Relationship between English Majors' Tendency to Use Modal Sequences and Time of Enrollment:A Corpus-based Study

QIU Shichun

Foreign Languages College
Jiangxi Normal University, JXNU
Nanchang, Jiangxi, PR China
e-mail: shichunqiu@gmail.com

*Abstract—***Based on the result of the trends of modal sequences in Chinese English majors' argumentation, this research focuses on the relationship between English majors' tendency to use modal sequences and their time of enrollment. With the aid of corpus analysis, it reveals that the tendency to use modal verbs, whether epistemic or deontic, is not related to their enrollment time. This study subsequently provides reference to the understanding of how Chinese students acquire modal verbs and gives suggestions of how to teach modal verbs as well.**

Keywords-modal sequences; modal sequence trends; deontic modality; epistemic modality; modal verbs teaching

I. INTRODUCTION

A. Modal Verbs, Modal Sequences and Their Meanings

Modal verbs are major forms of expressing modal meaning. It can be roughly categorized as deontic modality and epistemic modality (Biber et a1.1999; Coates 1983; de Haan 1997; Larreya 2004; Palmer 2001). Deontic modality, as Lyons (1977), is concerned with the possibility or necessity of acts performed by morally-responsible agents; whereas epistemic modality refers to the truth of proposition, matters of knowledge, or opinion rather than fact. The few prototype examples of expressing epistemic modality are "can", "could", "may", "should", etc., but most of them like "can", "may", "must", etc. may also be prototypes of expressing obligation modality. However, a research shows (Gabrielatos & T. McEnery 2005) that the application of epistemic modality of above modal verbs covers 83% of the expressions of epistemic modality.

Modal verbs typically combines with auxiliary verbs and notional verbs, and form the sequence "subject+modal verb+auxiliary or notional verbs", known as modal sequences. Susan Hunston (2004) argues that it is more of benefit to teach learners modality sequences than the meanings of modal verbs as auxiliaries. In the light of her research into modal sequences, she also believes that English learners will not master all kinds of meanings modal verbs express until they make a further study of modal sequences.

Sinclair (2004a) lists 14 meanings of "must", including phrases like "I must say", "it must be remembered that", "if you", "why must you" and "if you must know" etc. and classifies multiple meanings of "must" as "epistemic modality" and "deontic modality". By his research he concludes that learning modal verbs can not be limited to individual lexical items, but demands an exploration of multiple meanings based on its collocations. That is the method of modal sequences.

However, the method of learning modal sequences fails to attract adequate attention of scholars in China and abroad and lacks research achievements. To this day only Guo (2005) claims that learning modal verbs should be figured out from a perspective of phraseology rather than a single modal verb.

When it comes to the sense of modality, a research shows that there exists a strong correspondence between modal sequences and the types of modal meanings, thus the meaning of modal verbs can be predicted based on difference of sentence patterns of sequences (Biber et al. 1999; Coates 1983; Wärnsby 2003). The correspondence can be generalized as follows (table 1) (quoted in Liang, 2008), which is also the framework this author refers mainly to.

TABLE I. CORRESPONDENCE BETWEEN SENTENCE PATTERNS OF MODAL SEQUENCES AND MEANINGS OF MODALITY

Typical patterns	Examples	Epistemic Modality	Deontic Modality
VM+epistemic modal adv.	can probably do	Y	N
leading words+VM	it may be true that	Y	N
VM+perfective aspect	must have done	Y	N
VM+progressive aspect	must be working	Y	N
VM+stative verb	must be	Y	N
VM+dynamic verb	must work	N	Y
animate subject+VM	kangaroos can	N	Y
inanimate subject+VM	the weather may	Y	N
there+ VM+ be	there must be	Y	N

VM=modal verb, Y=yes, N=no.

B. Current Domestic Research of Modal Sequences and its Limitations

The systematical study of modal sequence of demestic English learners originates from Liang Maocheng (2008), who analyzes the usage characteristics of the modal sequence in written English of non-English majors in China. But his corpus is only adopted from College English Band 4 Learner (ST3) and College English Band 6 Learner (ST4) of Chinese Learner English Corpus (CLEC) collected before 2003. Liu Wenyan (2009) analyzes the usage features of modal sequence of English majors in written English, but her

research is only limited to College English Major Freshman/Sophomore (ST5) and College English Major Junior/Senior (ST6) of CLEC. Xie Jiacheng (2009) finds that Chinese English learners more freqently use the meaning of deontic modality and modal sequence of "must", while its meaning of epistemic modality and modal sequence are less used and even misused. He then points out (Xie, 2010) that modal sequence and its context should be blended in the teaching of modal verbs. However, this research is restricted to "must" of his self-built corpus.

C. Hypotheses of the Research and Reasons

This paper tries to explore the relationship between the tendency for the two groups of English majors to use modal sequences and their time of enrollment, and takes the following two hypotheses as a starting point:

- The tendency of using modal verbs of two groups of English majors varies to some degree. They understand and realize the modal sequences differently too.
- The approaches of college teachers remain unchanged after 11 years.

There are two reasons for the two hypotheses. Firstly, since 1999, the year the Expansion of University Enrollment began, most middle school students have felt, to some extent, more and more relaxed about learning a foreign language and more willing to communicate with people in English. Additionally, adapted in the years 1996, 2004 and 2007, course books for middle school which advocate advanced western teaching ideas like collaboration, exploration and learning grammar from practical language environment naturally shape the way they learn English. Secondly, in China's universities and colleges, nation-wide curriculum reform has not be carried out as frequently as in middle schools; Moreover, grammar courses for English majors who emphasize their spoken English do not clarify the basic concept of modal verbs; on the other hand, college teachers of all kinds find it difficult to apply the results of corpus research to implementation of explicit teaching and make students learn idiomatic sequences of native language.

II. RESEARCH DESIGN

A. Research Question

Based on the above analysis, this research tries to answer this question: are the features of English majors' overuse and underuse of modality sequences during 1996-2007 related to their enrollment time? Specifically speaking, here are three questions to be answered:

- Is the tendency of using modality sequences related to their time of enrollment?
- Is the tendency of using epistemic modal sequences related to their time of enrollment?
- Is the tendency of using deontic modal sequences related to their time of enrollment?

B. Preconditioning of Corpus and Analytical Method

The prophase conclusion of this research is made on the basis of tagging texts and adopting the analysis of Key word in context (KWIC), which is also the base of the research and the corpus analysis.

1) Corpus and its preconditioning

As mentioned above, the research use learner corpus, sub-corpora of English majors' argumentations of WECCL1.0 (Wen, et al. 2005) and WECCL 2.0 (Wen, et al. 2008). To examine whether a correlation exists between their using modal sequences and the time of enrollment, the author employs sub-corpora of English majors' tagged argumentation to build one learner corpora of argumentations, and names it as Group WECCL1.0. Similarly, for the purpose of this research, another learners' corpora is built with the name Group WECCL2.0, by using subcorpus generator to retrieve the texts with the heading STU1 from tagged WECCL2.0. The reference corpus is the Louvain Corpus of Native English Essays (LOCNESS). Table 2 shows the main text messages of the two corpora.

TABLE II.   BASIC INFORMATION OF TWO CORPUS

Corpus(Arguments)		Word token	Amount of text
Learner Corpora	WECCL1.0	1,005,607	3,053
	WECCL2.0	1,147,327	4,359
LOCNESS		264,041	297

Consulting the methods of Aarts & Granger (1998) and To-no (1999), the author processed the tagged texts, by means of text processor PowerGrep (see table 3), and thus syntactic features of modal sequences in the corpus come out into view.

TABLE III.   ILLUSTRATION OF TEXT PROCESS

	Active	Passive
Non-tagged text	We should set	The standard should
	the standard carefully.	be set carefully.
Tagged text	We -PPSI2 should-VM set -VVO	The AT standard-NN1 should-VM
	the-AT standard -NN1 carefully-RR.	be -VB0 set -VVN carefully-RR.
Processed	PPSI2 should VV0 AT NN1 RR.	AT NN1 should VB0 VVN RR.

By using the functions of WordSmith Tools like concordance, wordlist and cluster parse, the author finds it convenient to observe and retrieve all the modal sequences from the processed texts.

2) Corpus analysis method

a) Analysis of modal sequences

- Preprocess the corpora in accordance with the method of [II, *B,1)*].
- Make cluster lists of 2-8 words based on the two preprocessed corpora.
- List all clusters overused and underused by using KeyWords (KWs) analyzer to compare the cluster lists, and generate positive and negative KWs clusters of learner corpora tag sequences.
- Group KWs clusters lists into modal sequences overused and underused in learner corpus.
- Run independent sample test, using Statistical Package for Social Sciences (SPSS) to verify

frequencies of modal sequences learners overused (i.e. modal sequence A) (MSA) and frequencies modal sequences learners underused (i.e. modal sequence B) (MSB), in order to find out whether eminent difference exists in the frequency of learners overusing and underusing modal sequences in their compositions at different times.

b) Recognizing meaning type of modal sequences

To explore whether the tendency of students using epistemic and deontic modal sequences is related to their time of enrollment, the author examines only two aspects owing to space reasons:

- Is there significant difference between the frequency of "modal verbs + all sorts of stative verbs sequences" and their time of enrollment?
- Is there significant difference between the frequency of "all sorts of pronouns + modal verbs sequences and their time of enrollment?

In terms of Zhang Zhengbang's (2003, 163) definition and classification of stative verbs, the author tags 38 stative verbs with VB0, VBDR, VBDZ, VBI, VBR, VBZ, VH0, VHD, VBM, VVI, VBI, VHI, and then use modal verbs as context word to run concordance of the two preprocessed Group WECCL1.0 and Group WECCL2.0 respectively.

Meanwhile, based on typical sentence patterns of epistemic and deontic modal sequences listed on the table 1, the author merge and group learner's overused and underused modal sequences.

III. RESEARCH FINDINGS

A. Two Types of Modal Sequences

In this research, the author analyzes keywords in 2-8 word cluster lists based on the two corpus, gets numbers of keyword clusters and chooses top 20 positive keywords, meaning learners' overused modal sequences (see Long, 2009, 2011). Finally, in terms of the difference between members of sentences before and after modal verbs, the author divides the data of modal sequences learners' overused into two:

- Sequence "subject + modal verb", which can be summed up as "we +can/will/must", "you + can", "NN1/NN+can/should", and "they + can/should".
- Sequence "modal verb +verb", which can be further summed up as "can (should/will/may/must) (+adverb)+verb unmarked aspectually (+determiner(the/your)/(+ us+ verb unmarked aspectually)", and "should +verb without markup of tense and aspect + to(+verb without markup of tense and aspect)".

In the same way, the author gets top 20 negative modal sequences (table omitted, see Long, 2009, 2011).

Negative keywords reflect language features that native speakers often use in written English, while English learners relatively lack. The observation shows that many modal sequences consisting of "could" and "would (esp.)" rarely appear in learners' language.

B. Relationship between the Tendency of Using Modality Sequences and Students' Time of Enrollment

To verify the correlation between the tendency of learners' using modality sequences and students' time of enrollment, the author uses respectively MSA and MSB as search terms to retrieve two learners' argumentation texts, and lets the frequencies go through independent sample test. The result shows no significant difference in the variance of frequencies of MSA in the compositions of Group WECCL1.0, and Group WECCL2.0 (the homogeneity test of variances: sig=0.786>0.05, F=0.075, df=38, comparison of means: t=0.257, sig=0.799>0.05), which indicates that no significant difference exists between two groups in frequencies of MSA. The result also reveals that there is no significant difference in frequencies of MSB between two groups' compositions (the homogeneity test of variances: sig=0.877>0.05, F=0.024, df=38, comparison of means: t=-0.777, sig=0.442>0.05), revealing there is not a significant difference either in frequencies of MSB between two groups.

C. Relationship between the Tendency of Epistemic and Deontic Modal Sequences and Students' Time of Enrollment

1) Relationship between the tendency of using "modal verb + stative verb" and students' time of enrollment

To judge whether verbs (VVI) are dynamic from learners' overused CAN VVI, PPY CAN VVI, NN1 CAN VVI, PPHS2 CAN VVI, SHOULD VVI, PPIS2 SHOULD VVI, NN2 SHOULD VVI, WILL VVI, MAY VVI, PPIS2 MUST VVI, the author retrieve the verbs behind the modal verbs in the two corpora (native and non-native), and then classify the results according to their semantic features, which finds that learners mostly put dynamic verbs behind modal verbs. Below are learners' most frequently-used verbs (based on the occurrences, the top 20 are listed).

TABLE IV.   TABLE 4: THE MOST FREQUENTLY-USED VERBS BEHIND MODAL VERBS (TOP 20)

freq	verb	freq	verb	freq	verb	freq	verb
8020	**be**	1148	get	677	**find**	407	bring
1750	make	837	take	522	use	387	**become**
1589	**have**	770	give	511	say	321	**think**
1466	learn	754	help	484	**know**	313	tell
1460	do	681	**see**	483	pay	292	**feel**

Note: bold words are stative verbs

Of all the above 20 frequently-used verbs, only 8, namely "be", "have", "see", "find", "know", "become", "think" and "feel" can be used as stative verbs, while the rest are usually used as dynamic verbs. Compared with non-English majors who just view the verbs "be", "have", "see", "know" and "think" as the stative verbs, English majors have apparently increased a few more stative verbs (see Liang, 2008), frequency of which is low in the table.

Similarly, after the retrieval of the two corpora, the result shows that modal sequences like WOULD VVI, WOULD VBI, WOULD VHI, COULD VBI, WOULD RR, WOULD VHI VVN, WOULD RR VBI, WOULD XX, WOULD VHI TO, WOULD VVI AT, WOULD VBI AT1, WOULD VVI

II and COULD VBI VVN belong to typical sequences of epistemic modality, which are underused by learners. However, compared with non-English majors (see Liang, 2008), English majors boast more types of the sequences of "modal verb + stative verb".

The author retrieves the two sub-corpora preprocessed observing the previous method respectively, and then pretreats the frequencies, followed by independent sample test in order to find out whether there are differences between the frequency of the sequence of the "modal verb + stative verb" and their enrollment time. The result is that no significant difference is shown in the variances of frequencies of modal sequence "modal verb+stative verb" in their composition between Group WECCL1.0 and Group WECCL2.0 (the homogeneity test of variances: sig=0.853>0.05, F=0.034, df=38, comparison of means: t=-0.056, sig=0.956>0.05). It reveals that two groups have no obvious difference in the frequency of modal sequence "modal verb + stative verb".

2) Relationship between tendency to use "animate/inanimate subject + modal verb" and students' time of enrollment

An important basis to judge semantic features is modal sequence "animate/inanimate subject and modal verb". If the subject before a modal verb is animate, the modal sequence express deontic modality in general (Coates 1983;Warnsby 2003). An observation of learners' overused modal sequences may suggest that "personal pronouns ("I", "we", "you" and so on) + modal verb" are the most commonly used modal sequences by learners. Since personal pronoun is typical animate subject, it is believed that this structure is typical deontic modal sequence and heavy use of this structure is another prominent feature of learners' language (Liang, 2008), including English majors. Table 5 shows the differences of frequencies and frequency of using personal pronouns before modal verb in the two corpus.

TABLE V.   FREQUENCY CONTRAST OF "PERSONAL PRONOUNS + MODAL VERBS" BETWEEN NATIVE SPEAKERS AND LEARNERS

Modal Sequence	Native speakers		Learners	
	Frequencies	*Standardized Frequency (per 10,000)*	*Frequencies*	*Standardized Frequency (per 10,000)*
They +VM	287	5.43	6235	14.1
We +VM	202	3.82	12950	28.5
You +VM	111	2.10	5727	12.98
I +VM	126	2.38	1721	3.90
She +VM	68	1.28	79	1.79
He +VM	200	3.79	1357	3.07
Total	944	18.8	28096	63.6

Note: VM= modal verbs

The statistics in table 5 shows that in the learner corpora the frequency of using "personal pronouns + modal verb" is 3.382 times than in native speakers' corpora (i.e. 63.6:18.8). Frequencies of learners' using the modal sequence consisting of modal verb and "we", "you" are also several times than native speakers'. The differences of frequency of using "personal pronouns and modal verb" between learners and native speakers demonstrate again that native speakers use more epistemic modal sequences than learners. However, compared with the research of Liang Maocheng, an evident decline of this proportion of English majors using this sequence can be seen, particularly when they use sequence of the first person (I) basically the same as native speakers do. Anyway, that is a progress, revealing that English majors have a further comprehension and command of the concept of epistemic modality that modal verbs possess.

The author also compares the frequency of this kind of modal sequences in the argumentations of two groups of English majors through independent sample T test. The result shows that there is no difference between the frequency of "personal pronouns and modal verb" and the time of enrollment (the homogeneity test of variances: sig=0.767>0.05,F=0.093, df=10, comparison of means: t=-0.069, sig=0.946>0.05).

In terms of the relationship between the time of enrollment and the frequency of "modal verb + dynamic verbs" and "personal pronouns + modal verb" , two typical deontic modal sequences in learners' compositions, it can be decided that English majors' comprehension and demand of modal verbs are closer to native speakers than non-English majors.

IV.   DISCUSSION

Learners customarily use animate subjects a lot, especially first and second person preceded by modal verb, and followed by more unmarked dynamic verbs instead of stative verbs, which is a feature of using modal verbs. No different trend of English majors using modal sequences is found through 11 years' study, which is deemed as another eminent feature of using modal sequences. There is one explanation to this view based on the author's understanding of students and his teaching experience.

Students' time of enrollment is not related to the frequency of using MSA, the reason for which may be that learners with still poor language ability often choose to make ample use of this sequence so as to avoid possible mistakes, leaving monotonous sentences and phrasing disagreeing with argumentation style. As for MSB, the reason should be that this sequence is seldom used in argumentations of the two groups of learners. According to the statistics, only 3,205 instances of using MSB are found in the learner's argumentation corpora containing 7,412, 22 articles. Not a single case of using MSB is left among the rest of 4,771 articles covering 64% of the total, 22% higher than that of non-English majors. This result indicates in one aspect that even English majors are not capable of controlling the language freely, and fail to properly use modal sequences formed by "could" and "would" for unreal conditions. Furthermore, the result also explains that English majors' comprehension and command of epistemic and deontic modality of modal verb do not vary due to the time when they get enrolled, which denies the previous hypothesis.

The phenomenon could be interpreted as invariability. That is to say, learners' language study in middle school has lasted for over 11 years, during which teaching reform

continues, faculty and facilities remain unchanged, along with unchanged mode of textbooks compilation and rigidified teaching approaches of teachers in middle schools and colleges, which nearly tests the second previous hypothesis, therefore, the way a foreign language learner acquire knowledge of modal verb and modality remain nearly unchanged.

V. CONCLUSION

Although, deontic modality transforming to the epistemic modality embodies partly cognitive law, it seems that this kind of acquisition is inevitable with no need to learn in advance. But conscious teaching and explicit learning would help save learners from groping in the dark. This paper maintains that, on the one hand, textbook compilation, classroom teaching, even after-class summary and comprehension should obey the cognitive law; on the other hand, the role that syntax sequence of modal verbs plays should also be highlighted. Below are some specified suggestions.

- Modal verb expressing epistemic modality should appear in early senior textbooks when compiled. Revision sections may include detailed usage of modal sequences such as "could", "would" expressing unreal conditions and general usage. More examples are frequently given and repeated in later texts so as to help learners digest the usage.

- Students should exactly be informed in class native speakers' tendency to use modal verbs. Personal pronouns and inanimate subject, for example, tend to be used before the modal verbs, and after the modal verbs are most used verbs with markup of tense and aspect. Speakers' assessment of the proposition expressed is always made by using modal verb and stative verbs. Meanwhile, more drills or practice of a variety of sequences of modal verbs are intended for students in and after class.

- By means of examples of epistemic modal verbs that native speakers tend to use, teachers may develop students' awareness of modal verbs particularly in language-output practice like argumentative writing.

On the basis of the previous studies, this research makes a further analysis of the relationship between the sequence of modal verb and the time of enrollment of Chinese English majors, which shows that the tendency of using modal verbs/sequences, as well as of using epistemic and deontic modality, is irrelevant to their enrollment time. Still, this interpretation also can also play a part in English textbooks compilation and classroom teaching and, hopefully, will become an inspiration to the teaching of English modal verbs and the teaching of oral and written English.

REFERENCES

[1] Aarts, J. & S. Granger. "Tag sequences in learner corpora: A key to interlanguage grammar and discourse," in S. Granger (ed.).Learner English on Computer. London: Longman. 132-142. 1998.

[2] Coates, J. The Semantics of the Modal Auxiliaries. London: Croom Helm.1983.

[3] Biber, D., S. Johansson, G. Leech, S. Conrad & E. Fine-gan. Longman Grammar of Spoken and Written English. Harlow: Longman. 1999.

[4] de Haan, F. The Interaction of Modality and Negation: A Typological Study. New York: Garland.1997a.

[5] Gabrielatos C. & T. McEnery. 2005. Epistemic modality in MA dissertation. http://eprints:lancs.ac.uk/ (accessed 09/08/2010).

[6] Guo, X. 2005. "Modal auxiliaries in phraseology: A contrastive study of learner English and NS English". Paper presented at Corpus Linguistics 2005, Birmingham University, UK.

[7] Hunston, S. "Colligation, lexis, pattern, and text". In: Scott, Mike and Geoff Thompson (eds.), "Patterns of Text: In honour of Michael Hoey". viii, 323 pp. (pp. 13–33) 2001.

[8] Long, Shoyun, "Chracteristics of Using Modal Sequences in Argumentation by English Majors". Paper presented at the First National Symposium on Learner Corpora 2009, National Research Center for Foreign Language Education, Beijing Foreign Studies University.

[9] Long, Shoyun, 2009. A Study of Modal Sequences Features in English Majors' Argumentations. English Education in China. http://www.sinoss.net/qikan/2012/0207/12790.html

[10] Liu, Wenyan, A Corpus-Based Study on the Written Form of Chinese English Majors Use of Modal Sequences, Doctoral dissertation, Hunan University, 2009.

[11] Liang, Maocheng, 2008, "A corpus-based study of modal sequences in Chinese tertiary EFL learners' written production". Foreign Language Teachign and Research(3):51-58.

[12] Lyons, J. (1977). Semantics. 2 vols. Cambridge: Cambridge University Press.

[13] Sinclair, J. 2004a. "The lexical item". First published in E. Weigand(ed.) Contrastive Lexical Semantics. Amsterdam: John Benjamins.131-148

[14] Wen, Qiufang, Written Corpus of Chinese Learners (WECCL1.0), 2005, Foreign Language Teaching and Research Press.

[15] Wen, Qiufang, Written English Corpus of Chinese Learners (WECCL2.0) 2008, Foreign Language Teaching and Research Press.

[16] Xie, Jiacheng, 2009, "Study of Modals Based on Learners' E&C Parallel Corpus", Journal of Yangtze University (Social Sciences) (6): 58-61.

[17] Xie, Jiacheng, 2010, "A survey of modal verbs use of Chinese learners". Journal of Basic Education (1): 9-12.

[18] Zhang, Zhenbang, 2003, A New English Grammar Coursebook, (Students' book) , Shanghai: Shanghai Foreign Language Education Press.