

Qualification and Certainty in L1 and L2 Students' Writing

KEN HYLAND

City University of Hong Kong

JOHN MILTON

Hong Kong University of Science and Technology

A major problem for second language students writing academic essays in English is to convey statements with an appropriate degree of doubt and certainty. Such epistemic comments are crucial to academic writing where authors have to distinguish opinion from fact and evaluate their assertions in acceptable and persuasive ways. Despite its importance however, we know little about how second language writers present assertions in their writing and we often measure their attempts to master appropriate forms against the work of expert writers. Based on a corpus of one million words, this paper compares the expression of doubt and certainty in the examination scripts of 900 Cantonese speaking school leavers writing in English with those of 770 British learners of similar age and educational level. A detailed analysis of the texts reveals that these L2 writers differ significantly from the NSs in relying on a more limited range of items, offering stronger commitments, and exhibiting greater problems in conveying a precise degree of certainty. The authors highlight a number of issues raised by the research and make some pedagogical suggestions for developing competence in this important pragmatic area.

The ability to express doubt and certainty appropriately in English is a complex task for language learners, but one which is critical to successful academic writing. To be effective, writers need to make claims and assertions which academic readers judge to be warranted and which reflect appropriate social interactions. Statements must not only indicate the extent of the writer's conviction in their truth, which may range from uncertain possibility to confident prediction, but also convey a suitable degree of deference and modesty to the audience. These aspects of language use are generally referred to as epistemic: they enable writers to express their assessment of possibilities and indicate the degree of confidence in what they say (Coates, 1987, p. 112). Such epistemic comment is often seen as a principal means by which writers can use language flexibly to adopt positions, express points of view and signal allegiances (Lyons, 1977; Stubbs, 1986).

Direct all correspondence to: Ken Hyland, City University of Hong Kong, English Department, Tat Chee Avenue, Kowloon, Hong Kong <enhyland@cityu.edu.hk>.

While these forms can have a considerable effect on a reader's assessment of both referential and affective aspects of texts, they are generally acknowledged to be difficult for learners to acquire. This is because writer commitment can be expressed in an enormous variety of ways and these expressions can convey a wide range of meanings. Nor are NNS students assisted by the fact that their efforts to master these forms are often measured against an unrealistic standard of "expert writer" models such as academic research articles, a genre which is typically rigorously reviewed and revised before publication. As Hamp-Lyons (1991, p. 57) observes, students are frequently assessed in terms of the requirements of an academic community to which they do not belong and of which they may have little experience. Clearly, not all native speaker writers are equally competent in the manipulation of rhetorical aspects of argument. So, while teachers need a good understanding of L2 writing to help students effectively, we have no real idea how novice writers from different language groups differ in the ways they express doubt and certainty in English.

This study sets out to find how this important area of pragmatic competence is expressed in L2 students' essays by investigating the lexical items used to either "hedge" or "boost" writer commitment in two large corpora of written text. It focuses on how Cantonese speaking Hong Kong students modify assertions in their examination writing and whether this differs from the strategies employed by a group of British students of comparable age and educational level writing in a similar context. The purpose of the study is therefore to determine the typical forms and meanings used by the two groups to present claims in academic English prose.

MODAL EXPRESSIONS

The Complexity of Modal Expressions

The expression of doubt and certainty is central to the conventions of academic writing in English where writers are required to offer an assessment of the referential information they provide by commenting on the provisionality or definiteness of their statements. A large literature has demonstrated the clear pragmatic importance of modality as a discursal resource for negotiating knowledge claims and conveying a stance towards one's propositions and readers (Biber, 1988; Channel, 1994; Nash, 1990; Salager-Meyer, 1994). The construction of appropriate argument structures is a difficult task for NS learners (Connor & Lauer, 1988; McCann, 1989) and L1 speakers clearly have difficulties in qualifying relationships between their grounds and claims. Our experiences as EFL instructors, however, lead us to believe that L2 writers find the manipulation of degrees of probability particularly problematic.

A number of writers have commented on these difficulties. Skelton (1988) and Bloor and Bloor (1991), for example, observe that direct and unqualified writing is more typical of EFL students than native speakers, even of poor adult writers, and this view receives support in studies by Johns (1990) and Silva (1993). Hu, Brown, and Brown (1982) found Chinese L2 writers to be more direct and authoritative in tone and to make more use of strong modals than NSs, while Allison (1995) also expresses concern that ESL writers in Hong Kong frequently make unjustifiably strong assertions. Arab students also appear to have difficulties in this area (Scarcella & Brunak, 1981) and Robberecht and Peteghem (1982) refer to the problems of Dutch and French students in mastering modality. These problems persist for L2 writers at post graduate level where PhD supervisors are often required to counsel the need for appropriate degrees of qualification and confidence in expressing claims (Dudley-Evans, 1992, p. 47).

Modal expressions are complex for novice writers because they are polypragmatic, that is, they can simultaneously convey a range of different meanings. It is difficult to relate particular forms to specific functions on a one-to-one basis because a single linguistic form such as *could*, for example, can express ability and permission as well as possibility (Coates, 1983; Palmer, 1990). Particular items can therefore only be understood as expressions of doubt and certainty by attending to the contexts in which they occur. Meanings do not reside in the items themselves, but are assigned to utterances which contain them. Determining how an item is used therefore requires a pragmatic interpretation of actual instances of use.

The pragmatic use of epistemic devices is further complicated by the fact that they not only convey the writer's confidence in the truth of referential information, but also help contribute to a relationship with the reader. This affective dimension involves the need for cooperation and deference. Writers are expected to mitigate the definiteness of their claims in order to secure their ratification (Holmes, 1984; Hyland, 1996c). Statements of certainty, or categorical assertions, leave little room for feedback and relegate the reader to a passive, accepting role. The judicious use of hedges or other uncertainty markers can therefore help demonstrate a sensitivity to the views of an academic audience and its need to participate in a dialogue with the writer (Myers, 1989; Thompson, 1993). In other words, tentativeness can not only express doubt and reduce the author's liability by avoiding personal accountability for statements, but also help facilitate open discussion.

In addition to these complications, students experience difficulties because epistemic meanings can be signalled in many different ways. While the literature has tended to focus on the significance of modal verbs such as *will*, *may* and *would* as devices for expressing doubt and certainty, at least 350 other lexical devices are also used for this purpose (Holmes, 1988, p.27). These include epistemic verbs (*think*, *know*, *believe*), adjectives (*likely*, *perhaps*, *clear*), adverbs

(*indeed, probably, definitely*) and nouns (*doubt, possibility*). In addition, devices of imprecision such as *about* and *almost* can also modify the epistemic strength of statements (Dubois, 1987) while expressions used to manipulate definiteness, such as *frequently* and *usually*, also contribute to the “scales of probability and usuality to which the term modality strictly belongs” (Halliday, 1985, p. 86). Moreover, while the expression of writer “commitment” is mainly a lexical phenomenon; conditional clauses, questions, contrast markers, and tense can also be used to convey epistemic meanings (e.g., Perkins, 1983).

Students’ difficulties are compounded by the fact that the significance of these devices is largely ignored or misrepresented in writers’ handbooks, style guides and ESL textbooks (Holmes, 1988; Hyland, 1994). Often students are simply not taught sufficient expressions as alternatives to categorical assertion. This absence of adequate pedagogical materials is particularly important because inexperienced writers frequently lack a sufficient knowledge of audience (Connor, 1987) and, once again, these problems are increased for NNSs as different languages use different linguistic means for expressing modality. Bloor and Bloor (1991), for example, consider the degree of conviction permitted in academic writing a clear source of cultural difference. There is also considerable evidence from the IEA study of written composition in 14 countries that cultural variation characterises specific uses and conventions in essay writing tasks at the school leaver level (Purves & Purves, 1986).

A number of studies have identified variations in the certainty with which arguments are expressed in different languages. Academic writing in German and Czech, for example, appears more direct than in English (Bloor and Bloor, 1991, p. 9), while Finns (Mauranen, 1993), Japanese (Harder, 1984), Malays (Ahmad, 1995) and Chinese (Bloch & Chi, 1995) seem to favour a more cautious and indirect style when expressing opinions. These observations do not however predict the marked disparities between L1 and L2 argument paradigms, but they do suggest that students from different cultures may have preconceptions about the formal features of culturally and rhetorically appropriate writing which may differ from those which operate in English academic settings. The findings of cross-cultural rhetoric raise complex issues which cannot be explored here, but there is reliable evidence that the rhetorical preferences of different languages and cultures tend to manifest themselves consistently in the L2 writing of NNS students (Connor, 1995; Sötter, 1988). Such differences can make NNSs vulnerable to the risk of violating communicative norms as their writing may appear as too direct, running the risk of being considered as either brusque or dogmatic, or as too tentative, and therefore seen as equivocal, diffident or naive.

In summary, the need to present claims that are neither overstated nor understated in relation to evidence or reasonable assumption is an area of academic writing which many students, both native and non-native speakers, find seriously challenging.

Corpora and Methodology

The data for this study consist of two large corpora. The first is a collection of essays written by Hong Kong students for the high school matriculation General Certificate of Education (GCE) A level "Use of English" examination in 1994. It consists of about 500,000 words comprising 150 exam scripts in each of six ability bands. The grades in this exam range from A to F (Fail), an E roughly equating to a TOEFL score of around 450 and an A approximating to 600 (Hogan & Chan, 1993, p. 6). The second corpus, also of 500,000 words, was transcribed from 770 GCE A level General Studies scripts written by British school leavers of similar age and education level as the Chinese learners.

Clearly there are many possible educational and societal differences between these groups which prevents a direct comparison of their written work, and this should be borne in mind when considering the results. However, there are also considerable similarities, particularly in the subjects' ages, length of education and experience of a British curriculum structure. Similarly, the writing tasks in the two corpora are not identical, but they can be considered as being broadly comparable. Both consist of the timed examination scripts of school leavers on expository/argumentative topics concerned with contemporary social issues. The general equivalence of genre, text types and topics suggests little difference in the language requirements of the tasks in so far as these might effect the expression of assertion. Thus while we are unable to claim the empirically-valid equivalence of an experimental design, we believe the results reflect actual differences in how the two groups employ this pragmatic feature in authentic contexts of use.

To determine the range and frequency of lexical expressions of doubt and certainty in these corpora, a list of such items common to native speaker usage was prepared. The main source was Holmes' (1983, 1988) analysis of the 'learned' sections of the Brown and LOB (Lancaster/Oslo-Bergen) corpora of written English. This was supplemented by the research literature on modality (Coates, 1983; Hyland, 1996a; Perkins, 1983), reference grammars (Quirk, Greenbaum, Leech & Svartvik, 1972) and by studies of frequency expressions (Kennedy, 1987) which can be used to adjust the strength of claims (Halliday, 1994) and are said to characterise L2 students' assertions (Allison, 1995). From these sources an inventory of 75 of the most frequently occurring epistemic lexical items in native speaker academic writing was produced (see Appendix).

The corpora were then examined to determine the frequency of these words, including misspellings and other morphological variants, in each grade of the Use of English corpus and in the GCE data. Fifty sentences containing each of those items (if there were 50 occurrences) were then randomly extracted from each grade and from the NS sample using a text retrieval program in order to examine items in their sentential context. As we were principally interested in lexical expression, sentence analysis was sufficient to determine the function of terms

and establish the category used. All target items were systematically analysed by both authors working independently to ensure they expressed the writer's certainty or doubt. The interrater reliability for identifying epistemic functions was .89 (Kappa) and every disagreement was subsequently reconciled by discussion. Figures were then extrapolated for the entire sample.

RESULTS

Overall Frequency of Devices

The total number of lexical devices used to express epistemic meanings in the corpora is shown in Table 1. The analysis reveals remarkable similarities in the overall frequencies, with both student samples employing one device every 55 words. These figures are similar to those found in published academic writing (Adams Smith, 1984; Skelton, 1988), but may be only half as frequent as conversational uses (Holmes, 1988).

In addition to agreement in total frequencies, there are considerable similarities of usage, with *will*, *may*, *would* and *always* occurring among the top six most frequently used devices of both NS and NNS writers, although with strikingly different frequencies (Table 2).

Table 2 shows, for example, that epistemic *will* occurs twice as often in the NNS sample while *would* is represented twice as frequently in the NS data. As both forms can be used to refer to present or future probabilities, these distributions suggest conceptual differences, with L2 writers favouring confident prediction and native speakers more tentative expression. *May*, on the other hand, occurs about twice as often in the L2 essays and appears to be the preferred marker of possibility for NNSs. The use of *think* as an epistemic verb is almost exclusively employed to express the writer's certainty in both the NS and NNS data and occurs nearly three times as often in the latter. Available figures for "expert" academic writers (Holmes, 1988) show that *would*, *seem*, *will* and *may* figure most frequently and that no other items overlap with the student lists. *Will* is the only certainty marker in the top ten most frequent epistemic items in published academic writing.

TABLE 1
Total Lexical Devices Used to Express Epistemic Modality

	NNS	NS	Totals
Total devices	9,154	9,115	18,269
Tokens per 100 words	1.83	1.82	1.83

TABLE 2
Most frequent items expressing epistemic modality in rank order

Non Native Speakers			Native Speakers		
Rank	Item	Raw no.	Rank	Item	Raw no.
1	will	2731	1	would	1355
2	may	1215	2	will	1344
3	think	634	3	may	645
4	would	625	4	often	596
5	always	520	5	perhaps	439
6	usually	281	6	always	326
7	know	227	7	believe	297
8	in fact	225	8	seem	232
9	actually	201	9	indeed	206
10	probably	195	10	certainly	189
Totals		6854			5629

Overall, the L2 essays contain a more restricted range of epistemic modifiers, with the ten most frequently used items accounting for 75% of the total. In fact, the top five items constitute almost two thirds of the L2 sample, while ten devices are needed to reach this figure in the L1 scripts. Only 45 of the 75 items appear more than ten times in the NNS data and nine do not occur at all. The greatest differences are between *appear* (33 times more often in NS sample), *apparent(ly)* (10 times more) and *perhaps* and *possible* (each four times more). *About* and *think* each occur over 4 times more often in the NNS scripts. Among NNS students, there is a higher incidence of claim modification by students in the top three ability bands, averaging 2.01 devices per 100 words compared with 1.25 in the lowest band. The A grade essays demonstrate the greatest similarity to the L1 usage.

Grammatical Distributions

It is useful to categorise the items into grammatical classes for comparison. Figure 1 shows broad agreement in the use of lexical verbs, adjectives and nouns to express degrees of certainty and definiteness, but marked differences in adverbs and modal verbs.

Both groups make substantial use of epistemic modal verbs, particularly *will*, *would* and *may* to convey epistemic meanings, although L2 writers appear to depend far more heavily on these devices. The NNS preference for modal verbs relative to the NS sample may be due to L1 transfer or it could attest to the disproportionate attention they have received in L2 pedagogical writing materials (Hyland, 1994). L1 usage, on the other hand, exhibits a greater range and fre-

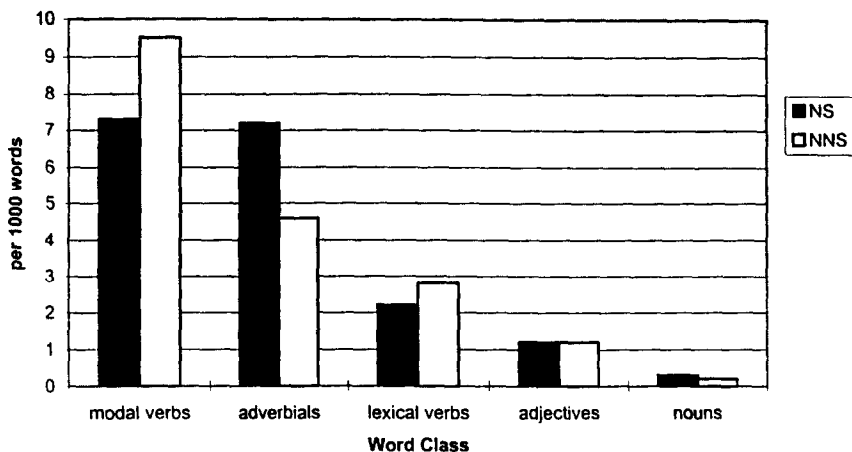


Figure 1. Relative Frequency of Grammatical Classes used to Express Epistemic Modality in the Corpora

quency of adverbials, with over 55% more occurrences. Both student groups favour adverbs over lexical verbs as strategies for expressing doubt and certainty, although the latter occur more often in published academic discourse (Holmes, 1983; Hyland, 1996a).

It is also interesting to compare proficiency variations of the grammatical categories. The top three NNS bands (discussed above) employ substantially more devices overall while the two lowest ability groups together use less than 18% of the total devices. This is considerably less than we would expect if usage was equally distributed across the bands, and the disparity remains even after scaling for the fact that average essay lengths decrease for each drop in grade. Here we projected the frequency of devices in each grade range by the same amount as the proportional drop in the amount of words between the size of each dataset and found the differences between ability groups remained. Low proficiency students also differ in their use of lexical verbs and adverbs. Figure 2 shows that the proportional use of adverbs tends to increase with proficiency while the use of lexical verbs declines, bringing the A group students closest to the NS figures. Over 41% of all epistemic adverbs in the corpus occur in the work of A and B students while adverbs and adjectives are under-represented in the F essays (14%).

The popularity of adverbs over semantically equivalent verbal forms may be due to uncertainty in how to employ lexical verbs appropriately in stating claims. Lexical verbs offer a more overt and precise means of signaling the writer's commitment to a proposition than adverbs, signaling relative degrees of assurance (1) and uncertainty (2):

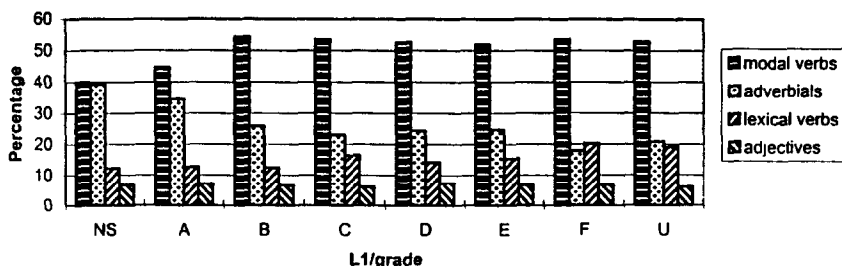


Figure 2. Frequency of Epistemic Items in Main Grammatical Classes by Ability Grade (%)

- (1) Moreover, I *think* reading comic books is a good entertainment because it is healthier than going to karaoke lounge or disco. (C)
- (2) To include Germany again in this argument we may *speculate* what the absence of a monarch may do to the state. (NS)

In addition to modifying the strength of a claim, verbs also indicate whether it carries a judgemental warrant, expressing the degree of conjecture involved (3), or provides evidential justification, indicating the reliability of the source of information (4 & 5) (Hyland, 1996a).

- (3) Personally, I *believe* that comic books are very useful if they are used for educational purposes. (B)
- (4) Psychologists *claim* that it is vital for people to release their stress. (B)
- (5) Young people *seem* to prefer reading comic books than textbooks which are full of words. (A)

The ability to produce effective warrants, which “authorise” claims by relating them to data, however appears to be a difficult task for both L1 and L2 writers (Carlson, 1988; McCann, 1989). Adverbs may be easier for novice writers to manipulate with assurance. Not only are they far more common in speech than writing (Holmes, 1983), but they are syntactically more mobile in clause structure and offer clear scalable distinctions (e.g., Quirk et al., 1972, p. 438ff). This allows writers to boost or hedge their commitment with some confidence that they are approximating to an academic discourse framework without running the high risk of error inherent in appropriate verb selection.

Equally importantly, the fact that novice writers’ seem to prefer epistemic adverbs to lexical verbs is related to the complex issue of appropriate academic tenor. Tenor relationships in expository genres are said to require the maximum distance between writer and reader of any genre and a high proportion of passive

constructions is often cited as a feature of good academic writing (Smith, 1986). Students often learn that academic writing requires an impersonal approach, but when writing this is complicated by the need to evaluate information and assess its validity. The use of an appropriate epistemic verb therefore involves critical lexical, tense and voice choices which can have significant rhetorical effects. Not only is the choice of item significant in conveying the writer's confidence in the truth of an accompanying proposition, but tense choices can also indicate a stance by manipulating proximity and distance (Swales, 1990), and voice can "modulate the degree of certainty on the author's part" (Lachowicz, 1981, p. 112). Adverbials offer a far simpler means for writers to express an attitude to their statements and adjust the strength of their claims without such grammatical and lexical complications.

On the evidence of our text samples, students respond to these difficulties by mixing informal spoken and formal written forms and transfer conversational uses to academic genres. Neither corpus is characterised by a rich use of lexical verbs: *think* and *know* in the NNS texts and *believe*, *seem* and *think* in the NS sample account for almost two thirds of all forms. The limited use of epistemic verbs and a preference for predominantly speech forms indicates the novice writers' uncertainty in how to use these items effectively in argumentation and their imperfect grasp of appropriate academic register. Apparently possessing only a rudimentary understanding of formal academic expectations, neither student group is able to employ "expert" forms in making claims. To avoid violating academic expectations, NNSs appear to seek a solution by employing more modal verbs and NSs by overusing epistemic adverbs.

Categories of Epistemic Commitment

Epistemic devices clearly express different degrees of certainty concerning the validity of the information asserted; *obviously*, for example, conveys greater certainty than *may*. To compare the use of modality in the two corpora we have therefore established discrete epistemic categories, locating particular patterns and lexical items on a scale extending from maximum to minimum certainty. While inevitably somewhat arbitrary in some cases, such a categorisation is compatible with those made by other researchers and widely employed in the literature. Similar categories have been adopted in reference grammars (Halliday, 1994; Leech & Svartvik, 1994; Quirk et al., 1972) and EFL coursebooks (Arnau-det & Barrett, 1984; Jordan, 1990; Weissberg & Buker, 1990) and have proved valuable in the empirical analysis of various written and spoken corpora (Holmes, 1983, 1984).

Analysis based on these categories reveals substantial differences in the degree of certainty and tentativeness employed by the two student groups (Figure 3). The

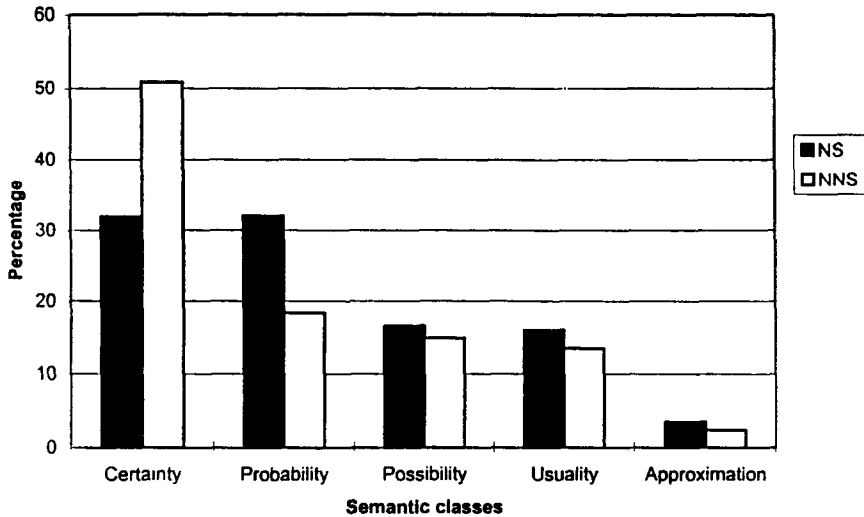


Figure 3. Distribution of Items Among Categories (%)

results seem to confirm the view that the academic writing of many L2 learners is characterised by firmer assertions, more authoritative tone and stronger writer commitments when compared with native speaker discourse. Overall, these NNS students employ about 60% more certainty markers than do their L1 counterparts in a similar size corpus while the L1 essays contain 73% more items expressing probability. Devices conveying possibility, usuality and approximation have similar distributions although, once again, almost half the 'usuality' (or "frequency") items in the NNS essays serve to express certainty (*always/never*) compared with only 20% of the devices used by NSs.

The fact that these Chinese students do not moderate their claims sufficiently could be due to inadequate linguistic knowledge, perhaps caused by different interpretations of equivalent semantic forms (Allison, 1995). Alternatively it could represent "sociopragmatic" violations resulting from an imperfect awareness of appropriate language use (Thomas, 1983). That is, while students may typically be more indirect when writing in Chinese, they overcompensate for what they perceive to be communicative conventions of explicitness and directness in English and fail to distinguish between the conventions of different text types. Thus students may select a strategy which they believe conveys a socially appropriate illocution but which actually carries an inappropriate degree of directness, deference or assertiveness for an academic register.

About half the epistemic devices in the NNS data serve to express the writers' certainty:

- (6) The trend from overseas *always* affecting Hong Kong people. (E)
- *Actually*, there has been tremendous concern on this issue. (A)
 - It is *certain* that Hong Kong will continue to develop prosperously. (B)
 - This will *definitely* improve your English. (D)
 - Buying expensive brand-name products is, *in fact*, a sumptuous activity. (E)

On the other hand, most items in the native speaker essays are markers of qualification, either expressing the view that the proposition is probably true (7) or conveying doubt in the validity of the accompanying proposition (8).

- (7) It brings the issue of racism to the forefront of the social agenda is a move which cannot be *essentially* condemned. (NS)
- In such cases, the press *appear* to have forced unnecessary actions. (NS)
 - This is *likely* to cause resentment in the poorer communities. (B)
- (8) It *may* be argued however that these people have chosen to be in their positions and are seen to be an example to others. (NS)
- More broadcasting equals worse broadcasting is *perhaps* too simplistic. (NS)
 - It is *possible* to argue, therefore, for a comprehensive Bill of Rights, with an inclusion on privacy, tied to a US-style Freedom of Information Act. (NS)

It should be pointed out, however, that while we have followed other researchers in establishing discrete epistemic categories, our results must be treated with some caution. The categories provide a useful basis for comparing the range and frequencies of items used by different groups, but there are serious difficulties in unequivocally mapping a precise degree of conviction onto particular terms in every instance. Modal expressions are rarely wholly determinate in meaning as the semantic value of any term is always subject to pragmatic constraints which can alter its meaning. The form *quite*, for example, can either hedge or boost the force of a word depending on its context (Hyland, 1996d). Every instance of the 75 forms selected for study was thus carefully scrutinised in context to ensure that it was being used to express epistemic meaning. This meant, for example, that homonymous examples of words such as *appear*, *possible* and *clear* were not counted and that many instances of modal verbs were excluded where they conveyed deontic (volitional) meanings. In addition, epistemic verbs such as *know*, *believe* and *think* were recorded as epistemic only if they followed a personal subject or were used to cite another author.

But context is not a disambiguating factor in all cases and linguists disagree on crucial matters. Epistemic *will*, for example, is said to express confident prediction about present, timeless or future events based on previous experience (Coates, 1983, p. 177). However, futurity always involves some uncertainty (Palmer, 1990,

p. 57) and Lyons (1977, p. 809) argues that introducing any modality into an utterance makes its factuality dependent upon the writer's, perhaps limited, knowledge. Our view is that on an epistemic scale extending from maximum to minimum certainty, *will* appears to convey an assessment that the accompanying proposition is valid *as far as the writer can be sure*. We have therefore followed Leech and Svartvik (1994), Quirk et al. (1972) and others in locating it in the certainty category, recognising that the general effect is to 'boost' the expression of writer conviction.

So, while our epistemic categories seek to capture the general semantic meaning of items from maximum to minimum certainty, we admit that categorisation is uncertain in some cases and precise quantification hazardous. These considerations should be born in mind when considering the data below.

Student Expression of Doubt and Certainty

Analysis of the epistemic categories again indicates an uneven distribution of items between ability bands, with higher grades approximating more closely to native speaker usage. Weaker students employ a significantly higher proportion of certainty markers while probability and possibility devices occur more often in the work of A and B learners. In fact, 40% of probability markers and 45% of possibility devices in the NNS corpus were found in the A and B essays. Qualification of claims using approximation devices such as *almost* and *a certain extent* are relatively constant within the bands.

This tendency of higher ability students to modify their statements with more tentative expressions is shown clearly in Figure 4, which combines all the items into two broad classes. Here we have categorised forms according to whether they act to strengthen or weaken the force of the claim being made, i.e., the extent to which they involve the expression of the writer's certainty or commitment in either "boosting" or "hedging" the illocutionary force of a statement (Holmes, 1984). Thus the high value modal forms *always* and *never* are included in the cer-

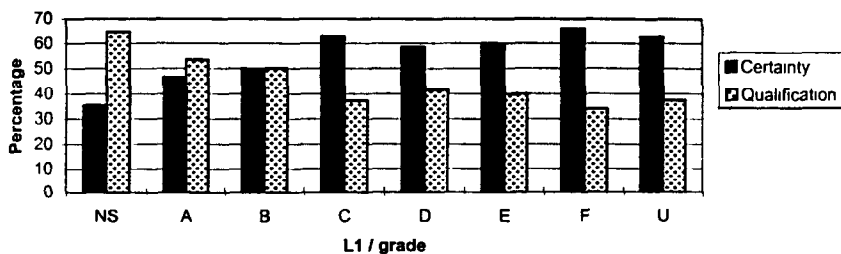


Figure 4. Distribution of Items Expressing Certainty and Qualification by Grade (%)

tainty group and all other, less determinate forms, in the qualification (hedging) category. It can be seen that while weaker students employ fewer devices overall, their writing is characterised by epistemically stronger statements. The work of A and B students on the other hand shows a more balanced distribution between qualification and certainty. Finally, the native speakers employ a higher proportion of tentativeness than any NNS group, with about two thirds of the modifiers serving to withhold full commitment to claims.

Despite differences in frequencies, there is considerable similarity in the items used by NS and NNS students. Table 3 shows the most frequently occurring items

TABLE 3
Most Frequent Items in Main Epistemic Categories

Item	Total occurrences		% of Category Forms by Grade		
	NS	NNS	A & B	C, D, E	F & U
A. Certainty					
actually	140	201	6.1	4.3	2.0
certainly	189	110	4.6	1.6	1.2
indeed	206	93	3.4	1.4	0.9
in fact	105	225	2.8	5.9	5.1
know	12	227	4.4	4.5	5.5
think	173	634	9.9	14.3	17.4
will	1344	2731	53.2	59.7	63.6
% of Certainty items	74.6	90.8	85.4	91.7	95.7
B. Probability					
believe	297	149	7.6	9.0	11.4
probably	164	195	11.5	11.3	12.9
quite	108	182	12.6	10.4	7.8
seem	232	76	5.1	3.8	5.1
would	1355	625	35.7	37.9	37.7
% of Probability items	73.8	72.9	72.5	72.4	74.9
C. Possibility					
may	645	1215	86.0	93.1	85.8
might	182	67	7.6	2.1	4.2
perhaps	439	37	2.5	2.3	5.0
possible (ly)	170	26	2.1	1.3	3.1
% of Possibility items	95.0	98.5	98.2	98.8	98.1
D. Usuality					
always	239	513	34.6	45.7	47.3
often	596	173	19.7	9.9	10.5
usually	146	281	21.7	23.5	22.8
% of usuality items	67.4	78.0	76.0	79.1	80.6

in the different modal groups and the proportions they constitute of each category. As can be seen, the same items comprise the overwhelming majority of terms used in each category, although NSs appear to employ a wider vocabulary of certainty and usuality terms. *Will* is the favoured marker of certainty for both groups but, together with *know*, *think* and *in fact*, is more frequently used by L2 writers, mainly by those in average and lower ability bands. The same forms also comprise similar proportions of probability and possibility terms.

There are a number of major proficiency differences, particularly in the use of certainty and usuality terms. Several predominantly native speaker forms such as *certainly*, *indeed*, *might* and *often* are far more common in the A and B essays, with over half of all occurrences in these bands. The weaker students' essays on the other hand are dominated by *will*, *think* and *may* which constitute 56% of all epistemic uses in E and F groups. *May* is the main marker of possibility for all learners but again is numerically more prevalent in the work of A and B students. It constitutes 17% of all epistemic devices used by higher proficiency students and 44% of all cases occur in their essays. In the usuality scale, the definite expressions *always* and *never* constitute over 50% of all forms in the E and F essays. More indeterminate markers like *often* and *usually* are significantly over-represented in the work of higher proficiency writers and items such as *generally* and *frequently* are almost exclusively to be found there.

Personalised vs Impersonalised Forms

One interesting aspect of how writers modify assertions is their use of personal and impersonal expressions (Leech & Svartvik, 1994). Personalised forms explicitly involve the writer in an assessment of propositional validity by use of a first person pronoun:

- (9) *I deeply believe* that this is not uncommon in our surrounding. (A)
 (10) *It seems to me*, to be safest to stick to the wind-farms, HEP's and solar panels, because fusion is a possibility not a certainty in the future. (NS)

Impersonalised forms on the other hand, avoid reference to the writer when commenting on the truth of a claim and typically conceal the source of epistemic judgments by use of sentence adverbs, impersonal pronouns or passive voice:

- (11) *It is certain* that reading too many comic books and neglecting the newspapers, our language ability will be inevitable slackened and undermined. (B)
 (12) *Apparently*, this trend is invading Hong Kong. (D)

As discussed above, person and voice choices with modal devices help establish an appropriate academic tenor with a high degree of formality between the writer and reader. As we might expect, the use of these features differ markedly between the two student corpora. The NNSs appear to transfer features of a more personal register to the expository genre and exhibit less consistency in their choices, often coding epistemic comment incongruently in the academic context. NNSs are far more likely to employ a first person pronoun with an epistemic verb than native speakers and this likelihood increases as proficiency declines. Personal pronouns occur frequently with epistemic verbs in published academic writing (Banks, 1994) where they typically function interpersonally to strengthen the force of commitment to an argument or to weaken a claim by hedging its generalisability. For many learners however, attempts to boost conviction in this way are fraught with hazards:

- (13) Supported by the above-mentioned arguments, *In my opinion I do have the confidence to believe* that wearing brand-name products is harmful to our youths. (B)
- (14) *As I know, I am quite sure* some parents are willing to pay whenever their children ask for. (C)

So while the presence of personalised forms does not in itself indicate inappropriate tone, their frequency, incongruity and relatively informality in the L2 essays suggests a comparative lack of control of this genre.

NNSs also appear to have greater problems in manipulating impersonal forms appropriately in academic writing. In particular, the injunction to suppress human agency creates a range of errors such as misuse of empty subjects and faulty tense choices (15-16), while initial adverbs to express certainty or doubt frequently confuses spoken with written registers (17-18):

- (15) *It is no doubt* that the students as well as the adults have this characteristic. (E)
- (16) *It is clearly showing* that these buyers usually lack confidence. (C)
- (17) *Obviously*, the fashion design is a clear and a far-reaching consequence. (D)
- (18) *Probably*, they can learn the importance of confidence when encountering barriers. (D)

A further means of disguising the epistemic source is to attribute a view to another:

- (19) *Many critics argue* that reading such books is disadvantages. (D)
- (20) *Many students claim* that the school text books are rather boring and hard to understand. (A)

Such 'depersonalised forms' (Holmes, 1988) are common in published writing (Thompson & Ye, 1991) and help protect the writer by shifting responsibility for the validity of a statement. Conveying a viewpoint on the status of information in this way, however, seems problematic for many L2 writers who rely heavily on "say" (Granger, 1994). L1 writers in the present study used depersonalised forms substantially more often than the Hong Kong learners and employed a greater range and frequency of epistemic verbs to comment on information. Clearly then, while such uses display an attempt to conform to the canons of objectivity and impersonality often exhorted by teachers and textbooks, the additional need to evaluate one's claims in academic genres can create serious problems of expression for L2 writers.

Epistemic Clusters

A final dimension of epistemic usage is the tendency of expert writers to use devices together (Banks, 1994). Lyons (1977, p. 807) employs the term "modally harmonic" to refer to contexts where two or more forms express the same degree of modality such that "there is a kind of concord running through the clause, which results in the double realisation of a single modality" (ibid, p. 808). The NS data suggests that about 25% of the modalised sentences contain at least two epistemic markers, principally functioning to weaken the strength of the claim being made:

- (21) By making such laws it *might* be *possible* to remove other prejudice and discrimination. (NS)
- (22) On balance it *would seem* that the only real solution to the problem *would* be to allow the papers to introduce an efficient self governing body. (NS)

A similar count of a representative sample of NNS's essays on the other hand revealed far fewer epistemic clusters and found that where they did occur their principal role was often to emphasise the strength of the accompanying proposition rather than weaken it:

- (23) *In my view*, I *think* that it is a good trend for students to follow only in the following sense. (B)
- (24) *Undeniably*, the influx of famous brand name is *certainly* an image of prosperity. (E)

Moreover, many L2 students have difficulty in combining epistemic forms correctly. Frequently learners appear to misjudge the combined value of two devices and invest their claims with inappropriate conviction:

- (25) It is *far beyond doubt* that Hong Kong is *certainly* an information- filled metropolis. (D)
- (26) I can tell with *confidence* it is *definitely* true that the most popular form of reading material for young people today are comic books. (B)

In addition, there are many examples of modally non-harmonic clusters in the L2 essays, where forms are collocated in ways which fail to achieve a congruent degree of certainty:

- (27) *Evidently*, they should bear a part of the responsibility *in fact*. (B)
- (28) *Indeed*, there are the advantages of reading the books, but *certainly*, of course, *to a certain extent*. (E)
- (29) If someone who cannot follow the atmosphere, he *might probably* be the laughing stock of others. (F)

The fact that many of these L2 learners combine excessively definite and certain forms and are unable to achieve an appropriate degree of assurance and probability in their writing demonstrates once again the difficulties that epistemic uses present for NNSs.

Some Pedagogic Implications

Clearly both samples display a reliance on a limited number of forms, but the inability to use epistemic devices appropriately is more apparent in the NNSs writing. The greater competence displayed by the more proficient students with this aspect of meaning however suggests they have acquired a better understanding of relevant discourse conventions, perhaps through more effective teaching (Milton, forthcoming) or by being more alert to the language patterns in their reading. A cross-sectional study such as this can say little about developmental issues, but it is uncertain whether learners are likely to acquire this aspect of pragmatic competence without first consciously "noticing" it (e.g., Schmidt, 1993). Explicit instruction may therefore help accelerate its acquisition.

Intervention should be designed to help learners develop strategies to distinguish observed facts from interpretation and to use a range of high frequency items to effectively express different epistemic meanings. The categories suggested above provide a useful starting point for raising awareness of different degrees of probability and can provide the basis for exercises whereby students identify particular items as conveying certainty, probability, possibility and approximation.

An active mastery of several forms and patterns for each epistemic category is essential, both for variety of expression and to convey the full range of meanings. Ideally, the lexical items selected for teaching will therefore include those which

can appear in various patterns (active, passive, dummy 'it' subject, etc.), occur as more than one part of speech (e.g., *possible, likely, certain*) and combine in useful collocations (*seems possible, might suggest, appears unlikely*). This study suggests some of the most frequently occurring, and most productive, items in each of the categories in relevant contexts of native speaker use. Students can thus be asked to discuss the epistemic effects of removing items from a text or of replacing them with items from other categories. Reflection on the significance of epistemic meanings can also be encouraged by allowing learners to judge the effects of reformulations which vary the level of certainty of a text.

Students also need opportunities to practice the use of different forms. Rewriting exercises which involve replacing certainty forms with hedges, for example, have proved to be useful in assisting L2 learners master the difficult notion of 'tentativeness' in academic writing (Salager-Meyer, 1994). Other productive tasks include learners paraphrasing texts of varying certainty, completing sentence frames and undertaking free writing activities which explore personal or political futures (Hyland, 1996b). Clearly there are a range of pedagogical techniques to assist students develop a repertoire of epistemic devices which can be confidently employed in academic essays. It is important however that teaching approaches must not only stress the importance and salience of epistemic items. Such methods should also provide students with alternative strategies to the complex syntactic patterns and over-reliance on modal verbs which often dominate textbooks and make modality so difficult to master.

CONCLUSIONS

Overall the data show that while both student groups are heavily dependent on a narrow range of items, principally modal verbs and adverbs, the manipulation of certainty and affect in academic writing is particularly problematic for the L2 students. The Hong Kong learners employed syntactically simpler constructions, relied on a more limited range of devices, offered stronger commitments to statements and exhibited greater problems in conveying a precise degree of certainty. We believe this lack of familiarity with a convention central to many expository genres in English may be detrimental to learners' academic and professional opportunities. This is because such errors often not only influence readers' judgments of coherence and comprehensibility, but can also effect the impact of the argument, and how the academic competence of the writer is evaluated.

Clearly this study refers to only one L2 language group, but the literature suggests it is unlikely that Hong Kong students differ greatly from other learners in the difficulties they experience in expressing doubt and certainty in English. English language teaching and research agendas however have largely over-

looked the importance of epistemic language, and the kinds of difficulties revealed in this study are partly attributable to this neglect. With greater numbers of NNSs attending courses at English medium universities it is critical that these features of academic discourse are made more conspicuous to students. Only then will they gain control of this area of rhetorical competence which is so critical to effective communication in such environments.

REFERENCES

- Adams Smith, D. (1984). Medical discourse: aspects of author's comment. *English for Specific Purposes*, 3, 25-36.
- Ahmad, U. K. (1995). *Academic language and culture: Some observations on scientific Malay and scientific English*. Paper presented at RELC Conference, Singapore.
- Allison, D. (1995). Assertions and alternatives: Helping ESL undergraduates extend their choices in academic writing. *Journal of Second Language writing*, 4, 1-15.
- Arnaudet, M., & Barrett, M. (1984). *Approaches to academic reading and writing*. Englewood Cliffs, NJ: Prentice Hall.
- Banks, D. (1994). *Hedges and how to trim them*. Paper presented at the 9th European symposium of LSP, University of Bergen.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bloch, J., & Chi L. (1995). A comparison of the use of citations in Chinese and English academic discourse. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 231-274). Norwood NJ: Ablex.
- Bloor, M. & Bloor, T. (1991). Cultural expectations and socio-pragmatic failure in academic writing. In P. Adams, B. Heaton & P. Howarth (Eds.), *Socio-cultural issues in English for academic purposes* (pp. 1-12). Basingstoke: Modern English Publications/British Council.
- Carlson, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 227-260). Newbury Park, CA: Sage.
- Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.
- Coates, J. (1983). *The semantics of the modal auxiliaries*. Beckenham: Croom Helm.
- Coates, J. (1987). Epistemic modality and spoken discourse. *Transactions of the Philological Society*, 35, 101-131.
- Connor, U. (1987). Argumentative patterns in student essays: Cross cultural differences. In U. Connor & R. Kaplan (Eds.) *Writing across languages: Analysis of L2 text* (pp. 57-71). Reading, MA: Addison Wesley.
- Connor, U. (1995). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press.

- Connor, U., & Lauer, J. (1988). Cross-cultural variation in persuasive student writing. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 138-159). Newbury Park, CA: Sage.
- Dubois, B. (1987). "Something in the order of around forty to forty four:" Imprecise numerical expressions in biomedical slide talks. *Language and Society*, 16, 527-541.
- Dudley-Evans, T. (1991). Socialisation into the academic community: Linguistic and stylistic expectations of a PhD thesis as revealed by supervisor comments. In P. Adams, B. Heaton & P. Howarth (Eds.), *Socio-cultural issues in English for Academic Purposes* (pp. 1-12).
- Granger, S. (1994). New insights into the learner lexicon: A preliminary report from the international corpus of English. In L. Flowerdew & K.K. Tong (Eds.), *Entering text*. (pp 102-113). Hong Kong: Hong Kong University of Science and Technology.
- Halliday, M. (1994). *An introduction to functional grammar*, 2nd ed. London: Edward Arnold.
- Hamp-Lyons, L. (1991). The writer's knowledge and our knowledge of the writer. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 51-68). Norwood, NJ: Ablex.
- Harder, B. (1984). Cultural attitudes in discourse analysis. *Canadian Journal of Linguistics*, 29, 115-130.
- Holmes, J. (1983). Speaking English with the appropriate degree of conviction. In C. Brumfit (Ed.), *Learning and teaching languages for communication: Applied linguistics perspectives* (pp. 100-121). London: British Association of Applied Linguistics.
- Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics*, 8, 345-365.
- Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 20-44.
- Hogan, D., & Chan, W. (1993). *Public Examinations in Hong Kong*, 3rd ed. Hong Kong: Hong Kong Examinations Authority.
- Hu, Z., Brown, D., & Brown, L. (1982). Some linguistic differences in the written English of Chinese and Australian students. *Language Learning and Communication*, 1, 39-49.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13, 239-256.
- Hyland, K. (1996a). Talking to the academy: Forms of hedging in science research articles *Written Communication*, 13, 251-281.
- Hyland, K. (1996b). Nurturing hedges in the ESP curriculum. *System*, 24(4), 477-490.
- Hyland, K. (1996c). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17, 433-454.
- Hyland, K. (1996d). "I don't quite follow": Making sense of a modifier. *Language Awareness*, 5(2): 91-109.
- Johns, A. (1990). Coherence as a cultural phenomenon: Employing ethnographic principles in the academic milieu. In U. O'Connor & A.M. Johns (Eds.) *Coherence in writing: Research and pedagogical perspectives* (pp. 209-226). Alexandria, VA: TESOL.
- Jordan, R. (1990). *Academic writing course*. London: Collins.
- Kennedy, G. (1987). Expressing temporal frequency in academic English. *TESOL Quarterly*, 21, 69-86.

- Lachowicz, D. (1981). On the use of the passive voice for objectivity, author responsibility and hedging in EST. *Science of Science*, 2, 105-115.
- Leech, G., & Svartvik, J. (1994). *A communicative grammar of English*, 2nd ed. London: Longman.
- Lyons, J. (1977). *Semantics*, Vol. 1 & 2. Cambridge: Cambridge University Press.
- Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English Economics texts. *English for Specific Purposes*, 12, 3-22.
- McCann, T. (1989). Student argumentative writing knowledge and ability at three grade levels. *Research in the Teaching of English*, 23, 62-76.
- Milton, J. (In preparation). The influence of instruction on the lexical and syntactic production of EFL learners.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, 10, 1-35.
- Nash, W. (1990). Introduction: The stuff these people write. In W. Nash (Ed.), *The writing scholar: Studies in academic discourse* (pp. 8-30). Newbury Park, CA.: Sage.
- Palmer, F. (1990). *Modality and the English modals (2nd Ed)*. London: Longman.
- Perkins, M. (1983). *Modal expressions in English*. London: Frances Pinter.
- Purves, A., & Purves, W. (1986). Viewpoints: cultures, text models and the activity of writing. *Research in the Teaching of English*, 20, 174-197.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English*. Harlow, Essex: Longman.
- Robberecht, P., & van Petegham, M. (1982). *A functional model for the description of modality*. Paper presented at the Fifth International Conference on Contrastive Projects, Jyväskylä, Finland.
- Salager-Meyer, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13, 149-170.
- Scarcella, R., & Brunak, R. (1981). On speaking politely in a second language. *International Journal of the Sociology of Language*, 27, 59-75.
- Schmidt, R.W. (1993). Consciousness, learning and interlanguage pragmatics. In G. Kasper & S. Blum-Kulka (Eds.) *Interlanguage pragmatics* (pp. 21-42). Oxford: Oxford University Press.
- Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, 27 657-675.
- Skelton, J. (1988). Comments in academic articles. In P. Grunwell (Ed.), *Applied linguistics in society* (pp. 98-108). London: CILT/British Association of Applied Linguistics.
- Smith, E. L. (1986). Achieving impact through the interpersonal component. In B. Cou-
ture (Ed.), *Functional approaches to writing: Research perspectives* (pp. 108-119). Norwood, NJ: Ablex.
- Sötter, A. (1988). The second language learner and cultural transfer in narration. In A. Purves (Ed.) *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 177-205). Newbury Park, CA: Sage.
- Stubbs, M. (1986). 'A matter of prolonged fieldwork:' Notes towards a modal grammar of English. *Applied Linguistics*, 7, 1-25.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

- Thomas, J. (1983). Cross-Cultural pragmatic failure. *Applied Linguistics*, 4, 91-112.
 Thompson, D. (1993). Arguing for experimental “facts” in science: A study of research article results sections in biochemistry. *Written Communication*, 8, 106-128.
 Thompson, G., & Ye, Y. (1991). Evaluation of the reporting verbs used in academic papers. *Applied Linguistics*, 12, 365-382.

APPENDIX

Most Frequent Epistemic Items in Academic Writing

Compiled from Holmes’ (1983 & 1988) and Kennedy’s (1987) analyses of the J sections of the Brown and LOB corpora; Hyland’s (1996a) biology research article corpus; plus grammars and the modality literature including Coates (1983, 1987); Leech & Svartvik (1994); Lyons (1977); Perkins (1983); Skelton (1988); Quirk et al. (1972).

Modal verbs

- could
- couldn’t
- may
- might
- should
- shouldn’t
- would
- wouldn’t
- will
- won’t

Adjectives

- always
- (not) always
- apparent
- certain
- a certain extent
- clear
- evident
- possible
- probable

Nouns

- claim
- doubt
- estimate
- evidence
- possibility

Lexical verbs

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

Adverbs

- about
- actually
- almost
- 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