

Corpus Linguistics and English Language Teaching Materials: A Review of Recent Research

Laleh Khojasteh,

Shiraz University of Medical Sciences, Shiraz, Iran

Nasrin Shokrpour

Shiraz University of Medical Sciences, Shiraz, Iran

1. Introduction

Many researchers have recommended the use of corpus-based findings to inform material writers as to L2 teaching materials (e.g. Biber & Reppen, 2002; Conrad, 1999; 2000; Carter & McCarthy, 1995; Frazier, 2003; Holmes, 1988; Harwood, 2005; Lawson, 2001, Romer, 2010, Kennedy, 2002). It is with the help of corpus-based studies that the “scope” of certain features can be investigated (Hulstijn, 1995), and according to Barbieri and Eckhardt (2007), “corpus-based analysis is 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”. Without this type of analysis, many believe that scripted textbook language models and dialogues are frequently unnatural and inappropriate for communicative language teaching because they depict unrealistic situations and oversimplify the language. This way, they argue that textbooks cannot adequately prepare students for the types of pronunciation, language structures, grammar, idioms, vocabulary and conversational rules, routines and strategies that they will have to use in the real-world (Cullen & Kuo, 2007; Gouverneur, 2008; Nitta & Gardner, 2005; Brazil, Coulthard & Johns, 1980; Levis, 1999). Although the empirical findings of research on the reported mismatch between ESL/EFL materials and native speaker corpus are very encouraging, only a handful of studies have summed up these significant results and their pedagogical implications. This gap in literature triggered us to delve more in the area of pedagogic corpus and compare and contrast the findings of related studies.

2. Corpus Linguistics and English Language Teaching Materials

Indeed, the computerized version of corpus linguistics started to inspire many researchers when some inconsistencies were found between lexical items and grammatical structures in the corpora and those found in traditional language textbooks that were often based on intuition of the material writers. Furthermore, some of the properties of individual words and phrases, such as their frequency and collocations, were not generally open to intuition, which means that corpus analysis can help the syllabus design process. As Ellis argues, “speaking natively is speaking idiomatically, using frequent and familiar collocations, and the job of the language learner is to learn these familiar word sequences” (1997: 129).

Traditionally, however, this naturally occurring discourse (empiricism approach) did not have any place in language theories because descriptions of language and resulting theories were solely based on intuition (rationalism approach). The suspicion about even the simplest statistical evidence in relation to language dates back to one of the most famous opponents of corpus linguistics, Noam Chomsky, who believed in introspection for any insights into the structural nature of language and argued that no collection of naturally occurring discourse can ever be substantial enough to be a true representation of a language because:

Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still others because they are impolite. The corpus, if natural, will be so widely skewed that the description would be no more than a mere list (cited in Tognini-Bonelli, 2001:51).

In fact, what can be inferred from here can be summarized by Derwing's (1979) theory, which states that “performance” – which is our use of language – does not accurately reflect our internal knowledge of language, which is “competence”. From Chomsky's viewpoint, grammar is assumed to be part of competence; hence, corpus-derived data that reflects performance cannot be relevant to questions of grammaticality.

It is also worth mentioning that when Chomsky made his criticism, text analysis was handicapped by lack of fully progressed computers; hence, researchers had to limit themselves to very small entities which were more manageably analyzed by hand. Therefore, we can see that for those small samples, Chomsky's criticism was not irrelevant. With the advances of technology and the development of machine-readable corpora, however, corpus studies have started to gain popularity, as they were able to show that native-speakers' intuition could be unreliable. For example, concerning the information about the occurrence frequency of linguistic features,

there are certain aspects of language that are simply not open to intuition. Although Howarth (1998) and Widdowson (1990) believed that more frequency does not necessarily mean more significance especially in the area of problematic processing of language, Romer (2004) claims that frequency serves as the key to words or structures that are central in a language. Without them, she notes, it is difficult to decide what should be included in teaching materials. For example, in the case of nine modal auxiliary verbs, corpus analysis identified which modals should be prioritized in EFL/ESL teaching. Likewise, Kennedy (1998) pointed out the need to concentrate initial teaching on high frequency items and to grade vocabulary and structures accordingly. Conrad (2000: 59), also emphasizes the importance of frequency information for teachers because it helps them decide which items to emphasize; for example, to provide low-level students with practice with the items they are most likely to hear outside class. However, despite more than two decades of language teaching aiming at fostering natural spoken interaction and written language, instructional textbooks still neglect important and frequent features of real language users (Barbieri & Eckhardt, 2007; Carter & McCarthy, 1995; Lawson, 2001; Harwood, 2005; Hyland, 1994).

Comparison of the language in textbooks with reference corpora has resulted in development of more effective pedagogical materials (Gabrielatos, 2005). In the past two decades, several researchers have advocated the use of corpus-based findings to inform material writers about L2 teaching materials (see, for example, Holmes, 1988; Carter & McCarthy, 1995; Conrad, 1999; 2000; Lawson, 2001; Biber & Reppen, 2002; Frazier, 2003; Harwood, 2005). Lawson (2001) suggests that there are four areas of language in which corpus linguistics can provide important insights. First, corpora can provide information about the frequency of linguistic features in naturally occurring language. For example, based on the four major corpus-informed studies on modal auxiliary verbs, the four most popular modal auxiliaries in spoken and written real language use are *will*, *would*, *can* and *could*, in descending order (Kennedy, 2002; Biber et al., 1999; Quirk et al., 1985; Coats, 1983). Such insights, according to Kennedy (1998), should directly contribute to the content instruction of language-teaching pedagogy and affect the selection of what to teach, the sequencing of pedagogy, and the weight given to items or parts of the language being taught. Second, corpora can provide information about register variation, i.e. about how the use of particular linguistic features varies across different contexts and situations of use. For example, “and” is used as a clause-level connector in conversations, whereas in academic prose it is typically used as phrase-level connector (Biber et al., 1999). Third, corpus-based analyses can provide information about the salience of particular features, or as Hulstijn and de Graaff (1994) as well as Hulstijn (1995) call it, the “reliability” and “scope” of certain features. The former refers to how reliable a grammatical feature

can be in English. For example, when it comes to two-syllable adjectives, we may face many exceptions that indicate low reliability as compared to three or four syllable adjectives or adverbs that are simply followed by “more” to make comparatives (except few exceptions) which makes them highly reliable. The latter, “scope”, refers to the number of times a rule is applied. The greater the number, the greater the scope is. For example, in order to pluralize a great number of nouns in English, we add –s to the noun, which is an example of a broad scope rule (Hulstijn & De Graaff, 1994). From the perspective of language acquisition, we can, hence, hypothesize that “...those rules with high reliability and broad scope are acquired before those with low reliability and narrow scope ... we suggest selecting [the former] prior to selecting rules with narrow scope in any syllabus” (Doughty & Williams, 1998: 225). Finally, corpus linguistics can provide information about the discourse properties of particular linguistic features (e.g. collocations, lexico-grammatical associations, etc.). In terms of collocations, Stubbs (1996) strongly believes that lists of collocates cannot simply be provided by intuition alone because they are unlikely to represent reliable facts about frequency and typicality.

In short, by providing accurate descriptions of naturally occurring language and important information about the occurrence frequency of particular linguistic features, corpus-based analysis is a useful 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. Powerful computers, robust software, and large electronic collections of actual language have enabled researchers to better identify and classify the otherwise elusive structures of English as well as of many other languages (Hunston, 2002; Sinclair, 2004; Thompson & Hunston, 2006; Stubbs, 1996). Nonetheless, a great deal of work remains to be done (Thompson & Hunston, 2006).

3. Discrepancies between English Language Textbooks and Real Language Use

In the last two decades, corpus research has allowed a growing number of researchers to identify a mismatch between the language used in textbooks and that found in corpora (Biber & Reppen, 2002; Carter & McCarthy, 1995; Frazier, 2003; Gilmore, 2004 to name a few). Many researchers have demonstrated that materials developers still largely base content selection on intuition and that they neglect important and frequent features of the language spoken or written by real language users (Thornbury, 2004). According to Barbieri and Eckhardt (2007: 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”. We will now describe some studies that have identified such mismatches.

Three studies conducted by Willis (1990), Mindt (1991) and Tognini-Bonelli (2001) looked into the use of *any* in real language and found that unlike what English learners are exposed to in their textbooks, *any* is widely used with positive structures. Tognini-Bonelli (2001) with 46%, Mindt (1991) with 51%, and Willis (1990) with 42% occurrences of positive structures illustrate well how easy it is for pedagogic grammars, if not to get their facts wrong, then at least to leave some common patterns drastically unaccounted for. Similar discrepancies have been found in the results of Kennedy (1991) regarding the prepositional and adverbial uses of *between* and *through* and Berry (1994) regarding *unless* and *if not* who came to the similar conclusion regarding the mismatch between prescriptive statements and actual language use.

Comparing the use of reported speech in seven textbooks and the Longman Spoken and Written English (LSWE) Corpus, Barbieri and Eckhardt (2007) reported that textbooks neglect important information on the use of this structure in real language. They further argued that by ignoring possible variation across different situational varieties of language (e.g. casual conversation, academic writing, newspaper writing, etc.), these textbooks implicitly portray reported speech as a monolithic phenomenon, which behaves in the same way regardless of different contexts and situations of use. They concluded that the books were not empirically based because it is not clear which principles informed the textbook authors about which reporting verbs to present.

Romer (2005) studied the similarities and differences between progressives in spoken register of British National Corpus (BNC) plus speech register of the Bank of English and only spoken language of two widely used textbook series (GLN & EG 2000) in Germany. The pedagogic spoken corpus was made by the author only from the dialogues, interviews, and speech bubbles from each series which consisted of six course book volumes used from grades 5 to 10. From her first corpus-driven study on two corpora, she analyzed the form and functions of most frequent verbs (100) and then compared the 100 selected -ing forms in her textbook corpus to see how conversational the school textbooks were. Out of 100 -ing forms retrieved from the two corpora, the results showed that in GLN and EG 2000 textbooks there were absolutely no incidence of 33 and 37 high frequent -ing verb forms, respectively. The findings also revealed that textbook writers showed a one-sided picture of present progressive forms which have been over-represented in textbooks at the expense of past progressive, present perfect progressive and past

perfect progressive ones. From the subject-verb type point of view, first person subject pronouns *I* and *we* are under-represented while *you* was dominantly used in both textbooks. At the end, Romer (2005) questioned the authenticity of the language presented in these textbooks and suggested that if learners were presented with appropriate grammatical structures in line with real language use, they would encounter fewer difficulties handling relevant structures in communicative situations.

Romer (2004) identified inaccurate description of modal verb usage in an elementary textbook series used in German elementary schools when it was compared with the one-million-word British National Corpus. With regard to frequency, semantic functions and co-occurrences, she showed that there were huge discrepancies between the use of modal auxiliaries in authentic English and the English taught in German schools. Syntactically, there were incidences of overused cases of modals like *will/'ll* and *can* and underused cases of *would/'d*, *could*, *should* and *might* as compared to the BNC. Semantically, the “ability” meaning of *can* and *could* was overused in textbooks while in the BNC *could* occurred more frequently to express a “possibility” rather than an “ability” notion. One of the most remarkable results, however, was that *shall*, in its meaning of “prediction” never appears in textbooks while this is one of the most frequent meanings in the BNC. She suggests that more corpus-based work is required to be done in order to enable pupils as well as teachers to learn and teach an English language which is more authentic and closer to that of native speakers.

Similarly, 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 modals followed by the bare infinitive were dominant for almost all modal auxiliaries in the textbooks, lower secondary learners were not really exposed to other verb phrase structures, particularly structures with passive, progressive and perfect aspects. In another study, Khojasteh and Kafipour (2012) reported that Malaysian English language textbooks at secondary level have overemphasized minor semantic functions at the expense of quite frequent functions in present-day English. In the case of many modal auxiliaries (*can*, *could*, *will*, *may*, *might*, *shall*), the treatment was heavily biased towards one of the meanings that the modal auxiliaries could have. Their report, along with similar findings with regard to prepositions in the same textbook series reported by Mukundan and Roslim (2009), indicates that some of the contents of the Malaysian lower secondary textbooks might have given the students an unrepresentative picture of the way modals and prepositions are actually used.

In a 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 Finnish 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 slanted view towards the "ability" sense of *can* and *could*. "Permission" meanings with less than 10 occurrences throughout the textbooks indicate that this meaning was biased at the expense of the meaning of "possibility". Similarly, there was a noticeable mismatch between the "obligation/ necessity" meanings as well as "volition/ prediction" meanings in the textbooks and their actual usage, showing that students were not exposed to the full range of meanings that the modal auxiliaries can have.

Aiming to raise the proficiency level of Hong Kong upper-secondary students of English for tertiary study and the workplace, Lam (2010) compared the coverage of one of the most frequently used discourse particles, *well*, in 15 Hong Kong English language textbooks with its usage in real language in use. According to this study, since the target language was developed within the community (of Hong Kong English speakers), it could not be compared with an external corpus, such as the BNC. For this reason, she used an intercultural corpus of Hong Kong spoken English (HKCSE) consisting of four sub-corpora (academic, business, conversation and public) spoken by primitively English native speakers who were living and studying in Hong Kong at the time of the study, as well as Hong Kong Chinese. As for the textbooks, the used pedagogic corpus comprised only spoken text-type in general textbooks (including all four skills) and the whole text-types for oral skill oriented textbooks. The results were discussed with regard to frequency, positional preference and discourse function of *well*. The findings showed that although *well* was highly frequent in lectures, and to a lesser extent in business presentations and speech in the Hong Kong spoken English corpus, this discourse particle was surprisingly rare in the textbook database. Furthermore, while both initial and medial positions were common in the reference corpus, textbook writers portrayed a one-sided picture of the particle by emphasizing mainly the initial positions.

In their study of interactional strategies used by speakers monitoring their interlocutors' understanding (e.g. *Do you see what I mean?*) and the interlocutors checking their own understanding with the speakers (e.g. *I'm afraid I don't get your point*), Cheng and Warren (2007) found similar mismatches between English textbooks in Hong Kong and real-world examples in the Hong Kong spoken English corpus. Based on their findings, the eight highly used forms in HKCSE were namely: *I mean, right, okay, you know, alright, yeah, you see* and *yes*.

However, none of the above-mentioned speaker-initiated forms was presented in the Hong Kong English textbooks except for *okay*. Moreover, while eight hearer-initiated forms in HKCSE are *wh-question, repeating and paraphrasing, you mean, so, yeah, summing-up, yes, and sorry*, many textbooks only emphasized “sorry”, “I’m sorry and “I’m afraid”. Conversely, there were incidences of examples that were present in the textbooks but were not found in the HKCSE, such as *you are suggesting that, are you talking about, are we discussing, I’m not really following*, to name a few. This finding also matches that of Seto (2009), who studied the expressions of agreement in 5 Hong Kong secondary textbooks and those in the Hong Kong spoken English corpus. The results indicated that out of 54 expressions of agreement used in the textbook corpus, only 7 occurred in real language in use. Although no reason was identified for this mismatch, it is likely that one important factor will have been the development of the teaching materials by intuition, rather than evidence.

Taiwanese English language textbooks have also been shown to exhibit similar issues. In a corpus-based study conducted by Wang and Good (2007), verb-noun lexical collocations in the three most often used series of textbooks in senior high schools were compared with their occurrence in the BBI dictionary of English word combinations and the Oxford collocations dictionary for students of English. The findings revealed that verb-noun lexical collocations in the English language textbooks in Taiwan had not been recycled enough for the students to lead to acquisition, with many only occurring three to five times in the entire series.

4. Conclusion

In a nutshell, what we observed in this overview was a simplified functional spectrum which needs to be expanded by means of a more complex and more complete picture. Although a simple pedagogical description of any grammatical feature is perhaps aimed by any textbook author, there is a certain danger of oversimplifying these grammatical structures at more advanced levels. As stated by Barlow (cited in Romer, 2005: 285), “less frequent patterns are important in moving the language learner from intermediate to more advanced levels of proficiency”.

Generally, this type of finding points 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 noted (1991), the traditional emphasis on the grammatical paradigm has to be revisited in favor of a more syntagmatic approach to be used contextually.

Misrepresenting linguistic facts, according to Tognini-Bonelli (2001: 25), frustrates 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”.

The findings derived from corpus evidence reported above lead to the conclusion that intuition is not comprehensively reliable as a source of information about the language and, as Sampson argues (1996: 25), “it is time to change [our] intellectual direction”. Sampson (1996) further states that:

If intuition could get the facts of language as wrong as this, there seemed little purpose in continuing to pursue abstract philosophical arguments for or against the existence of innate knowledge of language. There had to be some way of encouraging with the concrete empirical realities of language, without getting so bogged down in innumerable details that no analytical conclusions could ever be drawn.

5. Further implications of the study

With regard to the findings of comparative corpus-based studies, it is recommended that teachers and educators familiarize themselves with the textbooks' content, thus identifying the particular strengths and weaknesses in textbooks already in use. In this way, they can achieve a higher degree of authenticity and effectiveness in their textbooks. This level of awareness and familiarity with textbooks' content would also help teachers to supplement their textbooks and adapt their teaching materials. This has to be considered whilst writing their lesson plans. Also, as research has shown, when a structure is introduced to students, it is important for that structure to be featured repetitively in order to enhance the students' understanding before it fades away from their memories. For example, when reading, words stand a good chance of being remembered if they have been met at least seven times over spaced intervals (Thornbury, 2004). According to Celce-Murcia and Larsen-Freeman (1999), it makes sense to recycle various aspects of the target structures over a period of time. In other words, in terms of some semantic functions that are of insufficient volume and repetition in textbooks, it is crucial that teachers create an opportunity for students to engage with the items to be learnt as many times as possible. This repetition, according to Ur (2006), helps learners consolidate their learning. This “frequency of usage” is important if the aim is to make learners communicatively more competent (Romer, 2005: 283).

The use of corpus-based findings in order to inform L2 teaching materials writers has been emphasized by many researchers because it has resulted in development of more effective materials (Gabrielatos, 2005). However, this study does not suggest making drastic changes in the textbooks in order to create a textbook that mirrors exactly the language used by native speakers. According to Romer (2005: 275), it is not even “safe” to do that. Nevertheless, sometimes the attested differences found between the language in the textbooks and real language use are too significant to be ignored, so the most salient facts reflected from natural language corpora should not be ignored in the textbooks. Furthermore, if we want to enable pupils to communicate successfully, it is important not to leave out some of the grammatical structures; hence, it is important to emphasize the rules governing both forms and meanings. Based on Mindt’s (1995) communicative grammar sense, textbook authors should focus their prime attention on distribution of forms, functions, and contexts of any grammatical feature in real language use.

References and notes:

- 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.
- Berry, R. (1994). Using Concordance Printouts for Language Awareness Training. In C. S. Li, D. Mahoney, & J. Richards (Eds.), *Exploring Second Language Teacher Development (195-208)*. Hong Kong: City University Press.
- Biber, D., & Reppen, R. (2002). What Does Frequency Have to Do with Grammar Teaching? *Studies in Second Language Acquisition*, 24, 199–208.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse Intonation and Language Teaching*. Harlow: Longman.
- Carter, R., & McCarthy, M. (1995). Grammar and the Spoken Language. *Applied Linguistics*, 16(2), 141–58.
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar Book: An ESL/EFL teacher’s course*. 2nd ed. Boston: Heinle & Heinle.
- Cheng W., & Warren, M. (2007). Checking Understandings: Comparing Textbooks and a Corpus of Spoken English in Hong Kong. Special issue: Raising Language Awareness through Analyzing Discourse in Context. *Language Awareness*, 16 (3), 190-207.
- Coates, J. (1983). *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
- Conrad, S.M. (1999). The Importance of Corpus-based Research for Language Teachers. *System*, 27, 1–18.
- Conrad, S. M. (2000). Will Corpus Linguistics Revolutionize Grammar Teaching in the 21st Century? *TESOL Quarterly*, 34(3), 548–60.
- Cullen, R., & Kuo, I-C. (2007). Spoken Grammar is the Manifestation of Systematic Grammatical Phenomena. *TESOL Quarterly*, 4(2), 361-386.
- Derwing, B. L. (1979). Against Autonomous Linguistics. In T. A. Perry (Ed.), *Evidence and Argumentation in Linguistics* (pp. 163–189). Berlin: Walter de Gruyter.

- Doughty, C., & Williams, J. (Eds.) (1998). *Focus on Form in Classroom Second Language Acquisition*. Cambridge: Cambridge University Press.
- Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.
- Frazier, S. (2003). A Corpus Analysis of *Would*-clauses without Adjacent *If*-clauses. *TESOL Quarterly*, 37(3), 443–46.
- Gabrielatos, C. (2005). Corpora and Language Teaching: Just a Fling, or Wedding Bells? *TESL-EJ*, 8(4), 1-37.
- Gilmore, A. (2004). A Comparison of Textbook and Authentic Interactions. *ELT Journal*, 58(4), 363–71.
- Gouverneur, C. (2008). Phraseology in Foreign Language Learning and Teaching: a corpus-based study of EFL textbooks. (Unpublished Doctoral Dissertation). University Catholique de Louvain, Belgium.
- Harwood, N. (2005). What Do We Want EAP Teaching Materials for? *Journal of English for Academic Purposes*, 4, 149–161.
- Holmes, J. (1988). Doubt and Certainty in ESL Textbooks. *Applied Linguistics*, 9(1), 21–44.
- Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, 19(1), 24–44.
- Hulstijn, J. H., & de Graaff, R. (1994). Under What Conditions Does Explicit Knowledge of a Second Language Facilitate the Acquisition of Implicit Knowledge? A research proposal. *AILA Review*, 11, 97–112.
- Hulstijn, J.H. (1995). Not All Grammar Rules Are Equal: Giving Grammar Instruction its Proper Place in Foreign LanguageT. In R. Schmidt (Ed.), *Attention and Awareness in Foreign Language Learning* (Technical Report No.6) (359–386). Honolulu: University of Hawaii Second Language Teaching and Curriculum Center.
- 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.
- Kennedy, G. (1991). ‘Between and Through’, The Company They Keep and the Functions They Serve. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honor of Jan Svartvik* (95-110). 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* (73-90). Amsterdam: 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), 68-78.
- 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 Simpson, R.C. & Swales, J.M., (Eds.), *Corpus Linguistics in North America. Selections from the 1999 Symposium* (179–194). Ann Arbor, MI: The University of Michigan Press.
- Levis, J. (1999). Intonation in Theory and Practice Revisited. *TESOL Quarterly*, 33(1), 37-63.
- Mindt, D. (1991). Syntactic Evidence for Semantic Distinctions in English. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honor of Jan Svartvik* (pp. 182-1096). London: Longman.

- 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.
- Mindt, D. (2002). What is a Grammatical Rule? In L.E. Breivik, & A. Hasselgren, (Eds.), *Special Issues on Aligned Corpora* (197-212). Amsterdam: Rodopi.
- 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.
- Nitta, R., & Gardner, S. (2005). Consciousness-raising and Practice in ELT Course Books. *ELT Journal*, 59(1), 3-13.
- Nordberg, T. (2010). *Modality as Portrayed in Finish Upper Secondary school EFL Textbooks: A corpus-based approach*. (Master's Thesis). University of Helsinki. Retrieved from <https://helda.helsinki.fi/handle/10138/19357>.
- 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* (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-Fortuno & M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching* (18-38). London: Continuum.
- Sampson, G. (1996). From Central Embedding to Corpus Linguistics. In J. Thomas & M. Short (Eds.), *Using Corpora for Language Research*. London: Longman.
- Seto, A. (2009). I Agree With You: A Corpus-based Study of Agreement. *The Southeast Asian Journal of English Language Studies*, 15, 41-67.
- Sinclair, J. M. (2004). *Trust the Text: Language, Corpus and Discourse*. London, England: Routledge.
- Stubbs, M. (1996). *Text and Corpus Analysis: Computer Assisted Studies of Language and Culture*. Blackwell: Cambridge.
- Thompson, G., & Hunston, S. (2006). System and Corpus: Two Traditions with a common ground. In G. Thompson and S. Hunston (Eds.), *System and corpus: exploring connections* (1-14). 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.
- Ur, P. (2006). *Grammar Practice Activities*. Cambridge: Cambridge University Press.
- Wang, J., & Good, R. (2007). *The Repetition of Collocations in EFL Textbooks: A Corpus Study*. Paper presented at the Sixteenth International Symposium and Book Fair on English Teaching in the Republic of China, Taipei, November 10, 2007.
- Widdowson, H. G. (1990). *Aspects of Language Teaching*. Oxford: Oxford University Press.
- Willis, D. (1990). *The Lexical Syllabus: A New Approach to Language Teaching*. London: Collins COBUILD.

SUMMARY

Corpus Linguistics and English Language Teaching Materials: A Review of Recent Research

Laleh Khojasteh,

Shiraz University of Medical Sciences, Shiraz, Iran

Nasrin Shokrpour

Shiraz University of Medical Sciences, Shiraz, Iran

Motivation for using corpus linguistics in English language teaching is partly related to the inconsistencies found between the use of lexical items and grammatical structures in the corpora and those in traditional language textbooks that are often largely based on the personal judgments of the materials writers. This lack of fit between the language in the textbook and authentic language use has been reported in many studies; yet, an overview of this aspect has not been carried out. The purpose of this study is twofold: 1) to discuss the importance of corpus linguistics for the development of English language teaching materials, and 2) to present a survey of studies carried out in the last ten years, with particular reference to the mismatches found between the language in various corpora such as Longman Grammar of Spoken and Written English (LSWE) and British National Corpus (BNC) and that used in Finnish EFL textbooks, Malaysian ESL textbooks and Hong Kong secondary textbooks to name a few. The implication of this study is to provide L2 teachers with useful information about pedagogical corpus and the ways in which they can make optimal use of a textbook's strong points, recognizing the shortcomings of certain exercises, tasks, or entire texts and to show how they can improve the textbook and adapt their teaching materials accordingly.

Keywords: corpus linguistics, textbooks, English language, teaching, materials development