

How textbooks (and learners) get it wrong: A corpus study of modal auxiliary verbs

Laleh Khojasteh

(Assistant Professor, Shiraz University of Paramedical Sciences, Iran)

khojastehlaleh@yahoo.com

Hayo Reinders

(Professor, Chulalongkorn University, Bangkok, Thailand)

hayo@innovationinteaching.org

(Received: 09.01.2013, Accepted: 02.02.2013)

Abstract

Many elements contribute to the relative difficulty in acquiring specific aspects of English as a foreign language (Goldschneider & DeKeyser, 2001). Modal auxiliary verbs (e.g. could, might), are examples of a structure that is difficult for many learners. Not only are they particularly complex semantically, but especially in the Malaysian context reported on in this paper, there is no direct equivalent in the students' L1. In other words, they are a good example of a structure for which successful acquisition depends very much on the quality of the input and instruction students receive. This paper reports on analysis of a 230,000 word corpus of Malaysian English textbooks, in which it was found that the relative frequency of the modals did not match that found in native speaker corpora such as the BNC. We compared the textbook corpus with a learner corpus of Malaysian form 4 learners and found no direct relationship between frequency of presentation of target forms in the textbooks and their use by students in their writing. We also found a very large percentage of errors in students' writing. We suggest a number of possible reasons for these findings and discuss the implications for materials developers and teachers.

Keywords: Modal auxiliary verbs; Malaysia; corpus; textbook; material development.

Introduction

Materials play a key role in most language classrooms around the world and their evaluation is therefore of prime importance. Language learning materials can be evaluated at the pre-use stage, where they are seen as workplans or constructs, during use, when they are judged as materials in process, and retrospectively, which considers outcomes from materials use (Breen, 1989). Ellis (1997) suggests that predictive evaluation, which aims to determine appropriateness for a specific

context, is carried out either by experts or by teachers using checklists and guidelines. At the in-use stage 'long-term, systematic evaluations of materials . . . are generally considered to be successful' (Tomlinson, 1998, p.5). These include 'formative decisions for improvement through supplementation or adaptation and [sensitising] teachers to their own teaching and learning situation' (Nedkova, 2000, p. 210). In this study, we concern ourselves with retrospective evaluation in that we look at materials that were in use on a large scale,

by many thousands of language learners, at one given time, to learn about the type and quality of the language input contained in them. In order to do this we drew on corpora, the use of which in ELT and language learning research we will now discuss.

The role of corpora in ELT

The use of corpora for both teaching and research has increased significantly in recent years. The motivation for using a corpus approach in language learning research is related in part to the attraction of being able to offer a description of language *in use* and also to the fact that previous research on authentic texts have revealed significant inconsistencies between the use of lexical items and grammatical structures in corpora, and those found in traditional language textbooks that are based purely on introspective judgments (Campoy, Belles-Fortuno, & Gea-Valor, 2010). At the same time, corpus explorations can be carried out by learners themselves and can be used as an integral part of the learning process either directly or indirectly to both foster learners' and teachers' needs (Romer, 2010).

As a result of this growing interest, the use of corpora has resulted in the development of more effective pedagogical materials (Gabrielatos, 2005). Material writers can be informed of the differences between the language used in textbooks and of that used in the real world. Information about the frequency of occurrence of linguistic features in a reference corpus can also be very helpful when it is compared with prescribed pedagogical materials. While many linguists and researchers have focused on the advantages of corpus-informed materials, there are also limitations that need to be taken into consideration by textbook writers.

For instance, Howarth (1998) and Widdowson (1990) have questioned the pedagogical usefulness of frequency lists generated by corpora because in their view frequency does not equate to importance. However, this argument has been strongly rejected by many linguists such as Mindt (1995), Kennedy (2002) and Romer (2004) because, as they argue, frequency information leads to the identification of words or structures that are central in a language and that without this information it is difficult to decide what should be included in teaching materials. Kennedy (1998), among others, points to the need to concentrate initial teaching on high frequency items and to grade vocabulary and structures accordingly and Conrad (2000) emphasizes the importance of frequency information for teachers because it helps them to decide which items to emphasize, for example, to provide low-level students practice with the items they are most likely to hear outside class.

Lawson (2001) argues that insights from corpus linguistics cannot only provide information about the frequency of occurrence of linguistic features in naturally occurring language, but also about register variation, that is about how the use of particular linguistic features varies across different contexts and situations of use. This information, according to Kennedy (1998) can be of direct application to textbook writers. Furthermore, it is argued that corpus-based analysis can provide information about the salience or scope of particular features which otherwise are difficult to acquire (Lawson, 2001). Stubbs (1996) summarises:

There may be the illusion that they [lists of collocations] could have been provided, after a bit of thought, by intuition alone. But this is indeed an illusion. Intuition certainly cannot

provide reliable facts about frequency and typicality. And whilst a native speaker may be able to provide some examples of collocates (which may or may not be accurate), only a corpus can provide thorough documentation. (p.250)

In our study we use corpus linguistics not primarily to inform materials development, but to learn *about* materials, information which, subsequently, may be useful for further development

Methodology

The target structure

We chose modals for this study for several reasons. Firstly, modal auxiliary verbs are particularly challenging for language learners (Decapua, 2008) and also for Malaysian English learners (e.g. Manaf, 2007; Wong, 1983; De Silva, 1981). Perhaps as a result of this, they do not receive as much attention as part of the school curriculum as before. As De Silva (1981) observes: ‘the modal auxiliary system used in the Malaysian schools has been altered and functionally reduced through the continued use of fewer and semantically salient modals that serve multi functionally across notions (p. 12). Wong (1983) argues that the limited exposure of Malaysian learners to different forms of modal verbs and their functions has resulted in an overuse of one form or function over the others by teachers. As modal auxiliaries are so difficult, they are likely to be particularly influenced by the quality of the input and instructions learners receive on them and we were therefore particularly interested to see how this feature is presented to learners.

We also chose modal auxiliaries because they play an important role in learners’ language use. Many Malaysian learners aspire to study through the medium of

English and good use of modals plays an important role in successful social interaction (Celce-Murcia & Larsen Freeman, 1999). In other words, it is an important feature of the language, not just from a linguistic point of view, but also for the learners themselves, from a social-interactive point of view. Modal auxiliary verbs are also common and we therefore thought it would be likely that we would find many exemplars to analyse.

The final reason for the selection of modal auxiliaries is that previous studies conducted in other countries have reported that textbooks do not present this structure accurately (Hyland, 1994; McEnery & Kifle, 2002). In summary, modal auxiliaries are a difficult, common and important (to learners) structure that has often been misrepresented in English language textbooks.

Modal auxiliary verbs and Malaysian learners

Malaysian learners have been observed as having great difficulty with the modal auxiliary system. Examples (1) to (8) provide illustrative evidence for existing problems concerning the appropriate use of modal *can* with its various functions by Malaysian students (Wong, 1983, p.137):

- 1) You can have this book today. (“permission”)
- 2) You can drive? (“ability”)
- 3) Can lend me your bike or not? (“willingness”)
- 4) Can also/ Sure can. (“agreement”)
- 5) Can do. (“moderate approval”)
- 6) You come with me. Can or not? (“affirmation”)

Hughes and Heah (1993) made very similar observations based on learner data and report on problems Malaysian learners have

with the use of modals. The correct use of modals, according to them, was always among one of the most problematic areas for Malaysian learners (Hughes & Heah, 1993). Furthermore, in their study of students' errors in Form 4 students' composition, Rosli and Edwin (1989) found that verb forms and the verb aspects of modals are the most problematic for Malaysian learners. Twenty years since Rosli and Edwin's (1989) study, the same observation was made by Manaf (2009), who analyzed the modal auxiliary verbs in the Malaysian learner corpus (EMAS). According to her, students were not only uncertain about which modals to use to express modality (inaccuracies at the syntactic and semantic levels), but also had difficulty to use modals with appropriate verb form in a sentence (Manaf, 2009). Although the lack of equal counterparts between the English modal system and those in Bahasa Melayu might be the reason for this confusion for Malay learners, Romer (2005) believes that this problem is due to the teaching materials.

Modal auxiliaries in Malaysian grammar and textbooks

There are six modals which are required to be taught in *Kurikulum Bersepadu Sekolah Menengah (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 also investigated in this study in order to see how many times these modals are presented to students implicitly throughout the texts during four years of study. According to KBSM, in the Form 1 textbook, students are supposed to be exposed to and taught the three modals *must, will* and *should*. In Form 2 *can, will, must, may* and *might* are added and repeated in Form 3. In Form 4, *should* is added. The prescribed Malaysian English language textbooks used in schools are often reported as being prepared through a process of material development involving intuition

and assumption (Mukundan & Roslim, 2009; Mukundan & Khojasteh, 2011). Existing textbooks therefore appear to lack a broad empirical basis.

Corpus selection

In order to answer our research questions, we used two corpora; a pedagogic corpus and a learner corpus. A pedagogic corpus, as coined by Willis (1993) and defined by Hunston (2002), is a collection of data that 'can consist of all the course books, readers etc. a learner has used' in an ESL/EFL language learning program (p.16). In this study the population of our pedagogic corpus was sourced from four Malaysian English language textbooks currently used for secondary Malaysian students of Form 1 through Form 4, with a total of just under 230,000 words (Mukundan & Aneleka, 2007)¹. According to the researchers each page of the books mentioned above was photocopied and scanned and converted into a Tagged Image File (TIF) format. This was then saved and processed with Optical Character Recognition (OCR) software, which converted all TIF files into text files (.txt). The txt files were then checked for errors before saving and renaming them according to the respective units of the textbook.

The learner corpus we used was sourced from two written essays produced by Form 1 and Form 4 Malaysian students as part of a previous study (Arshad, Mukundan, Kamarudin, Rahman, Rashid, & Edwin 2002). In the study, approximately 600

¹ The original corpus consisted of 5 Malaysian English language textbooks used in the secondary cycle (311,214 running words). However, in order to suit the textbook data with our learner data we decided to only include Forms 1, 2, 3 and 4 and eliminate the Form 5 data from this pedagogic corpus. Hence, the remaining running words in this corpus consist of 229,794 running words.

Malaysian learners from across the country were required to write one essay on the topic of 'The happiest day of my Life' and another based on a given picture. Students were given one hour to write the essays and were not marked or given credit for them. Although perhaps not ideally representative of Malaysian learners' language proficiency, it was decided to use this corpus because of its very large size and the fact that it does give a broad indication of language learners' writing ability across the whole of the country.

Analysis

As our benchmark corpus we used the BNC, the British National Corpus. This corpus consists of 100 million word collection of samples of written and spoken language. Among all reference corpora available, the insights on modal auxiliary verbs were sought from BNC because the samples of written and spoken language used for this corpus were designed to represent a wide cross-section of British English (BrE) which is the closest English variety used in Malaysia (Mukundan & Roslim, 2009; Mukundan & Khojasteh, 2011). A previous study by Kennedy (2002) looked at the occurrence of modal auxiliary verbs and we draw on his findings here for our comparisons with the results from the textbook corpus and the learner corpus. In the latter two, we retrieved modal auxiliary verbs using the software package WordSmith and in particular its Concord tool to locate all references to modal verbs within both corpora. In order to examine the first research question, content analysis was carried to retrieve absolute frequencies of occurrences for nine core modal auxiliary verb forms from all written and spoken texts in the four Malaysian secondary English language textbooks. Then, the results were added up and compared with the frequency and rank order of the same modals in the

BNC in order to see if there were any discrepancies. Next, discourse analysis was carried out at the sentence level in order to examine the accuracy of the way in which the modals were presented at both syntactic and semantic levels.

In addition to looking at the frequency of use of modal auxiliary forms, we were also interested in looking at the grammatical accuracy of learners' use of this form. In order to do this, all sentences in the learner corpus that included modals were examined using Mindt's (1995) modal verb phrase structure framework. According to Mindt (1995), word categories can colligate with modals in five different structures:

- 1) modal + bare infinitive (e.g. You won't regret it!)
- 2) modal + passive infinitive (e.g. Something should be done)
- 3) modal + progressive infinitive (e.g. Define what you will be talking about)
- 4) modal + perfective infinitive (e.g. The number of the students will have increased)
- 5) modal + perfect passive infinitive (e.g. I know it must have been hard for her).

To this we added 'modal alone', a category suggested by Kennedy (2002).

Results

Here we present the results of our study. First we show the results of the analysis of the textbook corpus, followed by the analysis of the learner corpus. Finally, we present our analysis of the errors in the learner corpus.

Modal auxiliary verbs in the textbook corpus
Figure 1 shows the frequency of the modal auxiliary forms (including their negative

forms) in the four English textbooks in descending order.

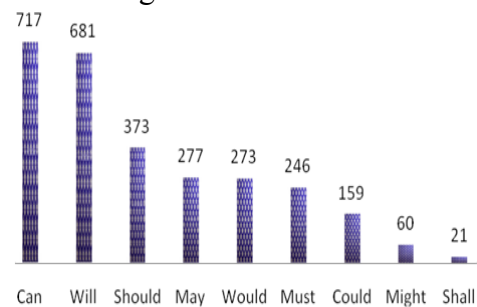


Figure 1: Frequency of modals in textbook corpus

There were altogether 2,807 instances of core modals in the textbook corpus. As can be seen above, there is a large frequency gap between *can* and *will* on the one hand and the other seven modals on the other. There are 1398 occurrences of *can* and *will* and a total of 1401 for *should*, *may*, *would*, *must*, *could*, *might* and *shall*. The most frequent modals *can* and *will*, therefore account for almost 50 % of all modal tokens in the corpus.

Modal auxiliary verbs in the learner corpus

Figure 2 shows the order of frequency in which students used modal auxiliary forms on the writing tasks.

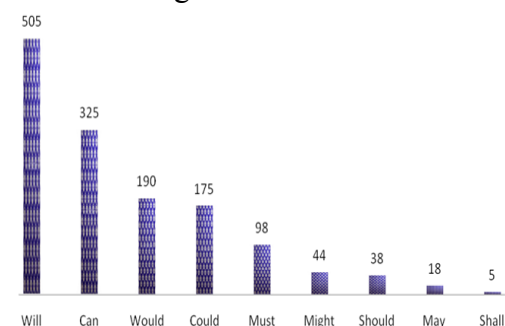


Figure 2: Frequency of modals in the learner corpus

The most frequently used modal auxiliary verb by Malaysian learners is *will* with 505 instances, or 36% of all modal tokens. The modal *can* represents 23% of all modal tokens. The modals *would* and *could* are used by Malaysian learners with 190

(13.59%) and 175 (12.51%) occurrences respectively.

Errors in modal auxiliary verbs in the learner corpus

Next, we analyzed the accuracy of learners' modal auxiliary use in their writing. Figure 3 shows the number of accurately and inaccurately produced modals.

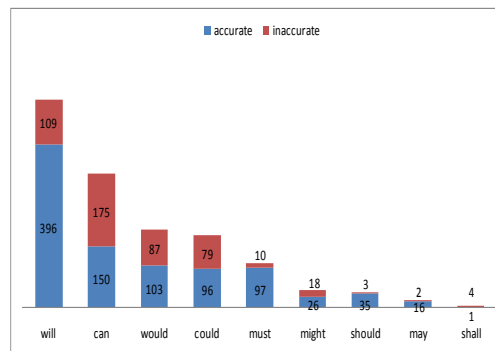


Figure 3: Frequency of syntactically accurate and inaccurate modal auxiliaries in the learner corpus

In descending order, the lowest percentage of syntactical inaccuracy was for *shall* (80%)², *can* (54%), *would* (46%), *could* (45%), *might* (41%), *will* (22%), *may* (11%) and *should* (8%).

Out of only five *shall* modals used by the learners, four were used with progressives or past tense forms of the verb. Examples (1) and (2) are sample sentences of inflected modals:

- (1) She also don't know how what she shall doing.
- (2) "Shall we invited John join with us?" I asked Ahmad again.

More than half of all *can* instances used by Malaysian learners were used inaccurately. 149 occurrences were used with structure one (modal + bare infinitive) but with the

² But note the small number of total occurrences

past tense of the verb. Examples (3), (4) and (5) are sample sentences of such errors.

- (3) I can saw many kind of tress.
- (4) He can spoke fluently in Malay language.
- (5) She hope that Raj, Ah Seng, and Ramlee can heard her.

There were also many incidences of the use of a non-English word after the modal and combining two modals. Furthermore, many of the negative sentences constructed by students using *can* were ungrammatical:

- (6) I hope I can will visit this place again.
- (7) She can't swam.

Would was used inaccurately 87 times by Malaysian learners. Although most sentences were still comprehensible, 81 of the inaccurate instances had the modal *would* followed by the past tense form. This was the same for those who had used this modal in structure 4. In only six cases was the verb after the modal *would* missing:

- (8) I felt something joyful would happened later.
- (9) If they call me, they would told me that the enjoyable day of their life was when they were in 3A1.
- (10) Probably they would have broke some records if we were to take the time.

The same tendency can be seen in the usage of *could* where in all cases the verb that follows the modal was in the past form:

- (11) and we could entered the semi-final because our compenen had a stomachache during the competition.
- (12) My heart beat was beating faster and faster as I could found nobody around.

Over-generalization of the past tense was also found in the use of *might*:

- (13) I didn't tell my husband because I scared that I might lost them especially my children.
- (14) One day, when I came back from school, my heart felt not very well and seemed that something might happened.

Ninety-nine out of the syntactically inaccurate uses of *will* were either followed by progressives or the past tense of the verbs. The rest were either preceded by the verb with the intervening *to* infinitive or a combination of two modals:

- (15) My parents will to stay with me for a few days.
- (16) I will can remember this party forever in my life.

May and *should* were the only modals in which students did not produce many inaccurate sentences.

Discussion

In the preceding section we presented the results of our analysis of the 1) frequency of modal auxiliaries in the textbook corpus, 2) the frequency of modal auxiliaries in the learner corpus, and 3) the errors in modal auxiliary usage in the learner corpus. In this section we will discuss and attempt to explain these findings.

The analysis of the textbook corpus showed that there were altogether 2,807 instances of core modals in the textbook corpus. Particularly noticeable were the large frequency gap between *can* and *will*, accounting for nearly 50% of all modals, and the other seven modals. We were interested to establish to what extent the order of occurrence of the modals matches

that found in native speaker corpora. To this end, we compared our findings with data from the British National Corpus (BNC), the corpus of Survey of English Usage (SEU), the Lancaster-Oslo/Bergen Corpus (LOB), and the Longman Grammar of Spoken and Written English (LGSWE) corpus. According to Kennedy (2002), the four most frequent modal auxiliaries in the native speaker corpora are *will*, *would*, *can* and *could*, accounting for 72.7% of all modal tokens. Similarly, Coates (1983) reported *would*, *will*, *can* and *could* as the most frequent modals, accounting for 71.4 % of all modal tokens. *Will* is therefore only the second most common form (Biber, Johansson, Leech, Conrad, & Finegan, 1999), while in the textbook corpus it is the first. Likewise, *can* is only the third most common modal in the above corpora, but it the most common in the textbook corpus. An even greater discrepancy is found with the modal *could*, which is the 4th most common modal in the above corpora, but the 7th most common modal in the textbooks. *Should* is over-represented as the 3^d most common modal in the textbook corpus but (according to Kennedy 2002, and Quirk, Greenbaum, Leech, & Svartvik, 1985) it is only sixth in the major corpora. *May* is more frequent in the textbook corpus than *could* and *would*, while in the native speaker corpora this is not the case.

In summary, the order of frequency of most modals in the Malaysian textbooks does not match that found in native speaker corpora. In some cases the differences are in fact quite significant. This points to the likelihood that the textbook development was not informed by corpus data but was based, at least in part, on the intuition of the textbook writers.

When looking at the frequency of modals in the *learner* corpus, we found that it did not match that of the modals in the textbook

corpus. A significant difference was found, for example, for the modals *would* and *could*, which were among the four most frequent modals in the learner corpus but which were not very common at all in the textbook corpus. What could explain these differences? One possibility is that the frequency of occurrence in the textbooks does not match the extent to which they are explicitly dealt with; in other words, although a modal might be used in many different texts throughout the book, perhaps there is no instruction in it, or vice versa. A previous study by Khojasteh and Kafipour (2012) looked into the amount and type of instruction on all nine modals in the textbooks and found that in the case of *would* and *could* these were not explicitly dealt with at all in the textbooks. That leaves two possibilities; teachers instruct learners in this modal in class, even though it is not part of the course book (which seems unlikely), or learners are exposed to this form elsewhere, which leads them to use it more often.

On the other hand, *should* did not appear much in the learner corpus, although it was somewhat common in the textbook corpus. One of the reasons for this may be that the nature of the writing topics that the learner corpus was drawn from (see above), which did not require students to use either the obligation or the logical necessity meaning of the modal auxiliary *should*. However, further research is needed to establish why we found these discrepancies.

When we looked at learners' errors in their use of the modal auxiliaries, we found that *shall*, *can*, *would* and *could* in particular proved to be difficult. Interestingly, *shall*, *would* and *could* were the only three modals out of the nine that were not dealt with explicitly in the textbooks. For *could* and *would* we have further evidence from Khojasteh and Kafipour (2012) that they

also not taught explicitly at primary and secondary levels in Malaysian textbooks³. All this may help to explain why learners struggle with these forms. In the case of *would* and *could* we speculate that, due to the lack of explicit instruction, students did not fully learn how to differentiate between the present and the past forms of these modals. The tasks given to the learners ('describe one of the best days of your life', and the picture story task) were more likely to require learners to use the past tense form of the modals, leading to a relatively higher number of errors. However, this does not help to explain why their comparative frequency in the learner corpus is so much higher than in the textbook corpus.

Conclusion and limitations

From this study we can draw a number of conclusions, each of which carries implications for further research as well as teaching practice. One of the most worrying observations is that the textbooks in our study expose learners to input in which the frequency of the modal auxiliaries simply does not match that found in native speaker corpora. Although there are sometimes sound pedagogical reasons for emphasising or reducing the focus on a particular form, that does not appear to be an adequate explanation here. The most common forms in the native speaker corpora are *will*, *would*, *can* and *could* and there is no apparent reason, for example, why *should* is a reasonable replacement for *could*. We believe instead that our findings point to the likelihood that the development of the four textbooks in this study was not informed by corpus data but was based, at least in part,

³ 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* are taught neither at lower nor higher secondary levels.

on the intuition of the textbook writers. Unfortunately, this is (still) not uncommon. 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 use (see also Hyland 1994, Harwood, 2005). Of course, our study only looked at one (albeit important) grammatical feature, and we need be careful not to generalise our findings to the rest of the textbooks. Nonetheless, if a central grammatical feature is handled in this way, it does raise concern and further research should be done to establish whether our findings apply to other grammar and vocabulary too.

For teachers, our findings point to the need to be vigilant and, where feasible, to extend coursebooks with other materials, to give students broad exposure to target language input. Many corpus tools are now freely and easily accessible (for example the BNC; <http://www.natcorp.ox.ac.uk/>), and these can help teachers to ensure appropriate weight is given to each grammar point. Another finding is that learners' production of modal auxiliaries does not match their presentation in the textbooks in terms of frequency. Some modals that are common in the textbooks are not frequently used in the learners' writing and vice versa. Why would this be so? At this point we are unclear and further research will need to be done, for example to establish the interaction between frequency, instruction, and learners' exposure to these features outside of class. Of course, as we have pointed out above, frequency of input is only one element contributing to L2 knowledge. The amount and type of instruction play an important role as well. Interestingly, we found that those modals that learners did not receive

explicit instruction in were the same ones they produced more errors on in their writing. What this shows is the relationship between instruction and accuracy in language production and the importance for teachers to be very much aware of what is and what is not covered in the textbooks they use, and to adapt or supplement this where necessary.

There are, however, a number of limitations to our study, which we would like to acknowledge here. Firstly, not much information is available about the methods for obtaining the learner corpus. For example, official publications do not specify the precise instructions that learners were given as part of the writing tasks. Similarly, little information is known about the students themselves. Nonetheless, we feel that the sample is sufficiently large to allow us to draw conclusions on the basis of the learner corpus.

A methodological challenge is the fact that learners of course only used one of the textbooks in their schools, but the textbook corpus is an average of all four state-selected books. In other words, we are not comparing individual students' writing against the specific textbook they learned from. Although it would have been interesting to make direct comparisons, our data did not allow us to do this as the original learner corpus did not include this information. Nonetheless, we feel that this issue is not of major concern given the fact that the learner corpus includes data from students who used all four books; in other words, the average of all students' modal usage is compared to the average occurrence of the modals in all four books.

Finally, the results allow us to draw a number of conclusions, but do not allow us to definitely explain why we found these results in the first place. For example, why

was students' performance so poor on the writing tasks? Although we have made some comparisons with the results from a previous study (Khojasteh & Kafipour, 2012) which may give some of the possible reasons, a more in-depth analysis of learners' exposure to the modals, not just from the textbooks, but also in class and beyond their schools, would be beneficial. We hope our study will be a starting point for such further research in future. Furthermore, to date, the focus of most pedagogic corpus-based research has been either on international type of textbooks (e.g. Meunier & Gouverneur, 2009), or on national textbooks mainly in EFL contexts such as Germany (Romer, 2004), Hong Kong (Lam, 2010) and Taiwan (Wang & Good, 2007), to name a few. Surprisingly, however, English for General Purposes in Iran has been the exception to this rule. Aimed at filling the existing gap, this study suggests doing corpus-based studies on tertiary Iranian English textbooks in order to provide better picture of the ways in which not only modal auxiliaries but also other grammatical structures are treated in each learning cycle in the Iranian context.

References

- Arshad, A. Hassan, Mukundan, F. Kamarudin, G. Rahman, Sh., Rashid, J. & Edwin, M. (2002). *The English of Malaysian school students (EMAS) corpus*. Serdang: Universiti Putra Malaysia.
- 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.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Breen, M. (1989). The evaluation cycle for language learning tasks. In P. Rea-Dickins & K. Germaine (Eds.),

- Evaluation*. Oxford: Oxford University Press.
- Campos, M.C., Belles-Fortuno, B. & Gea-Valor, M. L. (2010). *Corpus-based approaches to English language teaching*. London: Continuum.
- Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The Grammar book: an ESL/EFL teacher's course*. Boston: Heinle & Heinle.
- Coates, J. (1983). *The semantics of the modal auxiliaries*. London: Croom Helm.
- Conrad, S. M. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34 (3), 548–60.
- Decapua, A. (2008). *Grammar for teachers: A guide to American English for native and non-Native speakers*. New Rochelle: Springer.
- De Silva, E. (1981). Forms and functions in Malaysian English: the case of modals. *SARE*, 3, 11-23.
- Ellis, R. (1997). The empirical evaluation of language teaching materials. *ELT Journal*, 5 (1), 36-42.
- Gabrielatos, C. (2005). Corpora and language teaching: Just a fling, or wedding bells? *TESL-EJ*, 8 (4), 1-37.
- Goldschneider, J. M. & Dekeyser, R. M. (2001). Explaining the natural order of L2 morpheme acquisition in English: A meta-analysis of multiple determinants. *Language Learning*, 51 (1), 1-50.
- Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes*, 4 (2), 149–161.
- Hughes, R. & Heah, C. (1993). *Common errors in English: grammar exercises for Malaysians*. Shah Alam: FajarBakti.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, England: Cambridge University Press.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, 19(1), 24-44.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes*, 13 (3), 239–56.
- 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.
- Khojasteh, L. & Kafipour, R. (2012). Non-empirically based teaching materials can be positively misleading: A case of modal auxiliary verbs in Malaysian English language textbooks. *Journal of English Language Teaching*, 5 (3), 62-72.
- Lam, P. (2010). Discourse particles in corpus data and textbooks: The case of well. *Applied Linguistics*, 31(2), 260-281.
- Lawson, A. (2001). Rethinking French grammar for pedagogy: The contribution of French corpora. In R.C. Simpson, & J.M. Swales (Eds.), *Corpus linguistics in North America: Selections from the 1999 symposium* (pp. 179–194). Ann Arbor, MI: The University of Michigan Press.
- Manaf, U. (2007). *The use of modals in Malaysian ESL learners' writing*. (Unpublished Doctoral Thesis). Serdang: Universiti Putra Malaysia.
- McEnery, T., & Kifle, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Ed.), *Academic discourse* (pp. 182-195). Harlow:

- Longman.
- Meunier, F. & Gouverneur, C. (2009). New types of corpora for new educational challenges. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 179-201). Amsterdam: John Benjamins.
- Mindt, D. (1995). *An empirical grammar of the English verb: Modal verbs*. Berlin: Cornelsen.
- 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. & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *Journal of English Language Teaching*, 4 (1), 79-89.
- Mukundan, J. & Roslim, N. (2009). Textbook representation of prepositions. *English language teaching*, 2 (4), 123-130.
- Nedkova, M. (2000). Evaluation. In M. Byram (Ed.), *Routledge encyclopedia of language teaching*. London and New York: Routledge.
- Nordberg, T. (2010). *Modality as portrayed in Finish upper secondary school EFL textbooks*. (Unpublished Master's Thesis). University of Helsinki.
- Quirk, R. S., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- Romer, U. (2004). 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. (2005). *Progressives, patterns, pedagogy. A corpus-driven approach to English progressive forms, functions, contexts and didactics*. Amsterdam: John Benjamins.
- Romer, U. (2010). Using general and specialized corpora in English language teaching: Past, Present and Future. In M.C. Campoy, B. Belles-Fortunato & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18-38). London: Continuum.
- Rosli, T. & Edwin, M. (1989). Error analysis of form four English compositions. *The English Teacher*, XVIII, 110-124. Retrieved from <http://www.melta.org.my/ET/1989/main9.html>
- Stubbs, M. (1996). *Text and corpus analysis: Computer assisted studies of language and culture*. Blackwell: Cambridge.
- Thornbury, S. (2004). *How to teach grammar*. Malaysia: Pearson Education Limited.
- Tomlinson, B. (1998). *Materials development in language teaching*. Cambridge: Cambridge University Press.
- Wang, J. & Good, R. (2007). The repetition of collocations in EFL textbooks: A corpus study. Paper presented at the fourth corpus linguistics conference held at the University of Birmingham, 27-30.
- Widdowson, H. G. (1990). *Aspects of language teaching*. Oxford: OUP.
- Willis, J. D. (1993). Grammar and lexis: some pedagogical implications. In J. M. Sinclair, G. Fox & M. Hoey (Eds.), *Techniques of description* (pp. 83-93). New York, NY: Routledge.
- Wong, I. (1983). *Simplification features in the structure of colloquial Malaysian English*. Singapore: Singapore University.