Challenges of Using Corpora in Language Teaching and Learning: Implications for Secondary Education*

Shinwoong Lee

(Hanyang University)

Lee, Shinwoong. 2011. Challenges of Using Corpora in Language Teaching and Learning. *Linguistic Research* 28(1), 159-178. A number of advantages of using corpora in language teaching and learning have been identified by many corpus linguists and thereby its implementation into the language classroom has been highly recommended. However, the challenges and limitations of the use of corpora have not been extensively discussed, and without critically examining the use of corpora in language pedagogy it seems premature to urge teachers to use them in their classroom. In this vein, the purpose of the current paper is to provide a critical evaluation on the use of corpora in language teaching and learning and provide implications for their use in secondary school in Korea. It is argued that without a pedagogical mediation of corpora and resolving some practical problems, the pedagogical potentials of corpora may not be realized. It is also suggested that the integration of corpora into secondary school can be fostered by providing: (a) pedagogically relevant, level-specific corpora; (b) a Korean secondary learner corpus that can show the learners' common problems; (c) an online database of corpus-based resources; and (d) a corpus workshop for pre- and in-service teachers. It is concluded that the appropriate and effective use of corpora in the classroom is partly a technical issue, but primarily a pedagogical one. If the use of corpora in the classroom is not extensively discussed and researched to develop a pedagogical blueprint for the integration, the expected pedagogical outcomes that a number of corpus linguists simply expected may not accrue to learners and teachers. **(Hanyang University)**

Key Words corpora, language learning and teaching, secondary education

1. Introduction

As computer technology made it possible to store and analyze a large volume of language data efficiently and the web-based corpora are readily available, the potential benefits of corpora in language learning and teaching have been widely

acknowledged and even praised as a new language teaching/learning tool by many language professionals. The advocates of using corpora have argued that corpora can provide a powerful tool with which learners can explore and discover patterns of authentic language, providing such information as collocations, colligation, and semantic prosody that are hardly obtainable otherwise (Bernardini, 2004; Hunston, 2002; Meunier, 2002). It has been also contended that corpus-based language teaching has potentials to motivate learners and promote learner autonomy that are highly valued in pedagogy (Aijmer, 2009; Kaltenbock & Mehlmauer-Larcher, 2005).

Due to those potentials of corpora in language teaching and learning, a number of researchers (Aston, 1997; Braun, 2007; Conrad, 2004; Hunston, 2002; Tribble, 2001) presented them as a valuable resource and an innovative teaching tool, and their use has been considered somewhat trendy among language professionals. However, the reality of everyday teaching practice seems quite different from what many corpus linguists expected. Even though there have been a few cases reported on their actual use in the classroom at an undergraduate or a graduate level (for example, Chambers, 2005; Tribble, 2001), its application into secondary education has been a rare occurrence, yet to be reported in the literature. In this vein, Mukherjee and Rohrbach (2006) pessimistically noted that "we have the impression that in EFL countries like Germany there is a widening gap and a widening lag between on-going and intensive corpus-linguistic research on the one hand and classroom teaching on the other (p. 205, cited in Aijmer 2009). This may raise a question why there is a discrepancy between what can be done and what actually is done, and why teachers rarely use corpora in their classroom despite readily-available computers and some of undeniable merits of corpora in language teaching. Therefore, the purpose of the current paper is to provide a critical evaluation on the use of corpora in language teaching, identifying the problems that may hinder the integration of corpora into the classroom and also to suggest implications for the use of corpora in secondary school in Korea. The following questions guided the current paper.

1. What are the challenges of using corpora in language teaching and learning?
2. What are the pedagogical requirements for the integration of corpora into secondary school in Korea?

2. Challenges of Using Corpora in Language Teaching

In this section, I will discuss the challenging aspects of corpora in language teaching and learning in relation to the following questions: (a) Are corpora truly authentic?; (b) Are existing corpora pedagogically appropriate and relevant?; and (c) Are frequent lexical items always more valuable than less frequent ones in language pedagogy? Along with these questions, some of the practical problems that teachers may encounter when implementing corpora into the classroom will be discussed.

2.1 Are Corpora Truly Authentic?

It is often argued that corpora provide learners with 'authentic' or 'real' language, and since these words echo the key features of Communicative Language Teaching (CLT, hereafter) method that favors the use of authentic and real language over concocted ones, it is often assumed that corpus-based language materials are well-suited for CLT. However, some of the researchers have cast doubt on whether language data in corpora are truly authentic. Widdowson (2000; 2003) contrasted the concept of 'genuineness' and 'authenticity' and argued that 'genuineness' is the property of texts and is an absolute quality, while 'authenticity' is the characteristic of discourse interpretation. He claimed that language in corpora can be genuine, but it is not authentic because it is isolated from discoursal and communicative nature of language. In other words, language data in corpora do not tell us much about the authors of messages, their illocutionary intentions, the intended audience, and the circumstances in which the messages were produced. In line with Widdowson's argument, Cook (1995) emphasized the importance of context especially in spoken language and stated that "speech is often inseparable from [...] the circumstances of its production, and can only be apprehended in the context of the knowledge of the participants, their paralanguage and the situation" (p. 42, cited in Mishan, 2004). Along the same lines, Kaltenbock and Mehlmauer-Larcher (2005) provided an example from the British component of International Corpus of English (ICE-GB) to illustrate the problem of corpora in relation to their 'authenticity.'

A: Hello Track Records
B: Hallo Saucy It's Justin

A: Hi

B: You all right

A: Jesus   Give me a heart attack   Justin   who's Justin How you doing

B: OK   Is it all right to phone you at work

A: Yeah yeah

B: Oh right   I wasn't sure

A: Yeah   Oh dear

B: I I I need to make some phone calls from work I'll tell you about it
   when I see you

A: Do what

B: I need to make some phone calls from work you see but I'll tell you
   when I see you

A: Oh right   Oh right   yeah

B: Yeah   How are you ... (S1A-099-226 to 251)

(pp. 69-70)

As shown in the excerpt above, little information can be obtained about the details of speakers including their exact relationship, the purpose of phone call, the spatio-temporal situation in which this conversation takes place, let alone prosodic information. The communicative intents or purposes are missing in the text that can tell us why and how something is said and how it was understood. Supposed that language learning is closely related to producing and understanding discourse in a foreign or second language, a corpus itself may not be an adequate pedagogical resource (Braun, 2005).

To cope with the problem of corpora in relation to their 'authenticity,' it was suggested that they should be pedagogically mediated and authentificated when they are intended to be used in the classroom (Braun, 2005; Mishan, 2004; Widdowson, 2003). In other words, they must be re-contextualized in a pedagogical setting and be tuned to specific classroom purposes and thereby make them real and relevant for learners. In this light, Gavioli and Aston (2001) stated that "the question is not whether corpora represent reality but rather, whether their use can create conditions that will enable learners to engage in real discourse, authenticating it on their terms" (p. 240). Seidlhofer (2002) also maintained that the meaning of authenticity that most corpus linguists pursue (i.e., language description) is different from its meaning

in language pedagogy and stated that:

> [W]hile compilers of L1 corpora are interested in collecting authentic texts in the sense of 'attested': what native speakers have said. In language teaching, what is crucial is that learners should be capable of an authentic response to texts. This depends not so much on where texts originated but what learners can do with them in the way of performing appropriate activities which will engage their interest and stimulate the learning process. This is what is meant by authentification (p. 220).

This parallels to Widdowson's claim about the meaning of 'authenticity' in pedagogy, according to which 'authenticity' is not a matter of selection but of methodology. If 'authenticity' primarily resides in methodology, the authentification can be achieved by the enrichment of corpora with support materials such as corpus-based learning activities and exercises. For example, the corpus-based data-driven learning (DDL, hereafter) is well suited for focus-on-form activities in which learners are exposed to repeated instances of language forms. Then, as other researchers (Doughy & Williams, 1998; Joyce & Burns, 1999) showed, they can be integrated into a communicative task, leading learners to be involved in real discourse and to be socially situated, and consequently, re-contextualization or authentification of corpora can be accomplished.

Kaltenbock and Mehlmauer-Larcher (2005) noted that Breen's (1985) model may provide a general framework of authentification process of corpus language data. According to the model, (a) texts are used merely as input data, followed by (b) learners' own interpretation of the texts, (c) specific tasks based on the texts, and finally (d) a process of authentification through the social situation of the language classroom. Meanwhile, Flowerdew (2009) suggested that authentification of corpora can be assisted by including contextual information (e.g., MICASE has been marked up with socio-cultural information such as gender, age, academic position/role of interlocutor), having video clips (Braun, 2005), and providing didactic written hints (Milton, 2006) in corpora.

2.2 Are Existing Corpora Pedagogically Appropriate and Relevant?

A number of corpora in different languages have been created and their accessibility has been significantly improved as they are readily available on the Web. In particular, a number of English corpora (e.g, BNC, COCA) have been created and their utility as an English teaching tool has been widely recognized. However, corpora seemed to have little impact on language pedagogy despite their potentials as a language teaching tool and resource. One of the fundamental problems stems from how corpora were created on what purposes. Indeed, mega-sized general corpora make a perfect sense to linguists who believe that the balance and representativeness of corpora are critical. For example, a large-sized corpus is essential in lexicography because they need to obtain a sufficient number of occurrences of lexical or structural items in order to compare a relative frequency of occurrences. However, they do not seem particularly tailored to language pedagogy. For example, learners often end up with hundreds of concordance lines that are sometimes messy, ambiguous, and even misleading, and then they might become quickly overwhelmed by too much data that is not directly relevant to their learning. In this vein, Gavioli (1997) warned that the use of corpora in the classroom can leave learners too much alone, overwhelmed by information and resources. A more serious problem resides in the content of corpora. The words and sentence structures that collocate with key words are often way beyond learners' level of linguistic competence, and the retrieved instances are often from unfamiliar and widely differing contexts that make it more difficult for learners to interpret them.

Acknowledging the problems of large corpora in pedagogy, a number of researchers (e.g., Aston, 1997; Braun, 2005; Tribble, 2001) maintained that language learners and teachers may be better served by relatively small, homogenous, and domain-specific corpora (e.g., newspaper corpora). This is mainly because such corpora can be more relevant to learners' needs than large corpora which have limited direct relevance as a resource for students and teachers. However, it should be noted that advanced learners and teachers may benefit more from large corpora than from small ones. Learners may begin with smaller corpora that are limited in their size but that are more accessible, and then they can move on to larger corpora, developing learner autonomy gradually.

2.3 Are Frequent Items Always More Valuable Than Less Frequent Ones?

One of the distinguished merits of corpora is that one can access frequency data of lexical items using the built-in software and this can be substantially conducive to efficiency of learning. That is, by focusing on high frequency lexical items, learners can acquire the target language more efficiently, and this type of information is invaluable especially for EFL teachers and learners (Nation, 2001; Tsui, 2004) due to cost-effectiveness it can provide. In this light, some of the corpus linguists (e.g., Sinclair, 1991; Mindt, 1996) claimed that a high frequency of occurrence directly corresponds to a high effectiveness in language teaching and material development.

However, this view has been criticized by other researchers (e.g., Cook, 1998; Widdowson, 2003). Widdowson contended that even though quantitative corpus findings are valuable not only for language description but also for language teaching, frequency data should not be automatically taken as the sole criterion for pedagogical decision. Language teachers should also consider other crucial factors such as relative easiness (learnerbility: for example, primacy of the concrete over the abstract) and how much core or nuclear (generative value) words and structures are, together with the learning context in which learners are situated (e.g., learners' proficiency level, teaching objectives, and curriculum, etc.).

2.4 What Are the Practical Problems of Using Corpora in the Classroom?

One of the major barriers that make learners or even teachers less accessible to corpora is that few teachers and learners are knowledgeable about the use of corpora. They are often neither familiar with built-in software nor experienced in dealing with corpus data that involve a range of linguistic and metalinguistic knowledge. Tribble (2001) argued that teachers do not seem to use corpora very much in their classrooms and it is mainly because most of them do not have extensive experiences with corpora. The problem appears to be associated with a few practical issues. First, even though the number of readily-available corpora on the Web has been increasing and their interface has been somewhat improved, their user-friendliness still fell short of the standard. For example, the existing part-of-speech tagging system is too much complex and specific for non-linguists (e.g., UCREL tag sets include about 150 different tags), and it is even inconsistent among the corpora (e.g., COCA vs. BNC),

causing confusion.

In addition, teachers themselves are not familiar with the process of the interpretation and categorization of corpus data by which the data can be analyzed. Indeed, corpora entail not only quantitative analyses of data (e.g., frequency of occurrences) but also qualitative ones in which human decisions should be involved (e.g., what to include and exclude and how to regulate and interpret data), which makes the analysis highly demanding and time-consuming. Furthermore, as some researchers noted (e.g., Aston, 1997; Kaltenbock & Mehlmauer-Larcher, 2005), it is crucial for teachers to make corpus data 'palatable' for the learner by 'preselecting' and 'digesting' raw corpus data in order to make them pedagogically more relevant. In other words, the data should be often modified by selecting familiar contents, reducing the quantity of data, and simplifying the task, which can be another challenging task for teachers.

Last but not the least, the problem has to do with the curricular requirements of school. All of the materials and activities used in the classrooms should be compatible with the curricular requirements and many of the teachers are pressed for time to cover the textbook following the school curricular. Thus, creating corpus-based materials and activities can be too much demanding and unmanageable especially when they are handling things that they are not familiar with.

3. How to Foster Integration of Corpora into Secondary School

How can one implement corpora into the language classroom successfully? The answer to the question appears not that simple and it may vary according to the context in which teachers and learners are situated. However, the successful integration of corpora into secondary school in Korea seems dependent on at least a few pedagogical conditions/requirements. Therefore, in this section, taking into consideration some of challenges and limitations of existing corpora in language teaching, a few suggestions will be made for the integration of corpora into the classroom along with their rationales.

3.1 Development of Pedagogically Relevant Corpora

As discussed in a previous section, one of the questions raised concerning existing corpora was whether they are pedagogically appropriate and relevant, and it was argued that they are more tailored to linguistic research than pedagogical use. Thus, a number of researchers (e.g., Braun, 2007; Gavioli & Aston, 2001; O' Sullivan & Chambers, 2006) voiced a need of pedagogically motivated, local corpora that correspond to local conditions of learners. However, it is a very challenging task to develop such corpora for Korean secondary teachers and learners because they should meet a number of requirements that are needed for pedagogical mediation. As noted earlier, the content of corpora should be relevant and accessible to Korean secondary learners, while being complementary to the school curricular.

Taking all of those requirements into consideration, one of viable options would be to create textbook corpora. All of the textbooks used in secondary school in Korea are finely tuned to the National Curriculum of Korea and vocabulary and sentence structures are also controlled. Thus, if the textbook corpora are used in the classroom, pedagogical relevancy and learners' accessibility can be maximized. According to Chung (2007), the EFL teachers in secondary school in Korea voiced a need of such corpora, together with diverse corpus-based learning activities (e.g., DDL activities utilizing concordance lines). They also perceived that most of corpora available on the Web were not so much relevant to their teaching in terms of the contents and the students' language proficiency. Acknowledging those teachers' needs, Chung stated that a home-made small corpus such as a textbook corpus would better serve the Korean secondary teachers and learners, and provided a few example concordance lines retrieved from the textbook corpus that could be used in secondary school in Korea.

It's time to	*say*	goodbye now.
You can	*say*	that again.
I must	*say*	goodbye now.
Please	*say*	hello to your family.
How nice of you to	*say*	so!
	Say	cheese!
Don't	*say*	that!
People	*say*	that the earth is our home.

I enjoyed your	*talk.*	
Now he can	*talk*	to anyone in the world.
Can I	*talk*	to Inho?
Let's	*talk*	about them.
They	*talk*	with Nari's uncle.
I use it to	*talk*	to my friends in America.
May I	*talk*	with you for a minute?
My friends don't want to	*talk*	with me.
We have no chance to	*talk*	together.

Animals sometimes	*tell*	us about weather.
What does it	*tell*	us?
Now I'm going to	*tell*	you about my dream.
Please	*tell*	me more about him.
Here you can	*tell*	all your ideas about different topics.
They	*tell*	us an interesting story.
Can you	*tell*	me the difference between them.
Let me	*tell*	you a story about my short trip.

(pp. 56-57)

Figure 1. Concordance lines from 1st grade middle school English textbooks
'say', 'talk', and 'tell'

As shown above, the contents and level of vocabulary are very finely tuned to 1st graders of the middle school in Korea, resulting in high accessibility and relevancy.

If the level-specific textbook corpus (e.g., the middle school textbook corpus and its sub-corpus of 1st, 2nd, and 3rd grade) can be created and learners can choose the corpus that parallels to their language proficiency, learner accessibility can be significantly increased, promoting level-differentiated learning at the same time. Learners may move on to the next level of the corpus as their language skills improve. In the long run, they will be able to use a large native corpus once they reach a certain level of language proficiency.

Along with textbook corpora, it may be necessary to provide graded reader corpora. It is mainly because textbook corpora sometimes may not come up with sufficient examples for particular lexical items of interest, and thus they may need to be supplemented by other corpora that are relatively large enough to give adequate examples but that are still accessible by the learners in secondary school. That is exactly where graded reader corpora come into play to provide more examples, while giving similar data accessibility. Allan (2009) argued that the scale and type of lexical chunks in graded reader texts are sufficient to reflect authentic language when compared to BNC and stated that "graded readers may offer an acceptable balance of accessibility and authenticity" (p. 23). It was also suggested that graded reader corpora would be particularly appropriate for low- or intermediate-level learners and help teachers save their time and energy by making the filtering process (i.e., simplifying or deleting concordance lines) unnecessary.

Utilizing small corpora such as textbook and grader reader corpora would cost some authenticity. However, the advantages of using a small, but pedagogically relevant corpus do not seem to be outweighed by their limitations because learners are less likely to be overwhelmed by the data, and more likely to be able to understand it and draw conclusion from it, which is the core aspect of corpus-based DDL.

3.2 Development of Local Learner Corpora

Another important way of promoting the integration of corpora into secondary school in Korea is to develop a local learner corpus, and it is primarily because it begins with learners that would lead us to an important step closer to understanding local conditions of language learning. By examining the language use of particular learners, one can figure out what is particularly difficult for them and what features

should be especially emphasized in the classroom. Mukherjee and Rohrbach (2006) claimed that the development of a learner corpus for teachers' own students could be much more promising for both teachers and learners because "the focus on their own students' output will involve many more teachers in corpus-based activities" and "the exploration of learner data by the learners themselves will motivate many more learners to reflect on their language use and thus raise their foreign language awareness" (p. 228).

One of effective ways of using a local learner corpus is to make a comparison between learners' own language use and that of native speakers. This type of analysis can give information about qualitative (misuse) and quantitative differences (over- and underuse) in lexical, grammatical, and discourse features, helping teachers design and choose materials for their own students. As aforementioned, even though frequency is one of the most important criteria for pedagogical decision, it is necessary to strike a balance between frequency, difficulty and pedagogical relevance in the classroom. That is exactly where a local learner corpus comes into play to figure out the relative importance, identifying the forms which are problematic for particular learners. Research on native corpora would provide a native language description and be used when deciding the teaching agenda. However, research on a local learner corpus should be used to modify this agenda to meet the needs of the specific learner population.

From learners' perspectives, while comparing their language use and that of native speakers, they can notice the gap between them and be prompted to realize what they do not know and what they should know, initiating a process of restructuring of their linguistic knowledge. Mukherjee (2009) contended that corpus-based DDL activities can raise learners' awareness of language by enabling them to analyze both negative evidence (provided by a learner corpus) and positive evidence (provided by a native corpus).

For the reasons presented above, Korean EFL teachers in secondary school may need to develop a home-made, local learner corpus that consists of their own students' written/spoken texts. Then they can analyze the texts of their own students and reflect the result of analysis in material design and classroom activities. For example, DDL activities can be designed in such a way that learners are made more aware of their common mistakes and that knowledge reconstruction may occur, leading to grammatical rehabilitation (See an example below).

1) What grammatical structures appear to follow "accept"?
2) Do any grammatical forms only appear in learner data?
3) Discuss the differences in the use of "accept" between native and learner data and draw a conclusion about its use.
4) Carry out the same investigation of "possibility."

<div align="center">Native data</div>

not being able to	*accept*	that fulfillment of life is possible
be overcome? Why not	*accept*	the differences as an intentional
the act. Hugo cannot	*accept*	that they party line has changed
mothers and learn to	*accept*	their traditions
with their emotions and try to	*accept*	that diversity
If the peer group doesn't	*accept*	what the friend is wearing
of a woman, why not	*accept*	it and consider ways to use

<div align="center">Learner data</div>

think that women must	*accept*	that some differences exist
nor the children	*accept*	to recognize that
the parents	*accept*	that new visions of things may
don't always	*accept*	that their children also
women have to	*accept*	the other side of the coin
he could never	*accept*	to be inferior
Feminists have to	*accept*	to receive some viruses, some
Johnny will not accept	*accept*	to be treated as men

<div align="center">Native data</div>

from the two-fold	*possibility*	for joining the party
there is no	*possibility*	of his dominant position in
there seems every	*possibility*	that the present Queen will
there is no	*possibility*	of benefiting from that
mention the	*possibility*	that one of the motives for
popular because of the	*possibility*	for abuse. The second
but there is a	*possibility*	of entry for those

Learner data

there is a strong	*possibility*	that our society is still
the	*possibility*	for students to move from
argument against the	*possibility*	of an identity
already explored the	*possibility*	of forming other
culture and have the	*possibility*	to practice their
because we have the	*possibility*	to travel more freely all
the	*possibility*	to be in harmony with

(adapted from Granger & Tribble, 1998, pp. 202-203)

Figure 2. A comparison between native and learner data

A comparison between native and learner data can be made collaborately. Once concordance lines are provided to each group of learners, they can collaborate on interpreting, regulating, and drawing a conclusion about the data. In line with socio-cultural view of language learning, this can facilitate collective scaffolding and collaborative co-construction between advanced and less-advanced learners. Flowerdew (2009) argued that "in this scaffolding-type of activity, more proficient students are able to offer their insights and interpretations on the corpus data, thus assisting the weaker students to gradually develop more independence" (p. 404).

However, it should be also noted that corpus-based DDL activities are often time-consuming, thus the pedagogical decision about when to use and how to use in class should be made with caution, considering such factors as time constraints, students' most urgent needs, and relevancy to the objectives of class.

3.3 Corpus Training

In addition to the localization of corpora by designing a pedagogically relevant small corpus and a local learner corpus, teachers themselves should be cognizant of potentials of corpora in language teaching. They should be also knowledgeable about the use of concordance software and be trained to categorize and interpret language data retrieved from corpora. Indeed, one of the primary reasons for the discrepancy between what can be done and what actually is done is that many teachers are not aware of the potentials offered by corpora and DDL activities. Therefore, these subjects should be made a high priority in the teacher training program for pre- and

in-service teachers.

Maybe, it is necessary to integrate the corpus component into undergraduate- and graduate-level courses (e.g., *English Grammar*, *Language Teaching Methodology*, and *English Linguistics*). As would-be teachers become familiar with corpus-based research, they would naturally consult corpora as if they looked into a lexical item in a dictionary, and it will ultimately promote the integration of corpora into the classroom when they become a teacher. A corpus workshop should be also offered for pre- and in-service teachers on a regular basis. In the workshop, it is important to train teachers how to use concordance software efficiently and how to analyze corpus data through the process of categorization and interpretation. More importantly, it should be geared to help them learn how corpus-based materials/activities can be interwoven into a regular class in secondary school effectively. As noted earlier, corpus-based activities are time-consuming and also require a substantial amount of preparation on the part of the teacher concerning what to use, how to use, and when to use. Teachers also need to make sure that teaching material is interesting and thus motivating their students, while coping with the various learning styles and local conditions of their students. Some students may be reluctant to work inductively and expect that the teacher should play a major role in their language learning by explaining rules deductively. Indeed, all of these practical and pedagogical problems are still out there and they should be taken care of in such a way that the integration of corpora into the classroom can be fostered. To recapitulate, it is crucial to develop a methodological blueprint and provide it to teachers through the teacher training program in order to accelerate the integration of corpora into the classroom.

3.4 Development of an Online Database of Corpus-Based Resources

The future direction for the use of corpora in secondary school in Korea would be to develop an online database of corpus-based resources and make those resources available to the teachers. It may be suggested that teachers explore existing corpus-based resources on the Web and use them for their students in the classroom. For example, a number of corpus tools for DDL are available at Compleat Lexical Tutor (http://www.lextutor.ca) and numerous ready-made DDL activities are also presented at Tim Johns' site (http://web.bham.ac.uk/johnstf/hompage.htm). However,

the problem of those online resources is that they do not seem closely relevant to the needs of Korean secondary students and are not tuned to their English proficiency. Rather, they are designed for learners in college, focusing on a particular genre (e.g., Tim Johns' page was designed for teaching academic writing in college).

For these reasons, it is necessary to develop a localized online database of corpus-based resources and it can be a partial solution to the practical problems that most teachers have. That is, teachers can save their time and energy by using localized corpora and ready-made corpus-based activities on the Web. As suggested earlier, level-specific textbook corpora are already finely tuned to Korean secondary learners in many ways (e.g., proficiency level, curricular requirements, etc.). If these corpora are loaded on the Web along with graded DDL activities, the teachers will be able to choose the corpus and activities that correspond to their students' profiles (e.g., grade and proficiency level). Then, the use of corpus in the classroom can be considerably encouraged.

Furthermore, learner data can be collected across the secondary schools in Korea to construct a Korean secondary learner corpus, and systematic linguistic analyses of areas of English can be conducted in which the secondary learners experience difficulty. Then, the results of analyses can be made available on the Web and used to create consciousness-raising activities in the classroom. As in the case of Hong Kong (see, Allan 2002; Tsui, 2004, 2005), the learner database and its analyses can be also utilized for teachers' professional development whereby they can better understand the English of Korean secondary students and have implications for English teaching in secondary school.

In addition, it may be necessary to build an online community for the teachers who wish to use corpora in their classroom in which they can share their experiences in implementing and using corpora in their classroom. This type of collaboration will ultimately help create a collaborative culture among the teachers and enrich the database, providing practical advice for the integration of corpora into the classroom.

4. Conclusion and Implications

As claimed by a number of corpus linguists (Hunston, 2002; Meunier, 2002; Tribble, 2001), corpora have great potential to facilitate language teaching and

learning and the use of corpora is also in line with the prevailing beliefs in pedagogy that problem-solving and discovery learning would motivate learners and lead them to leaner autonomy and empowerment. However, as contended previously, there exist some limitations and problems that should be taken care of for the integration of corpora into the classroom; otherwise their pedagogical potentials may not be realized.

Thus, it was suggested that corpora should be pedagogically mediated by meeting a few essential requirements particularly for secondary teachers and students in Korea. Those requirements include that: (a) pedagogically relevant, level-specific corpora for secondary students in Korea; (b) a Korean secondary learner corpus that can show their typical problems; (c) an online database of corpus-based resources; (d) a corpus workshop for pre- and in-service teachers. It was also suggested that an online community for the teachers who want to use corpora in their classroom should be created in which they can share their experiences in implementing and using corpora in their classroom.

It has been argued that discovery learning using corpora may be most suitable for advanced learners who are filling in gaps in their knowledge rather than laying down the foundations (Hunston, 2002). However, through a certain level of pedagogical mediation, the benefits of using corpora can be extended to intermediate- or low-level learners who have limited language skills, and for this being the case, the pedagogical mediation of corpora seems critical in order to integrate them into secondary school in Korea.

It must be noted that corpus-based activities are not the replacement of the other teaching activities or methodologies, but a complement to them (Meunier, 2002). As in the case of other computer-assisted language learning tools, the use of corpora should not be technology-driven, but pedagogically driven. That is, we should not use corpora just because we have a trendy tool at hand and would like to apply it into the classroom. Their use should be "vindicated to the extent that it agrees with what we know about language and language acquisition, and can be shown to be an effective learning tool" (Johansson, 2009, p. 42). In this light, there is a strong need for empirical classroom-based action research conducted by teachers who are aware of the potential as well as the limitations of corpus (Seidlhofer, 2002). This type of classroom-based research would reveal more issues on the use of corpora and would be able to test their effectiveness in language learning in the classroom.

Indeed, the appropriate and effective use of corpora in the classroom is partly a technical issue, but primarily a pedagogical one. If the use of corpora in the classroom is not extensively discussed and researched to develop a pedagogical blueprint for the integration of corpora into the classroom, the purported educational outcome that a number of corpus linguists simply expected would not accrue to learners and teachers.

References

Aijmer, K. 2009. Introduction: Corpora and language teaching. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 1-10). Amsterdam: John Benjamins.

Allan, Q. 2002. The TELEC secondary learner corpus: A resource for teacher development. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 196-211). Amsterdam: John Benjamins.

Allan, R. 2009. Can a graded reader corpus provide 'authentic' input? *ELT Journal, 63*(1), 23-32.

Aston, G. 1997. Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 51-64). New York: Addison Wesley Longman.

Bernardini, S. 2004. Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Amsterdam: John Benjamins.

Braun, S. 2005. From pedagogically relevant corpora to authentic language learning contents. *ReCALL, 17*(1), 47-64.

Braun, S. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL, 19*(3), 307-328.

Breen, M. 1985. Authenticity in the language classroom. *Applied Linguistics, 6*(1), 60-70.

Chambers, A. 2005. Integrating corpus consultation in language studies. *Language Learning & Technology, 9*(2), 111-125.

Chung, H. 2007. *Korean EFL teacher attitude towards corpus use*. Unpublished master's thesis. International Graduate School of English.

Cook, G. 1995. Theoretical issues: Transcribing the untranscribable. In G. Leech, G. Myers, & J. Thomas (Eds.), *Spoken English on Computer*. New York: Longman.

Cook, G. 1998. The uses of reality: A reply to Ronald Carter. *ELT Journal, 52*(1), 57-63.

Conrad, S. 2004. Corpus linguistics, language variation, and language teaching. In J.

Sinclair (Ed.), *How to use corpora in language teaching* (pp. 67-85). Amsterdam: John Benjamins.

Doughty, C., & Williams, J. 1998. Pedagogical choices in focus on form. In W. Doughty & J. Willimas (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-262). Cambridge: Cambridge University Press.

Flowerdew, L. 2009. Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics, 14*(3), 397-417.

Gavioli, L. 1997. Exploring texts through the concordance: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 83-99). New York: Addison Wesley Longman.

Gavioli, L. & Aston, G. 2001. Enriching reality: Language corpora in language pedagogy. *ELT Journal, 55*(3). 238-246.

Granger, S. & Tribble, C. 1998. Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (199-209). London: Addison Wesley Longman.

Kaltenbock, G. & Mehlmauer-Larcher, B. 2005. Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL, 17*(1), 65-84.

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

Johansson, S. 2009. Some thoughts on corpora and second language acquisition. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 34-44). Amsterdam: John Benjamins.

Joyce, H., & Burns, A. 1999. *Focus on Grammar.* Sydney: National Center for English Language Teaching and Research, Macquarie University.

Meunier, F. 2002. The pedagogical value of native and learner corpora in EFL grammar teaching. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 119-141). Amsterdam: John Benjamins.

Milton, J. 2006. Resource-rich web-based feedback: Helping learners become independent writers. In K. Hyland, & F. Hyland (Eds.), *Feedback in second language writing* (pp. 123-139). Cambridge: Cambridge University Press.

Mindt, D. 1996. English corpus linguistics and the foreign-language teaching syllabus. In J. Thomas, & H. Short (Eds.), *Using corpora for language research: Studies in honour of Geoffrey Leech* (pp. 232-247). Harlow: Longman.

Mishan, F. 2004. Authenticating corpora for language learning: A problem and its resolution. *ELT Journal, 58*(3), 219-227.

Mukherjee, J. 2009. The grammar of conversation in advanced spoken learner English. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 203-230). Amsterdam: John Benjamins.

Mukherjee, J., & Rohrbach, J. M. 2006. Rethinking applied corpus linguistics from a lan-

guage-pedagogical perspective: New departures in learner corpus research. In B. Ketteman, & G. Marko (Eds.), *Planning, gluing, and painting corpora* (pp. 205-232). Frankfurt: Peter Lang.

Nation, I. S. P. 2001. *Learning vocabulary in another language.* Cambridge: Cambridge University Press.

Nesselhauf, N. 2004. Learner corpora and their potential for language teaching. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 125-152). Amsterdam: John Benjamins.

O'Sullivan, I. & Chamber, A. 2006. Learners' writing skills in French: Corpus consultation and learner evaluation. Journal *of Second Language Writing, 15*(1), 49-68.

Seidlhofer, B. 2002. Pedagogy and local learner corpora: Working with learning-driven data. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 214-234). Amsterdam: John Benjamins.

Sinclair, J. M. 1991. *Corpus, concordance, collocation.* Oxford: Oxford University Press.

Tribble, S. 2001. Small corpora and teaching writing: Towards a corpus-informed pedagogy of writing. In M. Ghadessy, A. Henry, & R. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 381-408). Amsterdam: John Benjamins.

Tsui, A. 2004. What teachers have always wanted to know and how corpora can help. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 39-61). Amsterdam: John Benjamins.

Tsui, A. 2005. ESL teachers' questions and corpus evidence. *International Journal of Corpus Linguistics. 10*(3), 335-356.

Widdowson, H. 2000. On the limitations of linguistics applied. *Applied Linguistics, 21*(1), 11-24.

Widdowson, H. 2003. *Defining issues in English language teaching.* Oxford: Oxford University Press.

[Applicable levels] tertiary and secondary education

Shinwoong Lee
Dept. of English Language and Literature, Hanyang University
17 Haendang-dong, Sungdong-ku, Seoul 133-791, Korea.
Tel: 070-7523-9393
E-mail: shinwoonglee@hotmail.com