

Modal Sequences in Chinese Senior School Students' English Compositions

LONG Shaoyun

Foreign Languages College, Jiangxi Normal University
Nanchang, China
steven8810469@sina.com

Abstract—This study examines the characteristics of modal sequences used in ST2 from CLEC, with A Level from LOCNESS as a reference corpus. Main findings reveals: Firstly, modal sequences led by *can*, *will* and *must* are overused while those constructed by *could*, *would* and *should* are underused. Secondly, Chinese senior school students tend to overuse personal pronouns as subjects before modal verbs, and overuse verbs with unmarked aspect and voice after, making it rather difficult for students to express relatively complex modal meaning. Consequently, they overuse deontic modality and underuse epistemic modality, making their English comparatively strong on narrative and short on evaluation of their statements.

Keywords—modal sequence, overuse and underuse, deontic and epistemic modality, Chinese senior school students

I. INTRODUCTION

Modal verbs are usually combined with auxiliaries and notional verbs in the form of “subject + modal verbs + auxiliaries or notional verbs”, which is acknowledged as modal sequences (Hunston 2004). In terms of semantics, interpersonal meanings can be conveyed by modals (Biber et al. 1999; Halliday 1985); most modals are being polysemy with their meanings overlapped (Aijmer 2002; Mindt 1993), and context is needed to clarify the intricate meanings (Bybee & Fleischman 1995; Hunston 2004; Kratzer 1981; Mindt 1993).

As the main carrier of modality, modal verbs have been studied by many scholars, but most of the previous studies focus only on the realm of modal verbs, leaving modal sequences aside (Hunston 2004). The syntactic and semantic characteristics of modal sequences are not given much attention.

With regards to Chinese English learners' acquisition of modal verbs, many scholars like Coates (1983), Biber et al. (1999) and Aijmer (2002), etc. have found out some characteristics: A. the subjects before some modal verbs are most probably personal pronouns, like *we*, *you*, *I*, or other animated subjects; B. some modal verbs are mostly followed by dynamic verbs and without tense/aspect marker at the most time, etc. Susan Hunston (2002) has also done massive researches on it.

Liang Maocheng (2008) analyzed the use features of modal verbs and modal sequences in non-English majors' English compositions. Results showed that the overused modal verbs were those introduced earlier in textbooks while the underused modal verbs were those that express euphemistic and subjunctive mood. He also noted that

learners overused personal pronouns as the subject of modal phrases and overused unmarked verbs after. However, the learner data he selected is limited to ST3 and ST4 from CLEC. There are some other scholars like Liu Wenyan (2009), Long Shaoyun (2011, 2012, 2013) who did such researches from different perspectives, but all were just limited to college students.

This thesis seeks to analyze the use features of modal sequences in Chinese senior school students' English compositions by making contrastive analysis between ST2 in CLEC and A level in LOCNESS.

II RESEARCH DESIGN

A. Purpose

This paper aimed to explore the use tendency of the modal sequences by Chinese senior school students by examining the use pattern in the written form of Chinese senior school students and English native grade 12- to grade 13 students. The specific questions are: What are the modal sequences overused or underused by Chinese senior high students? And why are the modal sequences overused or underused?

B. Methods

Corpora

This paper manages to find out the characteristics of modal sequences used in English writing, with CLEC as target corpus and native corpus LOCNESS as reference corpus.

CLEC (Chinese Learner English Corpus) is one of the first learners' English corpus in China, covering about 1 million words of English compositions written by Chinese learners of English with different levels of language proficiency, with their compositions named by ST2, ST3, ST4, ST5 and ST6 respectively. This study only concerns ST2, the free English compositions without restriction of time or resource, whose token is 208088.

LOCNESS (Louvain Corpus of Native English Essays) is a corpus of native English essays with a size of 324,304. The data used here are chosen from British pupils' A level essays extracted from LOCNESS, with a token of 60,209.

Almost all the essays adopted here are argumentations or expositions which are the writer's idea, values, arguments and so on. The writers are all senior students and most of them almost live in the same time. For the purpose of making comparisons, we use the standardized

frequency (occurrences per 1,000,000 words) throughout the whole study.

TABLE I. BASIC INFORMATION OF TWO CORPORA ADOPTED

Corpora		Tokens(per1,000,000 words)
Learner corpus	ST2	208088
Native corpus	A level	65841

Instruments

The frequency of modal sequences are concordanced by Antconc and the results are processed in SPSS to analyze the keyness of the target data. CLAWS is also used here to analyze the syntactic and semantic characteristics of modal sequences.

Antconc Tools is a freeware, multi-platform tool for carrying out corpus linguistics research. Here it is used to calculate the strength of keyness of modal sequences in ST2 and A Level. And by comparing the keyness of the modal sequences in two corpora, this study is to examine whether Chinese senior school students have a tendency to overuse or underuse some of the modal sequences.

Powergrep is a powerful Windows grep tool. In this research, the action type of Powergrep is set as “search and replace” with two regular expressions. The whole procedure of text processing is exemplified as said by Aarts & Granger(1998) and To-no(1999). After annotation, we can realize what we want with Powergrep. For example, there is an annotated sentence like this: “I_PPSI1 can_VM read_VVI_”. We can use “replace and substitution” method: first using “(\S+)_VM” to replace “\$1”, then we can get “I_PPSI1 can read_VVI_”; and then by using “(\S+)_(\S+)” to Replace “\$1”, we can get “The PPSI1 can VVI_”. Results shows like this.

TABLE II. AN EXAMPLE OF THE TEXT PROCESSING

Raw text	We should improve it.	It should be improved.
POS tagging	We_PPSI2 should_VM improve_VVI it_PPH1 ..	It_PPH1 should_VM be_VBI improved_VVN ..
Replaced	PPIS2 should VVI PPH1.	PPH1 should VBI VVN

The author uses Chi-Square Test in SPSS to examine all frequency differences (naming Keyness) between the NS and the NNS corpora with 95% as the critical level of confidence (P value $< .05$). Here, Keyness, linking two words or phrases usually assumed to be within a given span of each other, is a term used to describe the quality a word or phrase has of being “key” in its context. The set of keywords found in a given text share keyness, and they are co-key. Keyness can be reached through Chi-square in statistics by SPSS and the Log Likelihood Test (Rayson et al. 2004, Oakes 1998). As is shown by the equation of Chi-square (omitted here), keyness can be plus or minus, indicating over- and -under-use tendency comparing with the natives. And the larger the keyness,

the striking the difference.

C. Data Analysis Procedures

Keywords Analysis

Keywords are those whose frequency are unusually high in comparison with some norms. When keyness is very high, and “P” (the probability of the keyness being accidental) is very low, the word can be fairly safely called a keyword. This can be reached by Antconc, using the keyword button.

Clusters Analysis

Modal sequences can be obtained by all the overused and underused clusters, which are also called the positive and negative keywords clusters by building two-to-six-word cluster lists separately about the two corpora and then by comparing the keyness of those cluster lists. All can be done by using the cluster button of Antconc.

Arrangements of keyword clusters

First observing and then counting all the modal sequences in these two corpora, we can make arrangement of the keywords cluster lists by picking out the keywords clusters which starts and ends with some modal verbs separately and then by grouping them into different categories according to the varied forms of modal verbs, etc.

III RESULTS AND DISCUSSION

A. The Overuse of Modal Sequences

The following table shows the search results of the top 8 overused modal sequences.

TABLE III. THE OVERUSED MODAL SEQUENCES (TOP 8)

Modal sequence s	Examples	NNS		NS		Keyness	P
		Freq	%	Freq	%		
Will VVI	Will find	571	2.91	96	1.66	11.41	0.00
Can VVI	Can do	495	2.53	112	1.94	0.49	0.00
Must VVI	Must eat	384	1.96	20	0.35	54.98	0.00
PPIS must	I/We must	342	1.75	7	0.12	68.53	0.00
PPIS will	I/We will	325	1.66	2	0.03	74.80	0.00
NN will	Boy(s) will	232	1.18	128	2.22	58.48	0.00
PPIS can	I/We can	218	1.11	22	0.38	16.53	0.00
Will VBI	Will be	163	0.83	61	1.06	8.35	0.00

These modal sequences in the above table could be grouped into two main categories:

(1) “Subject + modal verbs” sequences: they could be generalized as “pronouns of singular and plural first person + modal verb” such as “I/we + can/will/ must”.

(2) “Modal verbs + verb” sequences: they could be divided into two kinds: A. “modal verbs (will, can, must) + verb with unmarked tense, aspect and voice (e.g. ‘modal verb + do’, not ‘modal verb + have done’)” ; B. “modal verb (will) + be” which could be further generalized as “will/can/must + do” and “will + be”.

The above results show that Chinese senior school students use more agents (especially first person pronouns

we and I) as subjects before modal verbs than the natives in most circumstances. What they talk about is mostly about responsibility and obligation of themselves or the student group they belong to. Their habits of language use make them less objective, which will greatly sabotage the arguing effects of the essays. the fact that almost all learners tending to use the simplest, most colloquial sequences like “personal pronouns+ modal verb” and “modal verb+ verb with unmarked tense and aspect” to avoid making grammatical mistakes is probably due to the fact that learners prefer to use the modal expressions which they acquired early and have had a good grasp of.

B. The Underuse of Modal Sequences

There are also some other modal sequences that are rarely used by Chinese senior school students. The top 8 modal sequences with negative keyness in learner corpus are listed as in the following table:

TABLE IV. THE UNDERUSED MODAL SEQUENCES IN NNS (TOP 8)

Modal sequences	Exam ples	NNS		NS		Keyness	P
		Freq	%o	Fre	%o		
NN would	book would	79	0.40	176	3.05	-396.39	0.00
Would VVI	would do	161	0.82	161	2.79	-189.57	0.00
Would VBI	would be	72	0.37	132	2.29	-264.14	0.00
NN should	desk would	36	0.18	109	1.89	-284.86	0.00
Should VBI	should be	21	0.11	101	1.51	-309.79	0.00
NN could	people could	51	0.26	87	1.51	-165.71	0.00
Could VBI	could be	7	0.04	75	1.30	-269.26	0.00
PPH would	they should	19	0.10	55	0.95	-140.93	0.00

These modal sequences reflect the use features which are often seen in native speakers' written English but hardly observed in Chinese senior school students' English compositions. They are modal sequences consisting of could, should and would (especially would).

And one of the obvious features is their seldom- used passive voice. Chinese students massively use modal verbs in active structures in which the agents play the role as subjects, resulting in pragmatic inappropriateness.

They seem to feel great responsibility when talking about problems. All the reasons are analyzed as follows:

(1) Learners tend to employ the “safe” application of language items. They choose to express with familiar words, phrases and sentence patterns. And so, they prefer to pick those simplest and easiest ones just to avoid making any grammatical mistakes. In that case, it becomes more likely and secure for them to get less minus marks.

(2) The teaching methods of modals leading to the overuse and underuse of modal sequences. Teachers may just give the students Chinese translation of modal verbs. However, most modal verbs are actually not

equivalent to their corresponding words in Chinese. Also teachers have the tendency to only instruct learners with the universal use of modals. Consequently, learners' frequently practicing these expressions cause the monotony in their English compositions.

(3) The cultural difference between China and English-speaking countries also contributes to the significantly divergent use of modal sequences. Native speakers put more emphasis on the objective description of any events while Chinese learners are more inclined to posture as masters with strong awareness of responsibility and obligation.

C. Major Findings

The characteristics of modal sequences in Chinese senior school students' compositions are as follows:

Modal sequences led by can, will and must are overused while those constructed by could, would and should are underused; learners tend to overuse personal pronouns as subjects before modal verbs and overuse verbs with unmarked aspects and voices after. Consequently, they overuse deontic modality and underuse epistemic modality.

IV. CONCLUSION

This research is about the use features of modal sequences among Chinese senior school students by way of corpus analysis. The sub-corpus ST2 in CLEC and the sub-corpus A level in LOCNEE are chosen as learner corpus or reference corpus. With the assistance of various software tools, the overused and underused pattern of modal sequences have revealed. The possible reasons are illustrated. Of course, this study has some pedagogical implications and suggestions for future researches (omitted here because of space).

REFERENCES

- [1]Aarts, J. & S. Granger. “Tag sequences in learner corpora: A key to interlanguage grammar and discourse,” In S. Granger, eds. London: Longman, 1998, pp.132-142.
- [2]Aijmer, K. “Modality in advanced Swedish learners' written interlanguage,” In S. Granger, J.H.& S. Petch-Tyson, eds. Amsterdam/ Philadelphia: John Benjamin publishing company, 2002, pp.55- 76.
- [3]Biber, Douglas et al. Longman grammar of spoken and written English. London: Longman, 1999.
- [4]Bybee, J.&S. “Fleischman. Modality in grammar and discourse: An introductory essay,” In J. Bybee & S. Fleischman, eds. Amsterdam: John Benjamins, 1995. pp.1-14.
- [5]Coates, J. The Semantics of Modal Auxiliaries. London: Croom Helm, 1983.
- [6]Halliday, M. An Introduction to Functional Grammar. London: Arnold, 1985.
- [7]Hunston, S Corpora in Applied Linguistics. Beijing: World Publishing Corporation /Cambridge University Press. .2002
- [8]Hunston, S. ““We can broke the forest”: Approaches to modal auxiliaries in learner corpora”. Paper presented at TaLC6, Granada, July 2004.
- [9]Kratzer, A. “The notional category of modality”. In H. Eikemeyer & H. Rieser eds.. Berlin: Mouton de Gruyter. 1981, pp.38 -74.
- [10]LIANG Maocheng, A corpus-based study of modal sequences in

Chinese tertiary EFL learners' written production[J],Foreign Language Teaching and Research (bimonthly)(1), 2008, pp.51-58.

[11]LONG Shaoyun, "A study of modal sequences in English major's argumentations",China English education,[EB-OL]<http://www.sinoss.net/qikan/2012/0207/12790.html>. 2011,vol 2, Dec. pp.1-13.

[12]LONG Shaoyun, "The relationship between English major's tendency to use modal sequence and their study time at school",Foreign language education, vol 4 , Dec.2012,pp.1-12.

[13]LONG Shaoyun, "A study of modal sequences' grade features in English major's argumentations" ,China foreign language education, vol 1, Mar.2013, pp.3-14.

[14]LIU Wenyan, "A Corpus-Based Study on the Written Form of Chinese English Majors' Use of Modal Sequences",Changsha,HuNan University. 2009.

[15]Michael P.Oakes, Statistics for Corpus Linguistics. Edinburgh University Press ,1998.

[16]Mindt, D. An Empirical Grammar of the English Verb: Modal Verbs. Berlin: Cornelson, 1993.

[17]Rayson, P., Archer, D., Piao, S. L., McEnery, T. "The UCREL Semantic Analysis System,". In proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with 4thInternational Conference on Language Resources and Evaluation (LREC2004), Lisbon, Portugal, pp. 7-12, 25th May 2004.

Tono, Y. "A Corpus-based analysis of interlanguage development: Analyzing part-of-speech tag sequences of EFL learner corpora". In B. Lewandowska-To-maszczyk & P. Melia eds. PALC 99, Frankfurt: PeterLang. 1999, pp.323-340.