Improving Modal Verb Treatment in English Textbooks Used by Chinese Learners: A Corpus-Based Approach

Lexi Xiaoduo Li

Abstract
This study demonstrates how native and learner corpora can enhance modal verb treatment in English-as-a-foreign-language (EFL) textbooks used in mainland China. Data analysis compares modal verbs in the textbook and native corpus by referring to distributional features, semantic functions and co-occurring constructions; and the analysis of the learner corpus refers to the use/misuse of modal verbs. A corpora comparison considers the authenticity of textbook language; and examination of learner use/misuse of modal verbs provides a learner perspective for textbook presentation that is applicable to learning difficulties. In this regard, this study highlights discrepancies between textbook treatment of modal verbs and authentic modal use in constructions, distributions and semantic functions, and its application of error analysis reveals inter/intra-linguistic interference in the use of modal verbs. This study can thus contribute to a corpus-informed and learner-centered design of grammar presentation in EFL textbooks that enhances the authenticity and appropriateness of textbook language for target learners.

Keywords
distributional patterns, EFL textbooks, learner corpus, modal verb, native corpus

Introduction
The semantic complexities of modal verbs have challenged semantic theory (Halliday, 1985, 1994) and descriptive grammar (Corbes, 1983; Palmer, 1979, 2001). Kennedy (2002) identifies a number of verb phrase structures that frequently co-occur with modal verbs, and highlights their variations across different genres in different usage domains. Researchers observe that these semantic and syntactic complexities can strongly challenge the second language (L2) and foreign language (FL) teaching and learning of modality (Celce-Murcia & Larsen-Freeman, 1999; Decapua, 2008; Manaf, 2007). Linguistic-specific obstacles that confront L2 learners are not overcome by textbooks, which are supposedly a primary source of modal verb input, especially in FL contexts. A lack of “fit” between the language descriptions of textbooks and real-life situations (Römer, 2004a, 2004b) means they are unreliable in presenting a faithful picture of modal verbs used in natural language.

To enhance the effectiveness of textbooks in supporting language learning, many studies (e.g., Biber & Conrad, 2010; Biber et al., 1994) suggest that native-speaker corpora can be a solution and the authenticity of textbook language (i.e., the accurate reflection of the target language use by the textbook language) can be greatly enhanced if the language presentations can be brought more in line with the corpus-based descriptions of language (Römer, 2005). So far, researchers have discussed the potential pedagogical applications of their corpus-based research, but actual implementation remains scarce (Aijmer, 2009). Only a handful of studies drew on corpus insights to inform modal verb presentations in EFL (English-as-a-foreign-language) textbooks (e.g., Nordberg, 2010; Römer, 2004a). Even though these studies brought the textbook language closer to native-speaker use, they overlooked learner needs. That is, the naturalness of textbook language does not always guarantee pedagogical appropriateness (e.g., Timmis, 2015; Widdowson, 1991). Leech (1998) provided an apt example for this point. Native speakers of English use “vagueness tags” such as “stuff like that” or “and things” more frequently than non-native speakers (p. xix). However, Leech argued that it is debatable whether such findings are suitable for materials design, as it may be not be pedagogically sound to...
teach “how to be vague in spoken English” (p. xx). To interpret corpus findings for pedagogical purposes, Granger (1996) called for the consideration of a learner angle in materials development. She stated that a combination of native and learner data can make sure teaching materials reflect both authentic native use and attested learner needs.

Up to now, only one study (Whitty, 2017) included a native corpus and a learner corpus in the investigation of EFL textbooks. Much more research is urgently needed to investigate how EFL textbooks can present linguistic features in the way native speakers use them and to take learner needs into consideration in the process at the same time. The present study is an effort to fill this gap by investigating whether the textbooks present modal verbs in the way that resembles native-speaker use and uncovering what difficulties that learners encounter in modal verb use through examining learner use and misuse of modal verbs in a learner corpus.

Regarding the investigation of modal verb presentation in EFL textbooks, studies mainly focused on their frequencies and distributions of forms and semantic functions (Nordberg, 2010; Römer, 2004a). As previously mentioned, modal verb use is complicated in the co-occurring structures (e.g., modal + be doing, modal + have done) and their variations across different genres. Therefore, the surrounding syntactic structures of modal verbs are worth research attention but largely neglected in the literature. This study, in addition to investigating modal-verb form and semantic functions, aims at filling this gap by looking into the textbook presentation of modal-verb constructions and learner difficulties in using these constructions.

The present study focuses on EFL learners of Chinese. As will be shown in the literature review, this context is severely under-researched. Also, secondary-level students are the target learners in this study because studies in the literature tended to target tertiary-level textbooks due to logistic convenience.

This study is guided by the following questions. It is expected that the answers to these questions can inform a textbook design that can both guarantee the naturalness of language and cater for learner needs.

1. Does the treatment of modal verbs in textbooks match authentic modal use in the native-speaker corpus?
2. How do Chinese secondary-school students use (distributions of occurrences, semantic functions and co-occurring structures) and how often they misuse (syntactic errors) modal verbs?

Background Literature

Modal Verbs

“Modality” refers to the speaker’s subjective attitude and opinion “towards the proposition the sentence expresses or the situation that the proposition describes” (Lyons, 1977, p. 452). Modal verbs are important exponents of English modality. This paper focuses on the nine central modal verbs (can, could, may, might, must, shall, should, will, and would), as identified by Quirk et al. (1985) and Biber et al. (1999).

In the past three decades, substantial corpus-based research has focused on the semantics of modal verbs but there is still no consensus on the framework of semantic analysis. For instance, some researchers (Collins, 2009; Palmer, 1990) adopt a tripartite framework and group modal verbs into three categories (deontic, dynamic, and epistemic) while others (Biber et al., 1999; Quirk et al., 1985) distinguish between intrinsic and extrinsic modality. This paper maintains that, in comparison with the three-way analysis (e.g., Palmer, 1990), the intrinsic/extrinsic dichotomy (Biber et al., 1999) is more simplified and, as a result, more pedagogically friendly.

In this framework, intrinsic modal meanings involve an aspect of human control over events, such as permission, obligation, and volition. Extrinsic meanings are typically concerned with human judgments related to likelihood, including possibility, necessity and prediction. This framework suggests that each modal has intrinsic and extrinsic uses. For instance, will has the intrinsic meaning of volition and the extrinsic meaning of prediction. The nine modal verbs are divided into three groups: (1) permission, possibility, and ability (can, could, may, might); (2) obligation and necessity (must, should); and (3) volition and prediction (will, would, shall). This framework is adopted in this paper as the framework for the semantic analysis of modal verbs.

Co-occurring verb-phrase is another important aspect of modal verbs that has been rarely explored in corpus linguistics applied in EFL textbook design. To my knowledge, only Orlando (2009) investigated five modal verb constructions adopted from Kennedy (2002):

Structure 1: modal alone (e.g., Can you do that? Yes, I can.)
Structure 2: modal + infinitive (e.g., You must come to my party.)
Structure 3: modal + be + past participle (e.g., He should be punished.)
Structure 4: modal + be + present participle (e.g., Mary will be giving a speech.)
Structure 5: modal + have + past participle (e.g., You must have seen it before.)

This paper also follows this framework for the analysis of co-occurring constructions of modal verbs.

Modal Verbs in EFL Textbooks

Only a handful of corpus-based/informed studies engage the treatment of modal verbs in EFL/ESL textbooks (Bouhlal et al., 2018; Khojasteh & Reinders, 2013; Khojasteh et al.,
difficulty for learners. In this regard, the existing literature does not give sufficient research attention to the investigation of learner difficulties. The present study fills this gap by conducting an error analysis of modal verb use in learner productions and using the findings to inform a textbook design that caters for learner needs.

To sum up, the present study contributes to the existing literature in several ways. First, this study identifies a serious methodological issue in previous studies. The studies were not conducted in the way that they could be built on each other due to the absence of a common framework of semantic analysis of modal verbs and the lack of actual investigations of an up-to-date native-speaker corpus. Second, while previous research focuses on how the insights from native corpora can inform textbook design and neglects the insights from learner corpora, the present study takes the learner angle into account to make textbooks a more effective resource for learning.

Corpora and Data Analysis

Corpora

Textbook corpus. Mainland Chinese secondary students in a middle school in Guangdong province use textbooks from the book series Go For It, which consists of five books that are each divided into 12 units, where each contains “topics,” “functions,” “structures,” and target languages. The books mainly present the topics in daily conversations and they are used by Grade 7 (two books), 8 (two books), and 9 (one book) students.

The compilation of the textbook corpus has three steps. First, the PDF version of the five textbooks is obtained from an online textbook supplier. The files are then processed with ABBYY FineReader Pro (version 12.0), an optical character recognition (OCR) software. The highlighting tool in ABBYY FineReader Pro selects all pages that contain words. These are then converted into txt files and manually checked for misspellings and missing words. These txt-format files are the textbook corpus, which contains 149,917 running words.

Native speaker corpus. The main criteria for native corpus selection are size and representativeness, which are closely linked. A larger corpus is more representative and generalizable. By this standard, BNC2014 (Love et al., 2017) is among the most accessible, representative and up-to-date corpora of British English. This study selects the spoken part of the BNCS2014 (Love et al., 2017), a 11.5-million-word spoken corpus that consists of 1,251 conversations and features 672 speakers.

The rationale for choosing the spoken part of BNC2014 (Love et al., 2017) instead of the written part (or both) is that it aligns with the textbook’s spoken language style. A cursory examination of the textbook contents shows that the texts are mostly in dialog format, and the topics are concerned with

Regarding modal-verb co-occurring structures, four studies (Bouhlah et al., 2018; Orlando, 2009; Römer, 2004a; Whitty, 2017) investigated verb phrase structures that surround modal verbs. Only one study (Khojasteh & Reinders, 2013) examined learner errors at the syntactic level. In Khojasteh and Reinders (2013), the learner corpus consists of 600 essays written by Malaysian secondary school students, which they used as a basis for syntactic error investigation. To improve the textbook presentation of modal verbs on the syntactic level, it is not only necessary to examine and compare the co-occurring structures of modal verbs in textbooks and native corpus, but to look into learner use of modal-verb constructions and find out particular areas of difficulty for learners. In this regard, the existing literature does not give sufficient research attention to the investigation of learner difficulties. The present study fills this gap by conducting an error analysis of modal verb use in learner productions and using the findings to inform a textbook design that caters for learner needs.

To sum up, the present study contributes to the existing literature in several ways. First, this study identifies a serious methodological issue in previous studies. The studies were not conducted in the way that they could be built on each other due to the absence of a common framework of semantic analysis of modal verbs and the lack of actual investigations of an up-to-date native-speaker corpus. Second, while previous research focuses on how the insights from native corpora can inform textbook design and neglects the insights from learner corpora, the present study takes the learner angle into account to make textbooks a more effective resource for learning.

Corpora and Data Analysis

Corpora

Textbook corpus. Mainland Chinese secondary students in a middle school in Guangdong province use textbooks from the book series Go For It, which consists of five books that are each divided into 12 units, where each contains “topics,” “functions,” “structures,” and target languages. The books mainly present the topics in daily conversations and they are used by Grade 7 (two books), 8 (two books), and 9 (one book) students.

The compilation of the textbook corpus has three steps. First, the PDF version of the five textbooks is obtained from an online textbook supplier. The files are then processed with ABBYY FineReader Pro (version 12.0), an optical character recognition (OCR) software. The highlighting tool in ABBYY FineReader Pro selects all pages that contain words. These are then converted into txt files and manually checked for misspellings and missing words. These txt-format files are the textbook corpus, which contains 149,917 running words.

Native speaker corpus. The main criteria for native corpus selection are size and representativeness, which are closely linked. A larger corpus is more representative and generalizable. By this standard, BNC2014 (Love et al., 2017) is among the most accessible, representative and up-to-date corpora of British English. This study selects the spoken part of the BNCS2014 (Love et al., 2017), a 11.5-million-word spoken corpus that consists of 1,251 conversations and features 672 speakers.

The rationale for choosing the spoken part of BNC2014 (Love et al., 2017) instead of the written part (or both) is that it aligns with the textbook’s spoken language style. A cursory examination of the textbook contents shows that the texts are mostly in dialog format, and the topics are concerned with
Table 1. Multidimensional Analysis (Tribble, 1999).

| Frequency Ranking (from high to low) | Textbooks | BNCS2014 |
|--------------------------------------|-----------|----------|
| 1                                    | The       | I        |
| 2                                    | To        | it       |
| 3                                    | A         | you      |
| 4                                    | I         | the      |
| 5                                    | you       | is       |
| 6                                    | and       | and      |
| 7                                    | in        | yeah     |
| 8                                    | is        | that     |
| 9                                    | it        | a        |
| 10                                   | of        | ?        |

daily-life situations such as “talking about abilities” or “talking about routines.” There is indeed a small part (about 10%) of texts dealing with writing skills (mostly about how to write emails).

Further to ensure the conversational style of the textbook language, the present study includes a “low-effort” Multidimensional Analysis (MDA) focusing on textbook language based on that of Tribble (1999); in addition, see Scott, 2001 for more details of MDA. MDA is useful in revealing whether a text’s language style is predominantly spoken or written. In the main, the method employed was to compare the top 10 most frequently used words with those in the spoken component of BNCS2014 (Love et al., 2017). Table 1 presents the comparison which indicates shows that the textbook language possesses features that are functionally related to the spoken language style in BNCS2014.

Learner corpus. This study uses a specific learner corpus, specifically “10 thousand English Compositions of Chinese Learners, version 1.1” (TECCL) (Xue, 2015). It is POSTagged and contains 1,817,335 words of 9,865 essay samples written by Chinese EFL learners from across China’s 32 provinces and autonomous regions. It was chosen because it is representative of the Chinese learners’ written productions due to the substantial sample size. Also, the essays, from the period between 2010 and 2015, provide the most up-to-date source for the investigation of Chinese EFL learners’ English. Further, the essays cover a large number of topics and text types. The texts were written in class, as practice or timed tests, and at home. This paper considered all essay samples in the entire secondary school section of TECCL, which consist of 2,701 essays and 961,084 words. All participants were aged between 13 and 15 years of age (Grades 7–9) at the time of data collection.

WordSmith Tools and CQPWeb

This study uses WordSmith Tools (Scott, 2008), version 5.0, which has three tools of analysis (Concord, Keywords, and WordList). It uses the Concord function of WordSmith Tools to retrieve and analyze data from the textbook and learner corpora. CQPweb is a new web-based corpus analysis system that belongs to the “fourth generation” of concordancers (McEnery & Hardie, 2012). The analysis options available in CQPweb include concordancing, collocations, distribution tables and charts, frequency lists and keywords or key tags. This study uses the standard query option of the concordancing function to retrieve the concordance lines of the nine modal verbs in BNCS2014.

Design

This study involves three corpora: a textbook corpus, a native corpus and a learner corpus. The aim of the research is to draw on relevant insights from both native and learner corpora in order to inform textbook presentation of modal verbs. The paper’s research design is inspired by earlier work undertaken by Biber et al. (1994) and Granger (2002). These researchers concur that native-speaker use and learner difficulty are both important for syllabus design (see Granger, 2002, p. 23). This indicates that EFL textbooks should reflect in the learning process both authentic language use in real-life situations and learner needs. When anomalies emerge between findings in the native and the learner corpus, the present study prioritizes learner needs over findings related to native-speaker data. This is because textbooks are designed for learners. A prioritization of a learner focus in corpus-based studies is fully endorsed in the writings, among many others, of Timmis (2015), Granger (2002), Meunier (2002), and Kennedy (1998).

To assess the quality of textbook language, the native corpus is compared with the textbook corpus in relation to modal verb form, semantic functions and co-occurring constructions. The comparison demonstrates the degree to which the textbook presentations of modal verbs deviate from native-speaker use. The findings are useful to help inform textbook writers which part(s) of modal-verb presentation should be adjusted in order to bring textbook language more in line with native-speaker use.

Further, the textbook corpus is compared with the learner corpus with respect to modal verb form, semantic functions and co-occurring constructions. The rationale for this comparison is to reveal the extent of textbook influence on learner productions. The findings will provide empirical evidence related to the necessity of modifying textbook presentations of modal verbs based on learner needs. Finally, the learner corpus is used to explore learner errors. The rationale is that once textbook writers are clear about where learner difficulties lie, they can take preemptive measures to remove such obstacles when designing textbooks.

Sampling and Data Analysis

The first research question addresses the distributions of forms, semantic functions and co-occurring structures in the textbook corpus and BNCS2014. The textbook corpus
is analyzed through the concordance functions of WordSmith Tools. To find out the distribution of forms in the first research question, it is necessary to obtain the overall frequencies of each modal verb, including all the full and contracted forms in affirmative and negative contexts. Modal frequencies are obtained from the textbook corpus, and all concordance lines are manually examined to eliminate cases where modal verbs are not used as modal auxiliaries. Some sentences are used to explain the meaning of modal verbs. Once the raw frequencies of the nine modal verbs are obtained, the concordance lines of each modal verb are saved in Excel sheets for semantic and structural analysis.

Biber et al.'s (1999) semantic framework is then applied:

- “Can” and “could” are analyzed for three meanings, specifically “ability,” “permission,” and “possibility.”
- “May” and “might” are analyzed for “possibility” and “permission.”
- “Will,” “would,” and “shall” are analyzed for “volition” and “prediction.”
- “Must” and “should” for “obligation” and “necessity.”

Raw and relative frequencies are calculated for each modal meaning. Finally, the concordance lines of each modal verb are examined with reference to the five co-occurring structures identified by Kennedy (2002). As with semantic analysis, the absolute frequency and relative proportion are calculated for each structure that co-occurs with the nine modal verbs.

After applying the standard query function provided by the CQPWeb online interface, raw frequencies of the nine modal verbs (full and contracted forms in positive and negative contexts) in BNCS2014 and calculated relative proportions are obtained. Five percent of all the concordances of each modal verb are sampled and saved in Excel sheets for semantic and structural analyses. The steps for function and construction analysis resemble those of the textbook corpus.

The results of the textbook corpus are systematically compared with those of BNCS2014 and specific reference is made to forms, semantic functions and co-occurring structures, which helps to highlight potential similarities and differences between the two corpora. The ultimate purpose is to assess the authenticity of textbook language.

The second research question addresses learners’ use/misuse of modal verbs. Analysis of the learner corpus comprises two stages. The first obtains frequency information for each modal form, along with different meanings and co-occurrences. The analysis of overall frequencies, semantic functions and constructions follows the same steps as the analysis in the textbook corpus. These findings on frequencies, semantic meanings and constructions are also compared against textbook corpus findings. This is justified by usage-based theories of language, which hold that usage results in frequently occurring form-meaning mappings to become entrenched in the learner’s mind, and claim that the degree of entrenchment is proportional to the frequency of linguistic knowledge (Ellis & Ferreira-Junior, 2009). The goal is not to establish a causal relationship between the textbooks and learner output, but rather to explore similarities and differences, so as to broadly indicate the potential influence of pedagogical materials on learner use of modal verbs.

The second stage of learner data analysis involves errors. Error identification was performed manually and was limited to formal syntactic errors. Error analysis was performed with reference to Richards (1974), who classifies errors into interlingual or transfer errors caused by negative interference from mother tongue habits and attributes intra-lingual or developmental errors to overgeneralizations caused by partial exposure to the target language. Inter-lingual errors occur when learners carry their existing knowledge of their native language over to the performance in the target language. The degree of interference depends in part on similarities and differences between the first and target languages. When the linguistic principles of the mother tongue differ greatly from the acquired language, learners tend to use native language rules in the learning process (Krashen, 1981, p. 65). Intra-lingual errors, as Richards (1974, p. 174) explains, are developmental and caused by a learner’s attempt to “build up hypotheses about the English language from his [sic] limited experience of it in the classroom or textbook.” These errors reflect general characteristics of syntax learning such as faulty generalization, ignorance of rule restrictions, incomplete application of rules and the hypothesizing of false concepts.

Overgeneralization involves the creation of one deviant structure in place of two regular structures. Ignorance of rule restrictions underpins the application of rules to incorrect contexts. *I made Sara to do the dishes* is a typical case of the learner ignoring the restriction on the phrasal structure of “make.” When incompletely applying the rules category, learners may use a deviant structure that represents the degree to which they have developed the rules required to produce correct utterances. Richards (1974) illustrates this when he observes that learners may add a question word to the statement form when it can alone be used as a question. The fourth type of error is hypothesizing false concepts, which results from the faulty comprehension of rules or principles in the target language. For example, the form *is* may be understood as the marker of the present tense. This misunderstanding may cause a learner to say *he is speaks Russian*.

**Results**

**Distributions of Occurrences of Modal Verbs**

In BNCS2014, *can, would* and *will* occur most frequently, followed by *could, should, might* and *must*, whereas *shall*
and may occur least frequently. In comparison with the native corpus, in the textbook corpus the sequencing pattern changes, although can remains the most frequently used modal. Although the ranking of can matches the native use, its relative proportion is remarkably high, accounting for 41.1% of all modal verb tokens in the textbooks, which creates a huge gap in the frequency of usage between can and will, should, and could. These last three modals account for 40% of total occurrences. Would, must, may and might follow and shall is the least frequent.

Besides raw frequencies, relative frequencies of modal verbs are also worth investigating as they show modal-verb distributional features. Comparison between BNCS2014 and the textbook corpus regarding relative frequencies reveals two significant patterns (Figure 1). First, some modal verbs (can, could, should, must, may) show higher relative frequencies in the textbook corpus than they do in BNCS2014. The remaining modals (might, shall, will and would) have lower relative proportions in the pedagogical materials. The large frequency gap between the two corpora is owed to can, should, will and would. Can and should are markedly more frequent, while will and would make fewer textbook appearances.

The ranking of the nine modal verbs in the learner corpus shows that can, will, and should are most frequently used (Figure 2). There is a sudden decline in frequency between should (N = 1,484) and would (N = 418). Would is followed by must, could, may, might, and shall. When compared with the ranking order in the textbooks, the ranking of modal verbs resembles the textbook corpus, with the exception of could, which ranks fourth in the textbooks but sixth in the learner data. Comparison of the distributional pattern with reference to relative frequency shows an even more striking similarity between the learner and textbook data. Figure 3 shows that although will increases and could declines more steeply in the learner corpus than in the textbook corpus, the overall increases and declines of the relative frequencies of the nine modals align in the two corpora.

These results highlight a possible connection between textbooks as input and learner output with regard to ranking and the relative proportions of modal verbs. The potential connection is manifested in specific cases of modal use. Learners’ use of would reveals the high frequency of would’d like to (34%) in the learner corpus, and this possibly correlates with its significant weight (51.1%) in the textbook corpus.
Modal Verbs in the Learner Corpus

![Bar Chart]

Figure 2. Raw frequencies of modal verbs.

Modal Verbs in Textbook Corpus and Learner Corpus

![Line Chart]

Figure 3. Relative frequencies of modal verbs.
Table 2. Raw and Relative Frequencies of Different Meanings.

| Modal verb | BNCS2014 | Textbook corpus | Learner corpus |
|------------|----------|-----------------|----------------|
|             | Raw freq. | Rel. freq. (%)  | Raw freq. | Rel. freq. (%) | Raw freq. | Rel. freq. (%) |
| Can         | 1,919    | 67.7            | 394       | 64.9          | 2,312     | 59.7          |
|             | 497      | 17.5            | 162       | 26.7          | 1,026     | 26.5          |
|             | 318      | 11.2            | 51        | 8.4           | 310       | 8.0           |
|             | 100      | 3.5             | 0         | 0             | 225       | 5.8           |
| Could       | 962      | 86.5            | 151       | 89            | 294       | 88.3          |
|             | 110      | 9.9             | 4         | 2             | 20        | 6.0           |
|             | 35       | 3.1             | 15        | 9             | 19        | 5.7           |
|             | 5        | 0.4             | 0         | 0             | 0         | 0             |
| Will        | 1,020    | 44.2            | 59        | 24.2          | 1,186     | 41.4          |
|             | 1,279    | 55.4            | 185       | 75.8          | 1,676     | 58.6          |
|             | 11       | 0.5             | 0         | 0             | 0         | 0             |
| Would       | 1,329    | 52.5            | 66        | 73.3          | 291       | 69.6          |
|             | 1,170    | 46.2            | 24        | 26.7          | 127       | 30.4          |
|             | 33       | 1.3             | 0         | 0             | 0         | 0             |
| May         | 66       | 97.1            | 44        | 86.3          | 311       | 99.4          |
|             | 2        | 2.9             | 7         | 13.7          | 2         | 0.6           |
|             | 0        | 0               | 0         | 0             | 0         | 0             |
| Might       | 515      | 100             | 36        | 100           | 36        | 100           |
|             | 0        | 0               | 0         | 0             | 0         | 0             |
|             | 0        | 0               | 0         | 0             | 0         | 0             |
| Should      | 394      | 74.6            | 167       | 99.4          | 1,477     | 99.5          |
|             | 126      | 23.9            | 1         | 0.6           | 7         | 0.5           |
|             | 8        | 1.5             | 0         | 0             | 0         | 0             |
| Shall       | 106      | 97.2            | 4         | 100           | 22        | 91.7          |
|             | 3        | 2.8             | 0         | 0             | 2         | 8.3           |
|             | 0        | 0               | 0         | 0             | 0         | 0             |
| Must        | 36       | 14.8            | 44        | 65.7          | 366       | 92            |
|             | 203      | 83.5            | 23        | 34.3          | 32        | 8             |
|             | 4        | 1.6             | 0         | 0             | 0         | 0             |

Distributions of Semantic Functions of Modal Verbs

A comparison of the native and textbook corpus shows similarities and deviations on semantic functions (Table 2). Among the nine modal verbs, the least problematic are might and shall, as the relative frequencies of different meanings of both modal verbs in the textbook corpus resemble those in the native corpus. This is consistent with Nordberg (2010) and Romer (2004a), and possibly reflects the fact that the two modal verbs are almost monosemous in native usage.

The relative proportions of different meanings of can and could roughly resemble those in BNCS2014. But this is not the whole story. First, in the textbook data, almost all occurrences of “possibility” (could) are in question form and perform a speech act function request. In the textbook corpus, 30% of could is in the structure of “could you…” accounting for only 2% of the native corpus. This suggests that the textbook corpus overrepresented one single structure of could at the expense of its major semantic functions, such as epistemic or dynamic possibility.

In can, the problem lies with the meanings of “ability” and “permission.” The textbooks strongly prefer the meaning of “ability” (32.6%) over “permission” (9.7%), but they have comparable relative proportions (ability = 17.5%; permission = 11.2%) in the native speaker corpus.

Volition/prediction (will and would) and obligation/necessity modal verbs (must and should) are the most problematic in the textbook corpus. When compared against the native corpus, the textbooks contain 20% more “prediction” meaning of will, but 20% “volition” meaning of will (Figure 4). The pattern of would in the textbooks is the precise opposite of will, and there are 20% more occurrences of “volition” and 20% fewer occurrences of “prediction” than in BNCS2014 (Figure 5).

Although the native corpus contains more examples of should as “obligation” (74.6%) than as “necessity” (23.9%), the textbook corpus shows an overwhelming tendency to represent the meaning of “obligation” (99.4%) (Figure 6). For must, the “obligation” meaning (66%) occurs more frequently than its “necessity” counterpart (34%) in the same corpus. But in BNCS2014, the relative frequency of
"necessity" meaning (84%) is higher than its “obligation” counterpart (15%) (Figure 7).

Semantic analysis shows learners exclusively prefer one meaning over others when using *may*, *might*, *shall*, *must*, and *should* (Table 2). They specifically prefer “possibility” (*may* = 99.4%; *might* = 100%) over “permission” (*may* = 0.6%; *might* = 0) when using *may* and *might*, and favor “volition” (91.7%) over “prediction” (8.3%) when using *shall*. When applying obligation/necessity modals (*must* and *should*), they strongly prefer “obligation” (*must* = 92%; *should* = 99.5%) over “necessity” (*must* = 8%; *should* = 0.5%).

Compared with the textbook corpus, the learner data exhibit a pattern quite similar to that in the textbooks, except for *will* and *must* (Figures 8 and 9). Figure 10 shows that learners use the “prediction” (58.6%) meaning of *will* slightly more frequently than the “volition” (41.4%) meaning, whereas the textbooks present much more instances of “prediction” (75.8%) than “volition” (24.2%). In the case of *must*, the difference between “obligation” (92%) and “necessity” (8%) in the learner data is much larger than that (obligation = 65.7%; necessity = 34.3%) in the textbook corpus (Figure 11).

**Table 3** shows the raw and relative frequencies of the five structures of the nine modal verbs in the three corpora. The
Structural distributions of the textbook corpus generally differ from the native corpus, and this is especially true of the Structures 1, 2, 3, and 5. Table 2 shows that the textbooks underuse Structure 1 across the nine modals when compared with BNCS2014, and this is especially true of can, could, will, would and should. Conversely, Structure 2 is, with the exception of shall, overrepresented in the textbook series for all modal verbs. The relative proportions of Structure 3
across the nine modal verbs in the two corpora are relatively balanced (the only exception is should). The textbook corpus contains a high proportion of this structure (16%) that co-occurs with should, and this is apparent when it is compared against the native corpus (4%). A closer examination of the data reveals that one single phrase structure (should be allowed) reoccurs on 59.3% of occasions. Finally, the relative frequencies of Structure 5, as it co-occurs with could, would, may, might, and must in the textbook corpus, are consistently lower than BNCS2014.

With regard to the five verb phrase structures that co-occur with modal verbs in the learner corpus, the general patterns resemble those of textbook corpus. Structure 4 (modal with be doing) is rare in both learner and textbook corpora. Moreover, in both corpora, it is noticeable that there are fewer occurrences of Structures 3 and 5 and the almost exclusive use of Structure 2. With the exception of should, this pattern applies to all nine modal verbs. For eight modal verbs, the relative frequencies of Structure 2 consistently exceed 95%, and the proportion of Structures 3 are less than 5%. In comparison, Structure 2 that co-occurs with should accounts for 91% and Structure 3 for 8.8%. This structural pattern could be related to the relatively high use of Structure 3 (16.1%) in the textbooks.

**Learner Errors**

The learner data contain instances where modal verbs are followed by noun, pronoun, adjective, preposition, adverb, and verb in ed/en form; verb with an inflected -s; and doing. For example,

- a. The teacher must done a good job.
- b. Students should not playing all the time.
- c. Animals can happy and in the wild.

Table 4 shows adjective is the most frequent (29.5%), followed by v-ed/en (25.1%), noun (14.2%), to do (9.8%), preposition (7.7%), v-ing (4.9%), v-s (3.8%), pronoun (2.7%), modal verb (1.6%), and adv. (0.5%). Table 5 displays the percentages of errors. Error percentages suggest learners are most likely to make mistakes when using must (5.8%), may (4.8%), could (3.3%), should (2.6%), and would (2.4%). For instance,

- a. You must back home immediately.
- b. My teacher may angry very much.

In learner writings, there are four types of errors caused by L1 transfer (i.e., the error cause by one’s first language). These sentences appear to be word-for-word translations from Chinese to English. L1 interference is present at the syntactic level, as English sentences have Chinese syntactic structures. When a modal verb is followed by an adjective (Type 1), the error derives from the absence of copular be after the modal verb. This is primarily because, in Chinese, an adjective can function as a predicate without the presence of a copular verb. For example,

- The teacher must patient.

The second type displays the same error of omitting copular be but for a different reason. In the example, the preposition “with” translates as “和…在一起” (he … zai yi qi), which is a verb phrase in Chinese that does not require a linking verb. For instance,

- a. The teacher must patient.
- b. My teacher must angry very much.

The teacher must patient.

The second type displays the same error of omitting copular be but for a different reason. In the example, the preposition “with” translates as “和…在一起” (he … zai yi qi), which is a verb phrase in Chinese that does not require a linking verb. For instance,

- a. The teacher must patient.
- b. My teacher must angry very much.

The teacher must patient.

He will with you.

He will with you.

The teacher must patient.

Lao shi bi xin.

He will with you.

Lao shi bi xin.
The third error type is characterized by the presence of double modals. Even though *must* and *can* are modal verbs in English, their Chinese equivalents “一定” (*yi ding*) and “也许” (*ye xu*) are adverbs that should, grammatically speaking, come before a modal verb. The following sentence illustrates this point.

I think you *must* *can* do it.

我 认为 你 *一定* 能 做到。

With regard to the fourth type, the error is caused by a noun (*comprehension*) following *can*. This could be

| Table 3. Relative Frequencies of Structures Co-occurring With Modal Verbs. |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                      | Can %               | Could %             | Will %              | Would %             | May %               | Might %             | Should %            | Shall %             | Must %              |
| BNCS2014             | 11.9               | 11.9                | 9.8                 | 12.5                | 1.5                 | 3.3                 | 10.2                | 0                  | 2.1                 |
| TEXTBOOK             | 2.8                | 1.8                 | 1.2                 | 1.1                 | 0                   | 0                   | 1.2                 | 0                  | 0                   |
| LEARNER              | 0.6                | 1.2                 | 0.1                 | 0.5                 | 0                   | 0                   | 0                   | 0                  | 0                   |
| BNCS2014             | 87.2               | 80                  | 86.8                | 75.6                | 89.7                | 85                  | 67.8                | 0                  | 65.8                |
| TEXTBOOK             | 95.6               | 96.5                | 98.8                | 98.9                | 94                  | 97.2                | 81.5                | 0                  | 89.6                |
| LEARNER              | 95.5               | 94.0                | 95.1                | 96.2                | 97.1                | 100                 | 91                  | 100                | 98.2                |
| BNCS2014             | 0.8                | 0.5                 | 1.1                 | 1.2                 | 0                   | 1.9                 | 4.0                 | 0                  | 2.5                 |
| TEXTBOOK             | 1.6                | 0.6                 | 0                   | 0                   | 6                   | 2.8                 | 16.1                | 0                  | 7.5                 |
| LEARNER              | 3.9                | 4.8                 | 3.8                 | 0                   | 2.6                 | 0                   | 8.8                 | 0                  | 1.3                 |
| BNCS2014             | 0                  | 0.8                 | 1.8                 | 1.6                 | 0                   | 1.7                 | 5.1                 | 0                  | 1.6                 |
| TEXTBOOK             | 0                  | 0                   | 0                   | 0                   | 0                   | 0                   | 1.2                 | 0                  | 0                   |
| LEARNER              | 0                  | 0                   | 0.8                 | 0                   | 0.3                 | 0                   | 0.2                 | 0                  | 0                   |
| BNCS2014             | 0                  | 6.9                 | 0.6                 | 9.0                 | 8.8                 | 8.0                 | 12.9                | 0                  | 28                  |
| TEXTBOOK             | 0                  | 1.2                 | 0                   | 0                   | 0                   | 0                   | 0                   | 0                  | 3                   |
| LEARNER              | 0                  | 0                   | 0.2                 | 0                   | 0                   | 0.1                 | 0                   | 0.5                |

| Table 4. Total Counts of Each Category Following Modal Verbs. |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                      | Can                 | Could               | May                 | Might               | Must                |
| Adj.                | 10                  | 1                   | 2                   | 0                   | 7                   |
| V-ed/en             | 11                  | 10                  | 4                   | 0                   | 3                   |
| Noun                | 10                  | 0                   | 2                   | 0                   | 4                   |
| To do               | 6                   | 0                   | 0                   | 0                   | 4                   |
| Prep.               | 7                   | 0                   | 1                   | 0                   | 1                   |
| V-ing                | 1                   | 0                   | 1                   | 0                   | 1                   |
| V-s                  | 3                   | 0                   | 3                   | 0                   | 1                   |
| Pron.               | 3                   | 0                   | 0                   | 0                   | 1                   |
| Modal               | 0                   | 0                   | 2                   | 0                   | 1                   |
| Adv.                | 1                   | 0                   | 0                   | 0                   | 1                   |
| Total               | 52                  | 11                  | 15                  | 0                   | 23                  |

| Table 5. Percentages of Errors of the Nine Modal Verbs. |
|---------------------|---------------------|---------------------|
| Modal verb          | Total frequency     | Total error frequency |
| **Must**            | 398                 | 23                  | 5.80                |
| **May**             | 313                 | 15                  | 4.80                |
| **Could**           | 333                 | 11                  | 3.30                |
| **Should**          | 1,484               | 38                  | 2.60                |
| **Would**           | 418                 | 10                  | 2.40                |
| **Can**             | 3,873               | 52                  | 1.30                |
| **Will**            | 2,862               | 34                  | 1.20                |
| **Shall**           | 24                  | 0                   | 0.00                |
| **Might**           | 36                  | 0                   | 0.00                |

The third error type is characterized by the presence of double modals. Even though *must* and *can* are modal verbs in English, their Chinese equivalents “一定” (*yi ding*) and “也许” (*ye xu*) are adverbs that should, grammatically speaking, come before a modal verb. The following sentence illustrates this point.

I think you *must* *can* do it.

我 认为 你 *一定* 能 做到。

With regard to the fourth type, the error is caused by a noun (*comprehension*) following *can*. This could be
explained by the absence of lexical morphology in Chinese. A Chinese word can function as a noun or a verb in a sentence without undergoing morphological change. The lack of derivational morphology in learner’s L1 could explain why learners are insensitive to the morphological change in L2 (from a verb to a noun or vice-versa). This can be exemplified by the following sentence.

I can comprehension.

我能理解。

wo neng li jie.

Intralingual interference might explain errors when a modal verb is followed by doing, to do, a verb with a past-tense marker or a verb with a third person singular inflected -s. For example,

(1) You could lost original thinking’s capacity.
(2) He can talks with the students patient.
(3) You can not to think how to go there.
(4) You can better learning Chinese.

The errors in these four sentences belong to the third type of errors identified by Richards (1974), specifically incomplete application of rules. For a learner, lost, talks, to think, and learning can be related to a verb in one way or another, and this suggests that they are aware of the rule that a verb should be used after a modal verb. But this rule is only partially applied because the “verb” does not take the form of a bare infinitive.

Both inter- and intralinguistic factors affect the use of each modal verb. But the dominance of a factor shows considerable variation (Table 6). The use of will and must is likely to be influenced by learners’ native language, and errors related to the use of may, could, and would are more likely to be due to developmental or intralinguistic factors. Should and can seem to be equally susceptible to inter and intralinguistic influence.

**Discussion**

The first research question aims at capturing the quality of the representation of the use of modals by native speakers of English in the selected textbooks? The above results on the comparison between the textbook language and authentic language use points to a large number of discrepancies in terms of ranking order, distributional features of forms, semantic functions and co-occurring constructions of modal verbs. The comparison between the pedagogical corpus and the native-speaker corpus on the above-mentioned aspects suggests that the teaching materials overemphasize some minor and less frequent aspects of modal verb use in the native-speaker corpus, while they underrepresent the features that are important and more frequently occur in the target language.

Based on the comparison, the answer to the first research question (Do the textbooks present modal verbs in the same way as native speakers use them in natural discourse in terms of forms and form-function connections?) is definitely no. The EFL textbook series fail to present modal verbs in accordance with actual language use This is consistent with the findings of previous studies assessing the authenticity of EFL/ESL pedagogical materials (e.g., Bouhlal et al., 2018; Nordberg, 2010; Orlando, 2009; Romer, 2004b).

The answer to the first research question illustrates the contribution of native-speaker corpora to ELT materials—namely that the frequency information drawn from a native corpus can be used to assess the textbooks’ authenticity. But frequency is only one criterion in ELT materials design. Textbook writers also need to consider the degree of difficulty that the target feature creates for a specific group of learners, and this leads to the second research question, which addresses learner needs by analyzing their use and misuse of modal verbs. In addition to learner use and misuse, the following section will also discuss the potential influence of textbooks on learner productions, with the purpose of linking learner needs with textbook design. A comment should be made at this point regarding the following section. That is, to the best of the author’s knowledge, this paper is the first study that compares the findings in a learner corpus with those in a textbook corpus. Consequently, there is no previous research with which to compare the present study. The discussion in the next section aims primarily to provide plausible explanations for the findings.

With respect to modal verb ranking, the textbook series and learner corpus are almost identical. Regarding relative
proportions, the general pattern is that the learner corpus matches the textbook corpus, and the only exceptions are that will has a heavier weight and could has a lower proportion in the learner corpus. There are two plausible reasons for learners’ relative overuse of will. The first is the positive transfer of the two meanings of will (volition and prediction) from Chinese equivalents (愿意/会去做 yuan yi/ hui qu zuo and 会/将会 hui/ jiang hui). Second, will does not pose any syntactic difficulty, as it does not frequently co-occur with complicated structures such as past participle (have done). This is further confirmed by the low error rate of will.

There are two possible interpretations for the underuse of could. One is that the explicit explanations of the modal meanings in the textbooks cannot clarify the complicated use of the three different meanings of could for learners. In the syllabus, although “permission” meaning is mentioned, the other two semantic functions (ability and possibility) of could are not addressed. In addition, could is frequently used with Structure 5, which is both formally and semantically complicated. The structure involves a present perfect aspect and conveys the speaker’s attitude toward a past event. This could be one of the reasons why learners avoid using could, and error analysis supports this explanation. Could is also among the most difficult modal verbs and the use of the ed/en form of a verb after could is a major source of errors (e.g., We could done so much with the time).

Regarding the distribution of the semantic functions of the nine modal verbs, learners use different meanings that resemble their native language, while errors related to the use of could, may, and would are more likely to be associated with developmental or intra-linguistic factors. Finally, can and should appear to be equally susceptible to inter- and intra-linguistic influence. These findings are key for the textbook series to ameliorate the presentations of modal verbs in the way that is appropriate for the target learners.

Taken together, the findings in the present study have several pedagogical implications. The first involves the relative weight given to individual modal verbs. Because must, may, could, should and would are the five modal verbs that are more likely to cause problems for learners, they should be given more space than less problematic counterparts. Second, regarding semantic functions, adjustments should be made for different meanings of will, would, must, and should. The two meanings (“volition” and “prediction”) of will and would should receive equal amounts of attention in the textbooks.

With regard to co-occurring structures, pedagogical materials need to decrease the proportion of Structure 2 (modal with bare infinitive) and increase the frequency of Structure 5 (modal with have done). An acknowledgment of the complexity of Structure 5 should be accompanied by explicit explanations that would help learners to comprehend the hypothetical meaning that this structure conveys, and this should be particularly emphasized when the structure is used with must, may, could, should, and would.

Textbooks also need to emphasize the avoidance of the random lexicalization of structures. Stranks (2013) notes that the authors of ELT materials do not sufficiently exemplify grammatical structures that use lexis that habitually co-occurs with the structures. This is also the case with the textbooks, and the unusually high proportion of would’d like to and should be allowed are cases in point.

The final suggestion on modal verb treatment is more qualitative. The textbook series should explicitly indicate L1 influence on the use of modal verbs for Chinese learners. In contrast to Chinese, adjectives cannot be used directly after modal verbs, and a copular be should come before an adjective. In addition, the canonical structure modal with bare infinitive accounts for over 90% of all five structures, but learners still tend to use erroneous forms (e.g., to do or doing) when a bare infinitive is required, which suggests that this structure might not be salient for Chinese learners. By implication, it will be necessary for the textbooks to highlight that modal verbs should not be followed by to do, doing, ed/en form of a verb or a verb with an inflected -s.
Conclusion

The overall aim of the current paper is to find ways to improve the presentation of modal verbs in a series of textbooks used by EFL secondary school students in China. This is done by drawing on insights from both a native corpus to assess the authenticity of textbook language and a learner corpus to find out learner needs in the use of modal verbs. The paper’s findings indicate a troubling deviation of textbook portrayal of modal verbs from native-speaker use of these verbs. Examining learner use and misuse indicates a clear picture of learner needs in the use of modal verbs (i.e., what is easy and what is difficult for learners in terms of modal-verb use). Based on these research findings, textbook writers should be able to design EFL textbook language (particularly for modal verbs) that more faithfully reflect both native-speaker use and target learner needs. It is hoped that the synergy adopted in this paper would help produce enhanced EFL teaching materials with more accurate linguistic description and thus more appropriately useful for learners. In addition, the research findings of the paper contribute more widely to research on learner corpora through an empirical demonstration of the role of learner corpora in syllabus design.

Future research should refine and advance the findings of this paper by pursuing different pathways and approaches. First, the current research used the spoken part of BNC2014 (Love et al., 2017) to assess the authenticity of the textbook language. This was because the written part was not available at the time this study was conducted. Although this paper argues that the spoken corpus fits the general language style of the textbooks, there are still some written texts involved in the textbook series. Future studies might usefully include both spoken and written parts of a native corpus in order to provide a more comprehensive assessment of textbook language. Moreover, researchers are advised to select the same framework of semantic analysis of linguistic items and make actual investigation of native-speaker corpora so that the research on corpus-based textbook design can be compared with previous literature and advance by building on each other. Second, due to the limits on scope, this paper concentrated only on the nine central modal verbs. Future research can focus on semi-modal verbs and other exponents of modality (e.g., adjectives, adverbs or structures of modal verbs). Moreover, time and scope limitations have prevented this paper from investigating the distributions of the four forms of modal verb (i.e., affirmative, negative, contracted, full). Further research could help learners ascertain the appropriate context for the use of modal verbs in their affirmative, negative or contracted forms. At the syntactic level, future research may explore a wider range of surrounding contexts such as pronouns that come before modal verbs or clause structures. In addition, the examination of learner use and misuse of modal verbs focused on a corpus of Chinese learners of English. It is hoped that future studies can extend to other EFL/ESL contexts. Also, a comparable L2 corpus of native secondary learners becomes available will provide a fuller picture of learner use and misuse of modal verbs.

In spite of the fact that the present study is limited in some ways, the findings point to a more effective way of improving the presentation of linguistic items in EFL textbooks. This paper contributes to the overall effort of advancing the field of EFL teaching and learning by harnessing the synergy of native and learner corpora.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Lexi Xiaoduo Li https://orcid.org/0000-0003-3427-7991

References

Aijmer, K. (2009). Corpora and language teaching. John Benjamins Pub. Co.
Biber, D., & Conrad, S. (2010). Corpus linguistics and grammar teaching. https://www.researchgate.net/publication/265403434_Corpus_Linguistics_and_Grammar_Teaching
Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. Applied Linguistics, 15(2), 169–189.
Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. Longman.
Bouhkal, F., Horst, M., & Martini, J. (2018). Modality in ESL textbooks: Insights from a contrastive corpus-based analysis. The Canadian Modern Language Review, 74(2), 227–252.
Celce-Murcia, M., & Larsen-Freeman, D. (1999). The grammar book: An ESL/EFL teacher’s course (2nd ed.). Heinle&Heinle.
Coates, J. (1983). The semantics of the modal auxiliaries. Croom Helm.
Collins, P. (2009). Modals and quasi-modals in English. Rodopi.
Decapua, A. (2008). Grammar for teachers: A guide to American English for native and non-native speakers. Springer.
Ellis, N. C., & Ferreira-Junior, F. (2009). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. Annual Review of Cognitive Linguistics, 7(1), 188–221.
Granger, S. (1996). Learner English around the world. In S. Greenbaum (Ed.), Comparing English Worldwide (pp. 13–24). Clarendon Press.
Granger, S. (2002). A bird’s-eye view of learner corpus research. In S. Granger, J. Hung, & S. Tyson (Eds.), Computer learner corpora, second language acquisition, and foreign language teaching (pp. 3–33). John Benjamins Pub.
Halliday, M. A. K. (1985). An introduction to functional grammar. Edward Arnold.
Halliday, M. A. K. (1994). An introduction to functional grammar (2nd ed.). Arnold.
Hinkel, E. (1995). The use of modal verbs as a reflection of cultural values. TESOL Quarterly, 2, 325–343.
Kennedy, G. (1998). An introduction to corpus linguistics. Longman Publishing.
Kennedy, G. (2002). Variation in the distribution of modal verbs in the British National Corpus. In R. Reppen, S. Fitzmaurica, & D. Biber (Eds.), *Using corpora to explore linguistic variation* (pp. 73–90). John Benjamins.

Khojasteh, L., & Kafipour, R. (2012). Are modal auxiliaries in Malaysian English language textbooks in line with their usage in real language? *English Language Teaching*, 5(2), 68–77.

Khojasteh, L., & Reinders, H. (2013). How textbooks (and learners) get it wrong: A corpus study of modal auxiliary verbs. *Applied Research on English Language*, 2(1), 33–44.

Khojasteh, L., & Shokrpour, N. (2015). The “permission/possession/ability” modals in Malaysian English textbooks: A corpus-based analysis. *Khazar Journal of Humanities and Social Sciences*, 18(2), 56–69.

Khojasteh, L., Shokrpour, N., & Rafatbakhsh, M. (2014). Semantic functions of “obligation/necessity” and “volition/prediction” modals in Malaysian English textbooks: A corpus-based analysis. *Khazar Journal of Humanities and Social Sciences*, 17(3), 5–17.

Krashen, S. D. (1981). Second language acquisition and second language learning. Pergamon Press Inc.

Leech, G. N. (1998). Preface. In S. Granger (Ed.), *Learner english on computer* (pp. xiv–xx). Wesley Longman.

Liang, M. C. (2008). A study of modal sequence in the writings of Chinese college students. *Foreign Language Teaching and Research*, 40(1), 51–58.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3), 319–344.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge University Press.

Manaf, U. (2007). *The use of modals in Malaysian ESL learners’ writing* [Unpublished doctoral thesis]. Universiti Putra Malaysia.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Meunier, F. (2002). The pedagogical value of native and learner corpora. In S. Granger, J. Hung, & S. Tyson (Eds.), *Computer learner corpora, second language acquisition, and foreign language teaching* (pp. 119–141). John Benjamins.

Milton, J., & Hyland, K. (1999). Assertions in students’ academic essays: A comparison of L1 and L2 writers. In R. Berry, B. Asker, & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 47–67). John Benjamins Publishing Co.

Römer, U. (2004a). A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 185–199). Benjamins.

Römer, U. (2004b). Comparing real and ideal language learner input: The use of an EFL textbook corpus in corpus linguistics and language teaching. In G. Aston, S. Bernardini, & D. Stewart (Eds.), *Corpora and language learners* (pp. 151–168). John Benjamins.

Römer, U. (2005). Progressives, patterns, pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics. John Benjamins.

Scott, M. (2001). Comparing corpora and identifying key words, collocations and frequency distributions through WordSmith tolls suite of computer programs. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT* (pp. 47–67). John Benjamins Publishing Co.

Scott, M. (2008). *WordSmith tools [computer software]*. Lexical Analysis Software.

Stranks, J. (2013). Materials for the teaching of grammar. In B. Tomlinson (Ed.), *Developing materials for language teaching* (pp. 337–351). Bloomsbury.

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice*. Routledge, Taylor & Francis Group.

Trible, C. (1999). *Writing difficult texts [Unpublished doctoral dissertation]*. Lancaster University.

Whitty, L. (2017). Exploring the complexity of ‘can’, ‘could’ and ‘be able to’ through corpus analysis and classroom- and coursebook-based investigation [Unpublished Doctoral Thesis]. Victoria University of Wellington.

Widdowson, H. (1991). The description and prescription of language. In J. Alatis (Ed.), *Georgetown university round table on languages and linguistics* (pp. 11–24). Georgetown University Press.

Xue, X. (2015). *Ten-thousand English compositions of Chinese learners (the TECLL corpus) [computer software]*. The National Research Centre for Foreign Language Education, Beijing Foreign Studies University.