

A COMPARATIVE STUDY OF LEARNER CORPORA OF SPOKEN AND WRITTEN DISCURSIVE LANGUAGE: FOCUSING ON THE USE OF EPISTEMIC FORMS BY JAPANESE EFL LEARNERS

Kenneth FORDYCE

Institute for Foreign Language Research and Education

Hiroshima University

INTRODUCTION

This study compares Japanese EFL learners' use of epistemic forms in spoken and written discursive language. The learners were asked to express their opinions both in writing and in speech, and this data was collected to create two small focused learner corpora, which could be compared both quantitatively and qualitatively. The main aims of this study were to find out: (1) the extent to which Japanese EFL learners use epistemic forms similarly or differently in discursive language in the written and spoken modalities; and (2) to look at individual patterns of use of these forms in order to uncover second language developmental sequences for epistemic forms.

Recent advances in corpus linguistics, including in particular the general availability of corpus tools, such as Wordsmith Tools (Scott, 1996), have enabled language researchers and language teachers to investigate learners' language use in a more empirically-valid manner. Computer learner corpus (CLC) research is a growing field of inquiry, which, as Granger (1998: xxi) states, has 'roots both in corpus linguistics and second language acquisition (SLA) studies, it uses the methods and tools of corpus linguistics to gain better insights into authentic learner language'.

The majority of studies in CLC thus far have focused on learners' written language. The obvious reason for this is that written data is much easier to collect and save on computer. This does not mean, however, that learners' spoken language cannot be examined using CLC methodology. It simply requires more work in transcribing learner speech. Some spoken corpora have been compiled. For example, there is the LINDSEI corpus¹⁾ (see Granger, 2002: 11), and some studies have also been carried out based on learners' spoken language (eg. De Cock et al. 1998; Housen, 2002).

Most research in CLC involves Contrastive Interlanguage Analysis (CIA). This can involve two types of comparison (Granger, 2002: 12-13):

- (1) non-native speaker corpus with a native speaker corpus
- (2) two different non-native speaker corpora

The second type of comparison has typically involved comparing data from two different L1 groups. However, two sets of data from the same group of learners could also be analysed.

This could, for example, involve a comparison of their spoken and written language, or language used in two different genres. One of the contributions of this paper is to compare discursive language by the same group of learners in the two different modalities of speech and writing.

Research into differences between spoken and written language has been carried out for native speaker language. Biber's (1988) work in this area used factor analysis of corpus data to identify dimensions of language use across different genres. Biber et al. (1999) investigated the British National Corpus to produce the first major grammar of English which described in detail the grammar of spoken language alongside a description of written language. Further to this, Leech et al. (2001) produced word frequency lists detailing the frequencies of words in spoken and written English.

All these studies have identified significant differences in the use of grammar and lexis between speech and writing. Acquiring these differences is also a challenge for language learners. Few studies, however, appear to have looked closely at this aspect of second language acquisition.

This study takes epistemic language as its focus. Palmer (2001) defines epistemic modality as being 'concerned with the speaker's attitude to the truth-value or factual status of the proposition' (p.24). Biber et al. (1999) contains a whole chapter on the grammatical marking of stance, which is defined as the expression of 'personal feelings, value judgments, or assessments' (p.966) by speakers and writers. These two terms, epistemic modality and stance, would appear to refer broadly to the same area of language, the mastery of which is crucial in order for speakers and writers to convey their point-of-view accurately and effectively. Epistemic stance can be expressed in English through a variety of grammatical forms:

- modal verbs: *may, might, will, etc.*
- lexical verbs: *think, guess, seem etc.*
- modal adverbs: *maybe, perhaps, probably etc.*
- modal adjectives: *possible, definite, certain etc.*
- modal nouns: *fact, opinion etc.*

It appears that this is a difficult aspect of language to master. Holmes (1988, p.21) writes that 'there is widespread agreement among both theoretical and applied linguists that modality is a complex and very important aspect of English which is not easy for first or second language learners to acquire.'

Some corpus-based studies have compared the use of stance in writing by learners as compared to its use by comparable groups of native speakers. Hyland & Milton (1997) and McEnery & Amselom Kifle (2002) compared the use of epistemic forms in discursive writing by EFL learners with that of British learners of a similar age and level of education. Both studies found that the learners used a smaller range of devices to express stance, although they differed with regard to the types of forms typically used. The Cantonese learners in Hyland & Milton's study tended to use stronger epistemic forms (i.e., those expressing more certainty) than the native speakers, whilst the Eritrean learners in McEnery & Amselom Kifle's study used more tentative epistemic forms.

In a further study along similar lines, Aijmer (2002) looked at more advanced learners of English. She focused predominantly on the argumentative writing of advanced Swedish learners, using the Swedish learner corpus from ICLE (see above) and LOCNESS (Louvain Corpus of Native English Essays, see Granger (1998)) as a reference corpus of native speaker argumentative essays. She also broadened her study by comparing the use of modality in Swedish learners' written interlanguage with that of German and French learners. She found that all three learner groups tended to overuse modal verbs compared to native speakers. Another of her findings was that the Swedish learners used lexical verbs conveying stance (e.g., *I think; I believe*) much more than the native speakers. This led her to suggest that learners tend to adopt 'a more speech-like style in their writing than the native writers'.

All of these studies have only looked at learners' use of epistemic forms in written language. This study endeavours to make a contribution in this area by comparing learners' use of epistemic forms in the two modalities of communication. By collecting both written and spoken production data in the same genre (discursive language) from a group of learners in order to create two corpora, it would be possible to find out whether learners used similar or different epistemic forms in speech and writing, and also to investigate possible acquisitional patterns.

Research Questions

The research questions for this study are as follows:

- 1) To what extent do Japanese EFL learners use epistemic forms similarly or differently in discursive language in the written and spoken modalities?
- 2) What individual patterns in the use of epistemic forms can be discerned which may offer insights into developmental sequences in second language acquisition?

METHODOLOGY

Subjects

A total of 41 subjects took part in this study. All of them were students in the Faculty of Education at Hiroshima University. They were all taking a second year vocabulary class and could broadly be defined as covering a range of English competence around the low-intermediate to intermediate level.

Data Collection

Spoken and written data was collected from each subject as follows:

(1) Written data:

This involved the learners writing a discursive essay on a topic which was considered relevant to their context as university students (see Appendix). For each task, learners were asked to discuss advantages and disadvantages, and express their own opinion. Three different tasks were created. The reason for this was that due to space limitations in the computer laboratory, not all subjects could be accommodated at the same time. By having

different tasks it was possible to decrease the likelihood that the learners could prepare in advance. All essays were written under controlled conditions. The use of dictionaries or peer-consultation was not permitted. Learners were asked to write at least 300 words, and given approximately 90 minutes to do so. Most learners were able to write more than the essay length requirement. The average length of essays was 322.6 words (range 223~484). These essays were saved as both Word documents and text files. I used the Word documents to correct the essays and provide feedback on them to the students (this task counted as a requirement for the class). The text files were used for Wordsmith Tools analysis (see below under 'Data Analysis').

(2) Spoken data:

This involved the students attending an interview of around ten minutes. For the first few minutes I engaged them in general conversation in order to try to relax them. This part of the interview was not recorded. Following that, I presented them with a topic card on which a statement was written. I gave them one minute to consider their ideas by themselves. After that, I asked them to state their opinion on the topic. I began recording them at this point. As far as possible I let them speak without intrusion. However, where necessary I would prompt them or ask further questions to try and elicit more opinions from them. I aimed to make this part of the interview last for between three and four minutes for each subject. In fact, the mean length was 3:31 minutes (range 2:04~4:47). Just as for the written data collection, three different topics were used to try to ensure that students would not prepare in advance (see Appendix). The topics were all different from those used for the written data collection.

Subsequent to the interviews, the recorded data was transcribed. Only the subjects' utterances were transcribed. Hesitators, such as 'erm', 'err' etc., were included in the transcription, but the length of pauses was not recorded as this was not considered essential to the data analysis. The mean number of words spoken by the subjects was 169.6 words (range 83~343).

Written data was collected from 40 students, but one student failed to save her essay. She subsequently rewrote the essay but it was decided not to include this data in the analysis as it could not be considered to have been written under the same conditions as the original essay. All 41 students took part in interviews but in one case I made a mistake with the recording equipment. I decided only to analyze data from students where reliable data had been obtained in both modalities. As a result, the final analysis involved data from 39 students.

Data Analysis

The first stage of data analysis involved reading through each of the essays and transcriptions and highlighting epistemic forms. Although time-consuming, this manual analysis enabled me to 'get to know my data'. I could identify broad patterns of use of epistemic forms, and also pick out cases where ambiguity of function would mean that decisions would have to

be taken about whether a form was epistemic or not. This is particularly the case with modal auxiliaries which can have both root and epistemic functions (see Coates, 1983).

The second stage of data analysis involved the use of Wordsmith Tools (Scott, 1996). The following procedures were carried out:

- two learner corpora were created: (1) a written corpus of discursive essays - 39 texts (12,583 words); (2) a spoken corpus of transcriptions of discursive language - 39 texts (6,615 words);
- wordlists were created for both learner corpora;
- the wordlists were checked for the frequencies of epistemic forms; for each form concordance information was analysed;
- where concordances contained both epistemic and non-epistemic forms, manual analysis of the data was carried out to identify the number of epistemic forms (theoretical support for choices made was based on Coates (1983));
- a keyword analysis was carried out between the two learner corpora to identify any stance forms which were used significantly more frequently in either corpus.

The data collected from this analysis was used to answer the research questions of this study.

RESULTS AND DISCUSSION

Research Question 1:

To what extent do Japanese EFL learners use epistemic forms similarly or differently in discursive language in the written and spoken modalities?

In order to answer this question the two corpora were analysed for the use of epistemic forms. Lists of common epistemic forms taken from Holmes (1988) and Hyland & Milton (1997), along with data on stance in Biber et al. (1999: 965–986), were used as the basis for deciding which forms to focus on. The frequencies for these epistemic forms in the two corpora, along with percentages which enable comparison to be made between the corpora, are shown below. For clearer analysis, the epistemic forms have been divided into grammatical categories.

Lexical verbs

Table 1 below shows the use of lexical verbs. As has been seen in other studies of learners' use of epistemic forms, there is an overwhelming reliance on the use of the verb 'think', particularly in the form 'I think'. Apart from 'think', only two other verbs ('know' and 'believe') occur more than three times in either corpus, and these three verbs are the only ones which occur in the spoken corpora. It is interesting to note that each of these three verbs occurs to an almost equal extent in the two corpora. This shows that learners do not vary their use of lexical verb according to the modality in which they communicating. It is interesting to compare this data with that provided by Biber et al. (1999: 982), see below:

Table 1: Frequency of lexical verbs conveying epistemicity

	think	know	believe	suggest	suppose	expect	guess	feel	imagine	tend
WRITTEN	189	30	9	3	3	2	2	2	2	2
	1.50%	0.24%	0.07%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
SPOKEN	84	18	3	0	0	0	0	0	0	0
	1.27%	0.27%	0.05%							

Table 12.1 Distribution of comment clauses as stance markers across dialects and registers; occurrences per million words

each * represents 10

	AmE CONV	BrE CONV	FICT	NEWS	ACAD
<i>I think</i>	***** *****	***** ***** *****	***		
<i>I suppose</i>	**	*****	***		
<i>I guess</i>	***** ****		*		
<i>I believe</i>	*		*		
<i>it seems</i>				*	*

Biber et al. state that what they term 'comment clauses' "are generally rare as stance markers; they are used with moderate frequency only in conversation". In this regard, the fact that the learners use 'think' as the most frequent marker of epistemic stance in their writing shows overuse²⁾ of this form. This may be due to negative transfer from Japanese, where the corresponding form ('to omoimasu') is used more often in written language. However, the fact that similar findings have been found for learners from various L1 backgrounds (Hyland & Milton, 1997; Aijmer, 2002; McEnery & Amselom Kifle, 2002) would seem to point towards 'think' having a primary function in the acquisition of epistemic stance.

Biber et al's data shows that 'guess' (in American English) and 'suppose' (more common in British English) are the next most common lexical verbs used in comment clauses in conversation. Neither of these verbs occurs in the spoken corpus of learner data, although they do occur in the written data. This provides evidence that the learners do not differentiate between spoken and written forms.

Modal verbs

The learners' use of modal verbs to convey epistemicity can be seen in Table 2 below. The first clear finding is that even the most commonly used modal verb, 'will', is used far less frequently than 'think' in both modalities. This finding is very much at odds with the data in Biber et al (1999) which shows, for example, that 'will' occurs around 5,600 times per million

Table 2: Frequency of modal verbs conveying epistemicity

	will (won't)	may	would(n't)	could(n't)	might
WRITTEN	35	17	9	7	4
	0.27%	0.13%	0.07%	0.06%	0.03%
SPOKEN	8	0	1	0	0
	0.12%		0.02%		

words in conversation, whereas 'I think' occurs around 160 times per million words. Furthermore, Biber et al. (1999: 489) give details of the use of modal verbs in different registers. The most popular choice of the learners, 'will', is almost three times as common in native speaker conversation as in academic writing, and although the learner data does not exactly correspond to these registers, it is still of note that in the case of the learners, 'will' occurs much more frequently in written language. The modal 'would' is not that much less frequent in native speaker language than 'will' but compared to this, the learners clearly underuse it. As regards 'may', learner use is more in line with native speakers in that 'may' is far more common in academic writing than in conversation.

The spoken corpus of learner discursive language is particularly lacking in examples of modal verbs used epistemically. This indicates that some of the learners are able to use modal verbs with an epistemic sense appropriately when given more processing time (ie, in the written mode). However, in the more pressurized environment of spoken production, they find it difficult to use modal verbs. The likely reason for this is that comment clauses, such as 'I think', form separate clausal units, which are easier to process. Modal verbs have to be integrated into a more complicated clausal structure which is more difficult to produce under the processing constraints of the spoken mode.

Epistemic stance adverbials

Another common way in which epistemic stance is realized, is through the use of stance adverbials. Table 3 shows the use of such adverbials by the learners in writing and speech. The most striking difference can be seen in the way in which the adverb 'maybe' is used. It is more than eight times more frequent in the spoken corpus than in the written one, and is the only epistemic form which is listed as a Keyword in the Wordsmith analysis of the two learner corpora. This matches the findings of Biber et al. (1999: 869) which show that 'maybe' is used much more frequently in conversation than any of the three written genres which they provide data for. In fact, 'maybe' is the second most common epistemic stance form used by the learners in speech, after 'think'. This finding accords to a certain degree with research (reported in Kasper & Rose, 2002) on the acquisition of modality, which found that the two most common ways in which lower-proficiency level learners expressed epistemic modality was through parentheticals such as 'I think' and modal adverbs such as 'maybe'

Table 3: Frequency of epistemic stance adverbials

	maybe	of course	actually	definitely	possibly	perhaps	probably
WRITTEN	5	9	5	2	2	1	0
	0.04%	0.07%	0.04%	0.02%	0.02%	0.01%	
SPOKEN	23	3	5	0	1	0	1
	0.34%	0.05%	0.08%		0.02%		0.02%

(Kärkkäinen, 1992). Totaling the data in Table 3 shows that these epistemic stance adverbials make up 0.51% of the spoken corpus and 0.20% of the written corpus. Kärkkäinen's study found epistemic adverbs to make up 0.36% of low-level Finnish learners speech acts. This increased to 0.59% for high-level learners.

These findings also agree to a certain extent with Biber et al. (1999: 867), who states that, "in general, conversation has the highest frequencies of the most common [stance] adverbials". However, the actual choice of adverbials differs from the Biber et al. (1999: 869) data (see below). Of particular note is the fact that the most common stance adverbial in Biber et al's data ('probably') occurs just once in the two learner corpora combined.

Table 10.15 Most common stance adverbials across registers and dialects; occurrences per million words

each ■ represents 100 □ represents less than 50 () marks semantically ambiguous occurrences which could also be interpreted as extent/degree adverbs

	CONV (AmE & BrE)	FICT	NEWS (AmE & BrE)	ACAD
epistemic—doubt/certainty				
<i>probably</i>	■■■■■■■■■■	■■■■	■■■	■■■
<i>maybe</i>	■■■■■■■■	■■■	■	□
<i>perhaps</i>	■■■	■■■■■■■■	■	■■■■■
<i>of course</i>	■■■	■■■■■■■	■	■■■
<i>certainly</i>	■	■■■	■	■
<i>definitely</i>	■	□	□	□
epistemic—actuality				
<i>really</i>	■■■■■■■■■■ (■■■■■■■■)	■■■■■■■	■(■)	■
<i>actually</i>	■■■■■■■■■	■	■	■
<i>in fact</i>	■	■	■	■

Other epistemic forms

Analysis of the corpora for other relatively common epistemic forms showed that the written data included fairly frequent use of the nouns 'opinion' and 'fact'. Table 4 shows the frequency of these nouns in the two corpora. These were often used in chunks. For example, 'in my opinion' occurred eight times and 'in fact' occurred four times in the written

Table 4: Frequency of 'opinion' and 'fact'

	opinion	fact
WRITTEN	18	13
	0.14%	0.10%
SPOKEN	0	1
		0.02%

data and once in the spoken data. It is notable that these forms are almost absent from the spoken data. Biber et al. do not give data on 'opinion' as such, but they do give data for the chunk 'in fact' (see above). This shows that it occurs with a similar frequency across different genres and modalities. However, as the actual frequencies are relatively small in the learner corpora, it is not possible to draw definitive conclusions from this data.

Summary

Overall, the use of epistemic forms in writing and speaking forms a rather unclear picture. The use of lexical verbs is very similar in the two modalities, with 'think' being the prominent epistemic form used by the learners in both written and spoken discursive language. However, modal verbs occur more frequently in the learners' written language, with modal adverbs much more frequent in spoken language. The low use of modal verbs in speech can probably be explained by the fact that they require a greater level of processing.

The higher use of modal adverbs, especially 'maybe', in speech might well reflect the fact that the learners look to more syntactically flexible lexical forms when under the pressure of communicating on-line.

Furthermore, the pattern of use of epistemic forms with regard to spoken and written language cannot be seen to match that of native speakers in any clear way, with the possible exception of the way 'maybe' is used in speech. It can be concluded therefore, that learners at this level do not distinguish between different modalities as regards epistemic language. This may well be due to their attention being taken up, especially when speaking, by processing demands.

Research Question 2:

What individual patterns in the use of epistemic forms can be discerned which may offer insights into developmental sequences in second language development?

The analysis above uses the two learner corpora to look at the learners as one group, reflecting the language production of low-intermediate to intermediate-level Japanese EFL learners. However, it is also important to look at the data at the individual level in order to try to discern patterns in the way epistemic forms are used. These patterns may enable conclusions to be reached about developmental sequences with regard to epistemic forms.

For this purpose, each learner's essay and spoken transcript were analysed in order to see the range of grammatical forms which the learners used to express epistemicity. Each text was marked for the use or non-use of the five categories below:

- 1) 'think'
- 2) epistemic stance adverbial (eg, 'maybe', 'of course')
- 3) modal verb used epistemically
- 4) an epistemic lexical verb other than 'think' (eg, 'believe', 'know')
- 5) an epistemic noun (eg, 'fact', 'opinion')

This categorization was finalized after preliminary analysis of the data. In particular, it was considered that it would be more revealing to separate 'think' from other epistemic lexical verbs. Table 5 below shows the data resulting from this analysis. The categories have been organized from left to right in order of the number of learners adopting that strategy over both modalities (it should be noted that if, for example, a learner uses 'think' in both speaking and writing, it counts as '2'). Furthermore, the table has been ordered from top to bottom in rank order based on the number of categories adopted by a learner over both modalities. By ordering the data in this way, it is possible to get a broad picture of the acquisitional pattern for epistemic forms. This analysis is supported by qualitative analysis of three learners. I will look at the learners at the bottom, in the middle, and at the top of Table 5.

Student B9 – Low user of epistemic forms

This learner only used one epistemic form in his essay, and it was written at the very beginning:

It is fact that mobile phones have many good things, but also they have bad things.

In his essay, he gives advantages and disadvantages of mobile phones but without presenting these views as his own opinion. For example, he writes:

'The first good thing is connecting. If we have mobile phones, we can connect other people any time.'

The use of this style could represent an authoritative and more objective approach to the issue rather than a lack of ability to use epistemic forms. However, this learner was one of the weakest learners in the class with a TOEIC (Test of English for International Communication) test score well below the average for the class. Furthermore, in the speaking task he produced just 96 words (well below the average), struggled to produce extended sentences, and had to be frequently prompted. There are no epistemic forms in his spoken discourse. Therefore, it is likely that his low use of such forms is due to a lack of competence in this area. He appears to represent an earlier acquisitional stage of epistemic forms in which there is no use of the verb 'think', epistemic stance adverbials or modal adverbs.

Table 5: Individual learners' use of epistemic forms in their written and spoken texts

student ³⁾	think	epistemic stance adverbial	modal verb	other epistemic lexical verb	epistemic noun	number of categories used across the two modalities
B8	WS ^{b)}	WS		W	WS	7
C4	WS	WS	W		WS	7
C5	WS	W	W	W	W	6
C11	WS	W	W	W	W	6
C14	WS	W	W	W	W	6
A4	WS	WS	W	W		6
A12	WS	S	W	W	W	6
B12	W	WS	WS		W	6
C3	WS	WS		W	W	6
A2	WS		W	W	W	5
C9	WS	W	W	W		5
C13	WS	W	W	W		5
A3	WS	WS	W			5
A7	WS	WS	W			5
A8	WS		W	WS		5
A10	WS	S	W		W	5
B5	S	S	W		WS	5
B6	WS	S	WS			5
A1	WS			W	W	4
A11	WS			W	W	4
B7	W	WS	W			4
C1	WS		W		W	4
C2	WS		W	W		4
C7	WS		W	W		4
C8	W	S		W	W	4
A5	WS	WS				4
B10		S	WS	W		4
C10	WS	WS				4
C6	W	W	W			3
A13	WS		W			3
C12	WS		W			3
B1	W		W			2
B2	W		W			2
B3	W			W		2
A6	WS					2
A9	WS					2
B11	WS					2
B4	S				S	2
B9				W		1

Student A1 – Mid-range user of epistemic forms

This learner used a total of six epistemic forms in her essay: think (3); suppose (2); opinion (1). The extract below shows examples:

'On the other hand, I can suppose the disadvantages about it. First we are forced to study a second foreign language. I think studying languages need a strong motivation.'

Unlike the learner discussed previously, she is able to show that the opinions expressed are her own. These sentences are taken from the beginning of the second paragraph of her three-paragraph essay. Each paragraph is framed by epistemic forms, thus making clear that it is her views which are being expressed. However, she does not use any modal verbs to convey this viewpoint.

In her spoken language she only used one epistemic form (think) at the very beginning. This was typical of many learners who may have used the word 'think' from the question on the task sheet (Which do you think is better?), but then didn't produce any more epistemic forms by themselves. This learner's level of language competence can be considered as higher than B9 from the fact that she only had to be prompted twice when expressing her views, and was able to produce 141 words. However, her spoken opinions are much less clear and organized than her written discourse, and her language is very hesitant:

'erm the countryside erm there is erm big place university and erm the student erm considerate erm consider erm the student can considerate to study'

It would appear that the on-line demands of producing extended spoken discourse causes learners with this level of language competence to struggle to adequately frame their discourse epistemically in the way they are able to do when writing.

Student B8 – High user of epistemic forms

In his written essay, this learner uses a variety of epistemic forms: think (2), definitely, actually, seems, tend to, in fact. The only type not used is a modal verb. This learner is also able to effectively frame his written discourse epistemically:

'Actually, it seems that mobile phone is a must for our life now. But we have to think judicious use of mobile phone.'

This learner's spoken language is much richer in epistemic forms than the 'mid-range' and 'low' users. Whereas learner B9 used no epistemic forms in his spoken language and learner A1 used just one, this learner uses 12 forms: think (7); maybe (3); believe (2). Furthermore, this learner's spoken language is much more fluent; he produced 244 words. Here is an example:

'..and I believe that practice makes perfect not perfect but gradually my skill is develop I believe - this is very important - yes..'

This learner's use of epistemic forms is typical of the learners at the top of Table 5 in that there is a tendency for them to use them much more in their spoken language than the learners in the middle and towards the bottom of the table. In fact, their spoken language is not that different from their written language with the exception that whereas most of them

used modal verbs and/or other lexical verbs in their written language, they did not use them when speaking.

Developmental Patterns in the Use of Epistemic Forms

In this section the developmental patterns for written and spoken language will be described separately as clear differences can be seen from the data analysed in this research.

Written Language

Stage 1: the use of epistemic lexical verbs, predominantly *I think*

Stage 2: the use of a further epistemic form – an adverbial, a modal verb, a noun

Stage 3: the use of a variety of epistemic forms from various grammatical categories

Spoken Language

Stage 1: almost no use of epistemic forms

Stage 2: the use of epistemic lexical verbs, predominantly *I think*

Stage 3: the additional use of an epistemic adverbial

Stage 4: the use of a further epistemic form – a modal verb or a noun

This clearly shows that the use of epistemic forms appears later in learners' speech than in their writing. A learner, who possesses knowledge of a range of epistemic forms, and can use several of them in writing, is likely to use only an epistemic lexical verb, and perhaps an epistemic adverbial, in speech. In particular, it is noticeable that modal verbs appear earlier or at the same time as adverbials in written language, whereas in spoken language, adverbials are used prior to modal verbs. This is almost certainly on account of the greater language processing required to produce modal verbs as compared to modal adverbials.

CONCLUSION

This study shows that the off-line nature of writing enables most learners to better demonstrate their knowledge of epistemic forms. Many learners found it difficult to produce the same range of epistemic forms when presenting their opinions on-line in the spoken modality. Only the better learners amongst the subjects were able to produce nearly as extensive a range of modal forms in speech as well as in writing.

From a second language acquisition viewpoint, the findings of this study agree to a certain extent with previous research that shows that learners tend to rely on lexis to express modality before using more grammaticalized forms (ie, modal verbs). However, previous research has suggested that 'maybe' is the primary term used to express epistemic modality.

However, in the case of Japanese learners it appears that 'think' is the primary term. As the previous studies were conducted with learners with other L1s, this may suggest that L1 transfer plays a role in the acquisitional profile of epistemic forms.

Further research in this area would ideally involve longitudinal research tracing learners'

development. This would enable the suggested patterns of acquisition of epistemic forms to be seen more clearly. It would also be useful to look at the use of epistemic forms in other genres which tend to require its use, such as descriptive language. Research of this nature, using learner corpora to compare learners' spoken and written English, could also provide insights into many other areas of second language development.

NOTE

- 1) Louvain International Database of Spoken English Interlanguage: <http://www.fltr.ucl.ac.be/fltr/germ/etan/celc/Cecl-Projects/Lindsei/lindsei.htm>
- 2) Within the field of CIA (Contrastive Interlanguage Analysis, see Granger, 1998) the terms 'overuse' and 'underuse' are used to denote quantitative differences in the use of linguistic features. No judgments about the quality of the writing are necessarily implied.
- 3) In the table, the letters A, B and C refer to the task which the students undertook. Letter A refers to the students who did Task 1 in both cases (see Appendix), B refers to Tasks 2 and C to Tasks 3.
- 4) W signifies that the student used this category of word in their writing; S signifies the same for speaking.

REFERENCES

- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55-76). Amsterdam: John Benjamins.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finnegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Coates, J. (1983). *The semantics of modal auxiliaries*. London: Croom Helm.
- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp.67-79). London: Longman.
- Granger, S. (Ed.). (1998). *Learner English on computer*. London: Longman.
- Granger, S., Hung, J. & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins.
- Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics*, 9, 21-44.
- Housen, A. (2002). A corpus-based study of the L2-acquisition of the English verb system. In Granger, S., Hung, J. & Petch-Tyson, S. (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 77-116). Amsterdam: John Benjamins.
- Hyland, K. & Milton, J. (1997). Qualifications and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6(2), 183-205.
- Kärkkäinen, E. (1992). Modality as a strategy in interaction: Epistemic modality in the language of native and non-native speakers of English. In L. F. Bouton & Y. Kachru (Eds.), *Pragmatics*

- and language learning* (monograph series vol. 3, pp. 197-216). Urbana, IL: Division of English as an International Language, University of Illinois, Urbana- Champaign.
- Kasper, G. & Rose, K. R. (2002). *Pragmatic development in a second language*. Oxford: Blackwell.
- Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in present-day British speech and writing*. Longman: London.
- McEnery, A. & Amselom Kifle, N. (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Ed.), *Academic discourse* (pp. 182-195). London: Longman.
- Palmer, F. (2001). *Mood and modality*. (2nd Ed.). Cambridge. Cambridge University Press.
- Scott, M. (1996). *WordSmith Tools*. Oxford: Oxford University Press: <http://oup.com/elt/global/isbn/6890>

APPENDIX

Discursive writing - tasks

- (1) Do you think that all Hiroshima university students should have to study a second foreign language (eg. French, German, Chinese, Korean etc.)?
- (2) Do you think that the availability of mobile phones is a good thing for society?
- (3) Do you think that the required age in Japan for voting, drinking and smoking (20) is appropriate?

Discursive speaking - tasks

- (1) Consider the advantages and disadvantages of going to a university in the countryside or in the middle of a city. Which do you think is better? Why?
- (2) Consider the advantages and disadvantages of doing a one year Study Abroad Program as an extra part of your university degree. Would you like to go on such a program? Why or why not?
- (3) Consider the advantages and disadvantages of doing a part-time job as a student. Do you think it is good for students to do a part-time job? Why or why not?

要 約

話し言葉と書き言葉による課題発話の学習者コーパス比較研究： 日本人英語学習者による認知的表現の使用に焦点を当てて

ケネス・フォーダイス
広島大学外国語教育研究センター

本研究の目的は、日本人英語学習者の話し言葉、書き言葉による課題発話に表れた認知的表現の使用を比較することである。データ収集のため、41名の学生に話し言葉および書き言葉で意見を述べさせ、それら2つの焦点化された学習者コーパスについて量的・質的比較を行った。本研究の主要な結果は、(1) 学習者は、書き言葉においてはそのオフライン的特徴から、同時的処理の必要な話し言葉よりもより多くの種類の認知的形式を使用したこと、(2) 学習者は、まず I think, maybe のような語彙レベルから、熟達度が上がるにつれて次第に法助動詞のような、より文法レベルの用法を使うようになったこと、である。今後、異なるジャンルや学習者の母語、そしてタスクの種類によって認知的形式の使用がどのように変化するかを精査しながら、本研究の結果をさらに検証していく必要がある。