27
	sans doute	'probably'	0.26	18	0.26	18
	peut-être	'perhaps'	0.16	12	0.16	12
	indiquer	'indicate'	0.06	4	0.35	24
	peut	'can/may'	0.06	4	5.24	360
	suggérer	'suggest'	0.06	4	0.09	6
	possible	'possible'	0.03	2	0.93	64
	pourrait	'could/might'	0.03	2	0.76	52
	probablement	'probably'	0.01	1	0.01	1
	probable	'probable'	–	–	–	–
	Totals		**1.9**	**130**	**9.1**	**626**
Medical	*sembler*	'seem'	0.75	46	0.75	46
	pourrait	'could/might'	0.31	19	0.52	32
	probablement	'probably'	0.28	17	0.28	17
	suggérer	'suggest'	0.28	17	0.28	17
	paraître	'appear'	0.13	8	0.13	8
	probable	'probable'	0.13	8	0.13	8
	possible	'possible'	0.11	7	0.49	30
	indiquer	'indicate'	0.09	6	0.51	31
	peut	'can/may'	0.09	6	2.58	158
	peut-être	'perhaps'	0.03	2	0.03	2
	sans doute	'probably'	0.03	2	0.03	2
	Totals		**2.3**	**138**	**5.7**	**351**

Table 4. Epistemic modality markers in the Norwegian research articles

Discipline	Marker		Epistemic		All	
			f/1000	no.	f/1000	no.
Linguistics	*kan*	'can/may'	0.70	64	8.25	747
	se ut	'seem/look /appear'	0.55	50	0.63	57
	anta	'assume/ suppose'	0.52	47	0.54	49
	synes	'seem'	0.36	33	0.40	37
	kanskje	'perhaps'	0.33	30	0.33	30
	mulig	'possible'	0.22	20	1.00	91
	trolig	'presumably'	0.17	15	0.17	15
	tyde på	'suggest/ indicate'	0.17	15	0.17	15
	antagelse	'assumption'	0.15	14	0.20	18
	sannsynligvis	'probably'	0.09	8	0.09	8
	muligens	'possibly'	0.06	5	0.06	5
	Totals		**3.3**	**301**	**11.83**	**1072**
Medical	*kan*	'can/may'	1.00	43	5.27	228
	synes	'seem'	0.46	20	0.56	24
	tyde på	'suggest/ indicate'	0.37	16	0.37	16
	mulig	'possible'	0.35	15	0.74	32
	sannsynligvis	'probably'	0.30	13	0.30	13
	anta	'assume/ suppose'	0.25	11	0.25	11
	muligens	'possibly'	0.19	8	0.19	8
	trolig	'presumably'	0.19	8	0.19	8
	se ut	'seem/look/ appear'	0.16	7	0.16	7
	kanskje	'perhaps'	0.09	4	0.09	4
	antagelse	'assumption'	0.02	1	0.02	1
	Totals		**3.4**	**146**	**8.14**	**352**

into two parts, one indicating the relative frequency per thousand words (f/1000), the other giving the absolute number of epistemic occurrences of the marker in question (no.). As mentioned above, many of the markers are polysemous, and all the occurrences therefore had to be classified according to meaning. The third column, *all occurrences*, gives the corresponding numbers and frequencies when all occurrences and all different senses of the markers are included. By comparing the two columns, we can determine the extent to which a marker can be seen mainly as an epistemic modality marker. If we look at the use of *may* in the English linguistics research articles, for

Table 5. Epistemic modality markers in the English research articles

Discipline	Marker	Epistemic		All	
		f/1000	no.	f/1000	no.
Linguistics	*seem*	0.79	136	0.80	137
	suggest	0.73	124	1.03	176
	assume	0.66	112	0.81	138
	may	0.46	79	1.64	280
	appear	0.43	74	0.64	109
	might	0.30	52	0.67	114
	perhaps	0.22	37	0.22	37
	indicate	0.11	19	0.45	77
	probably	0.08	13	0.10	17
	possible	0.06	11	0.87	148
	could	0.03	5	0.53	91
	Totals	**3.9**	**662**	**7.7**	**1324**
Medical	*may*	1.23	73	1.75	104
	suggest	0.81	48	0.93	55
	might	0.35	21	0.44	26
	could	0.27	16	0.72	43
	possible	0.20	12	0.39	23
	indicate	0.17	10	0.42	25
	appear	0.10	6	0.10	6
	assume	0.07	4	0.08	5
	perhaps	0.07	4	0.07	4
	probably	0.05	3	0.05	3
	seem	0.02	1	0.02	1
	Totals	**3.3**	**198**	**4.97**	**295**

example, we see that there are 280 occurrences, i.e. 1.64 per thousand words for this particular subcorpus; 79 of these were classified as epistemic, and epistemic *may* has a relative frequency of 0.46. This tells us that the epistemic use of *may* is only one of several uses.

So, what do all these numbers tell us? Before we turn to the question of frequencies, we shall briefly look at the variation across disciplines when it comes to the *type of markers* that are used. For practical reasons I cannot go into details here, but there are some interesting differences that merit mention. In the French corpus, the markers seem to be surprisingly discipline-specific. *Pourrait* for example (the conditional of *pouvoir*) is a frequent epistemic modality marker in the medical papers, while French linguists tend to avoid the use of epistemic *pouvoir*. The same goes for *suggérer* and *probablement*. *Sans doute*, on the other hand, is one of the most frequent epistemic modality markers in the linguistic texts but is almost absent in the

medical texts. An important exception to this 'discipline-specific marker tendency' is the semi-auxiliary *sembler* ('seem'), which is frequently used in both disciplines.

In the English corpus, there are also some interesting tendencies. *Could* is used rarely by the linguists but relatively often by the medical researchers, just like French *pourrait*. The use of *seem*, *assume* and *appear*, on the other hand, is almost entirely restricted to the linguistic papers. The case of *seem* is especially interesting, since the French equivalent (*sembler*) was one of the few markers that were frequently used in both disciplines.

In the Norwegian corpus, all markers are used in both disciplines, though there are, of course, some disciplinary characteristics. *Kanskje* ('perhaps'), for example, is a marker which is frequently used by linguists but which is rare in the medical articles.

There are some interesting similarities between the three languages: in all three, *perhaps* (Norw. *kanskje*, French *peut-être*) is used mainly by linguists and is rarely used by the medical researchers. The epistemic uses of *may*, *might* and *could,* like the epistemic uses of the Norwegian and French near-equivalents (*kan* and *peut*, *pourrait*, respectively), seem to be more frequent in medical discourse than in linguistic discourse. In French and Norwegian, markers expressing probability (*probable*, *probablement*, *sannsynligvis*) are far more frequent in the medical texts than in the linguistic texts.

I will now return to the question of which factors determine the frequency of epistemic modality markers in a research article. Let us have a look at the discipline factor first. A two-tailed Mann-Whitney test was run on all data sets to determine significance (alpha = 0.05). We see that in the French linguistic texts, the epistemic modality markers that I have studied constitute approximately 1.9 per thousand words. The corresponding number for the French medical research articles is 2.3. There is a slight difference between the two disciplines when it comes to the frequency of these markers, but the difference is not statistically significant (p = 0.561).

In the English corpus, there is also a difference between the disciplines, but the other way around. The relative frequency of the selected epistemic modality markers is 3.9 in the linguistic texts and 3.3 in the medical texts. However, linguists and medical researchers do not differ significantly from each other (p = 0.152).

In the Norwegian corpus, there are no considerable discipline-specific differences as far as frequency is concerned. The selected epistemic modality markers constitute 3.3 per thousand of the linguistic texts, and 3.4 per thousand of the medical texts. This difference was, of course, not statistically significant (p = 0.989).

Table 6 shows differences between disciplines, all three languages taken together. The relative frequency of the selected markers is 3.3 in the linguistic texts and 2.9 in the medical texts, suggesting that the selected markers are more frequent in linguistic papers than in medical papers. However, the difference between disciplines was not statistically significant (p = 0.589).

Table 6. Discipline-specific differences

Discipline	f/1000	no. of occurrences	no. of words
Linguistics	3.3	1093	330,287
Medicine	2.9	482	163,824

Table 7. Language-specific differences

Language	f/1000	no. of occurrences	no. of words
French	2.1	268	129,907
Norwegian	3.3	447	133,813
English	3.7	860	230,391

While the differences between disciplines are small, the results presented in Tables 3–5 show that there are considerable language-specific differences: epistemic modality markers expressing uncertainty are far more frequent in English and Norwegian research articles than in French ones. Table 7 shows language-specific differences, regardless of discipline.

The English texts have the highest frequency of epistemic modality markers, with a relative frequency of 3.7. The Norwegian texts come second with a relative frequency of 3.3. In the French corpus, the relative frequency is 2.1. The English and the French researchers thus constitute two opposite ends, with the Norwegian researchers situated between the two. It is clear, however, that in this respect the Norwegian researchers tend more towards the Anglo-American style of writing than towards the French style.

The differences between languages were clearly statistically significant (Kruskar-Wallis test run on all three languages, two-tailed Mann-Whitney on language pairs): the English authors differed significantly from the French authors ($p = 0.001$), the Norwegian authors also differed significantly from the French authors ($p = 0.008$), whereas the English and the Norwegian authors did not differ significantly from each other ($p = 0.465$).

Let us now have a quick look at the individual differences between articles. A certain degree of individual variation is expected, because the proportion of epistemic modality markers will vary according to subject matter, personality, age, professional stature etc. Figure 1 shows the relative frequency of epistemic modality markers (EMM) in each of the articles.

As can be seen from the figure (and from Table 7), the French papers have fewer epistemic modality markers than the Norwegian papers, and the Norwegian papers have somewhat fewer epistemic modality markers than the English papers. The French papers also have a smaller range of variation than the Norwegian and English papers. The relative frequency of epistemic modality markers in the French papers tends to be between 0 and 4 per thousand, whereas most of the Norwegian and English papers have a relative
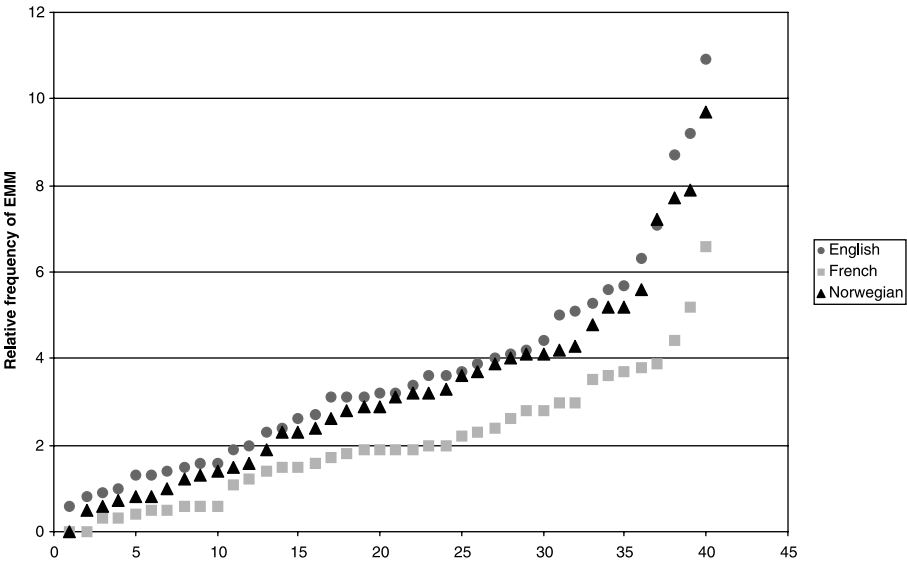
Figure 1. Relative frequency of EMM in English, French and Norwegian research articles
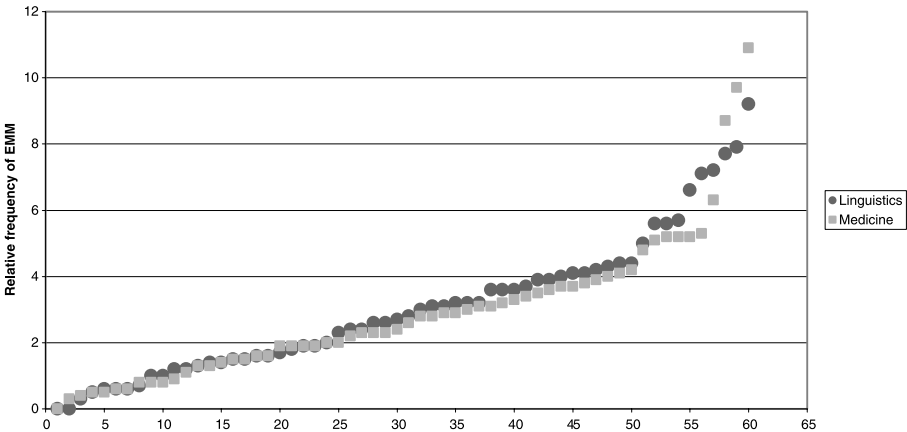
Figure 2. Relative frequency of EMM in linguistic and medical research articles

frequency between 0 and 6. In addition, there are quite a few Norwegian and English papers with an even higher relative frequency. The individual variation is thus particularly noticeable in the Norwegian and English corpora, while there is somewhat less variation among the French authors.

As can be seen from Figure 2 (and Table 6), there are no important differences between the two disciplines when it comes to frequency of the epistemic modality markers. On the contrary, the similarity is striking: the

Table 8. Gender differences (linguistics articles only)

Language	f/1000		no. of occur.		no. of words		no. of arts.	
	male	*female*	*male*	*female*	*male*	*female*	*male*	*female*
French	1.8	2.1	79	44	43,657	21,350	11	8
Norwegian	3.0	4.0	185	117	61,089	29,490	12	8
English	4.2	4.0	193	300	46,495	74,437	6	8
All three	3.0	3.7	457	461	151,241	125,277	29	24

figures for the two disciplines overlap almost completely, indicating that the individual variation is about the same in the two corpora. This is perhaps a bit surprising considering the fact that medical researchers are often said to tend to write in a similar manner, while linguists have a more personal style, suggesting that one would expect the individual variation to be more important within linguistics than within medicine.

Let us now pass on to the third factor I wanted to investigate: gender. Separating articles by the author's sex was possible only for the linguistic papers, because the medical articles in general have plural authorship. Six linguistic articles from the English corpus and one from the French corpus were excluded because of mixed authorship. This corpus thus consists of 53 research papers only. Table 8 shows differences according to the author's sex for the three languages taken together as well as for the languages individually.

As can be seen from the table, there are differences in the relative frequency of epistemic modality markers in articles written by women and men, but gender is not statistically significant in any of the languages (Fr. p = 0.364; Nor. p = 0.877; Eng. p = 0.439) nor for the languages taken together (p = 0.464) and therefore does not seem to be relevant.

Discussion and interpretation of results

From the results presented above, we can conclude that the most important factor when it comes to the frequency of this type of hedge seems to be language or nationality. The differences between disciplines are quite small compared with the differences between languages. These findings support Markkanen and Schröder's conclusion that ". . . the differences in the use of hedges between texts in different fields are not so great as has been often assumed" (1997: 10). It is important to remember, though, that I have compared two disciplines only. There might, of course, be bigger differences between other disciplines. For example, Varttala (2001: 154) found that the relative frequency of hedges was higher in the field of economics than in medicine and technology – the latter disciplines having approximately

the same frequency of hedges. Another important point in this respect is that a discipline does not, of course, constitute one unified culture. Rather, each discipline consists of several subdisciplines that might represent very different, even conflicting traditions and methodologies. In order to answer the question of disciplinary variation more fully, other disciplines, and even subdisciplines, need to be compared in future research.

What the results do indicate, and quite clearly so, is that the use of epistemic modality markers indicating uncertainty is more frequent in English and Norwegian research articles than in French ones. This is seen in all groupings: whether one compares linguists writing in these different languages, or medical researchers, or male vs female researchers, the English-speaking and Norwegian-speaking authors use more of these epistemic modality markers than their French-speaking colleagues. The English articles have the highest frequency of such markers, but the Norwegians are very close (see Figure 1). The similarity between English and Norwegian articles is not very surprising, since English and Norwegian are both Germanic languages and have similar language strategies in many respects. The image of English and Norwegian academic discourse on the one hand and French academic discourse on the other as opposite poles representing two different styles is corroborated by other findings from the KIAP project. These findings indicate that English and Norwegian research articles have a higher frequency of linguistic phenomena such as first person pronouns (Fløttum 2003a), textual metadiscourse (Dahl 2003, 2004), negation and concessive markers (Fløttum 2005a,c) and bibliographical references (Fløttum 2003b), suggesting that English- and Norwegian-speaking authors are more polemic and manifest themselves more clearly in their own texts than French-speaking authors (see Fløttum et al. in preparation). KIAP findings also show that English authors use more verbs like *argue* and *claim* than do French and Norwegian authors (Fløttum et al. in preparation). The high incidences of verbs like *argue* and *claim* might explain the high frequency of hedges, as the hedges then can be used to balance the impression of a strong arguer role (see Fløttum 2004, 2005b).

The fact that the selected epistemic modality markers are more frequent in English and Norwegian research articles than in French ones does not necessarily mean that French researchers hedge less than Norwegian researchers. I have only studied a specific group of hedges, and hedging strategies such as references to restricted knowledge or methodological limitations (see Hyland 1998: 141–8) have not been considered. Still, as epistemic modality markers are one of the most dominant and frequent type of hedges (see e.g. Hyland 1998: 149; Salager-Meyer 1994; Vihla 2000; Varttala 1999), there is reason to believe that English and Norwegian research articles are more heavily hedged than French ones. Hedging is often said to be a characteristic feature of academic discourse (see e.g. Lewin 2005: 164), but the findings here suggest that this phenomenon may be linked to the academic discourse of specific language groups rather than to academic discourse in

general. What are the explanations for these language-related differences in the use of hedges?

To answer this question it might be useful to take a look at what functions hedging strategies serve. As a starting point, we can talk about two major types of functions. The first group is what we can call *real hedges*. These are used to convey real uncertainty, for example when the nature of the research findings does not allow the author to make strong claims or draw clear conclusions. In such cases, the hedges serve to give an accurate picture of the level of certainty. Lewin (2005: 173) has shown that authors of scientific texts tend to see real uncertainty as the main motivation for their use of hedges. According to the authors in Lewin's study, they did not use hedges in order to be polite or modest but rather to be *precise* (2005: 169–70). This function is illustrated by the following examples:

13) Differences in treatment context **may** have contributed to divergent outcomes between the 2 studies, but it is also **possible** that our larger study simply provided a better estimate of acupuncture's "true" treatment effect compared with that of the 2 control conditions. (engmed14)

14) Il existe peu d'études épidémiologiques permettant de calculer l'incidence de la tuberculose ostéo-articulaire dans les pays développés, il **semble** exister une augmentation des cas dans ces pays à partir des années90 [3], [8] mais ceci ne **semble** pas être le cas dans d'autres séries [9]. (frmed15)

Hedging or expressions of uncertainty are also often associated with tentativeness, cautiousness, politeness, and a humble attitude (see e.g. Salager-Meyer 1994). These can be seen as *strategic hedges* (Mauranen 1997): they are not necessarily used to express real uncertainty; rather, they are part of the conventions for academic writing. These 'non-real' hedges fill a variety of functions. They may be used within a context that expresses possible opinions or interpretations, and thus the author anticipates potential criticism:

15) Quality-of-life data were missing at 1 or more time points in 19% of participants, and we cannot entirely exclude the **possibility** that differential loss of participants **may** have affected our results. However, we consider this **unlikely** because treatment assignment was not significantly different among women with complete vs incomplete data and our results were similar whether the analysis included or excluded patients with some missing data. (engmed11)

16) Une autre cause possible est l'augmentation du temps passé à l'intérieur, ce qui augmente l'exposition aux allergènes tels que les acariens et les animaux domestiques. Cette pointe saisonnière ne **semble** pas être causée par l'inclusion de patients souffrant de bronchopneumopathie chronique obstructive (dont les problèmes peuvent s'accentuer à l'automne) et ayant fait l'objet d'un diagnostic erroné d'asthme. (frmed02)

They may also be used as a politeness strategy in order to cautiously criticise fellow researchers, as in (17) and (18) where the authors comment upon other researchers' theories and work:

17) It **would seem** that to say that the conventional meaning of *but* is analysed in terms of a performative of the form *I suggest that P* is to say that there is a linguistically determined relation between *but* and the information that its utterance constitutes the performance of the act of *suggesting that P*. On the other hand, to say that this performative is tacit **would seem** to **suggest** that the utterance does not actually contain an expression which identifies the act being performed.

It **seems** that the apparent contradiction here stems from the assumption that tacit means implicit . . . (engling05)

18) Il a été noté dans ce travail 57% d'hypothyroïdies sub-cliniques (TSH de base comprises entre 5,4 et 7,6 uU/mL). Il nous a **semblé** que la fréquence de l'hypothyroïdie définitive dans cette affection est sous-estimée. (frmed12)

They may also be used to tone down statements and claims in order for the author to be less vulnerable to criticism. Sometimes their presence may even be caused by political or ethical discussions. A few months ago, I came across a comment from someone criticising a research report in genetic engineering – which is a very controversial area of study – for being too heavily hedged. She found that "the conclusions drawn . . . seem substansive and reasonably confident" (thus, she did not take the hedges to be real) and suggested that the report was heavily hedged because of the authors' concern about possible political implications of their findings (www.senseaboutscience.org.uk, Sept. 04).

Evidently, the interpretation of a hedged statement may vary from reader to reader (see Lewin 2005: 172). It is often impossible to decide whether a hedge should be considered a real or a 'conventional' hedge. One cannot read the author's mind. However, there is no reason to believe that there should be more 'real uncertainty' in English and Norwegian research than in French research. The language-specific differences that have been manifested are probably linked to the second category of hedges – hedging as a characteristic feature of a particular style.

But why should English-speaking and Norwegian-speaking researchers be more humble and polite and express themselves with more caution than their French-speaking colleagues? This may have to do with different academic cultures. Norwegian academic culture is to a large extent influenced by the Anglo-Saxon culture,[10] and cultural differences between French- and English-speaking researchers have been described before. For example, Salager-Meyer et al. (2003) have found that French scientists tend to be more critical and more authoritarian than Anglo-Saxon scientists. They point out that their findings are in accordance with those of Motchane (1990) and

Régent (1994), indicating that "French scientists are much more prescriptive, authoritative and categorical than their English-speaking colleagues" (Salager-Meyer et al. 2003: 232). These cultural differences may explain why Beaufrère-Bertheux (1997: 232) refers to the "*hypermodestie*" of Anglo-Saxon scientists and why Sionis (1997: 211) explains an international journal's rejection of an article written in English by a group of French academics by referring to the authors' "exaggerated self-confidence". To the extent that such cultural differences really exist, it is reasonable to believe that they are reflected in the linguistic choices that the author makes, and thus people of other nationalities with other cultural backgrounds might find the style somewhat inappropriate.

As regards the gender factor, we conclude that gender does not seem to influence the frequency of epistemic modality markers in these research articles. At first sight, the results seem to suggest that women tend to hedge their research papers more heavily than men do and that this difference is greater in the Norwegian corpus (Table 8). However, there is great internal variation within the two groups (in the female group the relative frequency of the selected markers ranges from 0.3 to 9.5, and in the male group it ranges from 0 to 7.2), and individual variation is especially high in the Norwegian corpus. Seemingly large differences between Norwegian men and women might therefore be due to one or two atypical articles. Although the results indicate that gender is not an important factor (the gender differences were not statistically significant for any of the languages), the limited size of the corpus suggests that more research is needed before one can draw definite conclusions.

Concluding remarks

In this article, I have examined the use of a selection of epistemic modality markers indicating uncertainty in research articles written in English, French and Norwegian from the disciplines of medicine and linguistics. The results show that English- and Norwegian-speaking researchers use significantly more expressions of uncertainty than French researchers do. Also, the individual variation as regards the frequency of such expressions is greater among the English- and Norwegian-speaking researchers. The French researchers constitute a more homogeneous group in this respect. No considerable discipline-specific differences in the frequency of the selected markers could be detected between the two disciplines. There were, however, important differences as regards the markers preferred, especially in the French and the English corpus, where many of the markers were used almost exclusively in one of the disciplines. The author's gender did not seem to be relevant for the frequency of epistemic modality markers used.

In my opinion, awareness of the differences between the "French" style and the "English" style of academic writing can prevent culturally based

misunderstandings and misjudgements. This kind of knowledge also makes it easier for researchers to adjust to the style of other languages when writing in a specific foreign language. A native speaker of French at the Department of Romance Studies at my home institution in Norway said that while proofreading her Norwegian colleagues' French papers, she always advised them to delete a lot of hedges. This does not amount to correcting errors but rather to adjusting the style of the paper to the academic style appropriate for the language in which it is written. This style was clearly not the same as the one for the authors' mother tongue.

Knowledge of such differences is important not only for researchers but also for students who strive to acquire the appropriate academic style. Hedging strategies have been taught in academic writing classes for several years, but students should also learn about disciplinary differences and language-related differences. This latter point is especially important for students who are planning to write research papers in a foreign language. As regards the disciplinary differences, we have seen here that different disciplines do not necessarily make use of the same type of epistemic modality markers. It is therefore important that the teaching of hedging strategies be adjusted to the disciplinary affiliation of the students. Students need to learn the vocabulary of 'their' field, including not only specialised terms but also which lexemes are generally used in the discourse of their discipline and which are not.

Notes

1. One example of this is the fact that, in several countries (e.g. Norway and Finland), many students in the humanities are still told not to use the first person singular pronoun when writing research papers, although recent studies (e.g. those undertaken by the KIAP project, see Fløttum 2003a) show that the first person singular pronoun is in fact used to a great extent.
2. For a very full account of the term *hedging* and its uses, see Varttala (2001).
3. The material in parenthesis refers to the article from which the example is taken: engling05 means English linguistics article no. 5, engmed01 means English medicine article no. 1, etc.
4. See more details about the KIAP corpus on http://torvald.aksis.uib.no/kiap/ mdcorpusdescr.htm. For practical reasons, I will use terms like "the English authors" or "the English researchers" to refer to the authors of the articles written in English, although the authors might of course be of other English-speaking nationalities.
5. For a translation of the French and the Norwegian markers, see Tables 3 and 4, respectively.
6. For practical reasons, I give examples from the English texts only here. The problems related to polysemous markers and difficulties in distinguishing related meanings exist in the other languages as well, of course.

7. *Indicate* and *indiquer* may also be used in the sense 'say, express', as in "All respondents (100%), regardless of company size or activity, indicated that graduate employees' ability to work in teams was important for job success" (engling13). This use is quite easily distinguished and therefore not discussed further.

8. In other examples, it may be impossible to decide whether the person responsible for the modalization is the author him/herself, the author of a reported text, or both. The question of who is responsible for the modalization is further developed in Vold (in preparation).

9. Note that the markers are listed in their canonical forms – they are lemmas that represent the lexeme in question in all its inflected forms. Thus, *anta* (see Table 3) represents all forms of the Norwegian verb *anta* (*anta*, *antar*, *antok*, (*har*) *antatt*), though not forms of the adjective *antatt*. For the French verb *pouvoir* the situation is somewhat different: it has been studied only in the conditional and the present indicative; *peut* covers all present indicative forms of the verb (*peux*, *peut*, *pouvons*, *pouvez* and *peuvent*), while *pourrait* represents all conditional forms (*pourrais*, *pourrait*, *pourrions*, *pourriez* and *pourraient*. Homonyms are excluded, of course, so that occurrences of e.g. the noun *May* (the month) have been excluded from the very beginning.

10. Since English is spoken in so many different parts of the world, it is of course not possible to say that all authors having English as their mother tongue represent one unified culture. Nevertheless, they do have something in common that can be used to compare and contrast them with other language communities.

References

Bazerman, C. (1988) *Shaping written knowledge: the genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Beaufrère-Bertheux, C. (1997) L'anglais de la recherche médicale: une grande diversité. *ASp (Anglais de Spécialité)* 15–18: 223–38.

Crompton, P. (1997) Hedging in academic writing: some theoretical problems. *English for Specific Purposes* 16: 271–87.

Dahl, T. (2003) Metadiscourse in research articles. In K. Fløttum and F. Rastier (eds.), *Academic discourse. Multidisciplinary approaches*. Oslo: Novus. 120–38.

— (2004) Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics* 36.10: 1807–25.

Fløttum, K. (2003a) Personal English, indefinite French and plural Norwegian scientific authors? *Norsk Lingvistisk Tidsskrift* 21.1: 21–55.

— (2003b) Polyphony and bibliographical references. In K. Fløttum and F. Rastier (eds.), *Academic discourse. Multidisciplinary approaches.* Oslo: Novus. 97–119.

— (2004) La présence de l'auteur dans les articles scientifiques: étude des pronoms *je*, *nous* et *on*. In A. Auchlin et al. (eds.), *Structures et discours*. Québec: Ed. Nota Bene. 401–16.

— (2005a) The self and the others – polyphonic visibility in research articles. *International Journal of Applied Linguistics* 15.1: 29–44.

— (2005b) Individual linguistic variation and possibilities for academic self-hood? *Akademisk Prosa* 3: 87–100. University of Bergen: Department of Romance Studies.

— (2005c) MOI et AUTRUI dans le discours scientifique: l'exemple de la négation NE . . . PAS. In J. Bres, P.P. Haillet, S. Mellet, H. Nølke, L. Rosier (eds.), *Dialogisme, polyphonie: approches linguistiques*. Brüssel: de Boeck. Duculot. 323–37.

—, T. Dahl and T. Kinn (in preparation) *Academic voices – across languages and disciplines*.

Hyland, K. (1994) Hedging in academic writing and EAP textbooks. *English for Specific Purposes* 13: 239–56.

— (1998) *Hedging in scientific research articles*. Amsterdam: John Benjamins.

— (2000) *Disciplinary discourses: social interactions in academic writing*. Harlow: Longman.

Johns, T. (2001) From evidence to conclusion: the case of 'indicate that'. In M. Hewings (ed.), *Academic writing in context: implications and applications*. University of Birmingham Press. 55–62.

Le Querler, N. (1996) *Typologie des modalités*. Caen: Presses universitaires de Caen.

Lewin, B.A. (2005) Hedging: an exploratory study of authors' and readers' identification of 'toning down' in scientific texts. *Journal of English for Academic Purposes* 4: 163–78.

Lyons, J. (1977) *Semantics*. Cambridge University Press.

Markkanen, R. and H. Schröder (eds.) (1997) *Hedging and discourse: approaches to the analysis of a pragmatic phenomenon in academic texts* [*Research in text theory*, vol. 24]. Berlin: Walter de Gruyter.

Mauranen, A. (1997) Hedging in language revisers' hands. In R. Markkanen and H. Schröder (eds.), *Hedging and discourse: approaches to the analysis of a pragmatic phenomenon in academic texts*. Berlin: Walter de Gruyter. 115–34.

Meyer, P. (1997) Hedging strategies in written academic discourse: strengthening the argument by weakening the claim. In R. Markkanen and H. Schröder (eds.), *Hedging and discourse: approaches to the analysis of a pragmatic phenomenon in academic texts*. Berlin: Walter de Gruyter. 21–42.

Motchane, J.L. (1990) Chercher, inventer, innover dans la langue. In B. Cassen (ed.), *Quelles langues pour la science?* Paris: Editions de la Découverte. 4–62.

Myers, G. (1989) The pragmatics of politeness in scientific articles. *Applied Linguistics* 10: 1–35.

Palmer, F.R. (1986) *Mood and modality*. Cambridge University Press.

Perkins, M. (1983) *Modal expressions in English.* London: Frances Pinter.

Régent, O. (1994) L'article scientifique: un produit culturel. *ASp (Anglais de Spécialité)* 5/6: 55–61.

Salager-Meyer, F. (1994) Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes* 13: 149–70.

— and M. Angeles, A. Ariza and N. Zambrano (2003) The scimitar, the dagger and the glove: intercultural differences in the rhetoric of criticism in Spanish, French and English medical discourse (1930–1995). *English for Specific Purposes* 22: 223–47.

Sionis, C. (1997) Stratégies et styles rédactionnels de l'article de recherche: les ressources de l'utilisateur non-natif devant publier en anglais. *Asp (Anglais de Spécialité)* 15–18: 207–23.

Skelton, J. (1997) The representation of truth in academic medical writing. *Applied Linguistics* 18.2: 121–40.

Swales, J.M. (1990) *Genre analysis: English in academic and research settings*. Cambridge University Press.

Varttala, T. (1999) Remarks on the communicative functions of hedging in popular scientific and specialist research articles on medicine. *English for Specific Purposes* 18.2: 177–200.

— (2001) *Hedging in scientifically oriented discourse. Exploring variation according to discipline and intended audience.* Electronic doctoral dissertation. Acta Electronica Universitatis Tamperensis 138 (http://acta.uta.fi/pdf/951-44-5195-3.pdf).

Vassileva, I. (1997) Hedging in English and Bulgarian academic writing. In A. Duszak (ed.), *Culture and styles of academic discourse* [Trends in lingustics studies and monographs 104]. Hawthorne, N.Y.: Mouton de Gruyter. 203–21.

— (1998) Who am I/who are we in academic writing? A contrastive analysis of authorial presence in English, German, French, Russian and Bulgarian. *International Journal of Applied Linguistics* 8.2: 163–90.

Vihla, M. (2000) Epistemic possibility: a study based on a medical corpus. In J.M. Kirk (ed.), *Corpora galore: analyses and techniques in describing English* [Papers from the 19th international conference on English language research on computerised corpora (ICAME 1998), Language and computers no. 30]. Amsterdam: Rodopi. 209–224.

Vold, E.T. (in preparation) The choice and use of epistemic modality markers in linguistics and medical research articles.

[Final version received 14/10/05]

Eva Thue Vold
Department of Romance studies
University of Bergen
Øysteins gt 1
5007 Bergen
Norway
e-mail: eva.vold@roman.uib.no