Corpus Linguistics: Influencing and contributing to
the EFL classroom

Robinson FRITZ

本論文は英語教育の中で Corpus Linguistics が、どのように教育学に貢献し、また影響を及ぼしているのかについて考察した。初めに、Corpus Linguistics についての大まかな理論と、Corpus Linguistics が新しい一つの手段として今までの規則を変化させる方法について述べている。次に、英語の教育や学習に Corpus Linguistics が様々な方法で利益をもたらしてるいこと、また教育者にとって役立つことについて述べている。最後に、Corpus Linguistics を用いて、語彙学習時にどのようにして EFL 学生を援助できるかについて述べている。

Abstract

This paper discusses how Corpus Linguistics (hereafter known as CL) can contribute and influence the language pedagogy in ELT. Firstly I will discuss the general theory behind CL and look at areas that have changed traditional language descriptions. The second section will deal with how current teaching and learning methodologies in the ELT field can benefit from the use of corpus-based research. In the final section I will expand on the ideas developed in this paper to show how corpus-based methods can be exploited to aid EFL students vocabulary acquisition methods.

1.　What are corpora and corpus linguistics?

Essentially, a corpus is a collection of authentic texts, spoken or written, that have been stored in a computer (McCarten, 2007: 2). Corpora range in size and language and are used in fields varying from ELT to translation and forensic linguistics (Hunston, 2009: 96). In ELT corpora are representative texts made up of selected discourse that allow a quantitative and qualitative analysis (O'Keffe et al, 2007: 1), and software is used to access and process data to give information about frequency, phraseology and collocation of a language. Today due to advancements in computer technology, corpora allow researchers to conduct lexical and grammatical studies with an "unprecedented scope and complexity" (Conrad, 2000: 548). However we have to consider that corpus data within a particular corpus is a representation of the various texts contributed to that corpus. The biggest corpora such as the British National Corpus (100 million words) and the Bank of English (450 million words in 2010) give an excellent representation of the English language, but the language within a corpus is not the "real language" (Gavioli, 2005: 18). Therefore we should consider corpus findings as an insight into language use. Nevertheless research has shown that corpora can also be a beneficial pedagogical tool (e.g. Hunston, 2009; Liu and Jiang, 2009; Schmitt, 2010), and this will be the main theme explored in this paper.

CL can be described as the study of language based on the results of corpora (Kennedy cited in Aijmer and Altenberg, 1992: 1). It is described as a methodology (McEnery and Wilson, 2006: 2) that has provided a "new way of thinking about language (Leech 1992, cited in Baker 2010: 106). CL shows convincing evidence of discrepancies between actual use and traditional descriptions of grammar and therefore allows the linguistic system to be viewed in new and different ways. Conrad (2000) from her research into CL suggests that grammatical choices are not simply decisions about formal accuracy, but these choices are register specific and highly systematic (Conrad, 2000: 558). This challenges Chomsky's "phrase structure grammar" which does not account for the probable choices that speakers make depending on context. Meyer (2002: 3) suggests that generative grammar descriptions present "highly abstract and decontextualized discussions of language" and that CL shows "complexity and variation as inherent in language".

The idea of language choice being register specific is expanded by Baker (2010). Sociolinguistics and CL share a number of fundamental tenets, as both areas of linguistics "…place a great deal of emphasis on language-in-use or social context" (Baker, 2010: 8). For the language teacher, real examples of lexico-grammatical items of English within their original context ready for analysis are a beneficial characteristic of corpora. This can aid explanations about difficult lexico-grammatical questions raised by students. In the next section I will look at new views and descriptions of the English language given by CL, and their implications for ELT.

2.　Corpus linguistics and its implications for ELT

The insights deduced from corpora have contributed to pedagogical materials such as dictionaries and grammar books. Hunston (2009: 97) points out the advantage of knowing the relative frequencies of words for dictionaries. Dictionary writers can use this information to decide which sense of the word to show first. Also explicit frequencies of words used in spoken and written language are shown in corpus-based dictionaries (e.g. COBUILD, 1995). These words are also presented to the user in the colloquial use of phrases, thus facilitating student's awareness of language use in context and different registers.

Another significant area of corpus-based research to ELT is that corpora have revealed patterns in language previously not detected by introspection (Sinclair,1997). Human intuition or introspection can only go so far to analyse language; however a corpus-based analysis can investigate in great detail language's lexical patterns, structural patterns, frequency and collocation details. Similarly, Hunston also argues that even though a native speaker has the experience of the language contained in a large corpus, much of that experience remains hidden from introspection (Hunston, 2009: 20). She concludes that human intuition is a "poor guide" for providing descriptions of language in four particular areas: frequency, collocations, prosody and phraseology. From this we can deduce that the meaning of words changes depending on the 'company' that words keep. Words do not exist in isolation, and this notion challenges ELT methodologies such as rote learning where words are taught in seclusion.

2.1　Corpus-based methods influencing existing ELT methodologies

If corpora based research is to influence ELT, corpus based methods must become popular with English language learners and teachers. So we have to think of how to integrate existing methodologies with corpus-based ones. This could be done in effectively in two different ways. First, students can be given direct access to a corpus and concordancing programs. This is known as data driven learning (DDL). Hunston points out that DDL allows learners to discover facts about language themselves and "…improves general skills of using context to deduce meaning." (Hunston, 2009: 170). Using DDL methods, Hunston reports how students successfully learned lists of vocabulary items by access to concordance lines, compared to using other methods (Hunston, 2009: 171). Other research (e.g. Cobb, 1997) has demonstrated that students using DDL methods acquired vocabulary items more efficiently by using concordance lines activities than learning short definitions of words. In this study, students learnt vocabulary by doing various concordance exercises that allowed them to understand the different senses of the target vocabulary used in varying contexts. A significant gain in acquisition was caused by their exposure to concordance use. We can therefore question that the exclusive use of bilingual dictionaries for vocabulary acquisition can produce a static or inactive lexical knowledge, thus not allowing students to produce the target word in a novel context.

The second way to utilize corpora in the classroom is for the teacher to pre-select concordance lines, take out difficult vocabulary and show students predesigned phrase patterns. This can help prevent students from being overwhelmed by numerous concordance results. Reppen (2009) points out that teachers using this method can make activities to raise student awareness of linguistic patterns and use of vocabulary in different registers. Also with this method the teacher can regain control over the contents of concordance results.

3.0　Vocabulary learning methods of EFL students

Vocabulary building is an essential part of English language learning for EFL students, and vocabulary learning strategies vary from student to student. Efficient vocabulary learning strategies of EFL students are often discussed in great detail and the need to learn vocabulary use in context is advocated by many (e.g. Carter, 1998; Hirata & Hirata, 2007; Schmitt, 2008). From questionnaires given to Japanese EFL students, Schmitt discovered that the preferred method for students to discover the meaning of vocabulary was by using bilingual dictionaries. Less popular methods with students were using monolingual dictionaries, asking the teacher to paraphrase and give synonyms and also guessing the meaning from context (Schmitt, 2008: 222).

In the same survey, Schmitt found the students preferred to consolidate meanings by written repetition, verbal repetition and consolidating the meaning by connecting words with synonyms and antonyms was much less popular. He acknowledges that this is "disappointing" and recognises that "[vocabulary] activities require a deeper, more involved manipulation of information to promote more effective learning" (Schmitt, 2008: 201). Hirata and Hirata discuss the disadvantages of Japanese university student's dependency on dictionaries as their

reference source for vocabulary. They argue that dictionaries don't provide enough "highly specific examples and meaningful contexts…to improve their lexical competence" and that students find it "difficult to understand how the [target] word is used in real-life situations and which words they usually occur with" (Hirata & Hirata, 2007: 33).

Therefore using memorisation, rote repetition and dictionaries to solely acquire vocabulary is to an extent ineffective. By using these methods, learners may be able to reproduce single vocabulary items from vocabulary word lists in a test environment, but they are usually challenged when he or she has to produce the item successfully in a particular context or register. Corpus research has helped us understand that words do not exist in isolation, and so teachers and learners must be aware of this when studying a language. Carter (1998: 200) points out that "…knowing a word involves knowing how to use the word syntactically, semantically and pragmatically". Corpora offer the potential for learners to understand these properties and by using corpus-based teaching methods alongside existing vocabulary acquisition practices, students have the capability to learn, understand and produce vocabulary effectively. This will be explored further in the following section.

3.1　　Japanese university EFL students and their motivations to learn vocabulary

The EIKEN and TOEIC tests in Japan require students to have an extensive knowledge of low frequency vocabulary items and this causes difficulties in successfully learning, understanding and producing these items meaningfully within context. The students use textbooks (e.g. Aizawa, Ishikawa and Murata, 2005; Okaura and Okaura, 2009) with vocabulary definitions presented by one or two senses of the word and a couple of example sentences. Using corpus-based learning methods described in previous sections could help students successfully acquire vocabulary; however, we can conclude that the corpus should be seen as a tool among the many in the learning of a language. In the next section I will look at the benefits of using a corpus based methodology.

3.2　Pedagogical benefits of using corpora-based methods in the classroom

It is suggested that learners efficiently learn vocabulary when adequate examples are noticed and processed by learners themselves (Cobb 2003, cited in Chen 2004). We have previously seen that corpus-based learning methods can allow students to retain lexical knowledge by showing vocabulary examples in context. For learners to retain vocabulary effectively, Richards (1994) points out a number of assumptions about the nature of lexical competence, three of the most relevant to this paper are;

- Knowing a word entails knowledge of the underlying form of that word and
  the derivations that can be made from it.
- Knowing a word entails knowledge of the network of associations between
  that word and other words in the language.
- Knowing a word means knowing the semantic value of a word.

(Richards 1994: 180-181)

By using concordance tools, students have the potential to learn for themselves about the Richards' "nature of lexical competence". Active student-centred learning activities for vocabulary acquisition can be created to help learners become conscious of the disadvantages of solely using inactive methods such as looking for word definitions, memorisation, repetition etc. Concordance lines present authentic language examples to students and thus helping them to understand the use of collocations, meanings of words in different contexts, the phraseology of words.

There are many ways in which concordance materials can be developed and used in the classroom. Using the word lists found in the EIKEN and TOEIC textbooks, the teacher can prepare materials for a predetermined investigation in order to allow students to find predictable results of the target vocabulary. I will implement John's approach of "Research-Practice-Improvisation" (Johns 1991, cited in Thurstun and Candlin, 1998: 271) as a methodology to demonstrate how to teach the target vocabulary in EIKEN and TOEIC textbooks.

3.3　　Concordance-based methodology: Research-Practice-Improvisation-Create

The main rationale behind Johns' concordance based methodology is that it can help learners to develop the ability to guess the meaning and use of unknown words from context (Thurstun and Candlin, 1998: 270). Learners are first presented with multiple examples of the target vocabulary items in context in form of one-line concordances. Learners then highlight word groups surrounding the target item to discover how they are used, then learners answer questions on their use in context. Further exercises are given to the learners to ensure that they are using the target items appropriately. Finally creative and improvisational activities based on the target vocabulary are carried out. The sequence of activities can be summarised as:

- LOOK at concordances for the key term and words surrounding it, thinking
  of meaning.
- FAMILIARIZE yourself with the patterns of language surrounding the key
  term by referring to the concordances as you complete the tasks
- PRACTICE key terms without referring to the concordances.
- CREATE your own piece of writing using the terms studied to fulfil a
  particular function of academic writing.
  　　　　　　　　　(Johns 1991, cited in Thurstun and Candlin, 1998: 272)

Using this proposed method, I will now present example exercises using target vocabulary from a student vocabulary textbook.

3.4　　Concordance materials to teach target vocabulary

Due to the constraints of the space for this paper, I chose only two vocabulary items from the students vocabulary textbook "TOEIC テストボキャプラ" (Okaura and Okaura, 2009). I chose the two low frequency items: "deal with" and "submit". These items had an inadequate explanation of the meaning and reproduction and so I produced corpus-based materials using Johns' procedure of "look, familiarize, practice and create" from the COBUILD Bank of English. The first step that I took was to select and develop one-line concordances including the target items. Figures 1 and 2 illustrate a sample of concordance lines that could be used as learning materials.

determining how the new nation was to deal with the problem of differences

because it's negative... You have to deal with difficult stuff.

The initial attempt to deal with this explosive problem came into danger

to make up his mind about how to deal with the new situation.

to apply what we know to help people deal with problems that are affecting

2194935661 is more difficult to deal with as a simple string than when it

establishing ways to deal with conflicts and problems.

Figure 1: concordance lines for "deal with"

she feels she must submit to the therapist's will

International travellers must submit to sometimes elaborate procedures

Tokyo will submit an anti-nuclear resolution

the British government will submit evidence

practitioners are required to submit reports on their work

an invitation to governments to submit information to the commission

Figure 2: concordance lines for "submit"

After checking MI scores for the target items, I decided to produce concordance lines featuring high frequency lexical and grammatical collocates. High frequency lexical items that collocated with "deal with" were: *problems, problem, difficult* and *situation*. Grammatical and lexical items that had a high frequency with "submit" were: *will, must, to, information, report* and *proposal*. In the concordance line materials, these high frequency collocates were underlined with the intention for students to become familiar with the patterns of language frequently surrounding the target vocabulary.

The next stage would be to do "familiarizing" exercises. A larger and varied sample of concordance lines could be given to students so that they can familiarize themselves with the phraseology, collocations and colligations

of the target word. For example, instead of asking students to comment on lexical items that frequently collocate with "submit" students could be asked to comment on the colligations that appear before and after "submit". This could help the awareness of lexical and grammatical words appearing together in phrases. Another activity would be for students to observe the situation or context that "submit" is used. Awareness exercises into the use of "submit" in different registers could be made to help students understand why it is acceptable to say "submit a report to my boss" but unusual to say "submit a letter to my friend".

Following the "familarize" stage, "practice" exercises would be implemented for the target vocabulary. At the simplest level, gap-filling exercises focusing on the exercises that the students were asked to do in the familiarization stage could be created. Pickard (1994) points out the usefulness of gap-filling exercises as "easy to prepare and students have a sense of familiarity with this format" (Pickard 1994, cited in Thurstun and Candlin, 1998). However more innovative tasks could be created for students to practise. Figures 3 and 4 show examples of these and they are based on Johns' "gapping using multiple contexts" (Johns 1991, cited in Thurstun and Candlin, 1998: 273). These exercises encourage students to form and test hypotheses while noticing meaning and collocation. Figures 3 and 4 are intended to show a small sample of practice exercises and so the word class choices in figure 3 are limited.

Decide which adjectives, nouns, or verbs to use in the blanks in the following examples. Choose from the lists below.

ADJECTIVES:　nationwide, social, difficult

NOUNS:　　　problems, Americans, setbacks

VERBS:　　　forced, learned, tried

When party leaders _____ to deal with the money question

You have to deal with _____ stuff

Many men learned another way to deal with _____.

Settle disputes and deal with local _____ promptly

British leaders were not ready to deal with _____ as equals

They naturally _____ how to deal with athletes

Congress sought to deal with many _____ problems

students will learn how to deal with _____ problems

She was therefore _____ to deal with a set of rules and procedures

---

Figure 3: Deciding word classes

Draw lines matching parts of the sentence on the left hand side with the appropriate part on the right.

We have no choice   to submit their proposals

The time has come to   law to submit a budget by January.

President Bush will formally submit   his demands

Writers are invited   the treaty to Congress later this year

He's required by   submit letters of resignation

We have no option but to submit to   to submit a complete story

America was asked   but to submit to their blackmail

Figure 4: Join the sentences

Figures 3 and 4 show that students need an extensive knowledge of vocabulary. The concordance lines could be edited by the teacher to make the exercises easier however the "realness" of the language used in these exercises would then come into question. Therefore these concordance based exercises would be better suited for advanced learners. The final stage would be to allow students to "create". Students would write a sentence or paragraph using the target vocabulary and teachers could give a specified context to help the students.

Two benefits to using DDL methodologies over using inactive acquisition methods such as using dictionaries, repetition and memorisation are apparent. Students are encouraged to problem solve and they are encouraged to experiment using the target vocabulary, thus helping them learn the language better. Needless to say, these methods will not suit all abilities of the students, but the teacher can offer varying levels of difficulty in the materials developed. From the teachers and students that Thurstun and Candlin tried this methodology with, they report that teachers and students thought it was a "…helpful and [a] very different innovative approach to vocabulary learning" (Thurstun and Candlin, 1998: 277). However they found that producing materials for target vocabulary was time consuming and that over-exposure to concordance lines tired the students (Thurstun and Candlin, 1998: 278). Students moreover had reservations about the amount of exercises and general vocabulary expected of them to learn a few vocabulary items. However we can argue that learning vocabulary extends far beyond just learning the one target word as corpus-research has shown us of the importance of context.

4.　　Conclusion

This paper has highlighted how corpus research has challenged generalised traditional grammar rules by presenting language structure as context specific. Also corpus research has allowed us to compare native intuitions of a language with actual use and thus enriched the understanding of frequency, collocations, prosody and phraseology. From this, beneficial pedagogical tools for ELT such as DDL have been developed to allow students to discover and understand patterns of use and contexualized grammar structures within a learner-

centred methodology. This gives learners a better chance to successfully acquire and produce language, compared with using inactive learning methods such as using dictionaries.

However DDL methods of teaching are not suited for all learner levels and these factors can affect the use of corpus-based methodologies in the classroom. The main disadvantage is that students need to have extensive vocabulary knowledge to understand the meaning of the target words in context. Also, the learner-centred methodology will not be ideal for all students, especially those used to teacher-centred methods. Teachers using 'tried and tested' methods in the classroom will also need to understand the usefulness of corpus based materials in the classroom and become acquainted with the ideology of CL. We can conclude that CL can help language teachers develop new methodologies in ELT and that its use in the EFL classroom is beneficial to both the student and the teacher.

References

Aizawa, K., Ishikawa, S. and Murata, M. (2005) **JACET8000 英語**
Tokyo: Kirihara Publishing.

Baker, P. (2010) Sociolinguistics and Corpus Linguistics.
Edinburgh: Edinburgh University Press.

Carter, R. (1998) Vocabulary: Applied Linguistics Perspectives. London: Routledge.

Chen, Y. H. (2004) "The use of vocabulary in the classroom". **The Internet TESL Journal** [online].
http://iteslj.org/Techniques/Chen-Corpora.html

Cobb, T. (1997) "Is there any measurable learning from hands-on concordancing?" **System** [online]. Vol 25, 2. pp.301-315. Available from www.sciencedirect.com

Cobb, T. (2003) "Do corpus-based electronic dictionaries replace concordancers?" In Morrison, B., Green, G. and Motteram, G. (eds.) **Directions in CALL: Experience, experiments, evaluation**. Hong Kong: Hong Kong Polytechnic University Publishing. pp.179-206.

Conrad, S. (2000) "Will corpus linguistics revolutionize grammar teaching in the 21st Century?" **TESOL Quarterly** [online], Vol 34, No 3. pp. 548-560. Available from www.jstor.org

Gavioli, L. (2001) "The learner as researcher: Introducing corpus concordancing in the classroom". In Aston, G. (ed.) **Learning with corpora**. Bologna: Athelstan. pp. 108-137.

Gavioli, L. (2005) Exploring corpora for ESP learning. Amsterdam: John Benjamins

Hirata, Y. and Hirata, Y. (2007) "Independent research project with web-derived corpora for language learning". **The JALT CALL Journal**. [online]. Vol 3, 3. pp .33-48. Available from http://jaltcall.org/journal/

Hunston, S. (2009) Corpora in Applied Linguistics. 6th ed.
Cambridge: Cambridge University Press.

Jiang, P. and Liu, D. (2009) "Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts". **The Modern Language Journal** [online]. Vol 93, 1, pp. 61-78. Available from http://onlinelibrary.wiley.com/.

Johns, T. (1991) "From printout to handout." **ELR Journal**. Vol 4. pp 27-46.

Kennedy, G. (1992) "Between and through: The company they keep and the functions they serve." In Aijmer, K and Altenberg, B (eds.) **English Corpus Linguistics**. New York: Longman. pp.95-110.

McCarten, J. (2007) Teaching Vocabulary, Lessons from the corpus, lessons for the classroom. Cambridge: Cambridge University Press.

McEnery, T. and Wilson, A. (2004) Corpus Linguistics: an introduction.
Edinburgh: Edinburgh University Press.

Meyer, F. (2002) English Corpus Linguistics: an introduction.
Cambridge: Cambridge University Press.

Okaura, Y. and Okaura, H. (2009) TOEIC テストボキャブラリー Tokyo: Nan'un-do

O'Keffe, A., McCarthy, M. and Carter, R. (2007) From Corpus to Classroom: language use and language teaching. Cambridge: Cambridge University Press.

Pickard, V. (1994) "Producing a concordance-based self access vocabulary package: some problems and solutions". In Flowerdew, L. and Tong, A. (eds.) **Entering text**. Hong Kong: Hong Kong University of Science and Technology. pp.215-226.

Reppen, R. (2009) "English Language Teaching and Corpus Linguistics: Lessons from the American National Corpus." In Baker, P. (ed.) **Contemporary English Linguistics**. London: Continuum International Publishing Group. pp. 204-213.

Richards, J.C. (1994) The Context of Language Teaching. 5th ed.
Cambridge: Cambridge University Press.

Schmitt, N. (2008) "Vocabulary learning strategies." 9th ed.
In Schmitt, N. and McCarthy,M. (eds.) Vocabulary: Description, Acquisition and Pedagogy. Cambridge: Cambridge University Press. pp. 199-227.

Schmitt, N. (2010) Vocabulary in Language Teaching. 11th ed.
Cambridge: Cambridge University Press.

Sinclair, J. (1997) "Corpus evidence in language description."
In Wichmann, A., Fligelstone, S. and McEnery, G. et al. (eds.) **Teaching and Language Corpora**. New York: Addison Wesley Longman. pp. 27-39.

Thurstun, J. and Candlin, C.N. (1998) "Concordancing and the Teaching of the Vocabulary of Academic English". **English for Specific Purposes** [online]. Vol 17. 3. pp. 267-280. Available from www.sciencedirect.com.

fritz@tc.nagasaki-gaigo.ac.jp