Corpus-based Studies in L2 Teaching: Implications for an EFL Context

Hyunsook Yoon  
(The Ohio State University)


Over the last decades, corpus linguistics technology has emerged as a promising tool for second language teaching and learning. However, few attempts have been made to incorporate the corpus approach into EFL classrooms. This paper reviews the recent growing body of corpus-based research on L2 teaching and explores their pedagogical implications for an EFL context. It starts with a brief overview of the basic tenets of the corpus linguistics approach and presents the benefits of the approach to L2 pedagogy. It then reviews major corpus studies that have been done in the field of L2 teaching. The studies are presented in two areas: 1) the development of teaching materials or teacher’s use of corpora for syllabus design, and 2) the exposure of students to corpora for their own use. Based on the themes and issues identified through the review of the literature, this paper examines the possible challenges as well as strengths of incorporating the corpus approach into language classrooms in an EFL context. In doing so, this paper attempts to offer valuable insights into feasible future directions for the use of the corpus-based approach in the teaching of L2 within the context.

I.  INTRODUCTION

Over the past decades, corpus linguistics methodology has become popular in language research with an enormous development of computer technology in the ability to store and process language. A “corpus” (plural: corpora) is a large collection of oral or written texts chosen to represent a language or a language variety. Put simply, corpus-based research analyzes a corpus by using a computer language processing program (i.e., concordancer) in order to identify patterns of language use (Conrad, 2000). As Granger (1998) puts it, “fundamental is the heuristic power of automated linguistic analysis, i.e. its power to uncover totally new facts about language” (p. 4). Language description, which is the basis of language teaching, has been grounded on human intuitions rather than on empirical observation (Johns, 1991). While such intuitions are often observed to be unreliable because of their subjective
nature, the corpus-based approach has been regarded as a viable source to represent actual language use since it produces objective data about language use (Biber, 2001; Biber & Conrad, 2001; McCarthy, 2001; McEnery & Wilson, 2001; Mindt, 1996, 1997). SLA educators have recently contended that the new insights of corpus studies into language can open a new era for language pedagogy (Aston, 1995, 1997; Biber & Conrad, 2001; Conrad, 1999, 2000; Johns, 1994; McCarthy & Carter, 2001; Wichmann, 1997).

This paper reviews the recent growing body of corpus-based research on second language teaching and explores their pedagogical implications for an EFL context. It starts with a brief overview of the basic tenets of corpus studies in order to provide some background in the corpus linguistics approach. It then presents several benefits of the corpus approach to L2 pedagogy, which illuminates how the approach can inform ELT. Next, it reviews major corpus studies that have been done in the field of L2 teaching. The studies are presented in two areas: 1) the development of teaching materials or teacher’s use of corpora for syllabus design, and 2) the exposure of students to corpora for their own use. While the first area largely draws upon text analysis to inform materials development and syllabus design, the second area explores empirical undertakings of the corpus approach in the classroom or in the individual student’s learning. Based on the themes and issues identified through the review of the literature, this paper examines the possible challenges as well as strengths of incorporating the corpus approach into L2 teaching particularly in an EFL context. In doing so, this paper attempts to offer valuable insights into feasible future directions for the use of the corpus-based approach in the teaching of L2 within the context.

II. AN OVERVIEW OF THE CORPUS LINGUISTICS APPROACH

1. Basic Tenets of Corpus Linguistics Research

The conspicuous development of computer techniques for text analysis has popularized corpus linguistics methodology in language research. The basic tenets of corpus linguistics research can be described in terms of three aspects (i.e., methodological, linguistic, and pedagogical), although they are closely related to each other.

From a methodological standpoint, the corpus approach assumes that human intuitions about language use are unreliable, and thus, empirical evidence is needed (McEnery & Wilson, 2001). Corpus linguistics can be described as “the study of language based on examples of ‘real life’ language use” (McEnery & Wilson, 2001, p. 1). While researchers have observed incorrect language descriptions based on intuitions or “armchair reflections” (Johns, 1991), computerized corpora have been believed to provide more accurate analysis
of language (Biber & Conrad, 2001).

From a linguistic perspective, one central notion of corpus linguistics is that vocabulary and grammar are interrelated rather than distinctive from each other (Sinclair, 1991). Instead of looking at words in isolation as the traditional approach does, corpus-based studies look for associations between grammar and vocabulary and emphasize the typical contexts of grammatical words, which is known as “lexico-grammatical connection” or “collocation” (Biber & Conrad, 2001; Conrad, 2000). As the combination of words and grammar is a vital notion of corpus research, the term of a “lexical approach” is often used interchangeably with a corpus approach (Jabbour, 2001).

The third basic tenet of corpus studies is closely related to the view of language in corpus linguistics. That is, as the corpus approach treats grammar and vocabulary interdependently, its pedagogical viewpoint states that “the most common words and their combinations should form the basis of instruction” (Jabbour, 2001, p. 298). Corpus researchers have posited that the most commonly used linguistic items are generally the most useful for learners, and thus, teaching should be based on the empirical frequency data of the items (Kennedy, 1987a; McCay, 1980). Since the primary concern of this paper is the pedagogical implications of corpus studies, the benefits of corpus approach to L2 pedagogy are elaborated in what follows.

2. Benefits of the Corpus Approach to L2 Pedagogy

The corpus approach can play a beneficial role in L2 teaching and learning in several ways. Corpus researchers have highlighted the usefulness of frequency information of linguistic items in language teaching. Kennedy (1987a) claims that if selection of linguistic items is unavoidable due to limited class time, the focus should be on the most common ways of language use that learners are most likely to encounter outside the classroom. For example, Mindt (1996) explains that irregular verbs in English can be ordered by the frequencies of their occurrences based on corpora, and if learners acquire the most frequent ten words of the list, then it means they come to know 45% of English irregular verbs in use. For Kennedy, “words are not useful because they are frequent, but frequent because they are useful” (1987b, p. 283). Thus, he continues, materials writers and teachers first need to know which items are frequent enough to be worth teaching on the basis of empirical data. Conrad (2000) also makes the same point that a frequency-based syllabus enables teachers to focus on the most useful items for learners.

Another good use of corpora is to provide students with a number of representative examples that have a variety of uses of target words. The multiple examples offer the students a rich experience of authentic language use and extend their understanding of the words in a wide variety of contexts. Of significance is that students who are exposed to
authentic texts can more successfully transfer their learning to real communication outside class (Little, 1997).

It is also noteworthy that the notion of “discovery” or “inductive learning” plays a pivotal role in the corpus approach to language pedagogy. Researchers claim that through the investigation of corpus data, learners can make their own discoveries of the language. Johns (1991) proposes a view of learners as researchers in his data-driven approach to learning:

research is too serious to be left to the researchers: that the language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data—hence the term “data-driven learning” (DDL) to describe the approach (p. 2).

According to Johns, in the task of DDL, learners are able to discover the rules of the language inductively for themselves, which is opposed to traditional rule-based learning. Thus he states that while a learner’s role is to “discover” the language, a teacher’s role is to provide a favorable context for the discovery. Thurstun and Candlin (1998) also make the comment that a concordancing program is one of the best ways to develop teaching materials to lead students to make direct discoveries about language.

In addition, some researchers have identified the use of corpus studies as the concept of connecting form and function. Tribble (1999) argues that the corpus approach provides learners with opportunities to identify forms as well as content that are appropriate to a given genre. Likewise, Jabbour (1997) claims that the text-based corpus approach raises learners’ language awareness, which in turn contributes to their understanding of functions of linguistic items in a variety of contexts. In a later work, Jabbour (2001) makes the same point that whereas the audiolingual method and communicative language teaching focus on either meaning or form, the corpus approach generally focus on both forms and functions in L2 pedagogy.

In sum, a great number of language educators have emphasized the value of the corpus approach for L2 teaching, while corpus analysis is based on authentic language use and fosters inductive learning in the learning process. Some even assert that the corpus approach will revolutionize language teaching (e.g., Conrad, 2000).

III. A REVIEW OF CORPUS-BASED RESEARCH ON L2 TEACHING

1. Research on the Development of Teaching Materials and Classroom Activities

Many L2 educators have argued that teaching materials should be informed by corpus
Corpus-based Studies in L2 Teaching: Implications for an EFL Context 289

studies (Carter et al., 1998; McCarthy & Carter, 2001). Considering that corpus linguistics provides new insights into language description, it is no surprise that its findings can offer valuable information for materials development (Conrad, 1999, 2000; Thurstun & Candlin, 1998). Gavioli and Aston (2001) claim that empirically-based corpus data can help material developers and teachers test often incorrect intuitions and focus on frequent uses of the language, thereby resulting in better-informed materials and syllabi in language teaching. Conrad (1999) makes the same argument – “corpus-based study can lead to more principled classroom materials and activities” (p. 1). In fact, the development of materials has been regarded as the area that can benefit most from incorporating the corpus approach into language teaching.

1) Comparative Analysis of Textbooks and Corpora

While corpus linguistics has a great impact on dictionaries and grammatical description of a target language, teaching materials have not been informed by corpus research (Mindt, 1996). Admittedly, English teaching materials, particularly textbooks, have long provided artificial examples of language use due to the strong prevalence of pedagogically contrived texts over authentic texts (Johns, 1994). However, there has been a growing interest in the use of authentic materials as communicative language teaching approaches have been emphasized in language pedagogy. The corpus has been presented as a valuable resource to provide authentic materials. Many studies have attempted to employ the actual corpus data in designing teaching materials so that the textbooks can reflect more accurate language use.

Several researchers have used corpus findings to critically examine English textbooks (Holmes, 1988; Kennedy, 1987a, 1987b; Ljung, 1990; Mindt, 1996, 1997). These studies used similar methods. They compared English textbooks with major (general) corpora to investigate whether the textbooks reflect actual language use as it is shown in the corpora. Their focus was on the examination of certain lexical or grammatical items of interest in the ESL/EFL textbooks. They analyzed the frequency and the treatment of those items in both sample textbooks and standard corpora of English. From their results, most of the studies discovered considerable differences in the presentations of the items between the textbooks and the corpora. Some textbooks emphasized the items that were found less frequently in real language use, while paying less attention to the items that were actually more often used. The researchers conclude that the findings of corpus-based studies should be incorporated into the textbooks in order to provide students with authentic language use.

Those comparative studies of textbook and corpus English are remarkable for their early attempts to apply corpus insights to materials development and broaden the scope of language teaching. Nonetheless, the studies were limited in their use of corpora to only
checking the usage of a certain item rather than making use of the corpora to generate corpus-based materials.

2) Sentence-based Corpus Analysis

With the development of various kinds of corpora, there has been another line of research that has used the corpora more extensively. The studies can be divided into two areas: a) sentence-based corpus analysis, and b) genre-based corpus analysis. Whereas corpus-based studies in the 1980s and early 1990s largely concentrated on linguistic analysis at the sentence level, corpus research has now expanded to embrace a broader genre-based analysis at the discourse level (Flowerdew, 2002). It is important to remember, though, that the studies share one common characteristic of corpus research, regardless of their use of a sentence or a genre as a unit of analysis. That is, they all look for lexico-grammatical patterning of texts, which is a central concept of any corpus research. In addition, they both aim to enhance materials development and syllabus design based on the insights and findings of corpus analysis (Conrad, 1999; Gavioli and Aston, 2001). In other words, their common objective is to employ more accurate descriptions of language use by corpus analysis to create teaching materials and activities.

The first area, sentence-based corpus analysis, focuses on linguistic patterns of a text at the sentence level, while mainly drawing upon general corpora for analysis. Thus it investigates linguistic characteristics across genres and attempts to obtain general descriptions of language. One common characteristic of the studies in this area is that they attempt to suggest a sample of concordance-based materials or tasks that can be used in the classroom. For a brief introduction, a concordancer (i.e., concordance program) is an essential tool for corpus analyses and searches. It produces not only quantitative information such as frequency of occurrence, but also KWIC (Keyword-in-context) format concordance in which “the keywords are arranged one below the other down the center of the page, with a fixed number of characters of context to the left and to the right” (Johns, 1991, p. 2). Concordances can be presented in two ways in the classroom: online hands-on concordancing activities (on computer) and off-line concordance-derived exercises (on paper). In fact, the use of concordances has so far been the primary method for applying the corpus-based approach to classroom teaching.

McCay (1980) was interested in teaching verbs by use of concordance to enhance learners’ lexical competence. She suggested a series of concordance-based tasks, such as forming generalizations from concordance lines, filing-in cloze tests, and matching appropriate collocates. For McCay, “lexical competence includes knowing the probability of a word’s occurrence in a particular context” (p. 18), and she noticed the usefulness of a corpus in teaching ‘semantic, syntactic, and pragmatic’ dimensions of verbs. However, she
implies that this strategy can be helpful for an intermediate and advanced level-student.

Tribble (1991) proposed concordance-based grammar awareness-raising activities on the belief that concordances can lead learners to discover rules for themselves. He designed off-line concordance examples in order to teach the uses of prepositions and articles. A series of classroom activities were designed to move from controlled exploration to independent undertaking. The activities included: consciousness raising with a gapped text, examination of concordance examples of grammatical items, gap-filling in concordances, and discussion of the results between groups. Tribble concluded that concordance-based teaching materials can encourage students to discover how language is used in contexts by providing authentic and intensive examples of linguistic features in syllabus. Tribble also tacitly remarks that intermediate or advanced learners may benefit this approach.

Thurstun and Candlin (1998) also explored a strategy of classroom concordancing through a small study. They aimed to develop a corpus-based teaching material to engage students in discovering patterns of vocabulary in academic writing. In an attempt to make the material useful for students from any disciplines, they focused on the most important and frequent lexical items of general academic English. By means of a concordancing program, they obtained a list of the most frequent words in academic writing and drew multiple examples of the chosen items. Students then worked on this material through a series of activities, such as exploring concordances and creating their own writing based on the sample concordances. Throughout the activities, the students were encouraged to become sensitive to collocation patterns of the words. The researchers reported that the students showed positive attitudes towards the new approach to vocabulary teaching. They concluded that “materials of this type provide a rich experience of the language with insights into collocations and contextualized grammatical structures linked to opportunities to develop students’ analytical abilities” (p. 277).

As seen, this line of research is largely concerned with a micro level of analysis, such as articles, prepositions, and certain words. However, although the analysis is at a micro level, it aims not only to enhance students’ awareness of contextualized grammar, but also to expand their understanding of functions of words in a variety of contexts. The studies of this area often demonstrate a strategy of designing corpus/concordance-based materials and tasks that can be used in the classroom. In so doing, they present possibilities of using corpus approaches in L2 instruction.

3) Genre-based Corpus Analysis

The second area, genre-based corpus analysis, has recently developed genre-based approaches into the corpus analysis of texts. The studies of this area look for identification of lexico-grammatical patterns of a particular genre rather than general descriptions across
Biber conducted an exhaustive analysis of spoken and written genres by use of large English corpora. His study was built on the idea that a single dimension cannot explain textual variation among genres and that a multi-dimensional approach is necessary. Biber thus developed a “multi-feature/multi-dimensional” model of which methodology embraces both quantitative and qualitative analysis. In particular, his analysis was based on patterns of collocation as well as frequency counts of linguistic features. The main assumption of his analysis was “particular sets of linguistic features co-occur frequently in texts because they are serving some common communicative function in those texts” (p. 13). As a result, he identified linguistic co-occurrence patterns that were grouped into factors, which showed underlying functional dimensions. His study contributed to establishing a linguistic basis for stylistic variation among genres.

Since Biber’s attempt, a growing number of studies have used corpus techniques in combination with genre analysis. Since corpus methodology and genre analysis are both based on text analysis, many researchers have argued for the usefulness of integrating the two approaches (e.g., Gledhill, 2000). The emphasis of this line of research is on linguistic specification of a given type of discourse by use of corpus analysis. It is, then, no surprise that the studies usually employ specialized corpora, which are compiled from texts of the target genre. With the emphasis on textual analysis of written language in use, the corpus studies of this area have been mainly concerned with literacy skills. In particular, those studies have argued that the teaching of academic writing needs to focus on the most frequent linguistic and rhetorical features that are specific within a discipline so that students can become aware of the function of the common collocational frameworks of the target genre and thus finally improve their writing skills (Gledhill, 2000; Hyland, 2002; Jabbour, 1997, 2001; Marco, 2000; Tribble, 1999, 2002).

Jabbour (1997) argues that language teaching needs to draw on the notion of genre as a way of framing syllabus. She analyzed a corpus of 61 medical research articles, and based on the findings, illustrated an ESP syllabus design for teaching reading. In the syllabus design, she highlighted the prominent linguistic patterns of the genre as the starting point of teaching. She argued that an EAP/ESP syllabus design should be based on frequency and collocation so that learners can benefit from exposure to typical examples of the target genre. In addition, she pointed out that learners’ exposure to authentic texts in reading is highly likely to improve their writing by using the texts as a resource for their own texts. In her later work, Jabbour (2001) more explicitly explores the notion of the lexical approach to teaching academic literacy. She presents sample materials for classroom instruction by using a corpus of medical research articles. She designs a series of tasks in which students are provided with an intensive exposure to a given genre and made become aware of word
associations in the genre. She claims that learners can identify characteristics of the target genre through examinations of large sample texts and finally engage in effective reading and writing.

Tribble (1999) shared the idea that genre should be adopted as a theoretical framework for writing syllabus. He developed a framework for writing instruction that combines a genre analysis and a corpus approach based on the empirical analysis of 14 project proposal corpus. Later Tribble (2002) continues to explore the potential of the corpus-informed approach to EAP writing instruction. He points out that corpora can be particularly useful for “context knowledge” and “language system knowledge” by helping students to extend their understanding of written academic literacy. He then presents a sample of contextual analysis as well as lexico-grammatical analysis, both of which contribute to identification of a particular genre. For Tribble, ‘difficult’ texts are ‘unfamiliar’ texts; learners regard certain texts as difficult texts because they are unfamiliar with the genres of the texts. He attributes the unfamiliarity to the lack of examples of the kinds of the texts. Thus, he argues for the need of exemplar corpora that can provide the learners with invaluable resources for raising language awareness and textual patterning of target genres.

Marco (2000) investigated a corpus of 100 medical research papers to analyze the most frequent collocational frameworks of the genre. She focused on the three collocational patterns: ‘the…of,’ ‘be…to,’ and ‘a…of.’ As a result, she found that the collocational frameworks occurred more frequently in medical research papers than in other genres, which shows a specific “phraseology” of the genre. She concludes that students’ writing can be improved if they are made aware of the function of common collocational frameworks in the target genre.

Hyland (2002) examined reporting verbs in academic writing. He compiled a corpus of 80 research articles from several disciplines. For the study, the reporting verbs were categorized into three functions: research acts (e.g., observe, discover), cognition acts (e.g., believe, assume), and discourse acts (e.g., ascribe, discuss). As a result of the analysis, he found that the research act verbs accounted for 57% of the reporting verbs in the corpus. In addition, he found that there were disciplinary preferences for particular forms in reporting practices, e.g., hard sciences prefer research act reporting verbs, while social sciences favor discourse act verbs. He concludes that the teaching of academic writing needs to help learners acquire discipline-specific literacy, i.e. particular linguistic and rhetorical features common to the specific disciplines.

In summary, the studies of this area often aim to provide insights into the teaching of a particular genre, which is commonly associated with the field of EAP/ESP. Consequently, the best use of those studies seems to enhance the development of materials for teaching the target genre, but, as Flowerdew (2002) notes, their findings have yet to be transferred to
2. Research on Students’ Own Use of Corpora

As presented earlier, corpus studies in L2 learning have mostly centered on materials development and syllabus design. In other words, they are mainly concerned about how to transfer corpus-based findings to teaching materials rather than giving students opportunities to use the corpus themselves. This section reviews the studies that shift the focus to learners’ own use of the corpus.

1) Experimental Investigation of the Effect of Corpus Use on Students’ Learning

In spite of the many arguments for using corpora in language teaching, few studies have examined the effect of use of classroom concordances on students’ learning. Only a couple of empirical studies have been done in this area, which makes corpus-informed pedagogy insubstantial (Flowerdew, 2002).

Stevens (1991) conducted the first empirical study that tested the effect of a corpus-based approach to vocabulary learning. He was interested in comparing traditional gap-fillers with concordance-based tasks in students’ performance on recalling words. A group of first-year Arab university students were tested on one gap-filler and one concordance-based vocabulary test. The contents of both tests were drawn from the students’ reading textbook. While the gap filler exercises had ten words blanked out, concordance-based exercises had ten sets of contexts for the same ten words. The results showed that the students generally showed better performance on concordance-based exercises.

Cobb (1997) replicated Stevens’ study in an online situation where students were able to manipulate a corpus-based computer program to learn new words, instead of recalling known words. The subjects were also first-year Arabic-speaking students in the same university where Stevens studied. Cobb created a computer tutor to teach a total of 240 new words over 12 weeks. He designed two versions of the program: an experimental version with concordances and a control version with no concordances. He then had the students access both versions on alternate weeks in order to measure their learning differences between the weeks with the experimental version and the weeks with the control version. The students were given pretest, weekly quizzes, and posttest. The results revealed that the mean score of the subjects’ vocabulary levels on the six weeks with concordances was much higher than the score on the six weeks with no concordance. Also, the students demonstrated favoritism to the corpus-based instructional materials.

These two experimental studies are remarkable for their innovative undertakings. However, this quantitative measurement limited to vocabulary learning does not show a
broader scope of what corpus technology does in students’ overall L2 learning. Besides, the studies did not examine a variety of learner variables that may have had an influence on the results. The effect of the corpus approach can be different depending on the students’ individual differences, such as their familiarity with concordance type of task and learning styles.

2) Students’ Attitudes and Responses to Corpus Use

Along with the test on students’ performance, it is also important to examine the students’ attitudes toward corpus technology to understand the aspects of corpus use that facilitate or impede L2 learning. Some research reports students’ positive attitudes toward corpus use in vocabulary learning, as supplementary findings (Cobb, 1997; Thurstun & Candlin, 1998). In order to determine the pedagogical potential of the corpus use in L2 teaching, it is essential to examine in greater depth the students’ responses to its use and any changes in their attitudes toward L2 learning as a result of their use.

A few studies have been conducted to explore students’ corpus use experiences in L2 learning. Sun (2000) conducted a study that focused on students’ attitudes. He examined EFL college students’ reactions to a three-week corpus-based lesson. The lesson was especially designed to teach the students about the corpus technique. Through a survey, Sun found that they were positive toward corpus-based learning, while some of them showed a concern about cut-off sentences in concordances and the limited corpus bank. This study presented a general understanding of the students’ responses to the use of the corpus, but it lacked a deeper qualitative insight into the topic by relying mainly upon one time quantitative data. In addition, because the study was based on a course especially designed for teaching the corpus technique, it failed to suggest how to incorporate the approach in a normal L2 classroom.

Turnbull and Burston (1998) conducted a longitudinal case study to examine the effectiveness of independent corpus uses by two ESL graduate students. The students used a corpus compiled from their own writings, which were corrected by the researchers. Turnbull and Burston used interviews and conference sessions, and also asked the students to write summative evaluations on their attitudes toward the use of the technique. The researchers found that unlike popular belief, the corpus approach did not always promote inductive learning strategy. Rather, individual differences, such as prior learning experience and learning styles, mediated the impact of the approach on the students’ learning. They concluded that learners need training to lead to independent corpus explorations. Besides, they suggested a further study about “the relationship between the use of concordancing strategies and language learning outcomes, and the relationship between varying degrees of concordance strategy training and learning outcomes” (p. 14). The study addressed the
importance of learners’ individual differences in using the corpus. Also commendable is their use of qualitative approaches that provided a deeper insight into the topic. Nonetheless, as the researchers admitted, the study has limitations in terms of the number of participants and corpus use. In addition, the study was conducted individually, not in a normal classroom setting, thus limiting its application to use in regular curriculum.

Unlike most corpus studies that used monolingual corpora, Fan and Xu (2002) used a bilingual corpus for their research. The purpose of the study was to evaluate the effectiveness of an online bilingual corpus (Chinese and English) in the teaching of legal English in a university of Hong Kong. In particular, the researchers were interested in looking at the usefulness of the use of the corpus in the students’ self-learning. The students finished two comprehension tests on legal English while working with the online bilingual corpus on a computer. A questionnaire and a group interview were then administered in order to examine their uses and perceptions of the corpus. The results showed that the bilingual corpus promoted the comprehension of the English text. The students responded that corpus use was helpful for enhancing their understanding of the English text by giving them an option to use the two languages. While the study explored the usefulness of bilingual corpora in language education, it has limitations: 1) as the researchers admitted, it did not explain why and when the students switched between the two languages, which is essential information to be examined regarding the use of bilingual corpora, and 2) more importantly in relation to this study, the students’ corpus experience was limited to only one time exposure for an hour and 40 minutes. Thus, it does not provide a fuller explanation of the impact of corpus use on students’ L2 learning over an extended exposure and experience.

In this regard, Yoon and Hirvela (2004) conducted a pioneering study that examined corpus integration into the regular curriculum over an extended period of time. While acknowledging the lack of classroom practice on corpus use, they attempted to answer the important question raised by Gavioli and Aston (2001) “how can corpus-based activities best be integrated with ‘normal’ language teaching, at different levels of proficiency? How can learners (and teachers) best be trained to profit from these resources?” (p. 245). The study investigated ESL college students’ actual use of corpora in an intermediate and an advanced level of academic writing courses. Based on a survey, classroom observations and follow-up interviews, the study examined the students’ corpus use behaviors, their perceptions of the strengths and weaknesses of corpora as a second language writing tool, and their reactions to using the tool in writing instruction. As a result, the researchers found that the students perceived the corpus activity to be beneficial for their English writing, particularly for learning common usage and collocates of words and for building confidence in their writing.

In summary, these previous studies have increased our understanding of corpus use in
L2 learning. However, the studies have been mainly conducted in ESL contexts, which raises a question of whether EFL situations would reveal the same results. Further studies are needed to expand upon these insights in order to more fully understand corpus pedagogy in different contexts. Teaching is extremely sensitive to a given context. It is hard to apply a universal principle across all contexts, so the evaluation of the corpus approach in a variety of contexts is greatly needed. The important question is how to adapt the corpus approach successfully to a particular learning context to enhance students’ L2 learning experiences.

IV. PEDAGOGICAL IMPLICATIONS: USING CORPORAS IN AN EFL CONTEXT

This section considers an EFL context of teaching (i.e., Korean context) and discusses the possible strengths and challenges of applying the corpus approach to the context. By doing so, it attempts to explore more practical aspects of integrating the corpus research into L2 instruction and feature future directions for using the approach in the context. While this section briefly reviews the potential benefits of the approach, the main focus is on the possible challenges in an attempt to pose a constructive criticism. This attempt is expected to help explore probable solutions to the challenging points and make suggestions to more successfully employ the corpus approach in L2 teaching.

1. Strengths of Applying the Corpus Approach to an EFL Context

One central advantage of the corpus approach is that learners can make their own discoveries of the language through investigation into multiple examples in a variety of texts. Whereas Korean students are presumably familiar with sentence-based grammar, the new approach can enhance not only their awareness of contextualized grammar, but also their understanding of functions of given types of discourse. In addition, the lexicogrammatical notion of the corpus approach can teach the Korean students how and where to put the words into sentences, for which the dictionary often has no explanation (Odlin, 2001). Therefore, their exposure to the multiple texts can ultimately produce improvements in their L2 learning.

It is also important to remember that corpus-based materials expose the students to authentic use of the language. In EFL settings, they seldom have access to English in natural contexts. In this regard, corpus can provide them with opportunities to encounter actual language in use. Furthermore, the corpus approach can become beneficial for EFL teachers as well. Non-native Korean teachers of English can access large language data
with which they can design their own teaching materials as well as they can employ the data to verify actual language use (Conrad, 2000).

The students’ familiarity with computer and computer texts can also work in their favor in using the corpus approach. If they favor computer texts over printed materials due to their strong exposure to a computer medium, they can benefit from using the computer in their learning of English. While enjoying exploring corpus data on the computer, they can have an increased number of encounters with English texts. In addition, the availability of on-line corpora can allow them easy access to valuable resources for their language learning (Conrad, 2000). In doing so, they can develop a way of self-learning and become independent learners outside class.

2. Challenges of Applying the Corpus Approach to an EFL Context

The first challenge of the approach is concerned with the suitability of the concept of inductive learning for the Korean context. Given that Korean students are believed to be familiar with a deductive type of instruction rather than inductive learning for themselves, it remains a question whether the students can acquire and benefit from the new learning style.

The issue of suitability is also related to the trainability and transferability of the new technique. One of the problems in the corpus approach is that a certain amount of time is needed for students to become accustomed to the new mode (Cobb, 1997). Thurstun and Candlin (1998) reported in their studies that “some students were puzzled by the cut-off sentences of the one-line concordances and daunted by the difficulty of the authentic academic texts” (271). Since the method is also quite new to the Korean students, it is uncertain whether and to what extent the strategy is trainable (Johns, 1994). Some students may not become easily adapted to the new style of presentation of language data in concordance lines. Furthermore, even though they successfully acquire the skills and strategies in certain types of texts, it remains a question whether the acquired expertise is automatically transferable to other types of texts or tasks (Jabbour, 1997). The issue of transferability can become more crucial in an on-line corpus approach. The central element is whether or not the acquired knowledge from on-line concordance can be transferred to the students’ general English skills.

Another possible challenge comes from a concern about time effectiveness and efficiency in classroom instruction. The corpus approach encourages students to investigate multiple examples for inductive discovery of the rules of the language. In this case, it is probable to encounter the same negative response as Thurstun and Candlin (1998) observed, i.e. “there are many exercises for just three words” (p. 277). In particular, considering that class is generally large in Korea, it may not be effective to devote a lot of class time to each student’s own discovery of the rules of the language. Moreover, one can
only wonder why the teacher does not simply use the efficient conventional way of teaching, i.e. teach them directly hard-and-fast rules about English, instead of letting them wander around a messy chunk of data.

In addition, the students’ linguistic proficiency may work as a variable in the use of the corpus approach. The approach may not work for all students at different levels of English proficiency. In fact, some researchers have assumed that the corpus approach benefit intermediate or advanced learners rather than beginners (Johns, 1991; Tribble, 1991), though the assumption has not been empirically examined. Then, a question arises as to whether a threshold exists for benefiting the approach in terms of linguistic proficiency. Carne (1996) also asks “how far the development of the specialized competencies is needed to undertake purposeful research” (p. 130). Furthermore, it can be a burden for classroom teachers to consider how differently the approach would work for each student while they are preoccupied with teaching subject matter.

As such, the use of the corpus approach also needs to be evaluated from a teacher’s perspective. As well noted thus far, the new approach gives primacy to learners, while teachers are regarded as providers of a context where the learners can conduct their own inquiries or research into the language (Johns, 1991). In this view, teachers are also ready to explore unfamiliar rules of the language with their students rather than try to be seen always as experts on that language. However, Swales (2002) is skeptical about the perception of teachers’ role in the approach:

Few of us, I suspect, are honest enough to admit in public that we would like time to prepare a proper answer, and most of us do our best to respond ‘on the fly’, sometimes with inspired results, but more likely with limp, evasive, or even downright misleading responses (p. 157).

In a teacher-centered context where teachers are seen as authority figures, it seems questionable whether the teachers are willing to negotiate their authority with the students in the classroom. This can be a critical issue for the teachers in teaching practice.

3. Several Comments about Future Applications

With such challenges in applying the corpus approach to classroom instruction, one can argue that the approach may be desirable, yet too ideal, and it cannot be incorporated into a real classroom. However, the posed drawbacks do not mean that the new approach should be abandoned. A closer look at the challenges can reveal that they are not indicative of serious structural problems of the approach; rather, they seem to address practical problems in applying the approach to a certain context of teaching. The challenges of practicality are
then, in part, due to the lack of empirical attempts to adopt and test the approach in a practical situation.

Given the increasing influence of corpus-based research on language pedagogy, it may be language educators’ responsibility to explore the new approach. The corpus techniques may not benefit all students, but it can play a beneficial role in a large number of students’ learning. The students need to be given the opportunities to learn with the new method, which is widely believed to be beneficial for their L2 development. Consequently, the appropriate question to be raised seems not whether to use the approach in teaching, but how to adapt it successfully to a specific context of teaching to enhance students’ learning experiences.

As noted in an earlier section, there are two ways to use concordances in classrooms: hands-on concordancing activities and offline concordance-derived handouts. While it may be often unavailable to engage students in using corpora on computer in the teaching context, students can be encouraged to use online corpora for their own English learning. There are free online corpora available for Internet users (e.g., Collins Corpus and BNC). More information regarding corpus-related resources and corpus use in English teaching is attached in Appendix A. Given the great rate of Internet use in Korea, the online corpora can be used as a valuable language resource for learners that can complement dictionaries. In this case, it is necessary that the students be taught the purpose of using the corpus and how to employ the corpus meaningfully in their own language learning.

Another way of using concordances in class is to present concordance-based data to the students on a handout in teaching English. This is still important in that students can be exposed to real language use. Teachers can use corpus work in general English classes (e.g., English writing) as a supplementary educational tool. A sentence-based corpus analysis task can be incorporated into a classroom task or activity, such as gap filler and cloze test, for vocabulary teaching as used in Tribble’s study (1991). This corpus work can also show the students the most common combinations and patterns of words in vocabulary learning, which is an important contribution of corpus linguistics to the field of teaching English.

While the sentence-based corpus analysis can expose students to the corpus approach directly, the genre-based corpus analysis can be perhaps the most appropriately applied to ESP materials development. For example, a corpus analysis of science-field journal articles can provide the students with the information about common structures and words used in the target genres that the students can use in their own science English writing. This type of the analysis is also an important application because it can not only help students with specific language data, but also enhance their awareness of contextual grammar and lexicogrammatical aspects of the language.

Here it is important to remember that few studies have investigated the effect of corpus use on students’ performance. In order to make corpus pedagogy more substantial, it would be necessary to examine whether and how corpus use actually affects students’ English learning.
One important question to be addressed for the successful implementation of the corpus approach is how we as L2 educators can meet the needs of students who are at variant levels in their development of English proficiency. Whereas many researchers have hypothesized that the corpus approach is suitable for more advanced students, Jabbour (2001) claims that the approach can be successfully used at all levels. If due attention is paid to syllabus design, the approach can be applicable at any level. For example, the length and difficulty of corpus texts to be used in class can be determined by considering the level of students’ English proficiency. It is pedagogically inadequate just to let students browse a large amount of texts without proper guidelines. Students need some principles to learn with the new approach. For more advanced learners, we can encourage them to explore corpus-based searches and analyses on their own. For beginner-level students, we can help them experience controlled corpus searches on the computer or start with concordances on the paper. On the other hand, it is also important to note that corpus findings can be beneficially used for teaching materials development for both levels (Kennedy, 1987a). CD-ROMs can be given to the students to manipulate the corpus data for themselves, or corpus-based materials can be prepared by the teachers to complement textbook materials.

More importantly, the corpus approach should be first introduced to teachers. While corpus-based studies have become popular among researchers, teachers who are at front in teaching seem to be less aware of the advantages of the corpus approach. But there is little doubt that if teachers are not aware of its usefulness, the new teaching method is hardly brought into the classroom. Therefore, it is the pressing demand to inform the teachers of the approach so that they can venture to use it in their classrooms. The first step can be to introduce and incorporate the corpus approach in teacher education programs for pre-service teachers as well as in-service teachers.

It may be impossible or, though possible, ineffective to replace the conventional way of teaching with the new approach. Here Tribble’s (1999) comment is noteworthy that the corpus approach should not be considered as a replacement of traditional, more familiar, teaching methods, but as a complement to inform teaching materials. For example, instead of completely abandoning their authority in class, teachers can be encouraged to use the approach while maintaining reasonable control of the class. Meanwhile, although the question of how they can enhance time efficiency in the classroom remains to be answered, the corpus approach can be used as a promising complementary method for L2 instruction.

V. CONCLUSION

This paper primarily aimed to survey corpus-based studies in L2 teaching. Based on the review, it also examined the possible strengths and challenges of applying the corpus
approach to an EFL context. A limitation of this paper, then, is the lack of the implementation of corpus work in real-life classrooms. Further studies are needed to test the usefulness of corpus use in English teaching and provide more context-grounded suggestions for corpus pedagogy. In fact, few attempts have been made to incorporate the corpus approach into EFL classrooms as opposed to the current popularity of corpus research in the field of L2 teaching. L2 educators are left to develop a framework for foreign language pedagogy to help learners by use of corpus linguistics techniques. In particular, the posed challenges need to be tested in a real classroom for more successful implementation. With further suggestions based on empirical research, the challenges can be turned into strengths for better pedagogical practice. When a new method arrives, it requires a certain amount of time and effort to specify its application to reality. Trial-and-error may be an inevitable process at this early stage of adoption of the new medium. A greater number of practical suggestions for a variety of contexts can encourage classroom teachers to take the first step to the new area of teaching English, and thus, contribute to the enhancement of students’ L2 learning experiences.

REFERENCES


in language teaching (pp. 67-86). Cambridge: Cambridge University Press.


APPENDIX

Corpus-related Resources

**American National Corpus**
http://americannationalcorpus.org/

**British National Corpus (BNC)**
http://www.natcorp.ox.ac.uk/

**Collins Cobuild Corpus Concordance Sampler**
http://www.collins.co.uk/Corpus/CorpusSearch.aspx

**Michigan Corpus of Academic Spoken English (MICASE)**
http://www.hti.umich.edu/m/micase/

**International Computer Archive of Modern and Medieval English**
http://nora.hd.uib.no/icame.html

**Journal Special Issues for Corpus Linguistics in Language Teaching**
Language Learning and Technology, 5(3), 2001 (http://llt.msu.edu/vol5num3/default.html)
TESOL Quarterly, 37(3), 2003

**Concordancing Software**
(http://www.lexically.net/wordsmith/)

**Useful Corpus Linguistics Pages**
Dave Lee’s devoted to corpora site (http://devoted.to/corpora)
Tim Johns’ site (http://web.bham.ac.uk/johnstf)
Michael Barlow’s corpus linguistics site (http://www.ruf.rice.edu/~barlow)

Applicable levels: college and university

Key words: corpus linguistics, corpora, concordancing

Hyunsook Yoon
BK International House A-dong #306,
Seoul National University, San 4-2, Bongchun 7-dong, Kwanak-gu, Seoul
151-818, Korea
Phone: (02) 881-9513, 010-7377-7691
Email: hsyoon3@hanmail.net

Received in February, 2005
Reviewed in March, 2005
Revised version received in May, 2005