

Epistemic Markers in NS and NNS Academic Writing

Tim Vandenhoeek
Musashi University

Abstract

The use of epistemic markers – words and phrases employed to show differing degrees of certainty and hesitation – is an important element of academic writing. Previous research has suggested that this is an area in which non-native speakers (NNS) struggle. Studies have indicated that NNS are prone to making overly strong statements and exhibit less range and sophistication in the devices that they do employ than native speakers (NS). However, there have as yet been few studies examining the use of such language by advanced NNS students in university academic writing. This corpus-based study examines the ways in which NS and advanced NNS students use epistemic markers in university academic writing. Two corpora totalling 31 659 words were formed from the discussion and conclusion sections of a written university assignment and analyzed for use of epistemic markers. The results indicate that though both groups employ the language at comparable rates and display similar levels of commitment in their writing, these NNS writers display some features of epistemic marker use found to be generally characteristic of other NNS writers.

Introduction

One of the main skills any university student must master is that of academic writing. Writing in a foreign language is viewed as challenging in and of itself (Hinkel 1997) and within that area lie several sub-skills, including the use of epistemic modality: the ability to show differing degrees of commitment to the truth of a proposition. Appropriate use is seen as both an important skill and one with which NNS writers struggle (Hyland 1995; Holmes 1988). This study uses two corpora to examine the ways in which one set of NS and advanced NNS students use epistemic markers – words or phrases used to express epistemic modality – in the same university academic writing task by analyzing the comparative frequency, range of markers employed and the degree of commitment shown by the items.

Literature Review

Within any discipline, there exists a common knowledge base comprised of propositions which are widely accepted as facts. Outside that, however, exist ideas which must be stated with care and caution (Hyland 1996b) and writers 'need to make a clear distinction' between the two (Crompton 1997: 274). McEnery and Kifle define epistemic modality as 'the different levels of commitment that a writer brings to his/her writing' (2002: 182); this definition will be used as the basis for this study ('epistemic marker' will be used to describe any word or phrase used to express epistemic modality).

Epistemic modality can be divided into hedging and boosting, i.e. the respective weakening and strengthening of a claim. Hyland defines a hedge as:

any linguistic means used to indicate either (a) a lack of complete commitment to the truth of an accompanying proposition or (b) a desire not to express that commitment categorically. Hedges are therefore the means by which a writer can present a proposition as an opinion rather than a fact. (1996a: 478)

Crompton states 'a hedge is an item of language which a speaker uses to explicitly qualify his/her lack of commitment to the truth of a proposition he/she utters' (1997: 281). Hinkel writes that 'in written text, hedging represents the employment of lexical and syntactic means of decreasing the writer's responsibility for the extent and truth-value of propositions and claims' (2005: 30). Hedging can be further divided into expressing two levels of hesitation, probability (seems, should, expect, etc.) and possibility (possible, may, sometimes, etc.). The opposite – the use of epistemic markers to show certainty (will, must, certain, etc.) and increase one's commitment to a proposition – is known as boosting. Boosting occurs less frequently than hedging (Carter and McCarthy 2006; Hinkel 2005) and is the subject of far less research, particularly as it is found in research articles (RAs). In the following statements from the corpora used in this study (emphasis added), the differing levels of commitment can be easily seen:

The Japanese society uses the same means of communication as anyone else in the developed part of the world but Mixi **clearly** shows also a form of glocalisation (NNS8).

There was **evidence** of ideoscapes on the website, as spreading the ideologies of environmentalism and sustainability takes a central role on the website in its press releases (NS19).

This **could** also give weight to the argument that the community on YouTube is a virtual community (NS18).

In the first case, *clearly* is a certainty marker used to strengthen the claim and is an example of boosting. The second and third examples are in contrast far more tentative and are examples of hedges to show probability and possibility respectively.

Corpus research is the basis of almost all enquiry into epistemic marker use in writing and all the research examined below is the result of corpus study. There has been considerable research into the area of how hedges in particular are used within academic literature, though little of this scholarship examines the use of boosting. Hyland examined a corpus of 26 biology journals and found articles within to be 'extensively hedged' with 2.0 hedges per 100 words (1995: 35). A high concentration of devices in the results and discussion sections was also noted, with 82% of all devices found in those sections. Malaskova (2011) examined five literary criticism and linguistics RAs and found a similar rate of 1.9 hedges per 100 words. Vasquez and Giner (2008) found that marketing articles are hedged far more frequently than two other scientific disciplines (2.54, 1.09 and .70 per 100 words respectively). Varttala (2001) compared the use of hedges in RAs from three fields (economics, medicine and technology) with usage in popular publications of the same disciplines. It was found that hedges occur most frequently in the results and discussion sections of RAs and overall rates of between 2.18 and 3.84 items/100 words. In nearly all the above studies, lexical verbs were the most often used grammatical category of item.

Epistemic modality is an area in which NNS student writers are seen to struggle. NNS writing in English is often seen as excessively direct (Hinkel 2005; Hyland and Milton 1997) and Hyland argues that even strong NNS writers display a 'failure to hedge statements adequately' (1995: 39). NNS writers often rely on a less varied range of items and are further hindered by what is seen as poor coverage of the subject in many ELT textbooks (Salager-Meyer 1994; Hyland 1994) and the fact that some style guides state that writers should avoid tentativeness (Hyland 1996a).

One of the largest studies into epistemic modality in student writing is that of Hyland and Milton (1997). The paper compared examinations written by Hong Kong high school students to those written by British NS students. It was found that while NS and NNS students used epistemic markers at essentially the same rate (NS- 1.82/100 words, NNS- 1.83/100 words) there were clear differences in the range and frequency of the items and NNS writers relied on a narrower set of items. NNS students included 60% more certainty markers, and these 'results seem to confirm the view that the academic writing of many NNS learners is characterised by firmer assertions' (Hyland and Milton 1997: 193). In another study, McEnery and Kifle (2002) studied a corpus of argumentative essays written by Eritrean university students and 16-year-old British students. Eritrean students used the devices at a rate of 1.99/100 words and the British students used 3.12/100 words. There was a large gap in the type and range of devices utilized by each group, with native speakers using far more lexical verbs than their NNS counterparts. They found that NS writers employed twice as many certainty markers as the Eritrean writers, a sharp contrast to most other studies. Nivales (2011) studied introduction and conclusion sections of referenced undergraduate theses written by NNS students in five courses (English language, political science, mass communications, biology and psychology). Nivales found that the writers in that study showed a 'tendency to be more direct and committed to statements' (2011: 41). Seskauskiene (2008) examined the introductions of 31 undergraduate theses by Lithuanian learners and found comparatively fewer markers. Writers in that study used epistemic markers at a rate of 1.45/100 words. Seskauskiene also examined the number of markers in each essay and found significant variation, from 0.29/100 words to a maximum of 2.5/100 words. Oh (2007) compared the essays of low to middle-intermediate Korean college students to those written by American and British university students from the Louvain Corpus of Native English Essays (LOCNESS). Using a larger list of markers than other studies, the investigation found that the groups used epistemic markers at nearly the same rate (2.81 and 2.83 items/100 words) and that NNS relied on a narrower range of devices and employed 9% more certainty markers than the NS group. Hu and Li (2015) compared the essays of Chinese learners at three levels- CEFR A2, B1 lower and B1 upper- with those of NS writers. Importantly, the corpora were comprised of essays replying to one of two possible prompts, so comparisons between the corpora could be seen as more meaningful than other studies. They found that the NNS employed slightly fewer markers in total and fewer boosters than NS writers with all three NNS groups showing broadly similar results. NS writers included noticeably more lexical verbs and all NNS groups showed less range in the number of different markers used. In a paper comparing epistemic marker use between Chinese NNS writers from five proficiency levels and expert British NS writers (published texts selected from the broader British National Corpus), Chen (2010) found, perhaps unsurprisingly given the different task types and backgrounds of the writers, a vast difference in the rates of epistemic marker usage and that the NNS groups used more certainty markers. Examining the writing of five NNS proficiency levels led to the conclusion that 'with increasing proficiency, the NNS writers perform more like the NS writers in terms of their linguistic behaviors in applying epistemic devices' (Chen 2010: 42). The general trend from the above studies is that, compared to NS writers, NNS students seem to use a narrower range of items and tend to show more certainty. Speaking broadly, both groups employ more adverbs and modals verbs than any other grammatical category of epistemic marker, with NNS writers particularly reliant on modal verbs.

Despite these studies, there remain several potential gaps in the research on the issue of epistemic markers in NNS writing. Previous studies comparing NS and NNS writing have mainly examined argumentative essays rather than authentic university academic writing assignments which include citations. Studies drawing data from NNS university writing often lack an NS complement for comparative purposes and in those that do, the tasks compared are generally similar rather than identical. The present study seeks to add to the scholarship on the issue by examining the written work of advanced NNS writers and comparing it to that of their NS classmates writing the exact same task.

Methodology

The data for this study is the written assignments of students in New Media, Language and Globalization, a fourth-year course in a university in Ireland. Prior to the study, the details of the proposed research were presented to the Research Ethics Committee and official permission to proceed was granted. Participation

was voluntary and participants received no compensation. Participating students completed a brief survey on their linguistic backgrounds (first language, years of English study, etc.). All assignments were anonymized and assigned a code. The assignment (guidelines and grading rubric in Appendix 1) is a study of a global website and has four sections: an in-depth description of the site and its linguistic landscape, a computer mediated discourse analysis, analysis in which students link observations from the first two sections to theories relating to globalization, language and mass media, and a conclusion. Only the latter two portions were analyzed as they are the sections in which propositional content and a degree of argumentation is found and may have potential for epistemic markers. The following examples are representative of the type of argumentation found in these sections and both show the use of epistemic markers to link observations to broader theories:

There is a clear demonstration of language hierarchy and also a threat towards many other languages (NS8).

Firstly, we can suggest Twitter functions as an Imagined Community as defined by Benedikt Anderson (NS17).

These sections were gathered together into two corpora, an NS corpus of 23 samples totalling 22 559 words and nine NNS samples totalling 9100 words. The combined size is 31 659 words.

The NS corpus (Appendix 2) is comprised of 21 Irish and two American students, all of whom are NS in English. The NNS section (Appendix 3) includes nine writers from different backgrounds who all identify a European language other than English as their first language. The NNS corpus could be considered representative of proficient users as they report having studied English for an average of 12 years. The corpora generally represent strong assignments; as shown in Table 1, over 30% received A grades and only one received lower than B3. The NNS component is particularly strong, with none scoring below B2. As part of the total grade, each received a mark for quality of writing, comprising 10% of the total mark. All participants scored between 7-10 out of 10, again suggesting that the NNS writers are quite proficient in academic writing.

Table 1. Grades

Corpus	A-band		B-band		C-band	
	Number	Percent	Number	Percent	Number	Percent
NS	7 (A1=2, A2=5)	30.4%	15 (B1=9, B2=3, B3=3)	65.2%	1 (C1=1)	4.3%
NNS	3 (A1=1, A2=2)	33%	6 (B1=2, B2=4)	66%	0	0%

The corpora were analyzed using AntConc concordance software (Anthony 2011) for the usage of 75 epistemic markers from the list devised by Hyland and Milton (1997) (Appendix 1) to create frequency lists. As several of those items have multiple meanings and functions, the use of each was individually examined to assess if it was epistemic in nature. For example, in the first instance below, *will* is used in a meta discourse function and such usages were not included in the frequency lists. However, the second example shows employment of *will* to make a strong prediction about the future and is counted.

I will concentrate on what I found the most relevant and of the most importance. (NNS2)

Society will expect from people to speak and understand English whether if they want to deal with people from other countries or not. (NNS1)

Those items were further analyzed by degree of commitment and classified into three categories – certainty, probability and possibility – based on the lists used in the Hyland and Milton (1997) and McEnery and Kifle (2002) studies.

Results

Use of epistemic markers

The NS corpus contains 260 epistemic markers, which averages 1.15 marker per 100 words. The NNS corpus has 91 items, averaging 1/100 words. From a total of 75 possible items, there are 54 different devices in the NS corpus compared to 40 from the NNS section. While NS writers use a wider range of markers, the size of the corresponding NNS corpus surely accounts for some of this difference.

Modal verbs feature heavily among the most commonly used markers in each corpus as seen in Table 2. Four of the first five items for each group are modals (*will*, *would*, *could*, and *may*). Five items (the above modals and *quite*) appear on both lists. The ten most common as a proportion of the total show an even distribution, forming 50.5% of NS markers and 49.8% of NNS markers.

Table 2. Most common items

NS			NNS		
Item	Uses	% of total	Item	Uses	% of total
1. would	23	8.2%	1. will	11	11.6
2. will	19	6.8%	2. would	9	9.5%
3. may	19	6.8%	3. quite	5	5.3%
4. often	14	5%	4. could	4	4.2%
4. could	14	5%	4. may	4	4.2%
5. suggest	13	4.6%	5. seem	4	4.2%
6. quite	11	3.9%	5. expect	4	4.2%
7. evident	10	3.5%	6. might	3	3.2%
7. think	10	3.5%	6. should	3	3.2%
8. almost	9	3.2%	6. claim	3	3.2%
8. always	9	3.2%			
Total (first ten only)		50.5%	Total		49.8%

Grammatical categories of epistemic markers

Figure 1 shows that both groups rely mainly on adverbs, modal verbs and lexical verbs, with these three forming nearly 90% of all epistemic markers in both corpora. Adverbs (Table 3) are the most common word form in both corpora, accounting for 37.5% of all items in NS writing and 36.2% in NNS writing. The wide range of possible items is partially responsible for their frequency compared to other categories; of a possible 37 items, NS writers employ 27 different adverbs and the NNS corpus contains 21. The writers use mainly different adverbs, with only two shared items among the five most common in each corpus (*quite*

and *always*). The five most frequent items form comparable proportions of the whole, 47.7% in the NS section and 46.9% in NNS work.

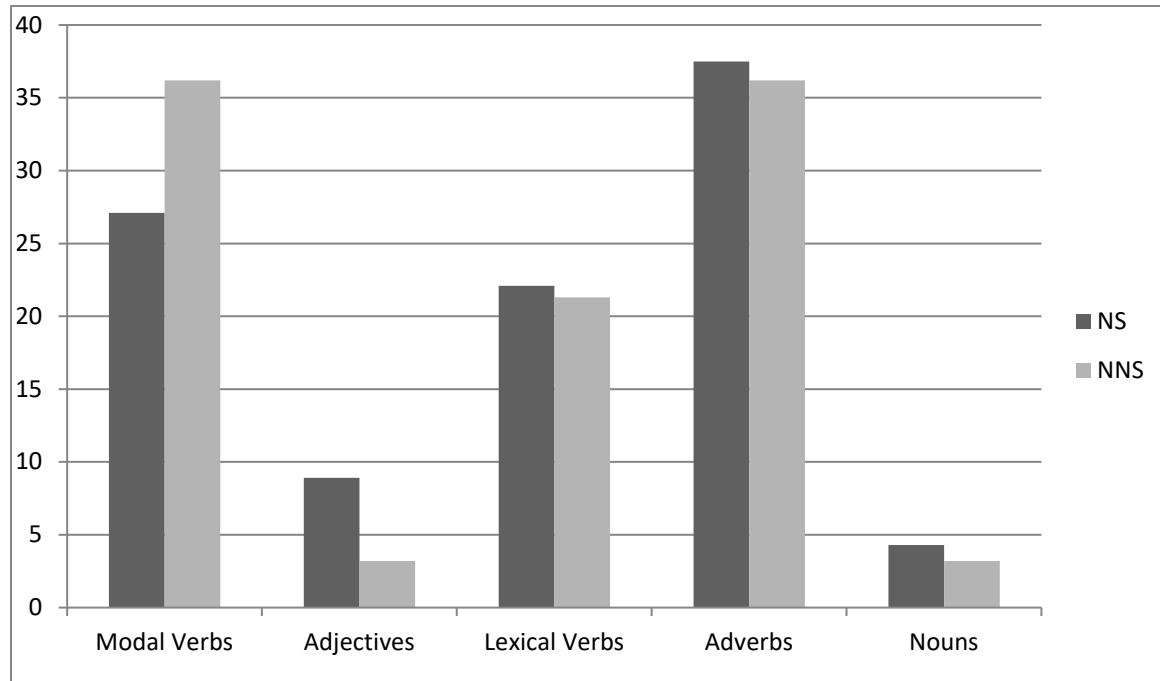


Figure 1. Grammatical category of epistemic markers (% of total)

Table 3. Most common epistemic adverbs

NS		NNS	
Item	Uses	Item	Uses
often	14	quite	5
quite	11	indeed	3
always	9	clearly	3
perhaps	7	never	3
in fact	7	probably	2

Modal verbs, shown in Table 4, represent a large proportion of devices used in both corpora. As noted, *would* and *will* are the most common overall markers and individual modals are the most frequently employed markers of probability (*would*), possibility (*may*) and certainty (*will*) for both groups. Modals account for 27.1% of all items used by NS writers, the second most frequently occurring category of epistemic marker in that corpus. The three most common items, *will*, *would* and *may*, make up just over 80% of modal use. The NNS group is more reliant on modals, equal with adverbs at 36.2% as the most

Table 4. Most common epistemic modal verbs

NS		NNS	
Item	Uses	Item	Uses
would	23	will	11
will	19	would	9
may	19	could	4
could	14	may	4
might	1	might	3

frequently found type. The NNS group shows slightly more range in use than the NS group, particularly *might* and *should*, which are actually employed more by NNS writers (three times each compared to one occurrence of *might* and none of *should* in NS writing).

Table 5. Most common epistemic lexical verbs

NNS		NS	
Lexical Verbs	Uses	Lexical Verbs	Uses
suggest	13	seem	4
think	10	expect	4
believe	8	claim	3
seem	7	tend	3
appear	5	think	2

Lexical verbs, seen in Table 5, are the third most common category of device for both groups and occur at almost identical rates, 22.1% of NS items and 21.3% of NNS items. There are differences in the range of devices and especially how some items are used. Of the five most common verbs in each corpora, only *seem* and *think* appear on both lists, and *suggest*, the most common verb in the NS section, is found only once in the NNS corpus. The most salient difference between the groups is how the items are utilized, particularly *think* and *believe*, the second and third most common lexical verbs for NS writers. These items generally (seven of ten for *think*, six of eight for *believe*) include self-reference in the NS section:

Although I am unsure as to their level of pay, I think it is safe to assume that they are not millionaires (NS8).

I personally believe that the dominance of the English language on the Vodafone site is not quite as detrimental to the other languages on the site as it first appears to be (NS1).

These combinations are absent from the NNS corpus.

Neither group make extensive use of adjectives or nouns. Adjectives account for 8.9% of items used by NS writers and 3.2% for NNS writers. Nouns are by far the lowest utilized category of epistemic markers in either corpus, 4.3% and 3.2% of NS and NNS items respectively.

Degree of commitment

Both groups show similar degrees of commitment and hesitation in their writing, in terms of the frequency of markers from each category of commitment (certainty, probability and possibility) and the relative proportions of each. As NS writers utilize more markers overall, Figure 2 shows they employ slightly more from each category, 0.59 items per 100 words compared to 0.51/100 for probability, 0.36/100 compared to 0.34/100 for certainty and possibility at 0.2/100 in the NS corpus and 0.15/100 in the NNS section. Figure 3 shows commitment proportionally and reveals that the writers show similar degrees of qualification in their written work. Both use probability markers more than any other category, 51% in both corpora. NNS writers use 3% more certainty and 3% fewer possibility markers than the NS writers.

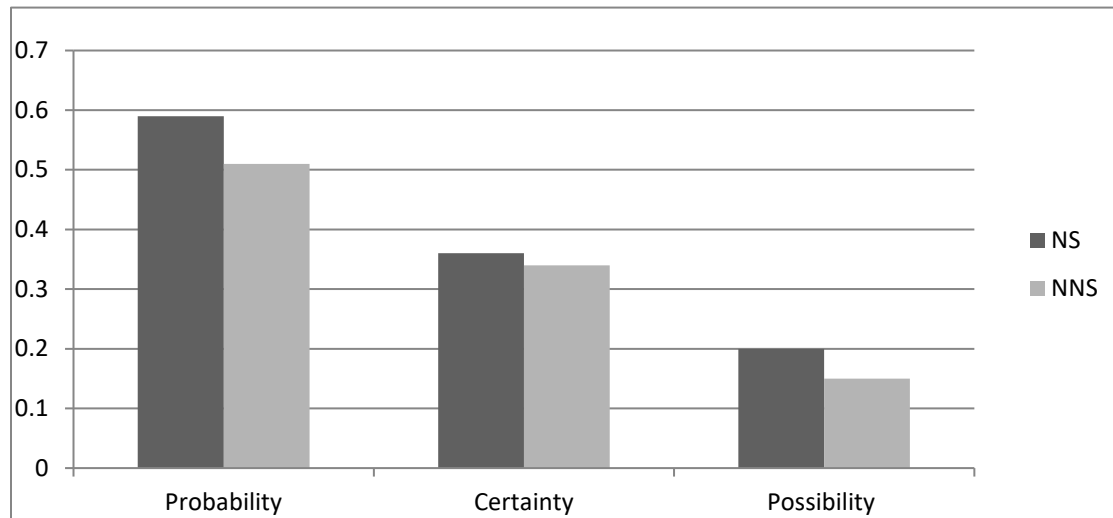


Figure 2. Degree of commitment by category (items/100 words)

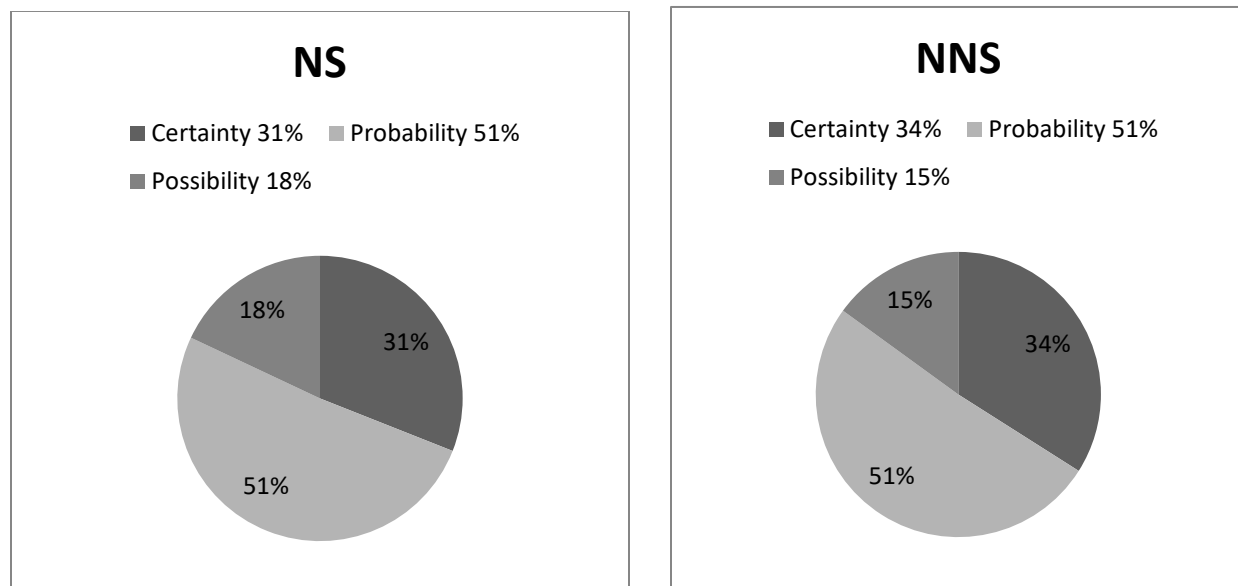


Figure 3. Epistemic category by proportion

Epistemic markers in high and low scoring papers

The data shows no link between either overall grade or style grade (quality of writing) and the number of epistemic markers used, or the degree of commitment expressed. This trend is most evident in the NS section. The three highest scoring assignments (Table 6) along with the two lowest scoring papers (Table 7) were individually examined for epistemic markers. The highest scoring assignment (NS9) received full marks yet includes only two epistemic markers, equivalent to 0.14 items per 100 words, far below the NS average of 1.15/100. NS18, which scored A1 overall and 9/10 for style, uses 1.2/100 (slightly above average) and NS19 (A2, style 9/10) has below average epistemic marker use at 0.89/100. An examination of the lowest scoring papers shows above average frequency of epistemic markers. Both writers employ numerous items, almost all hedges, at higher than average rates (1.73 and 1.59/100 words) and received overall grades of B3 and C1 and style marks of 7, the lowest in either corpus.

Table 6. High scoring NS assignments

Student Details	NS9 Grade A1 Style score 10/10 1412 words	NS18 Grade A1 Style score 9/10 1167 words	NS19 Grade A2 Style score 9/10 1234
Items	2	14	11
Items/100 words	0.14	1.20	0.89
Category of Items	Certainty 1 (50%) Probability 1 (50%) Possibility 0 (0%)	Certainty 5 (33.3%) Probability 5 (33.3%) Possibility 4 (33.3%)	Certainty 1 (9.1%) Probability 10 (90.9%) Possibility 0 (0%)

Table 7. Low scoring NS assignments

Student Details	NS22 Grade B3 Style score 7/10 809 words	NS16 Grade C1 Style score 7/10 313 words
Items	14	5
Items/100 words	1.73	1.59
Category of Items	Certainty 1 (7.1%) Probability 7 (50%) Possibility 6 (42.9%)	Certainty 1 (20%) Probability 1 (20%) Possibility 3 (60%)

The NNS results (Table 8) show that the three highest scoring assignments have a wide range in frequency of epistemic markers and degree of commitment without any accompanying variation in overall score or style grade. The assignments display marker use and certainty at rates above, equal to and well below the NNS averages. No NNS assignment scored beneath B2 which means analysis of low scoring NNS work is not possible.

Table 8. High scoring NNS assignments

Student Details	NNS4 Grade A1 Style score 8/10 1169 words	NNS1 Grade A2-B1 Style score 9/10 913 words	NNS5 Grade A2-B1 Style score 8/10 593 words
Items	12	10	2
Items/100 words	1.03	1.10	0.33
Category of Items	Certainty 5 (41.7%) Probability 5 (41.7%) Possibility 2 (16.6%)	Certainty 3 (30%) Probability 5 (50%) Possibility 2 (20%)	Certainty 0 (0%) Probability 2 (100%) Possibility 0 (0%)

Clusters

A final feature of epistemic markers in the corpora is the use of clusters, or combinations of two or more epistemic markers. Table 9 shows that there are 20 examples of clusters in the NS corpus but just two in the NNS section.

Table 9. Clusters

Structure	NS	NNS
modal + adverb (could possibly)	10	2
modal + lexical verb (may indicate)	4	0
lexical verb + complement with adverb (I think that it most certainly was)	2	0
lexical verb + lexical verb (tends to suggest)	1	0
lexical verb + complement with lexical verb (I think it is safe to assume)	1	0
adjective + complement with modal verb (possible that they could)	1	0
adjective + noun (clear indication)	1	0
Total	20	2

The most common structure is a modal verb with an adverb, of which there are 12 examples (emphasis added):

Many people think that the usage of English as a common language of communication on the internet is a threat to multilingualism and **could possibly** have detrimental effects on other languages. (NS10).

People already know how to deal with those things, so it's easier to handle the page and they **would more likely** buy something (NNS1).

Discussion

The data from this study reveals several similar traits and key differences in the use of epistemic markers by NS and NNS students. These writers use epistemic markers at comparable rates, 1.15/100 words by NS writers and 1/100 words by NNS writers. The difference that does exist is partially accounted for by NS use of clusters, as counting each cluster as one device lowers effective NS rate to 1.10/100 words. There is no clear trend in previous research as to which group employs more epistemic markers, but student writers in similar studies used noticeably more markers (1.82-3.3 items/100 words) than writers in this study, though each study was based on different lists of markers and task types. Several factors could account for this. One clear possibility is the difference in task types. The corpora in other studies were mainly comprised of argumentative essays and likely contain a higher amount of propositional content. In fact, such essays often include little other than arguments and supporting evidence which requires hedging or boosting. The task in this study would appear to require less epistemic marking as the propositional content is somewhat lower, resulting in lower rates of usage. Also, those studies primarily examined unreferenced essays. The degree to which writers make use of direct quotations or paraphrasing to support their argumentation is a research question not examined here, but could account for some of the variation. The difference between the writing styles and abilities of the participants in previous studies (some of whom are secondary school students or low-intermediate NNS) and mainly fourth-year university students and advanced NNS writers is another factor that almost certainly impacts the results.

Comparing the use of epistemic markers between these students and expert writers shows that writers in this study employ epistemic language less frequently. Previous research shows a wide range but a rough average of 2/100 words in published writing. Again this may be expected as expert writers often need to handle large quantities of empirical data in a manner which passes the rigour of peer review by displaying adequate hesitation. The writers in this study should have a basic awareness of the necessity of epistemic language; however, the assignments contain relatively little data (observations of one website) and thus less need to employ such language.

There is also notable similarity in the items most frequently used by the writers in this study. Of the 10 most commonly employed markers in the corpora, five are found on both lists and the 10 most common markers additionally form nearly identical proportions of total epistemic marker use. As found in previous studies, both groups rely on the same grammatical categories – adverbs, modal verbs and lexical verbs – to show epistemic modality. Given the number of possible items (37, nearly 50% of the total) and their range of functions, the finding that adverbs are the most common category of epistemic marker is not surprising. While adverbs are found more in total, individual modal verbs are four of the five most commonly used items in both corpora and the most common markers for each category of commitment. The frequency of modal verbs is a trend found in other studies and could be considered a feature of non-expert academic writing. In studies of expert writing, modal verb use was more limited, not exceeding 17.3% of total items, while these corpora both show higher frequency (NS= 27.1%, NNS= 36.2%). Perkins states that ‘modal expressions provide the least marked, and thus the most straight forward, means of expressing modality in English’ (1983: 104), a possible explanation for the high rates of usage by novice (NS and NNS) writers. The results for lexical verbs again align quite closely with previous studies in terms of relative frequency being substantially lower than adverbs or modal verbs. As lexical verbs are generally the most common epistemic markers in expert writer corpora, this is another potentially revealing difference between novice and expert writing. There is not extensive use of either nouns or adjectives for epistemic effect by either group, a trend found in previous studies exploring expert and student writing.

In these corpora, NS and NNS writers also exhibit largely similar levels of commitment. This finding does not support the generally held belief (supported by several empirical studies) that showing appropriate hesitation is a problematic issue for NNS writers. As each study surveys the written work of NNS students from various backgrounds, variables such as first language writing traditions, English learning history, learning materials and others could account for some or all of this variation. It is additionally possible that the perceived failings of teaching materials identified in previous research have been addressed and this group of students has benefited from those changes, as studies which found that NS and NNS writers show

similar levels of commitment tend to be more recent. Given that this group of NNS writers can be seen as advanced users, what the results from the current study do suggest is that as writing ability increases, use of epistemic modality by NNS writers approaches that of their NS counterparts, as previously found by Chen (2010).

Those similarities notwithstanding, there are three clear differences between NS and NNS utilization of epistemic markers, all of which suggest that although the NNS writers in this study are of an advanced level, there are still gaps between their language use and that of NS writers. First, as found in all previous studies, NNS writers in this study use fewer distinct markers than the NS group (40 compared to 54). Though surely the relative size of the NNS corpus accounts for some of this gap, the notion that NNS students have a smaller productive vocabulary is an expected result. Additionally, the NNS writers show more reliance on modal verbs, another finding which mirrors results from previous studies. Modals are indeed the only category in which NNS writers use more distinct items (*should* is absent from the NS corpus). The above comments relating to novice use of modals likely applies even more so to students writing in a foreign language and Hyland (1996) further notes that textbooks focus heavily on modal verbs as a means of expressing epistemic modality, another possible explanation for their frequency. The third difference is epistemic clusters. With only 22 occurrences across both corpora, usage is not particularly significant, but their near-absence from NNS writing is worthy of mention. Banks (1994, cited in Hyland and Milton 1997) considers clusters to be a feature of expert writing and their infrequency in NNS writing is another noteworthy variation in language use between the two groups.

The results indicating that rates of epistemic language use have no effect on either the overall grade or the style component should be treated with caution. This study only examines one type of language in two sections and, while the overall and style grades reflect the assignments in their entirety, no meaningful conclusions can be drawn from this data. Some aspects of the results are nonetheless interesting, particularly the variation in the highest scoring assignments as similarly performing assignments range from almost no use of epistemic markers to usage at or above average in both corpora. Hyland and Milton found that NNS students in lower scoring bands used 'considerably less' epistemic language than those receiving higher grades (1997: 190), but the lowest scoring NS assignments in this study include this language at higher than average rates, further indicating that adequate epistemic marker use in this context is not linked to performance.

Conclusion

Pedagogical implications

As both corpora in this study are comprised of what would generally be considered strong writing and as no connection between use of epistemic language and performance was found, there are relatively few pedagogical implications to this study and those that do exist are minor in nature. NNS writers use a narrower range of items than the NS group but a recommendation of expanding their productive vocabulary is one that applies to most NNS and many NS students, regardless of proficiency level or task type. One clear difference between the writers in the study is the use of clusters and NNS students may benefit from explicit instruction regarding these structures. A somewhat ancillary issue is personal reference in academic writing. *I think* and *I believe* are absent from the NNS corpus despite occurring with modest frequency in the NS section. There is little overall clarity on the issue and it is likely one that should be explicitly addressed by individual lecturers.

Suggestions for further research

There are several possibilities for further research on this topic. First, larger corpora should be used to test the findings of this study as they could produce more meaningful results, particularly if the NNS component were equal in size to the NS corpus. Also, as few studies seem to exist investigating epistemic marker use by European students writing in English as a second language, a corpus comprised of writers from a common first language background would provide more data relating to challenges or opportunities specific to that group. While several studies have examined this language in research articles and argumentative

essays, more work examining its role in different genres, particularly referenced university assignments, could also be of value. The assignment surveyed in this study is just one example of a university writing task and corpora such as the British Academic Written English Corpus (BAWE) or the Michigan Corpus of Upper Level Student Papers (MICUSP) provide access to the work of writers spanning several disciplines and task types. Of particular interest would be a study examining writing containing a higher level of propositional content, one in which, theoretically at least, epistemic language would play a more central role. Little research seems to exist on boosting and its use by either expert or novice writers, clusters in academic writing and any link between grades and epistemic modality, and all seem to be areas worthy of further enquiry.

References

- Anthony, L. (2011) *AntConc*. Version 3.2.2. <<http://www.laurenceanthony.net/software/antconc/>> [7 August 2018]
- Carter, R., and McCarthy, M. (2006) *Cambridge Grammar of English*. Cambridge: CUP
- Chen, H. I. (2010) 'Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence in L2 writing'. *Arizona working papers in SLA & teaching* 17, 27-51
- Crompton, P. (1997) 'Hedging in academic writing'. *English for Specific Purposes*, 16 (4), 271-287
- Hinkel, E. (2005) 'Hedging, inflating and persuading in L2 academic writing'. *Applied Language Learning*, 15 (1/2), 29-53
- Hinkel, E. (1997) 'Indirectness in L1 and L2 academic writing'. *Journal of Pragmatics* 27 (3), 361-386
- Holmes, J. (1988) 'Doubt and certainty in ESL textbooks'. *Applied Linguistics* 9 (1), 21-44
- Hu, C., and Li, X. (2015). 'Epistemic Modality in the Argumentative Essays of Chinese EFL Learners'. *English Language Teaching*, 8 (6), 20-31
- Hyland, K. (1996a) 'Nurturing hedges in the ESP curriculum'. *System* 24 (4), 477-490
- Hyland, K. (1996b) 'Writing without conviction? Hedging in science research articles'. *Applied Linguistics* 17 (4), 433-454
- Hyland, K. (1995) 'The author in the text'. *Hong Kong Papers in Linguistics and Language Teaching* 18, 33-42
- Hyland, K. (1994) 'Hedging in academic writing and EAP textbooks'. *English for Specific Purposes* 13 (3), 239-256
- Hyland, K., and Milton, J. (1997) 'Qualification and certainty in L1 and L2 students' writing'. *Journal of Second Language Writing* 6 (2), 183-205
- Malaskova, M. (2011) 'Hedging in research articles'. in *Academic Writing in Europe*. ed. by Schmied, J. Gottingen: Cuvillier, 143-156
- McEnery, T., and Kifle, N. A. (2002) 'Epistemic modality in argumentative essays of second-language writers'. in *Academic Discourse*. ed. by Flowerdew, J. London: Pearson Education Limited, 182-195
- Nivales, M. L. M. (2011) 'Hedging in college research papers'. *Asian EFL Journal* 52, 35-45
- Oh, S. Y. (2007). 'A corpus-based study of epistemic modality in Korean college students' writings in English'. *English Teaching* 62 (2), 147-175
- Perkins, M. (1983) *Modal Expressions in English*. London: F. Pinter
- Salager-Meyer, F. (1994) 'Hedges and textual communicative function in medical English written discourse'. *English for Specific Purposes* 13 (2), 149-170
- Seskauskiene, I. (2008) 'Hedging in ESL: a case study of Lithuanian learners'. *Studies About Languages* 13, 71-76

Varttala, T. (2001) *Hedging in Scientifically Oriented Discourse*. [online] available from <<https://tampub.uta.fi/bitstream/handle/10024/67148/951-44-5195-3.pdf?sequence=1>> [7 August 2018]

Vazquez, I., and Giner, D. (2008) 'Beyond mood and modality'. *Revista Alicantina de Estudios Ingleses* 21, 171-190

Appendix 1: Assignment Guidelines and Marking Rubric

CU4128 – Project Report Autumn 2011 – 60% of Final Mark

Please submit your report (2000 words) by Friday of week 13 (deadline is 4.00 pm on Friday 2nd December) to the School of LLCC (in the marked drop box).

Your report should follow this structure:

1. Title Page with Title of project, Name, ID number and Course
2. Introductory paragraph: (what are you doing, why is this important, what is the context, what are you studying) plus signposting of the report.
3. Section 1: Virtual ethnography (i.e. detailed description) of site.
4. Section 2: Linguistic landscape of sites or computer-mediated discourse analysis of selection.
5. Section 3: Analysis of what this tells us about globalization, new media and language, using sources referred to in lectures and on sulis.
6. Concluding paragraph: summarising the main points of the report
7. References: Use any style you want for your list of references, but make it consistent, make sure that a reader can find reference in the text in this list, make sure that only texts referred to in the project are in the list.
8. Appendices – if necessary.

To help you to prepare, here are the criteria I will use to mark your final report.

	A1	A2	B1	B2	B3	C1	C2	C3	D1	D2	F
Section 1. Virtual ethnography (in-depth description of site) 10%											
Section 2. Linguistic landscape analysis/ CMDA/Search engine efficacy etc. 25%											
Section 3. Analysis – linking 1 and 2 to concepts and theories discussed in the lecture and to the issues of new media, language and globalisation 30%											
Understanding of content and concepts and ability to explain 10%											
Evidence of reading beyond lecture notes 10%											
Relevance to module and to project brief 5%											
Style, format, quality of writing etc. of report 10%											

Appendix 2: NS Corpus Details

Code	Year of Study	Nationality	Grade
NS1	4	Ireland	A2
NS2	4	Ireland	A2
NS3	3	US	B3
NS4	4	Ireland	B3
NS5	4	Ireland	B1
NS6	4	Ireland	B1
NS7	4	Ireland	B1
NS8	4	Ireland	B1
NS9	4	Ireland	A1
NS10	4	Ireland	B1
NS11	4	Ireland	B2
NS12	4	Ireland	B1
NS13	4	Ireland	A2
NS14	4	Ireland	B1
NS15	4	Ireland	B2
NS16	4	Ireland	C1
NS17	4	Ireland	B1
NS18	4	Ireland	A1
NS19	4	Ireland	A2
NS20	3	US	B1
NS21	4	Ireland	A2/B1
NS22	4	Ireland	B3
NS23	4	Ireland	B2

Appendix 3: NNS Corpus Details

Code	Nationality	L1	Years of English Study	Year of Study	Grade
NNS1	German	German	12	4	A2/B1
NNS2	German	German	12	4	B1
NNS3	Czech	Czech	10	5	B2
NNS4	Polish	Polish	30	4	A1
NNS5	Swedish	Swedish	7	2	A2/B1
NNS6	Swedish	Swedish	10	2	B2
NNS7	Austrian	German	12	4	B2
NNS8	Japanese	Japanese/ French	10	3	B2
NNS9	Austrian	German	12	4	B1

Appendix 4: Epistemic Markers (Hyland and Milton, 1997: 205)

Modal Verbs

could
may
might
should
would
will

Adjectives

apparent
certain
a certain extent
clear
evident
possible
probable

Nouns

claim
doubt
estimate
evidence
possibility

Lexical Verbs

appear
allege
argue
assume
believe
claim
doubt
estimate
expect
indicate
know
predict
presume
propose
seem
speculate
suggest
suppose
tend
think

Adverbs

about
actually
almost
always
apparently
approximately
around
certainly
clearly

definitely
doubtless
essentially
evidently
frequently
generally
in fact
indeed
largely
likely
never
normally
obviously
of course
often
perhaps
possibly
presumably
probably
quite
rarely
relatively
sometimes
surely
undoubtedly
usually