A Corpus-based Study of Modal Verbs in Chinese Learners’ Academic Writing

Xiaowan Yang¹

¹School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Xiaowan Yang, School of English and Education, Guangdong University of Foreign Studies, No. 2 Baiyun Avenue (North), Baiyun District, Guangzhou, 510420, China. Tel: 86-20-3932-8080. E-mail: yangxwdd@163.com

Received: October 29, 2017   Accepted: January 16, 2018   Online Published: January 19, 2018

Abstract

While more Chinese students are going abroad to pursue their further academic study, how to help them improve academic writing competence has received wide attention. Modality, as one of the complex areas of English grammar, reflects the writer’s attitude and is extremely important in academic written discourse. Therefore, it is necessary to investigate how Chinese learners of English use modal verbs. For this purpose, a learner corpus (LC) with Chinese learners’ academic writing has been compiled and compared against a professional corpus (PC) which consists of published research articles. With the help of software Antconc 3.2.4w, the use of nine core modal verbs in both corpora has been explored. Findings indicate that compared with professional writers, Chinese learners tend to use modal verbs more frequently; they also tend to overuse can, will, could and would and underuse may. Based on an analysis of the two corpora, this study proposes possible reasons that account for these differences. This study provides some insights into the use of modal verbs by Chinese learners of English and thus informs teaching of modal verbs in the English classroom and contributes to the academic writing curricula design.

Keywords: modal verbs, Chinese learners, academic writing, learner corpus, professional corpus

1. Introduction

Modality is extremely important in academic written discourse as it conveys the writer’s attitude both to the propositions he/she makes and to the readers. It is only with the successful handling of modality can research findings be expressed accurately. The ability to use modality appropriately also contributes significantly to pragmatic aspect in English writing (Hyland, 1994; Myers, 1989) and may reflect an advanced level of both linguistic and pragmatic proficiency in the written mode (Chen, 2010).

However, past research has shown that learners of English seem to have difficulty in using modal verbs appropriately and that they frequently overuse or under-represent certain modal meanings or forms (Hinkel, 1995; DeCarrico, 1986). This leads to the need for an examination of modal verbs and their use by non-native speakers in the specific genre of research articles.

Unlike most research on learner corpora in which non-native speaker corpora are compared with native speaker corpora, this study sets out to compare the use of modal verbs in academic writing produced by Chinese learners of English and professional writers (including not only native speakers but also non-native speakers). It is hoped that the discussion of the use of modal verbs and patterns by learner writers which the learner corpus approach makes possible could give an understanding of non-native speaker use of modal verbs in this genre, contribute to the academic writing curricula design and thus help improve L2 learners’ academic writing competence.

The two main research questions in this study are as follows:

1) What are the differences in the use of modal verbs in the learner and professional corpora?
2) What are the possible causes that account for these differences?

2. Literature review

2.1 Modality and Modal Verbs

The study of modality in the English language is regarded as the most persistent and fascinating area of
philosophical and linguistic inquiry (Hoye, 1997). According to Quirk, Greenbaum, Leech, and Svartvik (1985), modality is “the manner in which the meaning of a clause is qualified so as to reflect the speaker’s judgment of the likelihood of the proposition of the sentence being true” (p. 219). Modals are general statements that represent the notion of the mind or events that may or may not take place in the future and reflect the speakers’ attitude about what he/she says (Palmer, 2001)

Modality can express a wide range of semantic meanings, like obligation, necessity, permission, request and so on (Quirk et al., 1985). A variety of devices can be employed to convey modality. Lexical devices, for example, use nouns (intention, determination, hope, presumption and expectation), adjectives (certain, doubtful, likely, conceivable, possible, and sure), adverbs (hardly, perhaps, possibly, probably, and evidently) and verbs (doubt, believe, think, predict, and suggest) to express modality (Hermerén, 1978). But still, modality is most frequently expressed by modal verbs, which will be the focus of this study.

Nine modal verbs have been identified in past literature as core modals, including can, could, may, might, shall, should, will, would and must (Quirk et al., 1985; Biber, Johansson, Leech, Conrad & Finegan., 1999). According to Halliday (1976), they are distinguished from lexical verbs by grammatical properties including a lack of non-tensed forms, no person-number agreement, occurrence with a following verb in bare infinitival form, non-occurrence in imperative clauses and so on.

Other modals have also been identified. Marginal modal verbs include need to, dare to, used to and ought to (Biber et al., 1999). Quasi-modals include had better, have to, have got to, be supposed to, be going to and can co-occur with modal verbs (Collins, 2009, p. 15).

2.2 Previous Research

The huge amount of research into English modal verbs reflects the complexity of their functions. Some influential publications are concerned with linguistic analyses of modal verbs in spoken and written discourse (Coates, 1995; Huebler, 1983; Hoye, 1997; Leech, 2005; Palmer, 2001). Others are motivated by how modal verbs represent a difficult area for learners of English and focus more on the investigation of how modal verbs are used.

Hoye (1997), for example, compared native speakers of English with Spanish learners and found that the speakers consistently neglected the potential for modals and adverbs to combine and that “there are points of contrast and equivalents between the L1 and L2 which may lead to negative transfer or interference and actively impede the learner’s level of performance in the L2” (p. 251). Kwachka and Basham (1990) and Basham & Kwachka (1991) concluded that in academic writing in undergraduate courses, native Alaskan students frequently “extended the standard functions of modals to encode their own cultural values” (Basham & Kwachka, 1991, p. 44). Hinkel (2002) also showed that the choice of modal verbs in L2 writing can be dependent on culture and topic.

Researchers have also found that Asian and Swedish learners both tend to overuse modal expressions (Hykes, 2000; Aijmer, 2002) and French learners use a much higher proportion of epistemic auxiliaries than of other epistemic devices (adjectives, adverbs and nouns) (Dagneaux, 1995).

Research has also been conducted on Chinese learners of English. Milton and Hyland (1999) examined doubt and certainty in Chinese non-native speaker writers’ essays and found an inappropriate overuse of directive and authoritative assertions compared with native speaker writers. This finding was supported by other researchers including Ma and Lu (2007) and Liang (2008).

3. Method

3.1 The Corpora

In response to the need for research on modal verbs used by Chinese learners of English, this study adopts a corpus-based approach comparing data from two corpora: the Chinese learner writers corpus (referred to as the learner corpus) and the professional writers corpus (referred to as the professional corpus).

The Learner Corpus (LC)

The LC is made of academic writings by Chinese 2nd-year full-time students in the International College at a southern University in China. These students are in a degree program jointly offered by this Chinese university and a university in the UK. Upon successful completion of their first 2-years’ study in China, they will continue their BA program in International Business and Trade in the UK. All students are native speakers of Chinese.

Their reports which make up the LC are all independent research projects written by the students as part of their coursework. Each of these reports is about 6,000 words and investigates a topic relevant to International business
and trade chosen by the students. The LC contains 246 files with a total of 1,637,722 tokens. The corpus therefore allows investigation of modal verbs used in academic writing by Chinese learners of English.

**The Professional Corpus (PC)**

As was mentioned in the literature review, most research on non-native speakers’ L2 learning included a comparable native speaker corpus. However, with English being spoken more by non-native speakers than native speakers (Graddol, 2007), English has become a lingua franca and is not any more owned only by native speakers. This is particularly true in the academic circles in which writers and audiences are becoming increasingly diverse linguistically and culturally (McIntosh, Connor, & Gokpinar-Shelton, 2017). Therefore, in this study, the native speaker data are not used for comparison; instead, a professional corpus consisting of published articles written by professional researchers all over the world is created. These published research articles reflect the actual use of English language by English users (native and non-native speakers alike) in the academic field and are used as the control baseline for the learner corpus analysis.

This professional corpus consists of 206 published articles which are all taken from the Journal of Business Research (JBR). These articles are chosen out of the following considerations:

- JBR is an international journal for business-related studies and all its volumes are electronically available online. This makes data-collection easier.
- All articles comply with the standards and expectations of the worldwide readership and represent the norms of academic writing discourse in business studies.
- These articles have topics similar to those in the learner corpus.

To create the PC, articles are first downloaded from the official website of ScienceDirect. Using the “Download PDF” function as is shown in Figure 1, the first 20 articles from 11 issues of JBR (Volume 67 Issues 10-12 and Volume 68 Issues 1-7) available to the researcher are downloaded and put in a file as raw data for the PC. However, all 11 editorials from the 11 issues are decided not to fulfil the need for the analysis, as they are not research articles, and are therefore deleted. The last 3 of the rest 209 articles are deleted as well so that the professional corpus has similar number of tokens as the learner corpus. 206 articles are finally converted to text files via PDF converter software and the professional corpus is created with a total No. of 1,637,460 tokens (only 262 tokens less than the learner corpus).

![Collection of articles for professional corpus](image-url)
Table 1. A description of the learner and professional corpora

<table>
<thead>
<tr>
<th>Corpus</th>
<th>No. of files</th>
<th>No. of tokens</th>
<th>No. of word types</th>
<th>Topic areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td>246</td>
<td>1 637 722</td>
<td>40 329</td>
<td>International business and trade</td>
</tr>
<tr>
<td>PC</td>
<td>206</td>
<td>1 637 460</td>
<td>46 404</td>
<td>Business-related studies</td>
</tr>
</tbody>
</table>

3.2 Procedures of Analysis

The purpose of this study is to examine the use of modal verbs in the learner corpus by comparing it with the professional corpus. To achieve this purpose, quantitative analysis was first conducted to investigate the frequency of modal verbs to identify L2 learners' ability to use modal verbs in the academic written discourse.

Due to the constraints of time, energy and other resources, only nine modal verbs were surveyed: *can, could, may, might, shall, should, will, would* and *must*. Apart from *must*, these are usually paired as present and past tense counterparts of single lexemes (*can/could, may/might, shall/should, will/would*), although the relationships between the counterparts are complex. For present purposes it is useful to treat them as individual items, as each has its own function (Bowie, Wallis, & Aarts, 2013).

AntConc 3.2.4w designed by Anthony (2007), specifically concordance features of the program, was used to capture all of the instances from the learner and professional corpora of the nine modal verbs. Counts were made of total as well as individual modal use in both corpora and were compared in order to see the general frequencies of use and differences in these frequencies. Specific modal verbs with the greatest differences in counts – *can, will, could, would and may* – were identified and then discussed in more depth.

It is hoped that the findings of this study could complement language learning research on Chinese students and thus inform teaching of modal verbs in the English classroom and contribute to the academic writing curricula design.

4. Results

This section presents the results of the corpora analysis and shows the use of modal verbs in both corpora.

4.1 Overall Counts

Table 2 shows that all nine modal verbs in discussion are found in both corpora, with *can* used most frequently and *might, must, shall* used least frequently.

Table 2. Frequency counts of nine modal verbs in both corpora

<table>
<thead>
<tr>
<th>Modal verbs</th>
<th>The Learner Corpus</th>
<th>The Professional Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frequency</td>
<td>Ranking</td>
</tr>
<tr>
<td>can (cannot, can’t, can not)</td>
<td>8 253</td>
<td>1</td>
</tr>
<tr>
<td>will (will not, won’t)</td>
<td>4 408</td>
<td>2</td>
</tr>
<tr>
<td>could (could not, couldn’t)</td>
<td>2 621</td>
<td>3</td>
</tr>
<tr>
<td>would (would not, wouldn’t)</td>
<td>1 918</td>
<td>4</td>
</tr>
<tr>
<td>may (may not)</td>
<td>1 347</td>
<td>5</td>
</tr>
<tr>
<td>should (should not, shouldn’t)</td>
<td>1 263</td>
<td>6</td>
</tr>
<tr>
<td>might (might not)</td>
<td>522</td>
<td>7</td>
</tr>
<tr>
<td>must (must not, mustn’t)</td>
<td>486</td>
<td>8</td>
</tr>
<tr>
<td>shall (shall not, shan’t)</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Total count</td>
<td>20 829</td>
<td></td>
</tr>
</tbody>
</table>

However, the learner writers and professional writers show a remarkable difference in the total frequency of modal verbs they use. Modal verbs appear in the learner and professional corpora 20 829 and 10 504 times respectively, which means that the learner writers employ modal verbs almost twice more than the professional writers. Besides, each of these nine modal verbs is used more frequently in the learner corpus than in the
professional corpus, with only *may* as an exception.

### 4.2 Counts of Individual Modal Verbs

While there are differences in the frequency of occurrence in all of the nine modal verbs, the modals *can, will, could, would* and *may* show the greatest difference in the corpora and thus account for a large part of the difference in counts between the two corpora.

From Table 2, we can also find out that while *can, will, would* and *could* are examples of the greatest overuse of modals, *may* is the only modal verb which is consistently underused by the learners. This finding can be further complemented by Figure 2, which shows a comparison of the percentage of modal verbs *can, will, could, would* and *may* between the two corpora. Four modal verbs *can, will, could* and *would* occupy the majority of modal verbs used by the learner writers (83%) while they only take up half of modal verbs used by the professional writers (56%). The results demonstrate that the learner writers use these four modal verbs much more frequently than the professional writers. Figure 2 also shows a percentage of 6% of the use of *may* in the LC in contrast with 25% in the PC. This result indicates that professional writers tend to use *may* more frequently than the learner writers.

![Figure 2. Percentage of modal verbs can, will, could, would and may in both corpora](image)

### 5. Discussion

In the present study, the comparison of the two corpora demonstrates several differences including the total frequency of modal verbs employed, the overuse of *can, will, could* and *would* and underuse of *may* in the learner corpus. In this part, reasons for these differences between the learner writers and professional writers will be analyzed.

#### 5.1 Overuse of Modal Verbs

From the results, we can see that Chinese learner writers use modal verbs twice more frequent than professional writers (20,829 VS 10,504 counts). This striking difference coincides with the findings in other studies. In an investigation of the Swedish learner corpora, Aijmer (2002) has identified an overuse of the categories of modal expressions investigated. Ma and Liu (2007) and Liang (2008) also find a higher total frequency of modal verbs in Chinese learner corpora than in native-speaker corpora.

According to Biber et al. (1999), modal verbs are more frequently used in spoken rather than written texts. It is thus possible that the more frequent use of modal verbs indicates a stronger tendency among Chinese learners to transfer conversational uses of modal verbs to academic genres (Wen, Ding, & Wang, 2003, pp. 268-274), which is supported by findings by Hyland and Milton (1997, p. 192) in which they suggest that some learner writers cannot distinguish between informal spoken and academic written forms.

Another possible reason is related to how English words are taught to students in China. In the course books in...
China, words are presented to the students with their Chinese translation rather than in the context in which they are used. As a result, students’ understanding of certain modal verbs may be quite different from the actual meaning of them in their real use and thus leads to the misuse of these words. Take should as an example:

[Business owners] should invite an interpreter to help if they are not confident with their own language competence. (File 12, LC)

Enterprises not only should be familiar with multinational, cross-regional customer needs of the market ... (File 38, LC)

In English should is usually used to indicate (weak) obligation or necessity. However, it is translated as 应该 in Chinese course books of English, which can indicate either obligation or suggestion/recommendation. In the above two examples, should is used to express suggestion or persuasion rather than obligation or necessity and can be replaced by “it is suggested that business owners invite …” and “it is important that enterprises are …” respectively. The misuse of should may make the writers sound bossy and rude and may be offensive to some people.

It is also the case in classroom teaching in China that more focus is put on the accuracy of forms rather than appropriateness in pragmatics. As modal verbs do not share some grammatical properties of other verbs, like tensed forms and person-number agreement, students may find them easy and safe to use and thus have a tendency to overuse them in their writing.

Limited linguistic resources also contribute to the lack of other means to express modality by learners. As is mentioned in Literature Review, modality can be conveyed not only through modal verbs, but through the use of nouns, adjectives, adverbs and other verbs as well. However, although these words are taught to the students, their use as modality devices are seldom introduced and made clear in the language teaching classrooms. Therefore, learners may need to rely mainly on modal verbs to convey modality.

5.2 Overuse of Can, Will, Could and Would

Can and will are both found to be overused in a number of studies, including Aijmer (2002), Ma and Liu (2007) and Liang (2008).

can

Three meanings of can can be easily identified in both corpora, indicating ability, logical possibility or permission.

1. Ability

... claim that t-knowledge can reduce the duplication of services and overhead costs of providing them. (File 121, PC)

The ideal Chinese import and export enterprises can be easily found in ... (File 149, LC)

2. Logical possibility

... joining an organization with similar values can be important for some people. (File 181, PC)

... online payments can be helpful when banks and credit card companies participate in ... (File 47, LC)

3. Legitimacy

... 35% to 50% of the variation in IPO initial day returns can be explained by publicly available information. (File 184, PC)

Thus, it can be suggested that managers in hospitality organisations need to maintain a good level of communication with their staff in order to find out what they expect from them and from the role; when this expectation is found out, managers can maintain a good relationship between themselves and the staff. (File 2, LC)

However, Biber et al state that “can is especially ambiguous in academic prose, since it can often be interpreted as marking logical possibility or ability” (1999, p. 492). This is particularly true in the learner corpus as besides the above three functions, can is sometimes used with ambiguity between legitimacy and ability or as a chunk. For example:

From the above result, we can see that a sound business plan may contribute hugely to the long-term development of a company. (File 201, LC)

In LC, can is also used frequently as the substitute for may. This accounts for the overuse of can and underuse of
may and will be discussed later in this research.

**will, could and would**

Previous research suggests that learners tend to use modal verbs that are first taught to them (Ma & Liu, 2007), which may help explain the overuse of will as it, together with can, is first modal verb taught to in Junior high school English classes in China and may thus be preferred by Chinese learners. However, overuse of could and would is quite confusing and it contradicts with findings by Ma and Liu (2007) and Liang (2008) with Chinese learners in which these two modal verbs are underused. A closer investigation into the concordences suggests that although could and would can be used to realize interpersonal metafunction (Halliday, 1985) and express more tentative and polite tones (Biber et al., 1999; Quirk et al., 1985), in the LC they are most frequently used as past forms of can and will and thus correlate with the overuse of these two modal verbs.

5.3 Underuse of may

This finding is in accordance with the results found in Milton and Hyland’s (1999) study that the Chinese learner writers use firmer assertions and more authoritative tones in argumentative writing compared with native English writers. It also coincides with findings of Aijmer (2002) and Chen (2010) in which Swedish and Chinese learner writers fail to use may properly in their academic writing. However, it contradicts with the findings in which may is found to be overused by Hong Kong learners of English (Hyland & Milton, 1997, p. 189) and properly used by Chinese learners (Liang, 2008).

The underuse of may may be partly due to the variety of substitutes students have for it, for example, can, possible/possibility (214 times), probable/probability (110 times) and so on. As this study is only about modal verbs, I will only focus on the substitute can.

The overuse of can may in a way help explain the underuse of may as it frequently functions in the LC to propose uncertainty and hedging. The following example shows that in the LC learners sometimes use can to indicate uncertainty or tentativeness.

*Despite of some defects, the research will make a contribution to the investigation of brand, the differences between men and women in the modern society and it can be useful for company owners respond positively to the marketing and their consumers.* (File 145, LC)

As epistemic use of may is most common, the underuse of it may suggest that unlike professional writers who often hesitate to make a full commitment to propositions or to readers in writing, learners do not make tentative statements to allow room for questions or doubt in their essays. Instead, they write overly positive, confident statements of fact, which is less confident or assertive, would be more appropriate.

*Psychological control may lead to the temporary external compliance of children to their parents, but will fail to help them internalize their own value system ...* (File 49, PC)

*Although contradicting with the previous studies, the findings suggested that strong ties with the local communication can lead to the success of businesses.* (File 107, LC)

In the above comparison, the use of can in the LC indicates too much certainty while the professional writer uses may to indicate less assertiveness in a similar context.

Cultural differences may have also played an important role here. The Chinese culture views certainty as a sign of strength and hedging as a sign of weakness, perhaps because certainty signals one’s assertiveness and self-confidence when presenting propositions. This cultural ideology is not only reflected in the Chinese learners’ L1 writing, but transferred to their L2 writing.

6. Conclusion

As the use of modal verbs explicitly signals the writer’s attitude to propositions and to readers, it is important in the written discourse (Hyland, 1998). Although differences in text type or discourse mode have been found to influence the frequency and use of modal verbs (Carretero, 2002), the findings here demonstrate the universality of modal verbs used in academic writing by Chinese learners of English.

The present study has identified some differences in the use of modal verbs between the learner corpus and professional corpus. The results suggest that Chinese learners of English tend to use modal verbs more than professional writers. They are also found to overuse modal verbs including can, will, would and could and underuse the modal verb may. Based on an analysis of the two corpora, this study proposes possible reasons that account for these differences.

However, it is necessary to admit the limitations of this study. First, although effort has been made to match
every aspect of the two corpora as closely as possible, they are still quite different. For example, the professional corpus includes more varied topics, more word types and has longer texts. However, it is decided that their similarity in genre and area of study is more important for this study in which an examination of the use of modal verbs is at the core. It should also be noted that some core modal verbs discussed here can function not only as modal verbs. Therefore, an accurate analysis of these words requires detailed manual examinations to isolate true modality uses from other entries. Manually tagging both corpora could also make it possible for the researcher to investigate pragmatic functions of certain modal verbs, which would be an interesting topic for future exploration. Finally, although the lack of other linguistic devices to express modality is assumed to be a major reason of the overuse of modal verbs in the LC, this study does not provide evidence for it as the core of this study is only on modal verbs. It would therefore require a combination of manual and automatic analysis before any firm statements could be made about global use of modality in learner writing.

The findings presented in this study can be especially useful to the teaching of modal verbs to Chinese learners in the EFL context. When teaching modal verbs, the teacher/course book designer can provide the context in which they are used and allow more real communication rather than controlled practice to take place among students. The focus of instruction should not only be put on accuracy and literal meanings of modal verbs but on the pragmatics in discourse as well. Devices to convey modality other than modal verbs should be introduced and made clear so that students have more linguistic resources that they can make use of. Comparison can be made between the different roles modal verbs play in research writing and more informal, subjective types of writing or speaking so that students can understand the use of modal verbs in a variety of discourse. Focus can also be put on how L1 use of modality is different from that of L2 to minimize L1 interference when learning L2.

References


language (pp. 101-135). London: Oxford University Press.


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).