Language Corpora for EFL Teachers: An Exploration of English Grammar through Concordance Lines

Supakorn Phoocharoensil*

Thammasat University, English Language Center Building, 2 Phrachan Road Phranakorn Bangkok 10200 Thailand

Abstract

As it is clearly seen that English is used as an international language across the world, learning the language has great significance and value (Crystal, 2003). English learning and teaching can be enhanced to a greater extent by the incorporation of data representing authentic English derived from language corpora. Learners of English as a foreign language (EFL) are often found to be far more successful when grammar is introduced through corpus-informed data, which evidently promotes inductive learning in such a way that learners not only acquire grammar by language data observation and self-discovery of rules, but also find it entertaining and exciting to make grammar rule generalizations on their own (Cheng, 2012; Hunston, 2002). Furthermore, exposure to corpus data of real English also provides EFL learners with some evidence of how grammar is used in actuality, which may be different from textbook-based rules. The researcher found that graduate Thai EFL learners were very satisfied with learning grammar through corpus-informed data.

Keywords: language corpora; English grammar; EFL teachers; concordance lines; corpus linguistics

1. Introduction

Students of English as a foreign language (EFL) can easily become bored with learning grammar in a traditional style, i.e., in a classroom setting where teachers primarily focus their instruction on presenting grammatical rules through a series of lecturers. Likewise, to some modern EFL teachers, such a teacher-
centered approach to grammar instruction often turns out to be dispreferred. These teachers seem to support learner-based instruction where students are encouraged to try their best to analyze real language data and make exciting discoveries of grammar rules from their own observation, possibly under teachers’ proper guidance (O’Keeffe, McCarthy & Carter, 2007).

Not only will EFL students enjoy learning grammar via self-discovery, they may also find that computer-based evidence, referred to in this paper as corpus-based data, can reveal certain grammatical patterns or occurrences that are unavailable in most traditional grammar textbooks (Cheng, 2012). Certain findings may contradict what has usually been accepted as a rule (Hunston, 2002). Furthermore, native speakers’ intuition on grammar is not always reliable, as there appear to be some actual uses of grammar that even native speakers are unaware of unless a large amount of language evidence is consulted.

In the next section, relevant literature regarding how corpus data can be used effectively in grammar lessons will be reviewed.

2. Review of Related Literature

2.1. Language corpora and concordance lines

Language corpora have been regarded as one of the most effective, accurate, and modern tools in language research and applied linguistics (Cheng, 2012; O’Keeffe & McCarthy, 2010). “It is no exaggeration to say that corpora, and the study of corpora, have revolutionised the study of language, and of the application of language…” (Hunston, 2002, p.1). According to Linquist (2009), corpus linguistics is viewed as a methodology, comprising several relevant methods used in the investigation of language in different aspects, e.g. sociolinguistics, syntax, semantics, pragmatics, discourse analysis, etc.

- **Defining corpora**

A corpus (plural corpora) refers to “a collection of texts, written or spoken, which is stored on a computer” (O’Keeffe, McCarthy & Carter, 2007, p. 1). Likewise, Reppen (2009) defines a corpus as “a large principled collection of naturally occurring texts (written or spoken) stored electronically” (p. 2). A corpus basically tells us what language is like. It serves as a more reliable guide to language use than native speakers’ intuition. In language teaching, a corpus can supply information about how language works, e.g. the relative frequency of grammatical patterns. For example, Mindt (2000) revealed that although the future time in English can be represented by different verb forms, e.g. be going to + V.infinitive, will + V.infinitive, present continuous (is/am/are + Ving), present simple (V1), the most frequent form in conversational English is will (or other modals) + V.infinitive. Such frequency evidence is rarely available from native speakers’ intuition (O’Keeffe et al., 2007).

Apart from frequency, corpora also provide a lexico-grammatical profile, i.e., the relationship between vocabulary and grammar. Language researchers can look at word partners, i.e., collocates, which occur most frequently and with statistical significance. For example, those who are interested in past-participle psyche-verbs, e.g. pleased, excited, surprised, interested, satisfied, scared, etc., can gain more information on these verbs as to the preposition often occurring right after them. That is, they will find that a verb like interested usually co-occurs with the preposition in, whereas shocked has by or at as its common collocates.

Corpora also indicate syntactic restrictions, i.e., grammatical patterns in which a word normally appears. For example, according to a corpus-based dictionary like the Oxford Advanced Learner’s Dictionary (2010), the verb indict is commonly used in a passive structure in the pattern be indicted for + noun, as in The senator was indicted for murder (p. 790).
Thanks to the advantages of corpora as grammatical evidence suppliers, major grammar references are nowadays corpus-informed (e.g. Biber et al., 1999; Carter & McCarthy, 2006; Carter, McCarthy, Mark & O’Keeffe, 2011), containing information that is helpful to both teachers and learners of English for clear explanations and real examples. For instance, most of the examples in English Grammar Today (Carter, McCarthy, Mark & O’Keeffe, 2011) are drawn from the Cambridge International Corpus (CIC), developed to describe English and provide examples for learners.

- Concordancing and concordance line

Concordancing, as an essential tool in a corpus study, refers to using software, known as a concordancer or a concordance program, to find all the occurrences of a particular word or phrase (O’Keeffe, McCarthy & Carter, 2007). Although there have been attempts to manually find and record every example of certain words, this endeavor has often run into problems, e.g. inaccuracy caused by human errors. At present, utilizing a concordance program enables researchers, teachers, and learners to search for a word or phrase in context with perfect accuracy in seconds (Lindquist, 2009). The search word or phrase, referred to as the node, is presented in concordance lines and appears in the center of the line with seven or eight words. There are many grammar textbooks using concordance information to present grammar to learners, encouraging them to observe the data and then find the rules accounting for the occurrences (Thornbury, 2004).

2.2 Data-driven learning (DDL)

According to Johns (2002), data-driven learning concerns the way learners develop an ability to see patterning in the target language and to make generalizations about language form and use. Put another way, acting as language detectives, learners are confronted with language data, i.e., authentic examples in the form of concordance lines, from which they are expected to formulate rules governing usage. As Hunston (2002, p. 170) remarked, this learning style is beneficial and supportive since students “are motivated to remember what they have worked to find out”.

DDL normally involves two main kinds of data presentation. The first pertains to using a raw corpus where both students and teachers will explore together. In this way, neither student nor teacher knows what discovery they will make. It is possible that a student may notice some grammatical pattern overlooked by teachers or not included in textbooks. However, Hunston (2002) pinpointed some drawbacks to this form of DDL in that teachers have very little control over what occurs in the classroom. Due to the possibly sizable amount of data, students can find certain grammatical occurrences the usage of which even teachers are not sure of, which could result in teachers’ loss of confidence as well as face.

The second type of DDL lies in a more controlled data presentation. Teachers prepare teaching materials based on selected concordance lines, so they know exactly what should be discovered in the lesson, i.e., what grammar rules learners will acquire through the corpus-based data. It is clearly seen that this particular DDL form is suitable and often effective for classrooms with time limits since the grammar lesson is taught in a controlled manner.

While arguments against DDL use in classrooms of low-level students exist, many studies have so far shown positive results of the application of DDL to high-level students (e.g. Cobb, 1997; Kennedy &Miceli, 2001; Lenko-Szymanska, 2002). As suggested by Bernadini (2000), students, especially those with high proficiency levels, reap the benefits of designing their own corpus investigations as they have the freedom to learn any language aspect that interests them. This notion is consistent with
Sripicharn (2004), who remarked that “it is important to point out to learners, especially those who have high language proficiency or native-like language competence, that they many use concordance data as a basis for their generalizations or at least to test their intuition against the authentic data” (p. 243).

Hunston (2002), in a similar vein, maintained that DDL is more beneficial to advanced learners who want to fill in gaps in their knowledge than to basic learners who need to lay down a target-language foundation. Cobb & Horst (2002) provided support for the application of DDL to English pedagogy. Through an experiment presenting a great amount of English vocabulary to EAP (English for Academic Purposes) students using concordance lines based on texts from the students’ language course, they discovered that the students were more successful in vocabulary acquisition in comparison with other teaching methods.

Cheng (2012) evidently also supports the incorporation of DDL into language studies because doing so helps link up language theories and the facts revealed by natural language. What is more, conducting language projects through corpus-based data can also enable learners as researchers to learn how to develop important learning skills, e.g. analytical reasoning, critical thinking, and problem solving, all of which are highly valued in any university program. To put it simply, students with corpus-based information are expected to think in a logical way and support ideas with well-reasoned arguments and evidence. Moreover, learning how to evaluate information and evidence critically also plays a crucial role as this is related to the development of critical thinking. As for problem-solving, Cheng (2012) claims that language projects allow learners to understand the problem, explore answers, and choose the best solution to the problem.

2.3 Grammar observations through concordance lines

English grammar seems to involve some rules that prescriptive grammarians have strictly formulated and forced others to follow. Put differently, there are rules to which even native speakers do not always conform, particularly in spoken or informal language. What is put in traditional grammar references is occasionally not what native speakers actually use in everyday life. One of the outstanding benefits of corpus-based data is that it provides real examples and evidence of English as it is really used. This paper will discuss some grammar topics that are often in apparent conflict between what is contained in traditional grammar references and what is found in language corpora.

The first English grammar topic to be discussed here is the conditional or if-clause. It has been discovered that conditionals occurring in native speaker corpora are not limited to only three classic types, commonly known as first, second, and third conditionals (Carter & McCarthy, 2006). In actuality, a wide range of patterns, i.e., over 30 possible patterns, of conditionals have been noticed. The different structures of if-clauses are adaptable to conditions of use. In particular, the most frequent pattern is the zero conditional (if + present simple + present simple/progressive) (Carter & McCarthy, 2006, p. 749). Corpus-based data also demonstrate that the occurrence of the three traditional types account for less than 50% of the entire if-clause tokens. In addition, of all the if-clause types found, the third conditional is used with the lowest frequency. The examples below in (1)-(4) represent non-traditional if-clause patterns evidenced in the corpus data:

(1) If you’re good at organising things make sure your discussions are organised and that will suit you better.

(O’Keeffe, McCarthy & Carter, 2007, p. 129)

(2) If I criticised her unfairly, then I always apologised.

(Carter & McCarthy, 2006, p. 749)
(3) If you’re suffering from hay-fever, you need one of these sprays.  
(Carter & McCarthy, 2006, p. 749)

(4) If you’re going to buy a house, then you’re going to need a lot of money. 
(Carter & McCarthy, 2006, p. 749)

In (1), the pattern of if + present simple + imperative is employed, while the use of if + past simple + past simple is represented in (2). As for (3), the present progressive occurs in the if-clause and the main clause is constituted by the present simple. As can be seen in (4), be going to is used in both the main clause and the if-clause. Carter & McCarthy (2006) claim that these occurrences of conditionals in English refer to real situations, meaning “tenses are normally used in the same way as in other kinds of sentence. Present tenses are used to talk about present and future events as well as about general truths and facts, and past tenses are used to talk about past events” (p. 749).

The corpus-informed data above clearly show that English speakers actually produce many different kinds of if-clause constructions, depending on the situation being referred to. Hence, teaching EFL students only three traditional patterns may not be adequate due to the fact that other conditional structures occur in authentic English.

The next grammatical structure to be covered here is the English relative clause with an emphasis on the relative pronouns who and whom. It is generally stated in several traditional prescriptive-grammar textbooks that the relative pronoun who occupies a subject position within a relative clause, as in (5), whereas whom has to be used as a relative clause object, as in (6).

(5) We like the lady who is very sociable.

(6) The teacher whom my friend fears is irritable.

Nevertheless, even native English speakers are often not sure about when to use the subject form who and when to use the object form whom (Linquist, 2009). Despite the rule that requires the use of whom as a relative clause object, native speaker corpora demonstrate that the object relative pronoun whom, as in (7), is often replaced by who, as in (8).

(7) …the man whom we have called the last Muddletonian,…

(8) …the woman who I meet up at you know who works in the coffee shop…

(Linquist, 2009, p. 132)

Linquist (2009) noticed that the frequency of whom decreases around the 1940s and then stays fairly stable. In contrast, “there is a slow but steady increase in who over the decades” (p. 133). It seems that who is increasingly replacing whom in the object position of the relative clause. Still, Linquist is careful in noting that in order to confirm the substitution of who for whom, it is necessary to conduct a further corpus-based study on the context in which who is used in the relative-clause object position (pp. 133-134).

3. Research Methodology
**Research question**

What are Thai EFL students’ attitudes towards corpus-based grammar teaching?

**Participants**

The participants of the present study were 17 Thai graduate students in an English for Careers program at a university in Thailand. The students were enrolled in an M.A. course entitled *Corpus Linguistics and English Language Analysis* taught by the researcher himself from 9.00-12.00 a.m. on every Sunday during June-September 2011. The participants of mixed gender were from different educational and vocational backgrounds and only one of them was working as a high school English teacher. All of them speak Thai as L1 and had more or less similar English proficiency.

**Research instruments**

The data were collected with two research tools: a questionnaire and an interview. Before both tasks were launched, the students were presented with grammar topics, i.e., conditionals and *who* vs. *whom* in relative clauses, through concordance lines. After corpus-based grammar instruction, the researcher administered the questionnaire, asking them to complete it in 15 minutes and then interviewing each of the participants for approximately 10 minutes. The questionnaire sought their opinions and attitudes towards the corpus-based teaching of English grammar. Additionally, the interview aimed to obtain more information about what the participants thought about this particular teaching method.

**4. Results**

The students taking part in this study mainly perceived the benefits of using corpus data to learn grammar. Most of them apparently had a very positive attitude towards this concordance-based information as it is considered to be a major contributing factor to success in grammar learning through an inductive method. The survey results, according to Table 1, show that they were very pleased with corpus-based grammar instruction. In particular, the majority of the EFL students in this class were of the opinion that language corpora are a very useful tool to teach grammar. Moreover, most of them thought that learning English grammar through concordance lines is better than other learning methods. They also seemed to enjoy making grammar rule discoveries based on the corpus data provided, as this enabled them to be proud of their own learning outcomes. In addition, many took the view that they could enhance their knowledge with corpus-based teaching of grammar.

| Table 1 Participants’ satisfaction with corpus-based grammar teaching and learning |
|---------------------------------------------------------------|
| 1. **Content** | **N** | **Mean** | **SD.** |
| 1.1 Corpus-based grammar teaching is useful. | 17 | 4.94 | 0.24 |
| 1.2 Corpus-based grammar learning has enhanced your knowledge. | 17 | 4.24 | 0.83 |
| 1.3 You enjoyed discovering grammar rules through concordance lines. | 17 | 4.71 | 0.47 |
| 1.4 You are proud of your grammar rule discoveries. | 17 | 4.59 | 0.51 |
| 1.4 Learning grammar through concordance lines is better than other methods. | 17 | 4.76 | 0.44 |
| **AVERAGE** | | **4.65** | **0.50** |
In addition to the quantitative results, the present study also reveals positive opinions of the participants regarding the utilization of corpus-based data in teaching English grammar. By interviewing these students, the researcher also found that they were pleased with learning and teaching grammar through concordance lines, which is consistent with the questionnaire results. Some of their opinions are presented below:

Participant 1
*I never knew there is such a useful modern tool like corpora. I learned a lot from the presented data. It was also fun to form a grammar rule using the information in concordance lines. The data suggest some aspects or uses I was unable to find in any grammar reference.*

Participant 2
*Learning how to use concordance lines in search of grammatical patterns and usage is so interesting and entertaining. I really like to look at the context of use in which each word occurs. I believe other English teachers should apply corpora to their grammar teaching rather than the old boring techniques.*

Participants 3
*Corpora allow me to dig deeper into some areas that I haven’t investigated before. Certain new grammar points are not available in textbooks used in my undergraduate level. At least I became aware that there are more than three if-clause patterns.*

Participants 4
*I never liked English grammar until corpora were introduced to me. In the past, my teachers sort of forced me to memorize tons of rules. I hated this indeed! But with concordance lines, I think I began to learn something I ignored before. I have been able to gain some new interesting grammar points through corpora.*

5. Conclusion

As indicated by the findings of this study, the EFL learners evidently enjoyed learning grammar using concordance lines, considered a trendy approach in grammar pedagogy. Not only were the students satisfied with the exploration into language data offered by corpora, they also seemed to begin to understand some grammar points they had been unsuccessful in learning through traditional means, e.g. memorization. In fact, grammar is not always something to be taught in a deductive way, i.e., rule-based teaching; on the contrary, it appears to be more challenging and worthwhile for students to come up with the rules by themselves. Teachers can serve as facilitators in the preparation of appropriate corpus-based lessons as well as by providing further explanations to students regarding English grammar.

References

Bernadini, S. (2000). *Competence, capacity, corpora.* Bologna: CLUEB.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.
Carter, R., McCarthy, M., Mark, G., & O’Keeffe, A. (2011). *English grammar today: An A-Z of spoken and written grammar*. Cambridge: Cambridge University Press.

Cheng, W. (2012). *Exploring corpus linguistics. Language in action*. Oxon: Routledge.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*(3), 301-315.

Cobb, T. & Horst, M. (2001). Reading academic English: Carrying learners across the lexical threshold. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 315-329). Cambridge: Cambridge University Press

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

Johns, T. (2002). Data-driven learning: The perpetual challenge. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus linguistics* (pp. 107-117). Amsterdam: Rodopi.

Kennedy, G. & Miceli, T. (2001). An evaluation of intermediate students’ approaches to corpus investigation. *Language Learning and Technology, 5*(3), 77-90.

Lenko-Szymanska, A. (2002). How to trace the growth in learners’ active vocabulary: A corpus-based study. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus linguistics* (pp. 217-230). Amsterdam: Rodopi.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

Mindt, D. (2000). *An empirical grammar of the English verb system*. Berlin: Cornelson.

O’Keeffe and M. McCarthy (2010), *The Routledge handbook of corpus linguistics*. London: Routledge.

O’Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.

*Oxford Advanced Learner’s Dictionary* (2010). Oxford: Oxford University Press.

Reppen, R. (2009). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.

Sripicharn, P. (2004). Examining native speakers’ and learners’ investigation of the same concordance data and its implications for classroom concordancing with ELF learners. In G. Aston, S. Bernardini & D. Stewart (Eds.), *Corpora and language learners* (pp. 233-345). Amsterdam: John Benjamins.

Sripicharn, P. (2010). How can we prepare learners for using language corpora? In A. O’Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. (pp. 371-384). London: Routledge.

Thornbury, S. (2004). *Natural grammar: the keywords of English and how they work*. Oxford: Oxford University Press.