Non-empirically Based Teaching Materials Can be Positively Misleading: A Case of Modal Auxiliary Verbs in Malaysian English Language Textbooks

Laleh Khojasteh (Corresponding author)

Department of Foreign Languages, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Tel: 60-123-804-248   E-mail: khojastehlaleh@yahoo.com

Reza Kafipour

Department of Foreign Languages, Shiraz Branch, Islamic Azad University, Shiraz, Iran

Tel: 98-917-315-8394   E-mail: rezakafipour@gmail.com

Abstract

Using corpus approach, a growing number of researchers blamed textbooks for neglecting important information on the use of grammatical structures in natural English. Likewise, the prescribed Malaysian English textbooks used in schools are reportedly prepared through a process of material development that involves intuition. Hence, a corpus-based study with the population that was sourced from five Malaysian English language textbooks (Forms 1-5) was adopted to identify modal auxiliary verbs' order and ranking in both whole text-types and spoken text-type of these textbooks. The WordSmith Tools 4.0 was used almost entirely to support quantitative and qualitative data analysis in this research. This study has revealed that for almost all modal auxiliaries, there is a discrepancy between frequency order in the textbook corpus and natural English. The findings of this study also show that the currently used pedagogical language in Malaysian textbooks is mainly based on written rather than spoken English.

Keywords: Modal auxiliary verbs, Prescribed textbooks, Spoken and written English, Pedagogic corpus

1. Introduction

High-powered computers and large electronic corpora have enabled researchers to provide insightful information about the frequency of occurrence of particular linguistic elements and render more accurate descriptions of naturally occurring language features which would otherwise be quite elusive to ESL/EFL language learners and practitioners (Hunston, 2002; Sinclair, 2004; Stubbs, 2001; Thompson & Hunston, 2006). Accordingly, corpus-based analysis is recognized as an ideal tool to re-evaluate the order of presentation of linguistic features in textbooks, and to make principled decisions about what to prioritize in textbook presentations. However, over past decades, it has been frequently reported that those reference materials and syllabuses that have scarcely scratched the surface of corpus linguistic, have ignored all the insights needed for the content of language teaching. In this regard Malaysian ESL textbooks were not exceptions. The prescribed Malaysian English language textbooks used in schools are reportedly prepared through a process of material development that involves intuition and assumption (Mukundan, 2004; Mukundan & Roslim, 2009; Mukundan & Khojasteh, 2011). If such is the case, present-day textbooks might lack a broad empirical foundation which leads us to the first reason for carrying out such a study; because non-empirically based teaching materials can be positively misleading. For this particular study, modal auxiliary verbs were chosen to be analyzed in five Malaysian English textbooks because they are reported to be one of the most troublesome grammatical structures for Malaysian learners. It is argued that the limited exposure of Malaysian learners to different forms of modal verbs might be one of the reasons that resulted to an overuse of one modal form or function over the others (Manaf, 2007; Wong, 1983). Hence the leading question for this study was:

How extensively are the modal auxiliary verb forms presented in all text types (written and spoken) as well as spoken-text types in Form 1-5 Malaysian English language textbooks in line with the modal forms used in natural English?

2. Discrepancies between English Language Textbooks and Real Language Use

Using corpus approach, over the past two decades, a growing number of researchers started to blame the textbooks

for neglecting important information on the use of grammatical structures as well as lexical items in real language use and provided ample information about the mismatch and lack of fit between language used in the textbook and real language in use (Biber & Reppen, 2002; Carter & McCarthy, 1995; Frazier, 2003; Gilmore, 2004; Glisan & Drescher, 1993; Harwood, 2005; Holmes, 1988; Hyland, 1994; O'Connor Di Vito, 1991; Lawson, 2001; O'Keeffe, McCerthy & Carter, 2007; Mukundan & Roslim, 2009; Mukundan & Khojasteh, 2011; Romer, 2004a; Romer, 2004b). Surprisingly, all of these studies indeed demonstrate that although frequency information exhibit in computer databases has improved a lot, syllabus designers still tend to operate by hunch and neglect important and frequent features of the language spoken or written by real language users (Thornbury, 2004). According to Barbieri and Eckhardt (2007, p. 321) textbooks "present a patchy, confusing, and often inadequate treatment of common features of the grammar of the spoken language, and ... do not reflect actual use".

Romer (2004a) has identified the inaccurate description of modal verb usage in an elementary textbook series used in German elementary schools when it was compared with one-million-word British National Corpus (BNC). As regard to frequencies, semantic functions and co-occurrences, she made it clear that there are huge discrepancies between the use of modal auxiliaries in authentic English and in the English taught in German schools. Syntactically, there were incidences of overused cases of modals of *will/'ll* and *can* whereas underused cases of *would/'d, could, should* and *might* as compared to BNC. Semantically, the ability meaning of *can* and *could* have been overused in textbooks while in BNC *could* more frequently express a possibility than an ability. The striking results though, according to Romer (2004a), is that *shall* with its prediction meaning is never used in textbooks while in BNC this is one of the most important meanings. At the end, she suggests that more corpus-based work needs to be done in order to enable pupils as well as teachers to learn and teach English which is more authentic and closer to that of native speakers. This has been supported by Ellis (1997, p. 129) who believes that "speaking natively is speaking idiomatically using frequent and familiar collocations, and the job of the language learner is to learn these familiar word sequences".

Following similar approach as Romer's (2004a) in the comparative study of textbooks and BNC, Mukundan and Khojasteh (2011) reported that for certain modal auxiliaries, there was a mismatch between modal frequency order in lower secondary Malaysian English textbooks (Form 1-3) and the BNC. They also revealed that there were great differences in the relative frequency of verb phrase structures in which modals could occur. For instance, whereas modal followed by the bare infinitive was overwhelmingly dominant for almost all modals in the textbooks, lower secondary learners were not really exposed to other verb phrase structures, particularly structures with passive, progressive and perfect aspects. Their report along with similar findings as regard to prepositions in the same textbook series reported by Mukundan and Roslim (2009) indicate that there are incidences of unsoundness of some of the content of the Malaysian lower secondary textbooks which might have given the students an unrepresentative picture of the way modals and prepositions are actually used.

In another study conducted by Nordberg (2010), it is reported that Finnish upper secondary schools EFL textbooks portrayed a one-sided picture of the semantic functions of modal auxiliary verbs. Although the frequency and ordering of nine core modals in Finish EFL textbooks is reported to be in line with the ordering of modals in real language use, these textbooks portrayed a biased picture of modals' semantic functions. For instance, among all "permission/ possibility/ ability" modals (*may, might, can* and *could*), textbook writers portrayed a monolithic view towards the "ability" sense of *can* and *could*. "Permission" meanings with less than 10 occurrences throughout the textbooks indicate that this meaning was being massively biased at the expense of the "possibility" sense. Similarly, there was a noticeable mismatch between the "obligation/ necessity" meanings as well as "volition/ prediction" meanings in the textbooks and their actual usage which indicate the extent students are disadvantaged to be exposed to the full array of meanings that the modal auxiliaries can have.

This type of findings point to the fact that a lot of mismatch between traditional descriptions and actual language usage stems from the fact that the strict interconnection between an item and its environment is more or less ignored. As Kennedy (1991) himself noted the traditional emphasis on the grammatical paradigm has to be revisited in favor of a more syntagmatic approach to use in context. Misrepresenting linguistic facts, according to Tognini-Bonelli (2001), results in frustration in most language learners because they cannot apply what they have learnt when they are about to produce the language themselves partly because "the rule is not sufficient to guarantee a good linguistic production".

3. Methodology

3.1 Population and Sampling

For the purpose of this study, two corpora were used in order to answer the proposed research question. The population for the English language corpora was sourced from Malaysian English language textbooks used for

secondary Malaysian students of Form 1 to Form 5. The main corpus (all text types) used in this study consists of 280,000 running words and can be classified as a "pedagogic corpus" coined by Willis (1993) and defined by Hunston (2002) as a collection of data that "can consist of all the course books, readers etc. a learner has used" (p. 16). The spoken mini-corpus, however, was compiled because a) there were no ready-made computerized collections of spoken part of Malaysian English textbooks available and it would have been a rather time-consuming to go over each and every dialogue or speech bubble to look for nine modal auxiliary verbs in five textbooks and b) based on the findings of empirical studies on modal auxiliary verbs, different varieties of English and different genres of text-types (spoken vs. written English) plays an important role in the distribution of modal auxiliary verbs (Coates, 1983 cited in Kennedy, 1998; Biber, Conrad & Reppen, 1998; Mindt, 1995). Altogether, this corpus of spoken-type texts from textbooks has a size of a bit more than 50,000 tokens. Although this mini-corpus does not have an impressive size as compared to the all text-type pedagogic corpus (written and spoken), we should bear in mind that this mini-corpus is a specialized corpus which only represents a type of language used in Malaysian textbook materials.

3.2 Instrument

The WordSmith Tools 4.0 was used almost entirely for the purpose of this research, because it has been recognized as a capable and suitable tool to support quantitative and qualitative data analysis by many researchers (Baker, 2006; Bondi, 2001; De Klerk, 2004; Flowerdew, 2003; Henry & Roseberry, 2001; Menon, 2009; Mukundan, 2004; Mukundan & Menon, 2006; Mukundan & Roslim, 2009; Nelson, 2001; Scott, 2001, and many more).

4. Results

There are six modals which are required to be taught in KBSM syllabus for lower and upper secondary students namely: *must, will, should, can, may* and *might.* The frequency of *could, would* and *shall*, however, is investigated in this study in order to see how many times these modals are presented to students implicitly throughout the texts during five years of study. According to KBSM, in Form 1 textbook, students are supposed to be exposed and taught three modals of *must, will* and *should.* The number of modals that students need to learn increases to *can, will, must, may* and *might,* in Form 2 and the exact same modals, *can, will, must, may and might* are stipulated for Form 3. In Form 4, however, this number dropped to only one modal of *should* and in Form 5 modals of *may* and *might* are repeatedly assigned for the third time. Table 1 shows the distribution of six modal auxiliary verbs explicitly featured to Malaysian students (symbolized by a star*) plus the other three that have been presented implicitly throughout the Malaysian English language textbooks Form 1 to 5.

As it can be clearly seen from Table 1, *can* and *will* are the most dominant modals in all the Forms of 1 to 5. In Form 1 textbook, for instance, of all 717 modal auxiliary verbs, modal *can* accounts for34% followed by *will* (24%) and *should* (14.64%). In this Form, *would* (9.20%), *could* (6%), *may* and *must* (5%) are moderately frequent throughout the textbook with *might* and *shall* at their least frequency occurrences (less than 1%). In the same way, *can* (36.67%) and *will* (22.63%) are the most frequently occurring of all modal forms (698) in Form 2 textbook, ranked ahead of *must* (11%), *may* (9%), *would* (6.5%) and *should* (5.7%). Although in Form 2 *might* (3.5%) occurred with slight majority compared to Form 1, there is still a paucity for modal *shall* (0.8%) in this Form. In Form 3, following the similar trend, *can* (33.53%) and *will* (20.54%) are still dominantly used throughout the textbook. Furthermore, the modals that yielded a much lower frequency occurrences in Form 3 are *should* (12%), *would* (9%), *must* (7.4%) *may* (6.93%) and *could* (6%). Out of 875 modal tokens, *can* (27.54%) and *will* (21%) are consistently the most frequent modals in Form 4 textbook; outstripping *should* (14.62%) and *may* (13.37). Maintaining similar frequency occurrences as compared to its previous level (From 3), *must* (7.77%) and *coul*d (5%) are relatively more common than *might* (0.91%) and *shall* (0.3%) in Form 4. Not surprising at this stage, Table 1 shows the predominance of *can* (26.42%) and *will* (24.42%) over the other modal auxiliary verbs throughout Form 5 textbook. *Would* (12.16%) is almost as frequent as *should* (12%) while *shall* is the least frequent modal auxiliary verb (1 instance) after *might* with 25 hits in Form 5 textbook.

Some crucial observations could also be made in the analysis of modal auxiliaries and negation in both written and spoken parts of the textbook corpus. In the following, some of the most interesting findings are listed. As it can be seen in the Table 2 the highest percentage of negations were found with *can* (34.91%) for the Forms of 1 to 5. In addition to that, the highest occurrence of any modal verb in negation is *can* with 53 hits in Form 2. Contracted forms (e.g. can't, 42%) are in all cases throughout all Forms of 1 to 5 much less frequent than full forms (e.g. cannot, 58%). The next favored modal in negation in Malaysian textbooks is *should* (accounting for 15% of all modal tokens in negation) which in Form 4 has the highest occurrences (26 hits) in comparison with other Forms (1-5). This rank order is followed by *will* which is approximately as equal as *could* with 65 and 63 occurrences respectively. *Will* with 23 hits is dominantly frequent in Form 5 and *could* with 16 hits in Form 3 is in its highest

position. *Must* and negative form is moderately frequent in Forms 2 and 5 (14 and 16 instances respectively) while in Form 1, 3 and 4 there are only 5, 9 and 4 instances of *mustn't/must not* respectively. Another observation that could be made is that *would* in negation form is not really frequent throughout the textbooks. *Wouldn't/would not* only occurred 4 times in Form 1 and Form 5, with 6 and 5 occurrences for Form 3 and 4 respectively. No instances found for *would* and negation in Form 2 textbook. Similarly *may not* and *might not* is the least frequent modals in negation before *shall* which is the least modal auxiliary verb in negative form throughout the textbooks.

Concordance queries were also done on frequency count of each modal auxiliary verb in dialogues, interviews and speech bubbles in five Malaysian English language textbooks. The results can be seen in Table 3.

As it can clearly be seen in Table 3, the number of modal auxiliary verbs that occurred in written English part of the Malaysian English textbooks is far more than the number of modals that occurred in spoken one. In Form 1, *can* is dominantly used in written English with 231 instances while only 12 hits occurred in spoken English. The gap between written and spoken form is still extreme in case of *will* with 163 and 10 instances respectively. *Might, must, shall* and *may* are the least frequent modals occurred in spoken corpus in Form 1. In Form 2, *can* is still the most used modals in both written and spoken English although the gap between the numbers is still great. In spoken English *will* (12 instances), *should* (11) and *must* (14) are moderately used modals in Form 2 compared to the least frequent modals of *would* (6), *could* (6), *may* (5), *might* (1) and *shall* (0). The distribution of modal auxiliaries in Form 3 indicates that *can* and *will* with 44 and 34 instances are the most used modals in dialogues and speech bubbles while the gap between modals in written and spoken English is less dominating than the previous Forms (1 and 2). Except for *shall* that its frequency occurrences seem more balanced in written and spoken (5 and 3 respectively), *could, would, should, must, may* and *might* are dominantly used in written rather than spoken English. In Form 4, modals are noticeably used in written English while in spoken corpus there is a very low occurrences of *should* (5 hits), *would* (8), *could* (3), *must* (7), *may* (3) and absolutely zero instances for *might* and *shall*. *Can* and *will* are still the most frequent modals in both written and spoken English. In Form 5, the gap between the frequency occurrences of all modals except for *might* is noticeably extreme. In terms of *can*, for instance, of all *can* tokens in this Form (278), only 22 instances occurred in spoken English while 256 instances occurred in written English. Similarly, the frequency occurrence of *will* in written English (232 hits) outweighed the occurrences in spoken English (25 hits). Interestingly though, *will* is the most frequent modal used in spoken English. Table 3 also shows the predominance of *should* and *woul*d in written English with the scarcity of their use (13 and 15 instances respectively) in spoken English. Must (3 hits) and shall (0) are the minor modals used in spoken English in Form 5 textbook.

5. Summary and Discussion

The first phenomenon was looked at in the context analysis of modal auxiliary verbs was the distribution of nine modal auxiliary verbs throughout Form 1 to 5 Malaysian English language textbooks. This section summarizes the findings reported earlier and discusses the results.

Figure 1 illustrates the results of the overall frequency counts of the analyzed modal auxiliary verbs in textbook corpus. As it can be seen in Figure 1 the modal auxiliary verbs (including their negative forms) found in the five English textbooks of lower and upper secondary level are presented in a descending order: *can, will, should, would, must, may, could, might* and *shall*. There were altogether 4154 instances of core modals in textbook corpus. As we can see in this Figure, there is a huge frequency gap between *can* and *will* on the one hand and other seven modals on the other hand. There are 1289 frequency occurrences of *can* and 938 occurrences of *will* but only between 22 and 493 instances of *should, would, may, must, could, might* and *shall*. The most frequent modals, *can* and *will* accounting for almost 54 % of all modal tokens in the corpus, with the most frequent modal (*can*) accounting for almost 31 % of all modal tokens in the corpus. *Should* with 493 hits is almost half as frequent as *will* and *would* standing at the fourth place has 400 (9.6%) occurrences. *May* and *must* are followed by *would* with 344 (8.2%) and 340 (8.1%) hits respectively. *Could* was not far behind with 239 hits (5.7%), after which come the two least frequent modals *might* and *shall* with 89 (2.1%) and 22 (0.5%) occurrences respectively. Considering the pairs of modal auxiliary verbs, the past time members are less frequent than their partners in all cases except for *shall/should*.

Although, one should admittedly be careful when making comparisons between large corpora and small corpus like this pedagogic corpus, the results indicate that the frequency and ordering of the modal auxiliary verbs in textbook corpus do not correspond reasonably well to the values presented in major corpus-based studies on the modal auxiliary verbs. When this order compared to the order of modal auxiliaries ranked by frequency as they are presented in the British National Corpus (BNC), LGSWE corpus, and LOB and SEU corpora, it is understood that there is a discrepancy between the way modal auxiliaries presented in real language use and the way it is presented in Malaysian textbooks. This lack of fit between the order of modal auxiliary verbs in textbook corpus and the other

three major corpora can be seen in Table 4.

As it can be seen in Table 4, while there are modal verbs that show a balanced frequency of occurrence in the four corpora (e.g., *shall*, *might, may*), others exhibit greater degrees of divergence. As it can be seen in all these three major reference corpora the most frequent modal auxiliary verbs in descending order are *will*, *would*, *can* and *could*. According to Kennedy (2000), these four modals are considered the most frequent modals (they account for 72.7% of all modal tokens) in the BNC. Similarly, Coates (1983) reported that *will, would, can* and *could* as the most frequent modals accounts for 71.4 % of all modal token in LLC and LOB. However, as it can be seen in the Table 4, except for *may*, *might* and *shall* there is a mismatch between frequency order of the other six modals in textbook corpus. *Will* which is supposed to be given the most emphasis in a pedagogic corpus reaches second while *can* that is ranked third in three major corpora has been overused by standing as the most frequent modal used in the textbook. Indeed, *can* is well overrepresented throughout Form 1 to 5 textbooks because although it is among the top four used modal auxiliaries, it is well below *will* and *would* in terms of frequency occurrence (Leech et al. 2009; Biber et al. 1998). It is interesting to see that although based on KBSM curriculum modals *must, will, may, might* and *should* are the ones that are stipulated to be taught in Form 1, Form 4 and Form 5 textbook, still modal *can* is used more than any other modals. The most remarkably biased toward modals in the textbook is *could* that has lost its place from 4th to 7th in textbook corpus. Surprisingly, this modal (*could*) is not only underused in Malaysian textbooks but also is not taught explicitly neither at primarily level nor secondary level in Malaysia. Similarly, *would* is among the top four modals in the textbook corpus but it is not taught explicitly in any of the textbooks. Although Thornbury (2004) has indicated that the most frequently occurring items are not always the most useful ones in terms of teachability, and that they may be better delayed until relatively advanced levels, in the case of this textbook corpus the modals *could* and *would* neither taught at lower nor higher secondary levels. Barbieri and Eckhardt (2007) indicate that despite more than two decades of language teaching aimed at fostering natural spoken interaction and written language, instructional textbooks still neglect important and frequent features of real language users. This has been supported by other linguists such as Carter and McCarthy (1995), Harwood (2005) and Hyland (1994).

Among other overused modal auxiliaries we can refer to modal *must* that appears before modal *could* in the textbook corpus having modal *may* in between, while in BNC, LOB and SEU, and LGSWE not only the modal *could* appears before *must* but also there are two other modals (*may* and *should*) in between. Finally, *shall* as the lowest frequent modal is lopsided throughout Malaysian textbooks. Although *shall* has been reported by Biber et al. (1998) and Leech et al. (2009) to be obsolete in current English, according to Mindt (1995) and Romer (2004a) the prediction meaning of *shall* (31%) is among one of the most widely used meanings in spoken British English. In the ESL environment, students need to be exposed to the language as much as possible to gain sufficient input and exposure. For example rare occurrences of *might* and *shall* (less than five times) may not be enough to lead learners to notice and acquire these forms. Even in vocabulary studies, repetition of words is very important to ensure acquisition of new vocabulary (Mukundan & Anealka, 2007). One kind of repetition that is important is repetition of encounters with a word. It has been estimated that, when reading, words stand a good chance of being remembered if they have been met at least seven times over spaced intervals (Thornbury, 2002). According to Celce-Murcia and Larsen-Freeman (1999) it makes sense to recycle various aspects of the target structures over a period of time: revisit old structures, elaborate on them, and use them for points of contrast as new grammatical distinctions are introduced.

In terms of modal auxiliaries and negation we can say that in almost many cases of modals and negation such as *should* in Form 2, *must* in Form 1 and Form 4, *may* in Form 1, Form 3, Form 4 and Form 5, *would, might* and *shall* in all the textbooks (1 to 5) the context provided is extremely positive with low occurrences for negative forms. Full forms are much more frequent than the contracted forms in case of modal auxiliary verbs in all the textbooks. However, this is contradicted with the findings of Mindt (1995, p. 176) and Romer (2004a). Both studies have reported that contracted forms are more popular and more frequently used in terms of negations. Of all *can* tokens in negation, Romer (2004a) has reported 94% for *can't* and only 5.75% with *cannot*. An explanation for these discrepancies may lie in the fact that based on the findings of the same research question (spoken vs. written) reported next, modal auxiliary verbs are more frequent in written part of the textbooks rather than in conversations. Hence, it is hardly surprising that the occurrences of the full forms are much more frequent than the contracted forms in the textbooks.

The fact that modals have high frequency as grammatical items, especially in spoken English, makes the results meaningful even in the comparison of such small corpus. An analysis of the spoken part of five Malaysian English textbooks' coverage of modal auxiliary verbs reveals a mismatch between the corpus-based cross register studies on modal auxiliaries and what is covered in the textbook (Figure 2).

Contrary to what was assumed about the higher share of modal auxiliary verbs in spoken rather than written English (Coates, 1983; Kennedy, 2002; Leech et al., 2009; Mindt, 1955; Quirk et al., 1985; Romer, 2004a) the data indicate that in this spoken mini-corpus, speech contains much less shares of modal auxiliary verbs than writing. If we look at the frequencies of individual forms of modal auxiliary verbs in textbook's conversation, we can clearly see that there is a considerable difference between the two registers for all modals. While there are only 133 frequency occurrences of *can* in spoken texts, this number leaped to 1156 in written texts alone. Similarly, *will* with a lower frequency occurrence (101) in spoken texts soared to 837 in written texts. Surprisingly, we can see that the rest of the modals, *would, should, could, may, must, might* and *shall* are relatively infrequent in spoken texts.

The frequency distribution of the modals in spoken mini-corpus differs quite a lot from the one reported by Romer (2004a) in the spoken part of the BNC. As we can see in Figure 3, the modals *can, should, must* and *may* are overused in textbooks while there is an underuse of *will, would,* and *could*. This underuse is especially significant in the case of *would*. In BNC this modal accounts for 23.48 percent of all modal tokens in spoken BNC while this modal in spoken mini-corpus is half frequent as it should be.

The overuse is also significant in terms of *can* which although it comes third in BNC (22.68%), is dominantly frequent in spoken mini-corpus standing in the first place. Similarly, the frequency occurrences of *may* and *must* are approximately three times greater than what they are expected to be in comparison to BNC.

After the advent of corpus linguistics, statistical evidence provided by corpora indicated that grammatical patterns differ systematically across varieties of English and most importantly across registers and this suggested the fact that ignoring grammatical variants undermine the effectiveness of teaching materials (Conrad, 2004). However, the findings of this study show that Malaysian English language textbooks are usually based on written norms only, thus ignoring the spoken language. Forms 1 to Form 5 Malaysian English textbooks of course have many positive features; their coverage of modal auxiliaries in conversation is only a small part of the books. However, as Conrad (2004) posits, "by minimizing the importance of variation, we are misrepresenting language in materials that we use with students" (p. 69). All in all, modal auxiliaries used in writing are covered, but the most frequent modals in conversation is not covered in most of the textbooks.

6. Conclusion

The findings of this study have shown several valuable insights. Firstly, the frequency and ranked order of modal auxiliary verbs found in the English language textbooks used in Form 1, 2, 3, 4 and 5 in Malaysian Secondary Schools have been revealed. The data shows how many times modals are used in the textbooks and that either directly or indirectly students have been exposed to these modal auxiliaries in varying degrees. This study has revealed that for almost all of the modal auxiliaries, there is a discrepancy between frequency order in the textbook corpus and the four major reference corpora. For example, although *would* and *could* are among the most frequent modals in real language, it is both a surprise and a concern to see that the both modals are neither among the top four most frequent modals in the textbook corpus nor have been taught to secondary learners. The reason for this discrepancy is unknown but it might be because of the content of the all major corpora which includes various authentic spoken and written texts while our textbook corpus only contains prescribed pedagogical texts. On the other hand, this discrepancy may also signal a deficiency in the preparation of the textbooks. Apart from many criteria proposed for principled selection of syllabus designs, frequency and range have been highly recommended after the advent of corpus-based research (Kennedy, 2002; Koprowski, 2005; Mindt, 2000; Moon, 1997; Romer, 2004a; Sinclair, 1991 and many more). Nation and Waring (1997, p. 17) state that applying frequency information in textbooks ensures that students are exposed to the language they most probably meet again outside the classroom walls. Romer (2004a, p. 152) believes we should always make sure that the language students are exposed to in their textbooks is as close as the language they are likely confronted with in natural communicative situations.

The findings of this study also show that the currently used pedagogical language in Malaysian textbooks are mainly based on written English rather than spoken. A higher degree of authenticity can be achieved if modal auxiliary verbs are presented in the spoken text of textbooks which is the kind of context in which they typically appear in actual language use. This is essential if we assume that the goal of grammar to be taught is for "communicative purposes" (Glisan & Drescher, 1993, p. 24). Indeed, it is argued that when students are exposed to the structure in textbooks that is unlikely found in current-day native speaker discourse, they most likely encounter great difficulties communicating successfully with speakers of that particular language (Romer, 2004b).

7. Suggestions for the Improvement of Teaching Materials

This study does not suggest making drastic changes in the Malaysian textbooks in order to create a textbook that mirrors exactly the language used by native speakers. According to Romer (2005, p. 275) it is not even "safe" to do that. However, the most salient facts reflected from natural language corpora should not be ignored in the textbooks.

For example, now that the textbooks show a lack of fit between the weight given to the modal auxiliaries and their real rank order in natural English, it is for the teachers to adapt their lesson plans accordingly in order to expose students with the modal auxiliaries that they will more likely to be confronted with in natural communicative situations such as *would* and *could* that are ranked among the top four used modals by native speakers but not even taught in Form 1 to Form 5 Malaysian English language textbooks. Focusing on more frequent modals and delay teaching those that are not very common in real language use has also been emphasized by Cowan (2008, p. 321) in order to attain more pedagogically beneficial textbooks.

Since the findings of this study revealed that the language in Malaysian textbooks is mainly based on written rather than spoken English, it is important for the teachers to involve their students with communicative type of tasks that expose students with variety of communication contexts. This can be achieved by the tasks that require repeated reception and production such as role-playing, guessing, describing pictures, writing letters or alike that support successful communication across a variety of communication contexts.

Finally, since modal auxiliary verbs are implicitly and explicitly taught in an extremely positive contexts, it is vital for the teachers to expose their students with the modals that highly used in negative forms such as *can't* that accounts for 94% of all modal tokens in negation in BNC. By leaving out some of the modal forms that students most probably meet again outside the classroom walls, students in ESL contexts are going to be misled and misinformed unless teachers and textbook authors apply frequency information in their lesson plans and textbooks.

References

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum.

Barbieri, F., & Eckhardt, S. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech.*Language Teaching. Research*, 1(3), 319–346. http://dx.doi.org/10.1177/1362168807077563

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics*. Cambridge: Cambridge University Press.

Biber, D., & Reppen, R. (2002).What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition,* 24, 199–208. http://dx.doi.org/10.1017/S0272263102002048

Bondi, M. (2001). Small corpora and language variation: Reflexivity across genres. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT.* (pp. 135-174). Amsterdam/ Philadelphia: John Benjamons Co.

Carter, R., & McCarthy, M. **(**1995). Grammar and the spoken language. *Applied Linguistics,* 16(2), 141–58. http://dx.doi.org/10.1093/applin/16.2.141

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar book: an ESL/EFL teacher's course.* 2nd ed. Boston: Heinle & Heinle.

Coates, J. (1983). *The semantics of the modal auxiliaries*. London: Croom Helm.

Conrad, S. (2004). Corpus linguistics, language variation, and language teaching. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 67-85). Amsterdam: John Benjamins.

De Klerk, V. (2004). The use of 'actually' in spoken Xhosa English: A corpus study. *World Englishes*, 24 (3), 275-288. http://dx.doi.org/10.1111/j.0883-2919.2005.00410.x

Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.

Flowerdew, L. (2003). A combined corpus and systematic-functional analysis of the problem solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37 (3), 489-511. http://dx.doi.org/10.2307/3588401

Frazier, S. (2003). A corpus analysis of *Would*-clauses without adjacent *If*-clauses. *TESOL Quarterly,* 37(3), 443–46. http://dx.doi.org/10.2307/3588399

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal,* 58(4), 363–71. http://dx.doi.org/10.1093/elt/58.4.363

Glisan, E. W., & Drescher, V. (1993). Textbook Grammar: Does it reflect native speaker speech? *Modern Language Journal*, 7 (1), 23-33

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes,* 4, 149–161. http://dx.doi.org/10.1016/j.jeap.2004.07.008

Henry, A., & Roseberry, R. L. (2001). Using a small corpus to obtain data for teaching a genre. In M. Ghadessy, A. Henry, & R. Roseberry (Eds), *Small corpus studies and ELT* (pp. 93-113). Amsterdam/ Philadelphia: John Benjamons Co.

Holmes, J. (1988). Doubt and certainty in ESL textbooks. *Applied Linguistics,* 9(1), 21–44. http://dx.doi.org/10.1093/applin/9.1.21

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, England: Cambridge University Press.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes,* 13(3), 239–56. http://dx.doi.org/10.1016/0889-4906(94)90004-3

Kennedy, G. (1991). 'Between and Through', The company they keep and the functions they serve. In K. Aijmer, and B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London: 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). Amsterdam: John Benjamins.

Koprowski M. (2005). Investigating the Usefulness of Lexical Phrases in Contemporary Coursebooks. *ELT Journal,* 59(4), 322-332. http://dx.doi.org/10.1093/elt/cci061

Lawson, A. (2001). Rethinking French grammar for pedagogy: The contribution of French corpora. In Simpson, R. C., & Swales, J. M. (Eds.), *Corpus linguistics in North America. Selections from the 1999 Symposium*. Ann Arbor, MI: The University of Michigan Press.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in Contemporary English*. New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511642210

Mindt, D. (1995). *An Empirical Grammar of the English Verb: Modal Verbs*. Berlin: Cornelsen.

Mindt, D. (2000). *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.

Moon, R. (1997). Vocabulary connections: multi-word items in English. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary Description, Acquisition and Pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.

Mukundan, J., & Anealka, A. H. (2007). A forensic study of vocabulary load and distribution in five Malaysian Secondary School Textbooks (Forms 1-5). *Pertanika Journal of Social Science and Humanities*, 15(2), 59-74

Mukundan, J., & Roslim, N. (2009). Textbook Representation of Prepositions.*English Language Teaching*, 2(4), 123-130

Mukundan, J., & Khojasteh, L. (2011). Modal Auxiliary Verbs in Prescribed Malaysian English Textbooks. *English Language Teaching,* 4 (1), 79-89

Nation, P., & R. Waring. (1997). Vocabulary size, text coverage and word lists. In N. Schmitt, and M. McCarthy (Eds), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6–19). Cambridge: Cambridge University Press.

Nelson, M. (2001). *A corpus based study of business English and business teaching*. Oxford: Oxford University Press.

Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, California: Sage Publications.

Nordberg, T. (2010). *Modality as portrayed in Finish upper secondary school EFL textbooks: A corpus-based approach*. (Master's Thesis). University of Helsinki. [Online] Available: https://helda.helsinki.fi/handle/10138/19357

O'Connor Di Vito, N. (1991). Incorporating Native Speaker Norms in Second Language Materials. *Applied Linguistics,* 12(4), 383- 396. http://dx.doi.org/10.1093/applin/12.4.383

O'Keeffe, A., McCerthy, M., & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511497650

Quirk, R. S., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.

Romer, 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). Amsterdam: John Benjamins.

Romer, 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, and D, Stewart (Eds.), *Corpora and Language Learners* (pp. 151-168). Amsterdam: John Benjamins.

Romer, U. (2005). *Progressives, patterns, pedagogy. A Corpus-driven Approach to English Progressive Forms,*

Functions, Contexts and Didactics. Amsterdam: 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). Amsterdam/ Philadelphia: John Benjamins Publishing Co.

Sinclair, J. M. (1991). *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. London, England: Rutledge.

Stubbs, M. (2001). *Words and Phrases. Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishers.

Thompson, G., & Hunston, S. (2006). *System and Corpus: Exploring connections*. London: Equinox.

Thornbury, S. (2004). *How to teach grammar*. Malaysia: Pearson Education Limited.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins Publishing Co.

Willis, D. (1993). *Syllabus, corpus and data-driven learning*. IATEFL Conference Report: Plenaries.

Wong, I. (1983). *Simplification features in the structure of colloquial Malaysian English*. Singapore: Singapore University.

Table 1. Weight given to each modal in Form 1-5 textbooks

Modals	Form 1	Form 2	Form 3	Form 4	Form 5
Can	243	*256	*271	241	278
Will	*173	*158	*166	184	257
Should	*105	40	100	*128	120
Would	66	46	77	84	127
May	37	*67	*56	117	*67
Must	41	*77	*60	68	94
Could	*44	23	50	42	80
Might	4	*25	*23	8	*29
Shall	4	6	8	3	1

Table 2. Modals in negation within Form 1-5 textbooks

Modals in negation	Form 1	Form 2	Form 3	Form 4	Form 5	Total
Can't/cannot	6/17	16/37	13/21	11/18	14/16	169
Shouldn't/should not	4/10	2/5	4/4	7/19	2/16	73
Won't/will not	6/9	3/13	4/9	1/8	4/19	65
Couldn't/could not	3/6	1/8	2/14	4/11	2/12	63
Mustn't/ must not	1/4	2/12	2/7	1/3	-/16	48
May not	3	10	2	1	-	33
Wouldn't/would not	-/4	-/-	2/4	2/3	-/4	19
Might not	-	6	2	-	5	13
Shan't/shall not	-/-	-/1	-/-	-/-	-/-	1

Table 3. The distribution of modal auxiliary verbs in written English as well as spoken English parts of textbooks

Modal	Form 1		Form 2		Form 3		Form 4		Form 5	
	W	S	W	S	W	S	W	S	W	S
can	231	12	220	36	227	44	222	19	256	22
will	163	10	146	12	132	34	161	23	232	25
should	99	6	29	11	92	8	123	5	107	13
would	54	12	40	6	66	8	76	8	112	15
must	33	2	63	14	56	4	61	7	91	3
could	34	10	17	6	38	12	39	3	72	8
may	36	1	62	5	46	10	114	3	57	9
might	4	0	24	1	19	4	8	0	19	10
shall	2	2	6	0	5	3	3	0	1	0
Total	656	55	607	91	681	127	807	68	947	105

Table 4. Three major corpora and textbook corpus ranked by frequency

	LOB and SEU (Written and Spoken) Quirk *et al.* (1985)	LGSWE (Written and Spoken) Biber et al. (1998)	BNC (Written and Spoken) Kennedy (2002)	Textbook Corpus (written and Spoken) Mukundan & Anealka (2007)
1	*will*	*will*	*Will*	*Can*
2	*would*	*would*	*Would*	*Will*
3	*can*	*can*	*Can*	*Should*
4	*could*	*could*	*Could*	*Would*
5	*may*	*may*	*May*	*May*
6	*should*	*should*	*Should*	*Must*
7	*must*	*must*	*Must*	*Could*
8	*might*	*might*	*Might*	*Might*
9	*shall*	*shall*	*Shall*	*Shall*

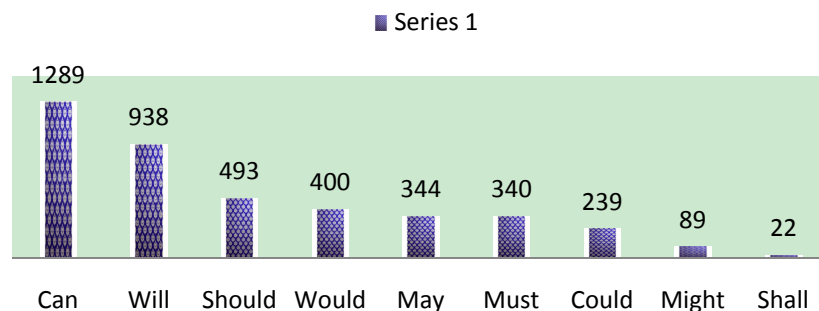## Form 1-5 Total Modal Tokens

■ Series 1

Figure 1. Frequency of modals in textbook corpus

The comparison between modal auxiliaries in written and spoken English in textbook corpus
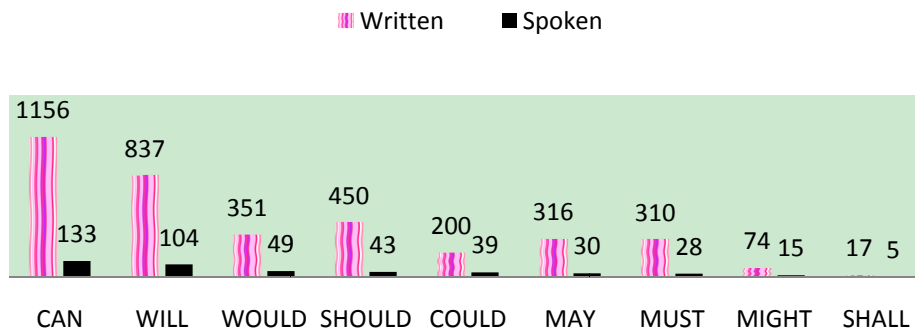

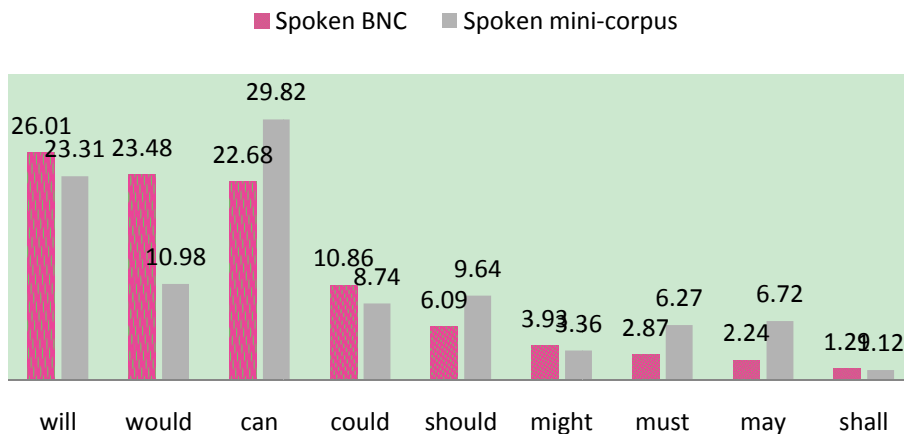Figure 2. The occurrences of modal auxiliaries in written and spoken parts of pedagogic corpus

Figure 3. Relative frequencies of modals in Spoken BNC and Spoken mini-corpus