Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts

DILIN LIU1 and PING JIANG 2

This article reports on a study that examined the effects of integrating corpus and contextualized lexicogrammar in foreign and second language teaching. The study was conducted in English as a foreign language (EFL) and English as a second language (ESL) courses at 1 Chinese university and 2 U.S. universities, involving 244 participants (236 EFL/ESL students and 8 instructors). A variety of data was collected, including students’ corpus search projects and reflection papers, teachers’ lesson plans and teaching journals, and a poststudy assessment survey. A close analysis of the data reveals several positive effects of the approach, such as improved command of lexicogrammar, increased critical understanding of grammar, and enhanced discovery learning skills. It also reveals some challenges of corpus-based lexicogrammar learning, including the daunting difficulty many students feel in corpus analysis. The study also identifies some variables influencing learners’ experience in using the approach, such as course content, student learning styles, and learning settings. Implications for pedagogy and further research are also discussed.

WITH NEW UNDERSTANDINGS ABOUT grammar arising from linguistics and applied linguistics and with the rapid advancement of educational technology, novel theories and teaching practices for grammar instruction have been proposed, including teaching grammar in discourse contexts, approaching grammar from a lexicogrammatical perspective, and using corpus data-driven learning. So far, however, there has been little empirical research on the effectiveness of integrating these theories as a unified approach (i.e., using corpus and contextualized lexicogrammar together) in grammar instruction in both foreign and second language contexts. This article reports on the study and its findings about, among other things, the approach’s benefits and challenges.

LITERATURE REVIEW

The past decade has witnessed a revived interest in grammar teaching in foreign and second language learning as evidenced by the many publications on the issue, including those demonstrating the importance and benefits of grammar instruction on students’ language acquisition (N. Ellis, 2005; R. Ellis, 2001, 2002; Philp, 2003; Yuan & Ellis, 2003) and those advocating new theories and approaches to grammar instruction (Conrad, 2000; R. Ellis, 1995; Francis, 1993; Hahn, 2000; Hinkel & Fotos, 2002; Hughes & McCarthy, 1998; Johns, 1994; Larsen-Freeman, 2002, 2003; Liu & Master, 2003; Thornbury, 2001). Of the proposed theories and approaches, three stand out: grammar teaching in discourse contexts, teaching grammar with a lexicogrammatical approach, and...
corpus or data-driven teaching. However, these three, as we will show and emphasize, are closely interrelated and should be implemented as a unified approach.

The call for teaching grammar in discourse contexts has its roots in functional grammar and is founded largely on the belief that grammar deals not only with forms but also with semantics (meaning) and pragmatics (context-appropriate use) (Celce-Murcia & Larsen-Freeman, 1999; Celce-Murcia & Olshtain, 2000; Halliday, 1994; Halliday & Hasan, 1989; Larsen-Freeman, 2002, 2003; McCarthy & Carter, 1994). Functional grammar focuses on meaning and treats grammar as a resource for language users in making meaning in a given social context. Thus, as Larsen-Freeman (2003) suggested, language form, meaning, and use should be approached as an integrated whole. The three aspects of grammar are intertwined because “a change in one will involve a change in another” (Celce-Murcia & Larsen-Freeman, p. 4). English language students should learn not only how to use correct grammatical forms but also how to use them in a meaningful and appropriate way. Traditional grammar teaching, however, focuses on grammatical forms while paying little attention to their discourse contexts, thus failing to address adequately when and why a grammatical form is used in a given context. As a result, students often do not know how to use grammatical forms meaningfully or appropriately. Teaching grammar in discourse contexts, in contrast, helps students examine and learn how given forms are used in contexts for meaningful communication.

Lexicogrammar views lexicon and grammar as two inherently connected parts of a single entity, challenging the traditional “wisdom of postulating separate domains of lexis and syntax” (Sinclair, 1991, p. 104). In this view, “a grammatical structure may be lexically restricted” (Francis, 1993, p. 142) and, conversely, lexical items often are grammatical in nature, for the use of a lexical item often has grammatical implications (Biber, Conrad, & Reppen, 1998; Conrad, 2000; Hunston & Francis, 2000). Many corpus studies have exhibited this close lexical and grammatical connection (Biber et al., 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Francis, Hunston, & Manning, 1996, 1998; Hunston & Francis). There also has been increasing evidence in applied linguistics showing the importance of contextual patterns in language use and learning (Hunston & Francis, 1998). In light of these findings, many scholars have argued for the use of a lexicogrammatical approach in language instruction (Aston, 2001; Clear, 2000; Schmitt, 2004, 2005; Sinclair, 1991). They base their argument on the fact that:

insofar as different words appear to have distinctive collocational, colligational, semantic, pragmatic and generic associations, … every word may have its own grammar in these respects, a grammar which can only be acquired through experience of its typical contextual patternings. (Aston, p. 15)

It thus follows that vocabulary learning and grammar learning often take place simultaneously and that the teaching of the two should be conducted jointly.

The suggestion of using corpus analysis in grammar teaching has resulted from rapid advancement in computer technology and corpus linguistics in recent years, which has showed unprecedented potential for language learning and teaching. Corpus concordance not only makes accessible an enormous amount of authentic language input but also creates various inductive and deductive language learning opportunities not available in the past (Aston, 2001; Conrad, 2000; Francis, 1993; Hunston, 2002; Hunston & Francis, 1998; Johns, 1994; Sinclair, 2004; Stevens, 1995). In terms of inductive learning, language learners observe grammar and vocabulary usages in concordance data, and then they discover and generalize findings about usage patterns and rules. In deductive learning, language learners use corpora either to test the rules and patterns they have learned or to classify concordance data by applying the rules and patterns. It has been argued that such learning activities, especially the inductive type, motivate students and promote discovering learning. And they are “particularly effective for the acquisition of grammar and vocabulary” because they help learners to notice and retain lexicogrammatical usage patterns better by engaging them in “deeper [language] processing” (Aston, p. 19). Francis and Johns (1998) also contended that conducting concordance analyses of recurrent collocational and colligational1 patterns leads to acquisition of more useful general grammatical rules. Furthermore, corpus data offer contextualized language, which enables learners to understand better what Larsen-Freeman (2002) called “grammar of choice” in language use, a choice that native speakers often make according to context. It is important to note, however, that corpus-driven learning may not be appropriate for beginning or low-level students due to their limited English proficiency (Aston, 2001). In fact, even for higher level students, a guided and controlled
corpus search is needed for effective learning (Aston, 2001).

Although arising from different theoretical domains, the three suggested practices (teaching grammar in context, approaching grammar from a lexicogrammatical perspective, and using corpora in teaching lexicogrammar) are inherently connected, as shown in many studies (see, e.g., Aston, 2001; Biber et al., 1998; Conrad, 2000; Hunston & Francis, 2000). Specifically, for example, lexicogrammar relies heavily on “contextual patternings” (Aston, 2001, p. 15). Identifying such patternings requires contextualized corpus analyses. Similarly, contextualized grammar teaching entails the analysis of contextualized grammar usage, an undertaking that calls for corpus study. In short, examined closely together, research findings have shown the need for a contextualized lexicogrammatical approach to grammar instruction and the useful role that corpora can play in such teaching. In other words, the integration of corpus use, lexicogrammar, and contextualization in grammar teaching is not a random proposition, but one motivated by the inherent connection and interdependency found among the three practices. However, although there have been quite a few publications introducing the use of corpora in language teaching (Aston, 2001; Flowerdew, 1996; Hunston & Francis, 1998; O’Keefe & Farr, 2003; Tribble & Jones, 1997), little empirical research has been conducted on the effectiveness of these novel theories and practices. The present study, therefore, was conducted to examine whether and to what extent these new theories and practices are applicable and effective when used as a unified approach in the classroom.

METHODOLOGY

Settings and Participants

The study was conducted at one large university in southeast China and two universities in the United States (one in the south-central and one in the southeast United States), lasting one semester at each school. English as a foreign language (EFL) and English as a second language (ESL) sites were included to allow a comparative study of the effectiveness of the new teaching approach in two distinct learning environments. The participants at the Chinese university were students enrolled in five sections of the Essentials of English course for second-year English majors (total number = 160) and their respective instructors (all Chinese). The course was designed to provide students with essential language skills, including vocabulary, grammar, reading, and writing. Participants at the southeastern U.S. university comprised two Level 5 reading and structure classes (21 students) at the school’s English Language Institute (where the highest level of classes is Level 6). Those at the south-central U.S. university comprised students enrolled in three English classes and their respective instructors. Two were composition classes for nonnative speakers of English (27 students); the other was a master’s in Teaching English to Speakers of Other Languages (MATESOL) grammar class, in which over 80% of the students were nonnative speakers of English (28 such students). The two composition classes differed somewhat in language content and skill, and the ESL students in the MATESOL class may have possessed a somewhat higher English proficiency. All participants were included in the study for two reasons: (a) They were the only ESL classes available, and (b) The number of participants in the ESL setting would otherwise have been too small, for even with these students included, the number was still significantly lower than that in the EFL setting (76 vs. 160). To limit the effect that the difference in language content might have on the study, efforts were made to align course content as closely as possible by making corpus-based lexicogrammar a focus of all the classes. Participants’ English proficiency level was generally within the intermediate to upper intermediate range, with some students in the MATESOL grammar class reaching the advanced level. No lower-level students were included because, as mentioned in the Literature Review section, corpus-based learning was deemed to be too difficult for them.

It is important to note that, before the study, no electronic corpora were used in English language teaching at any of the participating schools. The corpus used in the study was the British National Corpus (BNC; 2001) and the BNC Baby (2005). The latter was the major source for most of the students, although some classes and students in the United States had access to the BNC via a free online interface developed by Brigham Young University Professor Mark Davies.

Procedures

Prior to the study, the participating instructors underwent training on corpus use and issues such as lexicogrammar and the contextualization of grammar teaching. Since corpus use was an entirely new practice for the instructors, they received detailed, hands-on training on all of the basic corpus search functions, including
concordancing and collocation. To gain an understanding of lexicogrammar, the instructors examined many language examples in which a close connection between grammar and lexis was shown. In the process of analyzing the examples, they learned that the syntactic and semantic patternings of verb and noun phrases constitute the foundations of corpus-based lexicogrammar (Francis et al., 1996; Hunston & Francis, 2000). This knowledge, in turn, helped them understand that lexicogrammatical teaching should focus on such patternings, although not at the expense of other grammatical issues. Training also made the instructors aware of the limitations of corpus data and the danger of overgeneralizing corpus findings so they could guard against such pitfalls in their teaching. To help appreciate the need for contextualization of grammar teaching, instructors looked at many corpus examples that highlighted the importance of context in language users’ lexicogrammatical choices.

During training, researchers and instructors also spent a substantial amount of time discussing how to effectively incorporate corpora and lexicogrammar into the existing language curriculum and developing teaching strategies, sample classroom activities, and sample lessons. In the process, many meaningful exchanges occurred. Some instructors raised questions about specific technical and practical issues, as well as concerns about the general applicability of the use of corpus data in language teaching. As researchers, we made every effort to these questions and concerns seriously. We acknowledged the challenges involved in using the approach, especially the use of a corpus, and looked at existing publications on the topic to understand how other professionals had handled these issues. Finally, we explored possible solutions and/or strategies for dealing with them. The instructors commented that they found the discussions to be very useful. At the end of the training, even though we did not reach a complete consensus on every issue, we developed two complete sample lessons as well as a written framework for the new teaching approach. The framework consisted of “goals and rationales for the approach,” “guiding principles” for using the approach, and “recommended teaching strategies and techniques.”

At the end of the study, a questionnaire with a student and an instructor version (see Appendix) was given to the participants to obtain their assessment of the corpus-based lexicogrammatical teaching. Students at the Chinese university were allowed to answer the questionnaires either in English or Chinese so they could express themselves. Chinese instructors were given similar options for answering the questionnaires as well as writing their teaching journals/logs. Some students used Chinese or a mixture of both languages; all instructors, in contrast, answered the questionnaire in English, and only one wrote her teaching journal in Chinese. Data collected in Chinese were translated into English by the researchers (who were fluent in both languages).

**Data**

Major data for this study consist of the following: (a) students’ work, including corpus search assignments, grammar exercises, written reports about their corpus data analyses and findings, and reflections on their corpus studies of lexicogrammar; (b) instructors’ teaching logs, lesson plans, sample teaching activities, reflection journals, notes from instructor discussion meetings, and the authors’ discussions with the instructors; and (c) students’ and instructors’ poststudy questionnaires. The various sources of data were triangulated to enhance the validity and reliability of the results of the study. The questionnaire consisted of 10 open-ended questions on the students’ version (13 on the instructors’ version), as well as 5 Likert-scale questions. The open-ended questions sought to obtain information regarding the participants’ practices in and assessment of the use of corpora and lexicogrammar. The three extra open-ended questions on the instructors’ version dealt with their experiences using corpora as a teaching tool. With the exception of these three questions (Items 11–13), items on versions of the questionnaire were similar, although some were worded differently to reflect the instructors’ and students’ respective perspectives. The Likert-scale questions were intended to ascertain participants’ overall assessment of the new teaching approach using a 1–5 rating scale. These questions were similar on both student and teacher versions except, again, for slight wording differences (e.g., “How much have you learned from the use of corpora?” on the student version, versus “How much have your students learned?” on the teacher version). Of 242 total students participating in the study, 198 returned the questionnaire, with 148 out of 160 in China and 50 out of 76 in the United States. With the eight teacher responses (four in each setting) responses added, the total number of completed questionnaires was 206.

To help the reader better understand the study and the results, three teaching activities are presented here as additional data to show what was
actually done in the classes. Descriptions of the activities are based on observational notes and/or instructors’ teaching logs and journals and the activities are illustrated with samples of corpus data used and of work that students produced. The first activity relates to concordance data used by an instructor at the Chinese university to help her students understand the usage difference between “uninterested” and “disinterested.” Apart from a definition for the words (e.g., “disinterested” means “free from bias and self-interest”) and one example of the word “disinterested” used in the text, the students’ textbook did not give any other examples or information about the two words’ meanings and usages. Noticing that her students were having difficulty understanding the difference between the two words, the instructor decided to have her students examine some BNC data. She first put on the screen concordance data regarding the word “uninterested” (some examples are shown in Figure 1) and asked the students to identify any usage patterns. Several students stated that the word was often followed by prepositional phrases beginning with “in . . . .” The instructor acknowledged this observation and encouraged the students to look closely again for any other noticeable features. One student responded that the word was often preceded by a linking verb. The instructor confirmed the answer and then reminded the students that it was similar in structure to “be interested in,” an expression opposite in meaning that the students had already learned. Afterward, the instructor put on the screen some concordance examples of “disinterested” (see examples in Figure 1) and asked the students how this word differed from “uninterested” in its structural usage patterns. Students looked at the data for a couple of minutes with no answer. The instructor then asked the students to pay attention to the type of words that followed “disinterested” in comparison to the type typically following “uninterested.” With guidance, the students soon observed that “disinterested” was used mostly before a noun as an attributive adjectival (e.g., “disinterested observers”), whereas “uninterested” was used often as a predicative adjective (e.g., “is uninterested in . . . .”). Finally, the instructor asked the students whether they could usually find such usage information in a dictionary; many students answered “no.”

The second sample activity was one used in the composition classes at the south-central U.S. university to help students address the lexicogrammatical errors they made in their writing. When the instructors noticed errors in students’ writing they believed students should be able to correct with the help of corpus data, they would mark such errors and have students work on them during class in a computer lab. A scanned worksheet completed by a student is provided in Figure 2. As shown in this example, to complete the worksheet, a student needs (a) to list the lexicogrammatical problems that his/her instructor has marked; (b) to find examples from the BNC that use each lexicogrammatical item in the desired way and write one example down on the worksheet; and (c) to rewrite his/her original sentence using the information learned from the corpus. We can tell from

---

**FIGURE 1**
Sample of British National Corpus (2001) Concordance Examples for the Use of Uninterested and Disinterested

1. of ordinary Hong Kong residents usually uninterested in politics. In sharp contrast to Britain’s
2. this is now the main task. Mr Kinnock is uninterested in finding a visionary “big idea” for the p
3. Difference between the interested and the uninterested. At the same time the effect of television’s
terested (Chapter 2). Those who remained uninterested in politics reacted by avoiding the new
4. e Asian shop steward) which was not only uninterested in their struggles but actively opposed
dates are concerned, a conscientious and uninterested student may very well do better than
5. in his absence, he would have been upset and uninterested in eating. It was sufficient for him

1. Derrida’s admirers, and perhaps some disinterested observers, would reject such charges as
2. believe that literature was simply a matter of disinterested individual response; it was an inde
3. Well, A true gentleman, valorous in arms, Disinterested and honourable. Then fled: That was
get, for it can not stand up by itself to disinterested scrutiny. Almost everything is wrong with
5. Some of that information has to be gathered by disinterested investigators, not by politicians
6. Nikol’skaia Komsomol; Yakovlev was a more disinterested outsider. Makarenko wrote: “The lo
centrists and cross-benchers who had a belief in disinterested enquiry and collecting and weig
7. and if Prof Paulos is not a wholly disinterested party in matters of numeracy, you will certainly

---
FIGURE 2
Sample of Corpus-Based Lexicogrammatical Error Worksheet

<table>
<thead>
<tr>
<th>Computer Lab Worksheet (Assignment #2)</th>
<th>Name__________________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Composition 2</td>
<td>Spring 2006</td>
</tr>
</tbody>
</table>

A. Write down the problems noted on your paper.

1. crazy mobs who contributed to ostracize her
2. bored and boring
3. interested and interesting

B. Find a sentence on the corpus that uses the word/phrase (for each of the sentences you wrote above) in the desired way. Write the sentence below.

1. Léon makes a special case for Debord as a film-maker whose aim was to contribute to the ultimate destruction of cinema as a spectacularist medium
2. I’m bored. Wasn’t that a damn boring game!
3. Helen is not interested in making lists about her life. Informal admissions are also interesting.

C. Describe how this word/phrase is used.

1. After the "contribute to": Most of sentence has noun, not a verb phrase
2. If something needs to be bored, it should be "passive position. However, boring accompanies something active?"
3. Same pattern applies to "Interested" and "Interesting." "Interested" for a passive thing and "Interesting" for an active thing.

D. Rewrite your sentences using the information that you learned from the corpus

1. crazy mobs who contributed to ostracizing her.
2. The students are bored because of boring class.
3. The audiences were interested after the singer made the show interesting.

the sample worksheet the first problem that the instructor wanted this student to address was the incorrect use of “to” in “contribute to.” (The student had mistakenly used it as an infinitive sign instead of a preposition.) The second and third problems dealt with the difference between present and past participles used as adjectives. (This was an issue that, according to the instructor, many students in the class had difficulty with). It is clear from the student’s work that the problems were accurately addressed using the corpus information obtained.

The third activity was often conducted in a Level 5 reading and structure class at the southeastern
** Deplete (v) : line 139

* To reduce sth by a large amount so that there is not enough left

Example:

1. Food supplies were severely depleted.
2. The stock is expected to be depleted before new supplies can be found.

<table>
<thead>
<tr>
<th>REGISTER</th>
<th>SPOKEN</th>
<th>FICTION</th>
<th>NEWS</th>
<th>ACADEMIC</th>
<th>NONFIC MISC</th>
<th>OTHER MISC</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOKENS</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>15</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>SIZE (MW)</td>
<td>10.33</td>
<td>16.19</td>
<td>10.64</td>
<td>15.43</td>
<td>16.63</td>
<td>28.39</td>
</tr>
<tr>
<td>PER MIL</td>
<td>0.0</td>
<td>0.0</td>
<td>0.4</td>
<td>1.0</td>
<td>0.7</td>
<td>0.4</td>
</tr>
</tbody>
</table>

U.S. university's English Language Institute. In working on a new reading passage in class, students were asked to first find the meanings, usage patterns, and register information of the new words in the passage by using a dictionary and the BNC. Figure 3 shows a scanned copy of the work one student produced concerning the meaning and the usage patterns of the word "deplete" found in an article ("The Orphaned Swimming Pool") on which the class was working. In addition to a definition and two example sentences using "deplete," the student's work also included a printout of a chart generated from a BNC cross-register comparison regarding the use of the verb. Based on the student's work, it is clear that the student learned from the BNC data that "deplete" was used mostly in writing, especially academic writing (15 tokens), and seldom in speaking (0 tokens). Such register-specific usage information is very important for second language students learning how to use a word appropriately and in the right context.

RESULTS AND DISCUSSION

An analysis of all of the data has identified some beneficial effects, some challenges, and some useful strategies, in addition to factors that may influence learners' experience in corpus-based lexicogrammar learning. In order to avoid repetition and to make the discussion more cogent, the results of the data (except the Likert-scale questions) are discussed together and organized by themes of the findings rather than by the survey questions. The results of the Likert-scale questions are presented separately after the discussion of the other data, as these questions provided a general assessment and were quantitative in nature. As most of the open-ended questions on the student and instructor versions were the same with only slight wording differences, the discussion of the findings from the questionnaire will cover both the students' and teachers' responses with each type of source specified when necessary.

Positive Effects

Enhanced Language Awareness and Better Command of Lexicogrammatical Rules/Patterns. The first benefit of the use of the new teaching approach was enhanced language awareness and a better command of some lexicogrammatical usage rules and patterns, as reported by many of the participants. In answer to the first item on the questionnaire regarding what were the most useful and valuable things they learned in the course, a majority of the participants (118, or 59.6%)
mentioned word and structural usage patterns. More importantly, in answer to Items 3 and 4 concerning what they had learned from corpus searches and what aspects of language they found the use of corpus searches most helpful in learning, more than two thirds (138, or 69.7%) wrote that lexicogrammatical usage rules and patterns, especially differences between synonymous lexicogrammatical items, had been the most helpful. Some elaborated on their answer by saying that the corpus data helped them learn lexicogrammatical usage information they could not have gained otherwise. As one student put it, “I solved some [grammar] problems that I had for a long time [by using corpus searches].” Quite a few commented that they obtained a lot of information that they could not have found in dictionaries, a source they used to rely on heavily for lexicogrammar learning. In fact, effective learning of lexicogrammatical patterns was clearly shown in two of the three corpus learning activities mentioned (i.e., the ones on “disinterested” vs. “uninterested” and phrasal verbs vs. one-word verbs). As indirect evidence of the value of corpus research in learning lexicogrammatical patterns, the search for such patterns was the most frequently conducted type of search activity reported by the majority of the students. It was also reported as the type of search activity they would like to do more frequently in the future.

Greater Appreciation of the Importance of Context in Lexicogrammatical Choices. The use of corpus-based lexicogrammar also has led to a greater appreciation of the importance of context in language use. Over half of the participants mentioned that the study had made them understand better how lexicogrammar use is often affected by context. A similar number of the participants listed understanding the importance of context in lexicogrammar use as the most valuable thing they had learned. Furthermore, in answering the question of what they thought the role of context was in lexicogrammar use based on their learning in the course (Item 8), 182 (88.3%) participants wrote that it was “important” or “very important.” To elaborate, both on the questionnaire and in their reflection papers, some participants explained that examining lexicogrammatical rules and patterns gave them many opportunities to see firsthand how context determines people’s choice of lexicogrammatical items and how form, meaning, and use are interwoven, resulting in unique usage patterns. For example, one group of students in the south-central U.S. school examined the use patterns of “keep + noun + adjectival” versus “leave + noun + adjectival”; they found that “keep” is usually followed by a positive adjectival (i.e., used positively), as shown in examples like “kept someone alive/energetic,” whereas “leave” is often used negatively, as in “left someone dead/injured.” The finding appears to have opened the students’ eyes to the connection between form, meaning, and pragmatics and the importance of context. One member of the group wrote in the reflection paper, “Before I had only examined sentences out of context. By examining those sentences in context I was able to see why we make the word choices that we do when using ‘leave’ and ‘keep.’”

To give another example, one group’s corpus search project about the passive use of the verb “give” showed a very significant difference across BNC Baby’s four subcorpora. Out of the 1,455 tokens of passive “give,” only 98 (6.7%) were in the Spoken category, 230 (15.8%) were in Written Fiction, and 371 (25.5%) were in Written Newspaper. There were 756 (52%) in Written Academic. This finding clearly suggests the rarity of use of the passive form in spoken English but a high frequency in formal writing, and it led one member of the group to the following reflection:

I’d never given much thought to the passive voice in English before … Now I see that it is not only important to know how to use the passive voice, but also when to use it, which is something I had never considered.

The research of lexicogrammatical usage patterns across registers (i.e., subcorpora) that many of the students conducted was especially helpful in enhancing their understanding of the importance of context in grammar. Many students commented that they appreciated the BNC search capability because it provided them with the contextualized usage information they could not gain otherwise. As one student stated, “The four types [subcorpora] also can help me develop my sense for situational use which I have not been able to acquire from any book and instruction.” Corpora also provide contextual information that many students believe helps them understand language usage better. For example, one student conducted a corpus search about the phrasal verb “back up” in the different subcorpora and found that in the written language, especially academic and journalistic writing, it was used mostly to mean “support,” but in spoken language, it often meant “move backwards or reverse.” Referring to the value of contextual information that corpora furnish, one student wrote in the questionnaire, “It [corpus data]
Increased Critical Understanding of Grammar. Another positive impact of the use of corpora and lexicogrammar is an increased critical understanding of grammar on the part of both the students and the instructors. One student wrote, in answer to the question about the most useful and valuable things learned (Item 1):

Before the course, I learned traditional grammar from my Korean teachers. I followed their concepts [rules] and indications [explanations] without thinking why they [the rules] should be considered like that. . . . [Now I have developed] a good habit to judge whether a grammatical rule is correct . . .

Similarly, another student stated the following regarding the most valuable things gained:

Grammar can’t be taught like a math formula: applying a fixed formula to the sentences without the understanding of the meaning in context. Vocabulary can’t be just learned by memorizing without understanding how actual words are used in a sentence.

The participants’ enhanced critical understanding of grammar also can be seen in their responses to the question (Item 9) about whether their corpus research findings have challenged the traditional view about grammar being comprised of rigid rules that native speakers follow. A majority (68%) answered “yes,” 26% said “no,” and 6% did not respond. One student who answered “yes” explained, “Some traditional grammar rules tend to be broken, and especially [between the] Demographic Spoken and Written Academic [subcorpora], the rules seem to be different.” Another student wrote, “Our group [corpus] research results are somewhat different from [the description in] the textbook. As a result, I now know the grammar textbook is not always correct.” In short, a majority of the participants showed a more critical view in their response on this issue.

Promoted Discovery Learning and Made Learning More Interesting and Effective. In the questionnaire, reflection papers, and journals, quite a few students and instructors spoke highly of the discovery learning opportunities that corpus searches provided and promoted. More than a third of the students put down enhancement of discovery learning skills as one of the most useful and valuable things they learned in the course. Furthermore, quite a few students commented that they really enjoyed the discovery learning aspect of the corpus searches. One student wrote, “Comparing the different sentences and seeing how the speakers and writers used the same word or structure for varying effects was a surprisingly effective way to study not only English structure but usage by real people . . . .” Another student stated the point more explicitly: “I just can type words or phrases. Then a lot of examples come up on the screen in front of me. From the examples, I can figure out some rules of English” (emphasis added).

Quite a few students also mentioned in the questionnaire that they remember better the things they learned by conducting corpus analyses than things they learned without such analyses. A few instructors discussed this point in their teaching journals, too. One instructor at the Chinese university summarized the value of discovery learning this way:

I’ve found that using corpus searches allowed the students to infer the unique features and patterns [of the lexicogrammatical items they were learning] and then to compare their findings with the descriptions in the textbook. As a result, they gain a better understanding and better retention of what they were learning.

Several instructors at the Chinese university expressed great appreciation for the fact that corpus data could provide many authentic examples instantly for explaining difficult lexicogrammatical issues, something not possible through any other means. One Chinese instructor mentioned in her teaching log that, in the past, she had had great difficulty teaching students the use of “lest” and its following clause due to lack of good examples. Now, she said, the concordance BNC data made this task much easier. Her students were able to see a variety of examples and to learn, without too much difficulty, the different meanings of the “lest”-led clauses and the verb tense structure used in them. Another Chinese instructor also mentioned that corpus data made difficult lexicogrammatical points easier for students to understand. She said she used to have great difficulty explaining the different uses of “in time” as shown in “He arrived in time for class” and “In time, he will understand my point.” Now, instead of explaining the difference with a couple of fabricated sentences, she was able to have the students conduct a concordance corpus search of the phrase and go over the various examples to identify the difference from the two usages. After some discussion, the students seemed to have attained a good understanding of the difference.

Another positive effect of corpus searches for the instructors is that by working with students...
in their searches and reading their search assignments, they had the opportunity to observe and understand students’ discovery learning and thinking processes. For example, reflecting on how much she had learned from the way her students used corpus data to figure out the differences among modal verbs, one instructor in the south-central U.S. university wrote: “Interestingly enough, I think I learned as much as the students did, not about modals themselves but about how the students understand modals.” She then went on to elaborate on how her students viewed modals as tensed action verbs, something she said she had never previously considered.

Challenges

Our data analysis also reveals many challenges in corpus-based lexicogrammar learning. The greatest challenge to the students was how to effectively analyze concordance data to identify lexicogrammatical usage rules and patterns. Almost all the students mentioned this challenge either directly or indirectly in their answer to the question about what they considered as the greatest challenge(s) in using corpora (Item 5). Many of them stated that they often felt overwhelmed by the extremely large number of examples generated by their searches and the time required for going over and analyzing the data. The problem was sometimes further exacerbated when many of the generated examples were not relevant to their particular study question. In addition, many of the students were frustrated by the large number of unknown words in the data. Examined closely, three problems appeared as sources of the challenges. The first seems to be inadequacy in the training given to the students about how to conduct corpus searches effectively and efficiently. The second may have been a lack of more sophisticated search functions and capabilities in the existing corpus search engines and the need for more user-friendly search features. The third source of difficulty was the low level of English language proficiency or lack of vocabulary knowledge of some of the students. Yet another challenge was the lack of easy access to corpora, especially at the Chinese university, due to the limited number of computers with a corpus search program and Internet connection available to the students.

The challenges of corpus-based lexicogrammar learning are further evidenced by the fact that a few students indicated in the questionnaire that they did not feel they learned anything from the corpus searches or from their class as a whole. For example, 4 students answered “nothing,” “none,” or “having not learned anything,” in responding to the question about what they considered the most useful and valuable things they had learned (Item 1); 11 gave similar negative responses to the question about what they had learned from their corpus searches (Item 3). Obviously, the corpus-based lexicogrammatical teaching did not seem to have helped these students in their learning. It is interesting to note, however, that almost all of these students came from the same three (out of the nine) classes that participated in the study. This fact would seem to suggest that their very negative feedback could be the result of class-specific negative learning experiences. In other words, class settings and instructors might be one variable influencing students’ corpus learning. This issue will be addressed again later in discussing the Likert-scale questions.

Useful Practices and Strategies

The study has yielded some useful practices and strategies for helping students succeed in corpus-based lexicogrammar learning or corpus-based language learning in general. First, a lot of modeling on corpus searches by instructors is crucial for students’ success in their own corpus research. Many students and instructors talked about the need for such modeling based on both positive and negative experiences. When such modeling was lacking, students were found to experience great difficulty. However, although modeling is extremely helpful, it alone is not enough for students to become competent and proficient corpus users. Hands-on experience or “learning by doing” seems to be equally important. As a student research group reported, in their corpus project that compared the use of “maybe” and “perhaps” in the spoken corpus versus the written corpus, they first included the subcorpus of “fiction” as part of their written corpus. In their examination of the fiction data, however, they noticed that a substantial portion of the writing in fiction was actually dialogue, so they decided to exclude fiction from their written corpus—an informed decision based entirely on their own learning.

A second useful practice is to have students conduct search activities based on deductive learning before engaging them in inductive activities. The reason is that in deductive learning activities, the students are asked to test a lexicogrammatical rule or usage pattern that they have been taught or already know. In such an activity, all they need to do is to find examples
to confirm or reject the rule or pattern. Such an activity is much easier than an inductive one in which students must go through many examples to identify a rule or pattern by themselves. Furthermore, students’ success in deductive learning, as some instructors reported, would, in turn, give students more confidence in inductive learning, as some instructors reported, would, in turn, give students more confidence in inductive searches, thereby enhancing their interest and motivation.

Another useful practice is to have students conduct group corpus research assignments in addition to or in lieu of individual ones. There are several advantages to having students work in groups in corpus searches. First, identifying lexicogrammatical rules or usage patterns is a very demanding task. A group is generally more capable than one individual of handling the job. Second, corpus searches are time-consuming, especially the task of going through the many tokens or examples a search usually generates. When several people share the work, it becomes easier. Third, in group work, students have the opportunity to learn from one another. The perceived advantages of group work are evidenced in one student’s reflection on her corpus learning experience. She wrote, “When I was first introduced to corpus search, I found it overwhelming. There was so much to learn.” After working with her group members on a search project successfully, however, her feelings changed: “I was pleasantly surprised to find that every member of our group possessed a unique talent that allowed him/her to shine in his particular field... I actually enjoyed it [corpus research].” Of course, group work sometimes can be very difficult if there are problems with group chemistry or dynamics.

One other effective activity that can help students see the benefits of corpus use is to ask them to check their work on lexicogrammatical exercises against corpus data. There were many such exercises in the textbook used at the Chinese university, such as filling in blanks with appropriate verb particles and selecting the right synonyms. In the past, students found the exercises challenging and even frustrating because they often did not know whether their answers were correct. Now, by checking corpus data, they would often be able to figure out the right answers. Similarly, in the composition classes at the south-central U.S. university, the instructors marked their students’ lexicogrammatical errors and then had the students work on the marked errors themselves by checking how the given lexicogrammatical items were used in the BNC. Most of the time, the students learned to correct their own errors this way.

Another practical point some of the instructors learned for corpus-based teaching is that corpus searches are much more useful for working on lexicogrammatical items that have multiple meanings or functions with high-frequency use than for working on items with a single meaning and low-frequency use. The reason is that an explanation or a check of the dictionary would usually suffice for students to grasp items in the latter category. For items in the former category, however, students need many good examples to learn to understand the various meanings and functions in different contexts. Finally, to help students conduct corpus searches more effectively and learn lexicogrammar more successfully, it is a good idea to let them use dictionaries so they can check for words they do not know in the corpus and to compare their corpus findings against dictionary descriptions. Many students and instructors spoke favorably of such a practice.

**Findings From the Likert-Scale Overall Assessment Questions**

The results from the Likert-scale questions (summarized in Table 1), by and large, corroborate the aforementioned findings from the other data. Concerning Likert-scale Question 1 (how helpful the use of corpora was in lexicogrammar learning), 15 participants (7.3%) selected “Very,” 46 (22.3%) selected “Quite,” 106 (51.5%) selected “Somewhat,” 31 (15%) selected “Minimally,” and 8 (3.9%) selected “Not at all.” In other words, 80% of the participants found corpora use at least somewhat helpful, with 30% considering it quite or very helpful. Less than 4% viewed it as not helpful at all.

Regarding Question 2 (how much they felt they learned from the use of corpora), 5 participants (2.4%) selected “A great deal,” 75 (36.4%) selected “A good amount,” 100 (48.5%) selected “A little,” 18 (8.7%) selected “Minimal,” and 8 (3.9%) selected “Nothing.” Again, 87% believed they learned at least a little, with about 40% feeling they learned a good amount or a great deal. Less than 4% felt that they learned nothing. For Question 3 (whether they would use corpora in their future learning and teaching), 18 participants (8.7%) chose “Yes, very much,” 90 (43.7%) chose “Yes,” 65 (31.6%) chose “Not sure,” 24 (11.7%) chose “Probably not,” and 9 (4.4%) chose “No.” Although a majority (52.4%) of the participants said yes, the overall response to the question could not be construed as very positive because only slightly more than 30% were not sure they would use corpora again, and about 15% said
they would not use them at all. An informal interview of some participants about their answers to the question indicates two main reasons for the uncertainty and unwillingness expressed by some of the participants. First, as discussed in the section on the challenges of using corpora, concerns the amount of time and the effort that corpus analyses demand. The second reason relates to the relative lack of access to corpora.

With respect to Question 4 (the comparison of their previous and current views on how closely grammar and vocabulary are connected), 21 participants (10.2%) selected “Much closer,” 81 (39.3%) selected “Closer,” 63 (30.6%) selected “About the same,” 34 (16.5%) selected “Not quite as close,” and 7 (3.4%) selected “Not as close.” In other words, 50% now consider the relationship between the two closer or much closer than earlier. Just over 30% hold the same view as earlier, and 20% now view it as not quite as close or not as close.

The results are rather intriguing but simultaneously unsurprising. As indicated in the literature review, scholars (like the participants in this study) differ in their views on the relationship between grammar and lexicon, with one camp viewing them as two inherently connected parts of one entity and the other seeing them as two different domains. For Question 5 (in comparison with their previous view, what their current assessment of the importance of context in language use was), 38 participants (18.4%) responded “Much more important,” 73 (35.4%) responded “More important,” 70 (34%) responded “About the same,” 15 (7.3%) responded “Not quite as important,” and 10 (4.9%) responded “Not as important.” It is safe to say that the use of corpora and lexicogrammar did appear to enhance the majority of the participants’ understanding of the importance of context in language use because 54% of the participants now consider it more important than they previously thought. Only 12% thought it less important. Furthermore, of the 34% who held the same view as before the study, some did not change their view because they likely had understood the importance of context even before the study. In fact, quite a few participants stated as much in answering Item 8 in the open-ended question section, a question similar in nature to Question 5 on the Likert-scale section (see Appendix).

### Factors That May Influence Learner Experience in Corpus-Based Lexicogrammar Learning

To ascertain whether the learning environment affected the participants’ learning experience and assessment, a t-test was conducted on the EFL and ESL group’s total mean ratings of the five Likert-scale questions combined. The reason for looking at the combined mean instead of the means of the five questions individually is as follows. The five Likert-scale questions were all related (i.e., all about the effectiveness of some aspects of the corpus-based teaching approach), and they all used the same rating scale (1–5), with 5 indicating the most positive and 1 showing the least positive assessment. A look at the overall mean allows us to obtain basically the same assessment information by looking at only one single dependable variable instead of five related ones, and it leaves out some unnecessary statistical information and saves discussion space. This, in turn, allows us to examine and discuss a few other
factors (in addition to learning environment) that could have influenced the participants’ learning experience.

The results of the $t$-test (reported in Table 2 with the effect size) indicate that the ESL total mean rating is significantly higher than that of the EFL. There are three likely reasons for the more positive responses in the ESL setting. First, the students in the ESL setting had better access to corpora. Second, their classes were generally smaller, which would likely mean more individual attention from the teacher and more interactive learning opportunities. Third, one of the ESL classes was part of a master’s course in which students possessed higher English proficiency, appeared more motivated (as they showed better class attendance and participation and had a better rate of completion of the surveys than the undergraduate ESL students), and gave a higher overall assessment. In other words, the latter class’s assessment helped elevate the ESL setting’s overall mean rating and, as a result, might have been a confounding factor for the comparison between the two settings.

Regardless of whether there was a significant difference between the EFL and ESL students’ evaluations of the approach, there are obviously some important differences between EFL and ESL settings in general. In addition to variations in corpus accessibility and class size, the two settings also differ in their language environment and teachers’ experiences and backgrounds. Whereas EFL students have little exposure to target language outside of class, ample target language input is available in the ESL setting. Whereas most of the teachers in EFL are native speakers of English, the majority of ESL teachers are nonnative speakers.

Some of the characteristics of the EFL setting (e.g., lack of access to good quality corpora and large class size) pose special challenges in the implementation of the corpus-based approach, and others (e.g., lack of adequate target language input) point to the unique value of the approach because corpora offer EFL learners a source of authentic language data not available otherwise, a point commented on by many participants in the EFL setting in this study. Thus, how best to deal with the special challenge of making corpus-based learning more effective in EFL is a very important question for EFL teachers.

Based on findings from this study and previous research, there are several things that teachers can do. First, in dealing with the issue of lack of good access to corpora, a teacher sometimes can print out concordance lines about a lexicogrammatical issue he or she wants to talk about in class and give each student a hardcopy. In this way, students will have access to the data and be able to work on the problems in or out of class. Second, to help students in a large class become more engaged, small group work may be used because small group activities give students more opportunities to participate and interact with one another. In fact, as mentioned earlier, the findings of this study show that students generally prefer group corpus projects over individual ones. Third, most EFL teachers are nonnative speakers and they are often not quite sure about some lexicogrammatical usage rules and patterns because, as research has shown, nonnative-speaker teachers often do not feel confident enough about their own language abilities (Liu, 1998; Llurda, 2005). It is therefore a sensible practice for these teachers to thoroughly prepare before each class. We are not suggesting here that native-speaker teachers do not need to prepare well; rather, nonnative-speaker teachers may need to spend more time and make more detailed preparations by doing any necessary corpus searches before class on the lexicogrammatical points to be taught so they will not be caught off guard. In fact, the process of conducting such corpus analyses also gives these teachers an extra opportunity for language study, which will, in turn, further enhance their own language skills, something many nonnative-speaker teachers desire.

In addition to comparing the two settings’ responses, we ran an ANOVA to check whether there was a significant difference among the different classes. The results indicate a significant difference (see Table 3); a post hoc Tukey’s test (shown by subscripted letters attached to the class
A quick look at the results shows that four of the classes stood out, with three (EFL 2, ESL 3, and ESL 4) having noticeably higher means and one (ESL 2) showing a markedly lower mean. Of the three classes with the highest means, two (ESL 3 and 4) were grammar or grammar/reading classes, and the one with the lowest mean was a composition class. This would suggest that the language content or skill being focused on might have been a factor influencing the students’ experience.

To test whether that was the case, we did an ANOVA on the students’ mean ratings grouped by the three major language skills the classes respectively focused on: (a) composition (two classes with 22 students), (b) overall skills (the five EFL classes with 148 students), and (c) grammar and reading/grammar (two classes with 28 students). The results of the ANOVA and a post hoc Tukey’s test (reported in Table 4) indicate that the grammar group’s mean rating is significantly higher—and not significantly—than that of the composition group. Such a finding would suggest that the language skill focus of a class was a likely factor affecting the effectiveness of the corpus-based teaching approach.

Going back to Table 3, we will see that there is also a significant difference in the assessment mean among classes with the same language skill focus, as with the five EFL classes, for example. These EFL classes used the same textbook, syllabus, and facility, and the students were all second-year English majors with basically the same language proficiency. Obviously, there must have been other factors causing the significant difference among these classes, such as the teachers and students themselves. Scrutiny of the instructors’ responses to the Likert-scale questions against their own students’ seems to show a correlation between the two. The students whose teachers’ ratings were high also responded more positively on the Likert-scale questions than those whose teachers’ ratings were low. We also found the instructors’ ratings generally reflected their means) reveals where the differences lie (i.e., between which groups).
Tabelle 5 - Student Assessment by Groups Based on Instructor Attitude

<table>
<thead>
<tr>
<th>Group by Instructors’ Attitude</th>
<th>M</th>
<th>SD</th>
<th>df</th>
<th>F</th>
<th>Sig</th>
<th>$\eta^2$ (Effect Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (n + 39)</td>
<td>2.96a</td>
<td>0.68</td>
<td>2/195</td>
<td>33.65</td>
<td>0.000</td>
<td>0.31</td>
</tr>
<tr>
<td>Middle (109)</td>
<td>3.21a</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High (50)</td>
<td>3.90b</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Means (M) sharing a common subscript are not significantly different by Tukey’s test, where $p < 0.05$.

attitudes and views about the corpus-based teaching approach during the semester. The teachers whose Likert ratings were high also showed a high level of enthusiasm and very positive views about the approach in their teaching practice, journals, and the instructors’ meetings. It seemed sensible, therefore, to ascertain whether the instructors’ attitudes indeed had an impact on their students’ learning experiences and assessments of the teaching approach. We classified the instructors into three groups based on their Likert mean ratings: “Not very positive” (with a mean of 3 or lower), “Median positive” (with a mean between 3.1 and 3.99), and “High positive” (with a mean above 4). Two instructors fell into the low-positive-attitude group, 4 in the middle, and 2 in the high group. Then we ran an ANOVA on their students’ mean ratings. The test with a post hoc Tukey’s test (results reported in Table 5) reveals that the high-group students’ mean rating is significantly higher than those of the other two groups, and the mean of the middle group is higher—although not significantly—than that of the low group. The results appear to suggest that the instructors’ attitudes likely had an influence on the students’ learning experience.

Of course, the students themselves could have been an important factor in determining the effectiveness of the approach. In fact, the wide distribution of the students’ answers on the Likert-scale questions shown in Table 1 may serve as evidence of this point. On every single question, there were always a few students who selected the lowest rating (i.e., 1) and a few who chose the highest (i.e., 5). Based on the participating instructors’ observations, students who were usually more motivated responded to corpus use more positively than those who were often less motivated. Those who usually enjoyed discovery learning also appeared to like corpus searches better than those who preferred deductive learning. Such observations are also supported indirectly by the students’ survey responses to the open-ended questions. Like their responses to the Likert-scale questions, students’ responses to the open-ended questions also varied substantially, with some hailing the use of corpora as an excellent weapon for conquering their language learning problems and a few considering it a confusing and time-consuming program of little value. Typically, those who responded very positively about corpus use talked about how they enjoyed discovering rules and patterns themselves from their corpus searches. Several of those who stated they learned nothing from corpus searches, however, complained that the search results did not show or tell them anything. The latter’s complaints could have resulted either from their difficulty in seeing patterns or from their unwillingness to take the time needed to analyze their data.

**Conclusion**

This study has shown that the use of corpora and lexicogrammar can enhance students’ language awareness, improve their command of lexicogrammatical rules and usage patterns, increase their appreciation of context in language use and their critical understanding of grammar, and promote discovery learning, thus making learning more effective. The results of the study also have revealed challenges of corpus-based lexicogrammar learning, such as the daunting difficulty many students experience in sorting through large amounts of data to identify lexicogrammar rules and usage patterns, the large number of unknown words and the complexity of language found in some corpus data, the limited functions and capacity of corpus search engines, lack of access to corpora experienced by some learners, and low level of motivation shown by some learners in engaging in discovery learning. However, the findings also indicate that some effective practices and strategies can be used to overcome such challenges; for example, more modeling and group work may enhance student success and ability in conducting corpus searches, reducing the difficulty corpus data analysis presents.
Finally, the study has uncovered some variables that may affect learners’ experiences in corpus-based grammar learning, such as learning environment, the language skills being learned, instructor and student attitudes and motivation, and students’ learning styles and level of language proficiency.

**Pedagogical Implications**

Based on our research findings, language educators need to be fully aware of the challenges and variables in using corpora for lexicogrammar learning and should take them into consideration in designing and implementing corpus-based curricula. They will need to decide whether and to what extent they want to incorporate corpus-based learning according to their students’ learning objectives, language proficiency level, and so on. In addition, they should strive to lessen their students’ difficulties with corpus use by modeling and conducting well-designed training. Of course, some of the challenges highlighted in the study, such as the limited functions of search engines, will require advancements in technology. Still, teachers can make valuable suggestions for improvement in this area based on their students’ and their own experience in corpus research. In fact, some progress has been made in this regard already; for example, regarding the problem of too many difficult words in the corpus data for learners, Wible, Chien, Kuo, and Wang (2002) have developed a software program, “Lexical Difficulty Filter,” that can filter out examples that contain difficult words. Of course, more work in this area is still needed.

**Limitations and Research Implications**

This study has several limitations. First, there was a large difference in the number of participants between the two settings; also the classes in the United States did not quite match those in China with respect to the content and language skills taught. Second, due to limited resources and other factors, no formal face-to-face interview was conducted with the students or teachers, denying us the opportunity to gain some more in-depth understanding of some of the issues studied here. Third, the study did not employ language tests to measure, in quantitative terms, students’ language learning achievement. Therefore, it lacked a truly objective gauge of students’ learning gains. For future research, several lines of inquiry are helpful. One is more content- and language skill-specific research on the use of corpora to help determine their effectiveness in various specific language skill areas, such as lexicogrammar skills in reading or writing. Another is quasi-experimental research using language tests to more accurately measure the effect of the use of corpora and lexicogrammar on students’ language learning.

**ACKNOWLEDGMENTS**

We would like to thank The International Research Foundation for English Language Education (TIRF) for the research grant that made this study possible. We would also like to express our appreciation to all the students and instructors who participated in the study.

**NOTES**

1Colligational patterns refer to the grammatical environment(s) or structure(s) in which a lexical item is typically used. For example, the verb “situate” is used mostly in the passive voice form. As another example, whereas “surprising,” “astonishing,” and “shocking” are all common in the affirmative sentence structure “It’s surprising/astonishing/shocking,” only “surprising” is usually used in the negative structure “It is not surprising.” We seldom say “It’s not astonishing/shocking.”

2Created for English language learning and teaching, the BNC Baby is a subset or condensed version of the BNC. It contains 4 million word samples from the BNC and consists of four subcorpora: (a) spoken, (b) academic writing, (c) written fiction, and (d) newspapers.

**REFERENCES**


BNC Baby, version 2. (2005). Distributed by Oxford University Computer Sciences on behalf of the BNC Consortium.


APPENDIX

Poststudy Questionnaire

Note. Questions 11–13 were on the teacher version only.

I. Open-Ended Questions

1. What are the most useful and valuable things you have learned in the course?
2. What corpus search activities have you done and for what type of information?
3. What have you learned from the corpus searches that you and your class have done?
4. In learning what aspects of language have you found the use of corpora most helpful?
5. What do you think are the greatest challenge(s) in the use of corpora for English learning?
6. What types of searches would you like to do more in the future for English learning?
7. Based on your learning this semester, what do you think is the relationship between grammar and vocabulary (entirely different or closely related) and why?
8. Based on your learning this semester, what do you think is the role of context in our choice of words and grammatical structure in language use?
9. Traditional grammar (especially prescriptive grammar) views grammar as rigid rules that naïve speakers of the language follow. Have your corpus research findings challenged this view? By the same token, have your corpus findings changed your view about grammar? If yes, then how?
10. Besides what you have been provided, what additional help and resources would you like to have in the future in order to use corpora and contextualized lexicogrammar more effectively for English learning?
11. In what ways do you find corpora useful for you as an ESL/EFL teacher?
12. What are the challenges you have found in incorporating corpus-based lexicogrammar in your teaching?
13. What have you done in assessing your students’ learning of lexicogrammar?

II. Likert Questions

1) How helpful has the use of corpora been for your learning?
2) How much have you learned from the use of corpora?
   1. Nothing. 2. Minimal. 3. A little. 4. A good amount. 5. A great deal.
3) Would you like to include the use of corpora for your future English learning?
   1) Not at all. 2. Probably not. 3. Not sure. 4. Yes. 5. Yes very much.
4) Compared with your previous understanding (i.e., before the course), what is your current view about the relationship between vocabulary and grammar (i.e., how closely they are connected)?
   1. Not as close. 2. Not quite as close. 3. About the same. 4. Closer. 5. Much closer.
5) Compared with your previous understanding (i.e., before the course), what is your current view about the importance of context in determining language users’ choice of words/grammar?
   1. Not as important. 2. Not quite as important. 3. About the same. 4. More important. 5. Much more important.

Forthcoming in The Modern Language Journal, 93.2


Arja Piirainen-Marsh & Liisa Tainio. “Other-Repetition as a Resource for Participation in the Activity of Playing a Video Game.”

Andrea Castro & Cecilia Alvstad. “Conceptions of Literature in University Language Courses.”


Christopher Blake. “Potential of Text-Based Internet Chats for Improving Oral Fluency in a Second Language.”