

A Corpus-based Analysis of Korean EFL Learners' Word Usage

Park, Hui Kyung
(International Graduate School of English)

I. Introduction

Information technology made it possible to access authentic data easily and to analyze patterns of text through corpus. Corpus linguistics is descriptive linguistics aided by new technology (Kennedy, 1999). As corpora have become more widely available to teachers and learners, it helped to identify in what respects learner's language differs from each other and from native speakers' (Lee, 2006). This study will examine Korean EFL learners' word usage based on a learner corpus in order to investigate the learner's language.

Liszka (2004) studied the effects of first language influence on second language pragmatic processes among German, Japanese and Chinese. She found that the L1 might influence the L2 acquisition of the present perfect tense. Cross-linguistic influence explains that the significant role of prior experience in learning act and the huge influence of the native language as prior experience (Brown, 2000). In that sense, L1 may influence the target language learning in terms of word choice, collocation and expression. If the target language is always translated into mother tongue, the original meaning can be distorted. For example, the adverbial conjunction 'even though' and 'even if' are used as equivalents among Koreans even though there is a meaning difference (Kosofsky, 1990). Also, as shown in the example of 'see an exam', mother language can affect L2 learner's speaking habit.

This study mainly focuses on three aims of the learner corpora studies as Tan (2005) suggested. Learner corpora study is conducted 1)to correlate the findings of data with the learners' L1, 2)to investigate how a particular word, phrase or structure is used incorrectly by learners of one cultural groups compared to the natives of the target language groups, and 3)to recommend the pedagogical measures to help the teachers and learners pay more attention on the errors.

Research questions of this study are

- 1) Are there the usage differences of adverb 'actually' between Korean EFL learners and native speakers?
- 2) Are there the usage and frequency differences of adverbial conjunction 'even though', 'even if' and 'although' between Korean EFL learners and native speakers?
- 3) How can we implicate the result in the class?

II. Theoretical Background

1. Corpora

Corpus is defined as a collection of texts in an electronic database. Electronic corpora consist of whole texts or continuous text samples (Kennedy, 1999). A corpus is a collection or a subset of naturally-occurring language text assumed to be representative of a given language. The texts can be in spoken or written forms, and they are selected and ordered according to explicit linguistic criteria (as cited from Sinclair, 1991; Jan Aarts, 1991; Eagles, 1994). A concordance is "a list of extracts retrieved from a corpus" (Aston, 2001, p.47). The corpus-based approach starts with a set of explicit rule whereas the corpus-driven approach builds up the theory to generalize the rules of usage (Tognini-Bonelli, 2001).

1) Brown Corpus

Brown Corpus is known as the first computer corpus compiled for linguistic research. The samples of Brown Corpus are taken from various text categories, from both imaginative and informative prose. 75% of the corpus is from informative text and the other 25% of the corpus is from imaginative prose such as novels and short stories. (Kennedy, 1999) Brown corpus contains more than one million words, and the samples in each category were established in 1963 (as cited from Francis & Kucera, 1982).

2) Learner Corpus/Corpora

A learner corpus shows the characteristic patterns in student's writing so that it is useful both teachers and learners to identify the learners' common errors. In other words, it is useful for error analysis in order to examine where errors occur most frequently and to discover similarities and differences in native speaker and learners. Because learner corpus focuses on the teaching process, it can be a source of activity and materials (Tognini-Bonelli, 2001).

2. Adverb ACTUALLY

According to the Oxford University Press (2005), a word 'actually' can be used mainly in four different situations.

- i) used in speaking to emphasize a fact or a comment, or that something is really true: *It's not actually raining now.*
- ii) used to show a contrast between what is true and what somebody believes, and to show surprise about this contrast: *The food was not actually all that expensive.*
- iii) used to correct somebody in a polite way: *They're not married, actually.*
- iv) used to get somebody's attention, to introduce a new topic or to say something that somebody may not like, in a polite way: *Actually, I'll be a bit late home.*

3. Adverbial Conjunctions EVEN THOUGH, EVEN IF and ALTHOUGH

According to the Longman dictionary (2008), the conjunction 'even if' is used to emphasize that, although something may happen or may be true, it will not change a situation. On the contrary, the conjunction 'even though' is used to introduce a statement that makes the main statement coming after it seems surprising, unlikely, or unexpected. The conjunction 'although' is used to introduce a statement that makes your main statement seem surprising or unlikely. Kosofsky (1990) pointed out the meaning differences between 'even though' and 'even if' in his book, *Common Problems in Korean English*. 'Even if' is used for conditions that do not really exist at the moment in hypothetical situation whereas 'even though' is used for conditions that really exist. The typical problem sentences are following.

- 1)* Even though you lose all your money tomorrow, I will still be your friends.
- 2)* Even if today is a warm day, many people are wearing coats.

There are meaning differences between two conjunctions; however, Koreans often use them as an equivalent. First example is not semantically natural because it is a hypothetical condition. Therefore, it should be 'even if', not 'even though'. Second example is not correct because it is a factual condition.

III. Method

1. Data

In this study, two different sets of corpora were used to compare Korean learners' word usage: a learner corpus and Brown corpus. 42 graduate school students' journals were the source of learner corpus. Students are majoring in English education, and their proficiency level is upper-intermediate and advanced. Although they are fluent in English, some mistakes or errors are found in writing. When they wrote journals, only topic was given to the students without any limitation of the format and length. Suggested topics are various such as self-introduction, favorite book or song, successful writing experience, and simple description of objects. Students had to complete their journal writing for ten minutes each time. Each student wrote eight or nine journals so that approximately 340 journals contained the 54,656 tokens and 5,387 types. On the other hand, the Brown corpora were used for a native corpus to compare with the learner corpus because the size of corpus is relatively smaller than other general corpora. However, still the Brown corpus is even bigger than the learner corpus as shown at Table 1 below.

Since the journal writing is less formal and more prompt than purposeful writing, it shows learners' repetitive mistakes or errors. On the contrary, the formal writings are revised and carefully elaborated so that journals were chosen for the learner corpus. Because the genre of the journal writing can be informal essay it is close to

the spoken language rather than formal writing.

2. Tool

Wordsmith version 3 was used to sort out word list and concordance. The general description of data is shown at Table 1.

[Table 1]

| Text File | Learner Corpus | Brown |
|------------------|----------------|-----------|
| Tokens | 54,656 | 1,015,476 |
| Types | 5,387 | 42,576 |
| Type/Token Ratio | 9.86 | 4.19 |
| Sentences | 3,517 | 51,351 |
| Sentence length | 15.54 | 19.78 |

IV. Results

1. Adverb Actually

The first research question was if there are usage differences of the adverb 'actually' between Koreans and the native speakers.

1) Meaning Difference

The result is shown at table 2. Total 59 concordance lines of 'actually' were counted in learner corpus whereas 163 concordance was sorted in Brown corpus. Each sentence was classified into four different categories based on the Oxford University Press (2005).

[Table 2]

| | Learner Corpus | | Brown | |
|--------------|----------------|-------------|------------|-------------|
| | Number | Percentage | Number | Percentage |
| i* | 9 | 15.25% | 106 | 65.00% |
| ii* | 10 | 16.95% | 16 | 9.81% |
| iii* | 0 | 0 | 0 | 0 |
| iv* | 36 | 61.02% | 39 | 23.92% |
| etc.* | 4 | 6.78% | - | 0.00% |
| Total | 59 | 100% | 161 | 100% |

i* Emphasize a fact or a comment, or that something is really true

ii* Show a contrast between what is true and what somebody believes, and show surprise about this contrast

iii* Correct somebody in a polite way

iv* Get somebody's attention or say something that somebody may not like in a polite way

etc.* The name of the movie, Love Actually

Among 59 concordance lines, 36 were used according to the fourth definition in learner corpus. They began the sentence with the word 'actually' in order to get

others' attention. The examples are following.

[Excerpt 1 - Learner Corpus]

- she's first birthday is coming in september. Actually my English level is not good.
- She wrote the book after traveling around the world. Actually she sold her house to make travel expense.
- I recommend you to read is long historical Korean novel, Arirang. Actually, I haven't still finished it, but it's quite interesting...
- I've already read it more that 5 times. Actually I possess it.

It may be influenced by Korean because people often start the sentence with '솔직히' or '사실은' in Korean. Also, they learned the word 'actually' as equivalent of those above words in Korean. Because the journal writing is informal writing close to speaking, the word 'actually' appeared relatively often. As below examples are demonstrated, first definition was most common in Brown corpus.

[Excerpt 2 - Brown Corpus]

- Women actually began to appear unaccompanied in the...
- The Boston Association did not actually command Parker to leave the room...
- Is an arbitrary retirement age of 65 actually costing your plant money?
- Have we not actually developed idea worship?
- Only two principal storehouses were actually established- one at Mobile, the other at...

2) Frequency

Korean learners use 'actually' much more frequently than the native speakers. The rate of using 'actually' in the learner corpus was 0.10% which means 59 tokens among 54, 656 whereas the rate in Brown corpus was 0.016% which is 163 tokens among 1,015,476. Also, native speakers begin the sentence with the word 'actually' relatively less than Korean learners. Korean learners starts the sentence with 'actually' 40 lines out of 59 concordances, which is approximately 67.8%. On the contrary, Brown corpus shows only 38 lines out of 163 concordances begins with 'actually'. which is about 23.3%.

3) Collocation

Several differences were marked as shown at Table 3. For example, any collocation on the right was not investigated in a learner corpus.

[Table 3]

| No | Left collocation | | Word | Right collocation | |
|----|------------------|-------|-----------------|-------------------|-------|
| | Learner | Brown | | Learner | Brown |
| 1 | WAS | WAS | actually | - | THE |
| 2 | NOT | ARE | | - | HAVE |
| 3 | - | NOT | | - | ONLY |
| 4 | - | WERE | | | |
| 5 | - | THEY | | | |

2. Adverbial Conjunction Even though and Even If

The second research question was to see how Korean learners use the adverbial conjunction 'even though', 'even if' and 'although'.

1) Meaning

As Kosofsky (1990) advocated, it was found that students use two conjunctions, 'even though' and 'even if' as a synonym regardless the condition. The concordance from learner corpus is displayed at Excerpt 3.

[Excerpt 3 - Learner Corpus]

- My younger brother is still child for me even if he is a adult and make money, now.
- Even if I'm a bit old comparatively, I feel like a teenager.
- Even if the topic is used to me, I usually can't speak anything.
- His father already prepared dinner for us so we should eat again even if we are so full.

Because they are not hypothetical conditions but factual conditions, 'even though' is supposed to be used. The concordance of 'even if' was shown only five times whereas the number of 'even though' clause was thirty-seven. The reason why the 'even if' clause is rarely used may be because of the given topic. Since the given topic was not much related to the hypothetical condition, the students did not need to use that clause. Any prominent error was not found in 'even though' clause, but a minor mistake is like below. Since the students were at upper intermediate or advanced level, they rarely made a grammatical mistake.

- He couldn't stand up or sit down easily. Even though, he had to take care of our son.

The clear meaning differences were shown between 'even though' and 'even if' in Brown corpus. 'Even if' was used in hypothetical condition and 'even though' was mentioned in factual condition.

[Excerpt 4 - Brown Corpus]

- Even if he angry, Tim wouldn't hurt a woman.
- Even if he'd somehow missed seeing him, he wouldn't have gone off and left
- ...poisons would be enough to recommend the frequent use of such a fruit, even if its nutritional values were limited.
- We would take whatever action was appropriate to prevent this, even if we had to go it alone.

[Excerpt 5 - Brown Corpus]

- A satisfactory cloud was produced even though these nozzles were only about 5 per cent...
- ...can make you uncomfortable even though the temperature is still low.
- You may deduct these payments even though your child uses the money to purchase...

2) Frequency

As the Table 4 displays the frequency of each adverbial conjunction, there is a big frequency difference in use of 'although' between two corpora. The number in parenthesis means the number of errors per concordance.

[Table 4 - Frequency of each adverbial conjunction]

| | # of EVEN IF | # of EVEN THOUGH | # of ALTHOUGH |
|----------------|--------------|------------------|---------------|
| Learner corpus | 5 (5) | 37 (1) | 10 (1) |
| Brown corpus | 61 | 76 | 319 |

3) Structure

As shown at figure 2, the students tend to start from the subordinate clause. 34 concordances among 37, which is about 87%, has the same pattern that the 'even though' clause comes first. Although the order between clauses can be reversed, students tend to start with the subordinate clause. It could be L1 influence because it is more natural in Korean. On the contrary, only 13 sentences among 76 concordances, which is approximately 17%, start with the subordinate clause in Brown Corpus. The number of 'even though' concordance was 76, and the number of 'even if' clause was 61 among 1,015,476 tokens. Some of the examples are following.

| N | Concordance |
|----|----------------------------------------------------------------------------|
| 1 | n which the Southerners of 1787, even though they continued it, fou |
| 2 | 2 Brown Corpus racy even though it's only 3-1/2 inches |
| 3 | 3 list. Even though it was known that th |
| 4 | 4 rested in knowing that much, and even though with the fingerprint |
| 5 | 5 ally obligated to furnish him, and even though you may be entitled t |
| 6 | 6 aerial skills. This difficulty arises even though we can give exampl |
| 7 | 7 had handled immediately before, even though they were altogether |
| 8 | 8 plication. As his disciples boast, even though his emphasis is else |
| 9 | 9 ith religion as the dominant bond, even though spatially dispersed a |
| 10 | 10 found this immensely comforting, even though Mercer did not make |
| 11 | 11 n the protection of their company even though I had nothing to be pr |
| 12 | 12 , these works are not comparable, even though the same brain concs |
| 13 | 13 paid by a transferor corporation, even though the reorganization m |
| 14 | 14 the national game in those days, even though professional baseball |
| 15 | 15 e, anguished; her son was dead. Even though he would later be res |
| 16 | 16 if he qualifies as your dependent, even though he earns \$600 or mor |
| 17 | 17 sequential than their differences. Even though in most cases the c |
| 18 | 18 is at an initial cost disadvantage even though only 1 to 10 per cent |
| 19 | 19 e notion something is being done, even though it is something they |
| 20 | 20 ir voices and had closed her door, even though it was a very hot da |
| 21 | 21 rging from an economic downturn, even though the signs of resurgen |
| 22 | 22 company in return for free drinks, even though he made good mon |
| 23 | 23 bject to tax on his own earnings even though his parent may, unde |
| 24 | 24 tinuously with no apparent effort. Even though the bondage of his v |
| 25 | 25 mance of services as employees even though they are not away fro |
| 26 | 26 l to which he is denied entrance? Even though in civil rights legislati |
| 27 | 27 to the address computed from **f, even though **f does not match ** |
| 28 | 28 or influence, and he must go far, even though he may go fast, for s |
| 29 | 29 m up as the champion of freedom. Even though his theological thes |
| 30 | 30 n 'll come find you" "I gotta an Even though this is my rock you |

| N | Concordance |
|----|-----------------------------------------------------------------------------------------|
| 1 | 1 preparation class that I have had before. Even though I find it difficult to go to wo |
| 2 | 2 Learner Corpus 2g. Even though I don't often visit this Web |
| 3 | 3 tr. Even though we chose the wrong days, |
| 4 | 4 it there whenever I turn on the computer even though I don't have anything to bu |
| 5 | 5 ach and had wonderful dining every day. Even though I spent pretty big amount |
| 6 | 6 -page was sort of vogue in those days.) Even though it was so sloppy with lots |
| 7 | 7 ook is seems to travel Europe. Even though psychology is not her maj |
| 8 | 8 d me to watch movie " Love Actually ". Even though there was no translation, I |
| 9 | 9 share the first unforgettable experience. Even though you might think my name |
| 10 | 10 r brother or went out to meet my friends. Even though I seemed that I didn't do a |
| 11 | 11 seek, we were visited ancestor's graves even though traffics were terrible. My h |
| 12 | 12 human vegetable. He always talk to her even though she can't respond anything |
| 13 | 13 and especially maths. Thanks to him, even though I gave up maths, I came to |
| 14 | 14 was not that good at doing housework. Even though things didn't go well as I pl |
| 15 | 15 ot a Christian) to bless me with this joy. Even though I had little time for studyn |
| 16 | 16 She seems to feel everything about me even though I don't say anything. It so |
| 17 | 17 ve .Matrix' is one of my favorite movies. Even though I'm not quite fond of scien |
| 18 | 18 ve it'Matrix' is one of my favorite movies. Even though I'm not quite fond of scien |
| 19 | 19 many noblemen sold their family name. Even though they sold it, they could stil |
| 20 | 20 I had many other things to take care of. Even though I couldn't be together with |
| 21 | 21 time, I didn't know how to use that one even though I read the instruction. It's t |
| 22 | 22 " etc. to my children, friends and others even though Im all thum in singingWell, |
| 23 | 23 xtbook of horror movie by some people. Even though I don ' t watch thrillers mo |
| 24 | 24 s. My husband is warm-hearted person. Even though he is at same age as me, |
| 25 | 25 use, it seems that I get some presents even though I buy them with my credit |
| 26 | 26 student clubs and volunteer programs. Even though it was a very challenging p |
| 27 | 27 ing. I'm very afraid of speaking in public even though I'm a teacher. It's really ok |
| 28 | 28 ing after seven is one of her strict rules. Even though she has lunch at around t |
| 29 | 29 the same when we went back to Seoul. Even though I prefer 5th or 6th graders, |
| 30 | 30 y uncle's family come down from Seoul. Even though I have to do a lot of works |

[Figure 2 - Comparison between Brown corpus and Learner Corpus]

V. Conclusion

The results propose the application of new descriptive insights to the design of English teaching syllabuses and materials. If each word is translated into Korean in the class, the original meaning might be twisted. The three aims of the learner corpora, which were suggested in research questions, were achieved in this study.

First, the usage of adverbial conjunction 'even though' and 'even if' was investigated through the learner corpus, and some of common errors were found. The result may help the teachers to teach students correct usage of those two adverbial conjunctions. Second, the usage of the 'actually' was quite different from the natives'. It may be related to L1 translation and L1 speaking habit. Third, the results can suggest the pedagogical tool to help the teachers and learners pay more attention on the errors.

The limitation of this study is that the size of the learner corpus was not big enough to generalize. Also, the genre of journal is not clear if it is the spoken register or essay writing. In addition, there is a potential problem that errors might be made by certain few people because one person wrote eight or nine journals. Even though this research has several limitations, this study provides the actual examples which are written by learners and compares with the native's language usage. The specific instances of adverb and conjunction may provide valuable insight into utilizing corpora. Both the teachers and learners take it into account to improve the accuracy as well as authenticity through a corpus-based analysis.

Reference

- Actually. (2005). Oxford University Press. Retrieved September 30, 2008, from http://www.oup.com/oald-bin/web_getald7index1a.pl
- Aston, G. (2001). *Learning with Corpora*. Bologna: Athelstan.
- Even if. & even though. (2008). Longman dictionary. Retrieved September 30, 2008, from <http://pewebdic2.cw.idm.fr/>
- Kennedy, G. (1999). *An introduction to corpus linguistics*. New York: Pearson Education Inc.
- Liszak, S. (2004). Exploring the effects of first language influence on second language pragmatic processes from a syntactic deficit perspective. *Second Language Research*, 20(3), 212-231.
- Lee, S. (2006). A corpus-based Analysis of Korean EFL Learner's Use of Amplifier Collocations. *English Teaching*, 60(1), 3-17.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Tan, N. (2005). Authentic language or language error? Lessons from a learner corpus. *ELT journal*, 59, 126-134.
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Philadelphia: John Benjamins.