Modern English Education, Vol. 9, No. 1, Spring 2008

A Computer-Aided Error Analysis (CEA) of a Korean Learner Corpus

Dong-Ju Lee
Sanggye High School

Lee, Dong-Ju. (2008). A computer-aided error analysis (CEA) of a Korean learner corpus. *Modern English Education, 9*(1), 73-89.

This study describes a computer-aided error analysis (CEA) of a small corpus of writing in English produced by EFL Korean secondary school students. The analysis identified the most common errors the students made using an error-tagging software tool and subsequent analyses of the errors were performed with a view to obtaining information to design remedial corpus-based pedagogical materials to enhance the students' writing and grammar. Despite the error-tagging method being time-consuming, it provides teachers and materials designers with data on the most common error types students are likely to make, thereby making it possible to predict to some extent particular language problems at any given level of proficiency for any given learner group with a higher degree of accuracy.

[CEA/learner corpus/error-tagging/컴퓨터보조 오류 분석/학습자 코퍼스/오류표시]

I. INTRODUCTION

From the early 1990s, the computer learner corpus (CLC), a collection of machine-readable natural language data produced by language learners, has been used as a new source of data for second language acquisition (SLA) and interlanguage research. Many studies of CLC offer a new type of error analysis (EA), which is obviously different from the traditional EA. As Leech (1998) points out, although they are based on a data-oriented approach, CLC-based research differs from traditional EA because "the computer, with its ability to store and process language, provides the means to investigate learner language in

a way unimaginable 20 years ago" (p. xvii).

In CLC, once computerised, learner data can be analysed automatically mainly with tagging and parsing by a wide range of linguistic software tools. Granger and Meunier (1994) and Milton (1994) have argued for the importance of CLC-based learner language research in that the errors that foreign language learners make are widely different from those of native speakers, and it is worth having access to comprehensive lists of authentic learner errors and their respective frequencies.

The International Corpus of Learner English project (ICLE, Granger, 1993) on the computer analysis of learner language has been leading the CLC-based studies. Like the ICLE project, most of CLC-based research usually involves computer-aided error analysis (CEA) (Granger, 2002, 2004), focusing on errors in interlanguage and using computer tools to tag, retrieve and analyse them.

Dagneaux, Denness and Granger (1998) introduced a novel technique of CEA. They used the learner corpus of English written by French-speaking learners of intermediate and advanced level to examine progress rate of the students along a range of grammatical and lexical variables. In the study, they suggested that the CEA approach provides one way of discovering valuable features of learner writing, in particular areas of persistent grammatical difficulties and error-prone items, and sheds light on the development of materials and syllabus design. It shows a fully error-tagged learner corpus can be used as a diagnostic tool for identifying common errors made by particular groups of learners, thereby facilitating ELT materials designers to create useful pedagogical tools which could be expected to be more learner-friendly and learner aware.

II. COMPUTER-AIDED ERROR ANALYSIS OF A KOREAN LEARNER CORPUS[1]

1. Research Question

It has been argued that error-tagged learner corpora are a valuable resource for improving our knowledge of learner interlanguage and an aid to developing and adapting pedagogical materials tailored to meet learners' needs. Although many

1 This analysis was carried out in 2006 and the results and discussion are partly included in the author's unpublished doctoral dissertation in 2007.

corpus-based materials (e.g., Granger & Tribble, 1998; Tribble & Jones, 1990, 1997) provide a very comprehensive view of the authentic use of words, phrases, structures, etc., it is necessary to identify what particular problems learners have at their specific level of proficiency. With the help of CEA techniques, it is possible to identify the most frequent error categories and view the errors in context through concordances of specific error types.

In the current study, a computer-aided error analysis of a small corpus of written English produced by EFL Korean secondary school students was performed to explore the following research question: What are the most common errors Korean secondary school students make in writing?

2. Design of the Korean Learner Corpus (KLC)

The Korean Learner Corpus (KLC) compiled for this study consisted of 19,610 words of written English by Korean secondary school Grade 10 students who are at high beginner/elementary level of English. While the corpus is relatively small in size, this can be defended on the basis that there has been no uniformity of learner corpus size in CEA and a researcher must decide on the appropriate size for his/her corpus depending on his/her own research purpose (Granger, personal communication, April, 2006) Similarly, Harwood and Scholfield (personal communication, June, 2006) agreed that there is no 'magic number' as regards corpus size and the appropriate size will depend upon what the analyst wishes to study. Because the current CEA was to investigate relatively common and frequent error types (e.g., article errors), it can be argued that a corpus of less than 20,000 words is large enough for this purpose.

The data were collected from 290 students (126 boys and 164 girls) attending five different high schools in Korea. The students were asked to write an in-class essay on a given topic: "What do you want to do when you grow up?" They were required to write a free composition for about 50 minutes without the aid of dictionaries or reference books. The length of their writing varies depending on the students: for example, some students wrote only a couple of sentences, whereas some others wrote more than 20 sentences. However, on average each essay is between five and ten sentences and between 50 and 100 words.

In accordance with general principles of corpus linguistics, the data were compiled on the basis of strict design criteria. That is, variables pertaining to the learner (age, language background, learning context, etc.) and the language

situation (medium, task type, topic, etc.) were taken into account so as to compile a homogeneous corpus. All characteristics of the corpus are constant: age (16-17 years old), learning context (EFL), medium (writing), genre (essay type writing), length (approximately 50-100 words, unabridged), task type (in-class task without access to grammar books and dictionary), and topic ("*What do you want to do when you grow up?*").

3. Data Analysis

1) Error-tagging System of the CEA

The present CEA adopted the error system developed by Dagneaux et al. (1998). As mentioned above, their study has been considered as a particularly influential CEA study using the error tagging method.

In their study, Dagneaux et al. (1998) used the error tagging system developed at Louvain University. According to the *Error Tagging Manual version 1.1* (Dagneaux, Denness, Granger & Meunier, 1996) and *1.2* (Dagneaux, Denness, Granger & Meunier, forthcoming), the error classification and error-tagging are based on a descriptive system; the errors are described in terms of linguistic categories, not in terms of the source of error such as L1 transfer, overgeneralization, etc. The system is hierarchical; error tags consist of one major category code and a series of sub-codes. There are eight major categories and 56 sub-categories in the error-tagging system. The eight major categories are: formal (F), grammatical (G), lexico-grammatical (X), lexical (L), punctuation (Q), word redundant/word missing/word order (W), style (S) and infelicities (Z) (see Appendix for more details of the error-tagging system).

These major category codes are broken down into one or more sub-codes, which provide further information on the type of error. For the grammatical category, for example, the first sub-code refers to the word category: GV for verbs, GN for nouns, GA for articles, etc. This code is then followed by any number of sub-codes. For example, the GV category is broken down further into GVAUX (verb auxiliary errors), GVM (verb morphology errors), GVN (verb number errors), GVNF (verb non-finite/finite errors), GVT (verb tense errors) and GVV (verb voice errors). The system is flexible in the sense that researchers can add or delete subcategories to suit their research purposes.

2) Procedures of the CEA

The CEA of the KLC consists of the following steps. Firstly, the learner data is corrected manually by the researcher who inserts correct forms in the text. Secondly, the researcher assigns to each error an appropriate error tag according to the error tagging manual, and inserts the tag in the text file with the correct version. The error-tagging work can be aided by a specially designed editing software tool. The process of inserting error tags and corrections into the text files is very time-consuming and painstaking. Fortunately, the *UCLEE* (Hutchinson, 1996), a kind of MS Windows error editor, can speed up the process.

By just clicking on the appropriate tag from the error tag menu, the researcher can insert it at the appropriate point in the text. Using the correction box, the researcher can also put the corrected version into the text with the appropriate formatting symbols. Figure 1 shows an example of tagging an error (in this case, GVN for Grammar, Verbs, Number—subject-verb concord error—in the sentence: *Children is very cute.*) semi-automatically, with the researcher selecting the appropriate error tag from the error tag menu of the *UCLEE*.

Next, Figure 2 shows how to use the correction box to insert a correct expression and a sample case of a completed error-tagged and corrected text (in this case, *Children (GVN) is are very cute*.). The correction is indicated by the presence of dollar signs.

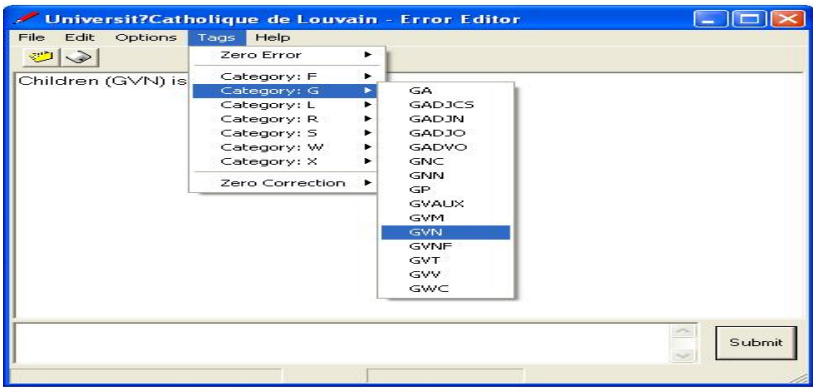

Figure 1. An example of using the error tag menu.

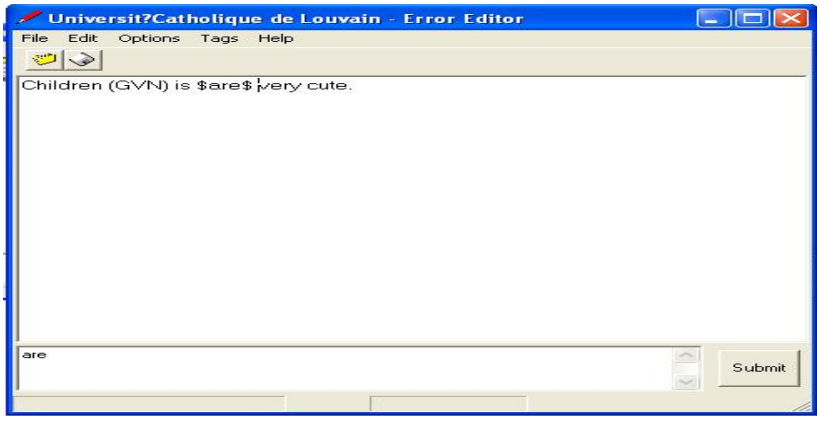

Figure 2. An example of using the correction box.

When the process is completed, the error-tagged data can be analysed using a text retrieval software tool such as *WordSmith Tools* (Scott, 1999), thereby making it possible to search errors just as one could search for any other element in the corpus, count errors, draw up comprehensive lists of specific error types and view errors in context, etc.

4. Results and Discussion

Following the steps illustrated above, the CEA of the KLC was carried out. In brief, the researcher looked at each text in the raw KLC corpus and inserted error tags and corrections following the *Error Tagging Manual* (Dagneaux et al., 1996, forthcoming). The error editor, *UCLEE* (Hutchinson, 1996), speeded up this semi-automatic process. Figure 3 gives an example of a sample of error-tagged text.

I want to be a history teacher because I like history. History is exciting and **(FS)** facination $fascination$ **(GWC)** facination $fascinating$. I feel happy when I **(XVPR)** study about $study$ history. History class is **(LS)** very $lots of $ fun because **(GNC)** history class of Mr. Hong $Mr. Hong's history class$ is very interesting.

Figure 3. Sample of error-tagged text.

Five errors were identified and tagged in this text. Firstly, the word '*facination*' can be seen to be erroneous due to its wrong spelling and was coded (FS) to signify the 'Form, Spelling' error. Thus, the correct form, *$fascination$,* was inserted. However, the word also violated another rule, (GWC)

(Grammar, Word Class), which involves inappropriate use of a word class. In this case, the noun '*fascination*' should be changed to the adjective '*fascinating*', so it was coded (GWC) again. Hence, in line with one of the basic principles of the error-tagging manual, this word was double tagged. Secondly, *'study about'* was coded as (XVPR), to signify an error involving 'Lexico-grammar, Verbs, Dependent Preposition'. In this case, '*about*' is not needed after the verb '*study*'. Thirdly, the word '*very*' was coded as (LS) which stands for a 'Lexical, Single' error category that is used for collocational lexical errors in single words. In this case, *'very fun'* was corrected as *'lots of fun'*. Finally, *'history class of Mr. Hong'* was coded as (GNC), 'Grammar, Nouns, Case' error, which involves genitive errors.

After the error-tagging and inserting work, error codes were searched using the text retrieval tool, *WordSmith Tools* (Scott, 1999). Figure 4 is an extract of the output of a search for errors bearing the code (GNN) (i.e. Grammar, Noun Number errors).

1 (FS) Best $best$ of (GA) 0 the (FS) Best $best$ **(GNN)** artist $artists$ in the (FS)
2 I have many **(GNN)** dream $dreams$ (QC). $,$ (LP) example $for
3 I like computer **(GNN)** game $games$ (QM)0 $.$ I will make (GA) a
4 $introduce$ (WRS) to 0 Korea (LS) for to **(GNN)** foreigner $foreigners$.
5 my (GNC) friends $friends'$ **(GNN)** house $houses$ some day. And I have a great

<More examples follow>

Figure 4. Output of a search for GNN.

A total of 4,479 errors were identified and classified into major categories as well as subcategories as a result of the analysis of the error-tagged KLC corpus. Of the 56 error categories of the *Error Tagging Manual version 1.2* (Dagneaux et al., forthcoming), 48 categories were found in the error-tagged corpus while seven categories (GDT, GPF, XADJCO, XCONJCO, XNCO, FSR, LSF and LPF) were not found.

In the analysis, in fact, the LSF (Lexical Single, False friends) and LPF (Lexical Phrase, False friends) error categories were excluded because there were no errors resulting from the influence of any formally similar word or phrase in the Korean students' mother tongue, Korean. In addition, the FSR (Form, Regional, Spelling) error category was not considered because both American and British English are seen as equally acceptable by most Korean teachers. Appendix

presents the details of the number of errors and proportions of errors found in each category and sub-category.

Based on a frequency count, it was found that grammatical errors **(G)** were the most common error type (37.5%). Punctuation errors **(Q)** were the second most common type of error (14.5%), followed by form **(F)** (12.4%), lexis**(L)** (11.5%), style **(S)** (7.3%), lexico-grammar **(X)** (6.3%), word redundant/word missing/word order **(W)** (5.3%) errors, and infelicities **(Z)** (5.2%). Figure 5 shows the breakdown of the major error categories.

As shown in Figure 6, a further analysis of subcategories revealed that GA (Grammar, Articles: 663 occurrences, 14.8%: *I will make (GA) a 0 many computer games*.) was the largest sub category of all errors in the whole corpus. The second largest sub category was FS (Form, Spelling: 542 occurrences, 12.1%) followed by QM (Punctuation, Missing: 346 occurrences, 7.7%), GVT (Grammar Verb Tense: 304 occurrences, 6.8%), LS (Lexical Single: 254 occurrences, 5.7%), Z (Infelicities: 233 occurrences, 5.2%: *(Z) My hobby is*

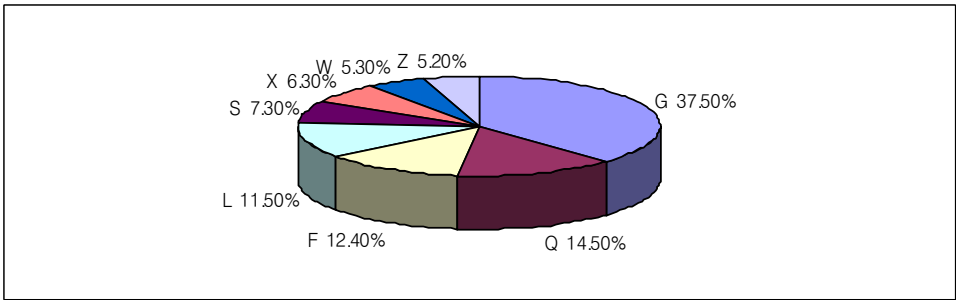

Figure 5. Breakdown of the eight major error categories.

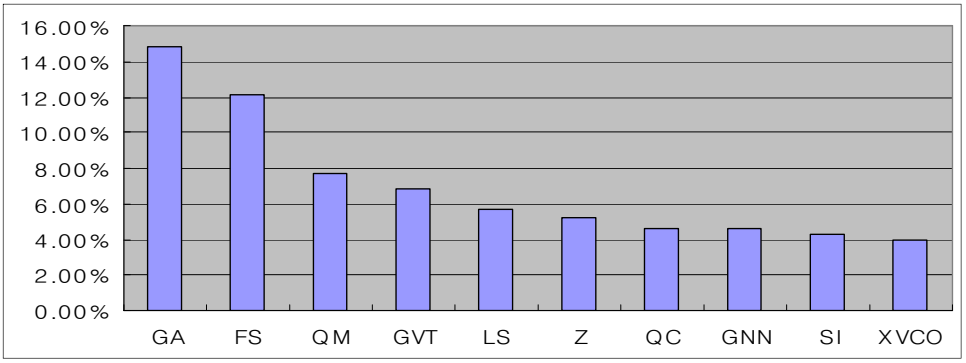

Figure 6. Ten most frequent error categories.

computer $My hobby is playing on the computer$.), QC (Punctuation, Confusion: 208 occurrences, 4.6%), GNN (Grammar Noun Number: 204 occurrences, 4.6%),

SI (Style Incomplete: 194 occurrences, 4.3%), and XVCO (Lexico-Grammar, Verbs, Complementation: 177 occurrences, 4.0%). Figure 6 presents the ten most frequent error categories.

It was a little surprising that the students made so many grammatical errors (1,681 occurrences, 37.5%) even though it is judged that grammar would have been the major focus in class, based on the researcher's experience as a teacher and textbook writer in the Korean context and others' arguments (e.g., Byung-Eun Cho, 2004).

A careful examination of the subcategories reveals that three main areas of grammatical difficulty—articles, verbs and nouns—were found. As can be seen from Figure 7, these three categories account for about 92% of the total number of grammatical errors (39.4% for articles, 29.9% for verbs, and 22.4% for nouns, respectively).

A look at the various sub-categories provides us with a more detailed picture of each of these categories. The breakdown of the GV (Grammar Verbs) category (Figure 8) shows that GVT (Grammar Verb Tense: *Since I was a little boy, I*

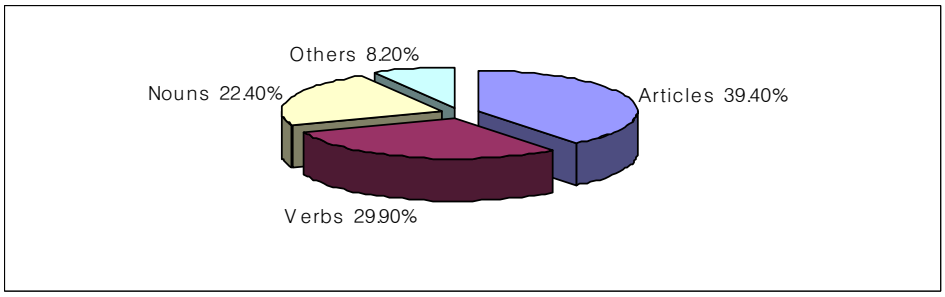

Figure 7. Breakdown of the grammatical error (G) category.

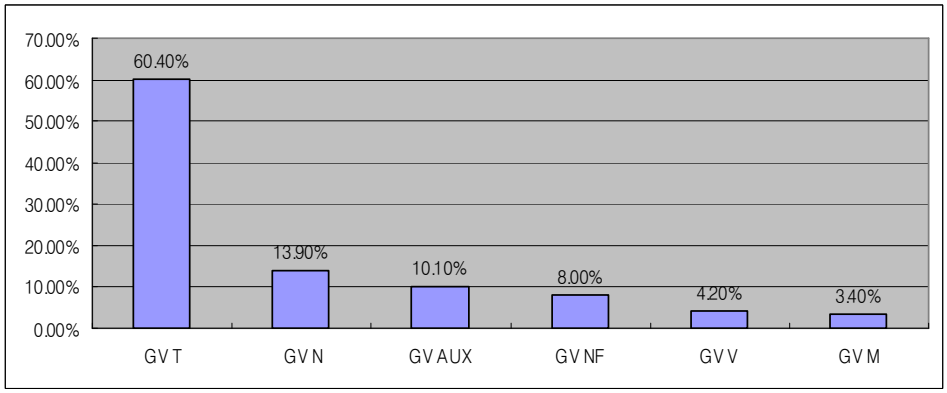

Figure 8. Breakdown of the verb error (GV) category.

(GVT) have $have had$ this dream.) is the most frequently occurring subcategory (60.4% of GV errors). The second largest category is GVN (Grammar Verb Number, 13.9%: *Children (GVN) is are very cute*.) followed by GVAUX (Grammar Verb Auxiliary, 10.1%: *So I (GVAUX) do $will$ keep studying*.), GVNF (Grammar Verb Non-finite/Finite, 8%: *I will (GVNF) to be be a teacher*.), GVV (Grammar Verb Voice, 4.2%: *I want to (GVV) respect $be respected$ by students*.) and GVM (Grammar Verb Morphology, 3.4%: *I will (GVM) made $make$ my parents very happy*.).

The breakdown of the GN (Grammar Nouns) category indicates that GNN (Grammar Noun Number: *I can't remember many (GNN) rule $rules$*.) is the most frequent error subcategory (54.1% of GN errors) followed by GP (Grammar Pronouns, 25.5%), GD (Grammar Determiners, 15%), and GNC (Grammar Nouns Case, 4.5%: *I was interested in (GNC) human's body $human body$*.). The GP category has seven sub-categories and GPD (Grammar Pronouns Demonstrative: *I like sports. I enjoy (GPD) it $them$*.) was found to be the most frequent error type (36.5%) of the GP errors followed by GPP (Grammar Pronouns Personal, 27.1%: *(GPP) I My hobby is designing*.), GPR (Grammar Pronouns Relative/Interrogative, 13.5%: *Now I don't know (GPR) that $what$ I want to be*.), GPU (Grammar Pronouns Unclear References, 10.4%: *And I want to be respected by (GPU) them $students$*.), GPI (Grammar Pronouns Indefinite, 7.3%: *I'm not good at (GPI) something $anything$, so ...*), GPO (Grammar Pronouns Possessive, 5.2%: *This is what (GPO) my I really want*.) and GPF (Grammar Pronouns Reflexive/Reciprocal, 0%). Of the GD errors, GDI (Grammar Determiner Indefinite: *So I will make (GDI) many $a lot of$ money*.) proved to be the most frequent error type (55%) followed by GDO (Grammar Determiner Possessive, 33.3%: *I want to be a teacher because (GDO) 0 my English teacher is very pretty*.) and GDD (Grammar Determiner Demonstrative, 11.7%: *I like (GDD) 0 $these$ kinds of music*.), while GDT (Grammar Determiner Other) errors were not found at all. Figure 9 summarizes the results for noun error subcategories.

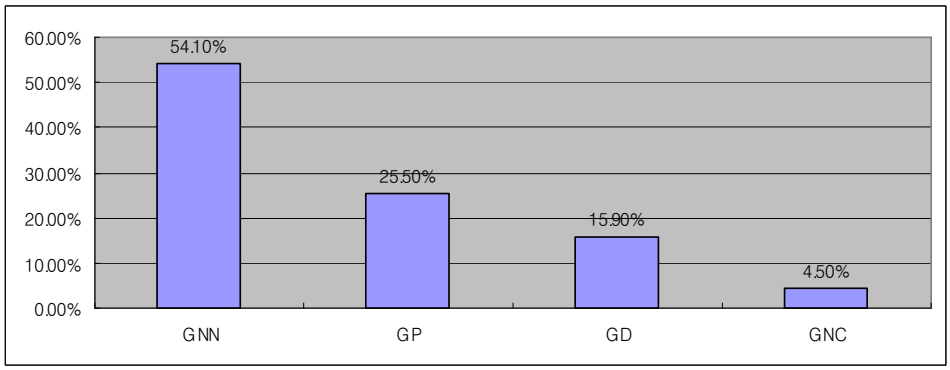

Figure 9. Breakdown of the noun error (GN) category.

In this way, further analyses of each major error category were performed. As regards the second most common error category Q (Punctuation), QM (Punctuation Missing: *I want to be a fashion designer (QM) 0 $.$ So, I'll ...*) was the largest subcategory (53.4%) followed by QC (Punctuation Confusion, 32.1%: *When I was young (QC) . $,$ I was very ...*), QR (Punctuation Redundant, 11%: *I want to be a doctor (QR) . 0 because I ...*) and QL (Punctuation Lexical, 3.5%: *I like science (QL) , and music.*). Figure 10 indicates the breakdown of the punctuation error (Q) category.

Of the F (Form) errors, FS (Form Spelling: *I will not give up being a doctor (FS) easilly $easily$.*) errors were remarkably dominant (97.5%) and there were some FM (Form Morphology, 2.5%: *I (FM) teached $taught$ many friends.*) errors.

As seen in Figure 11, the breakdown of the L (Lexis) category showed that LS (Lexical Single: *I want to (LS) do be a soccer player.*) was the most common error type (49.4%), followed by LP (Lexical Phrase: *(LP) From now*

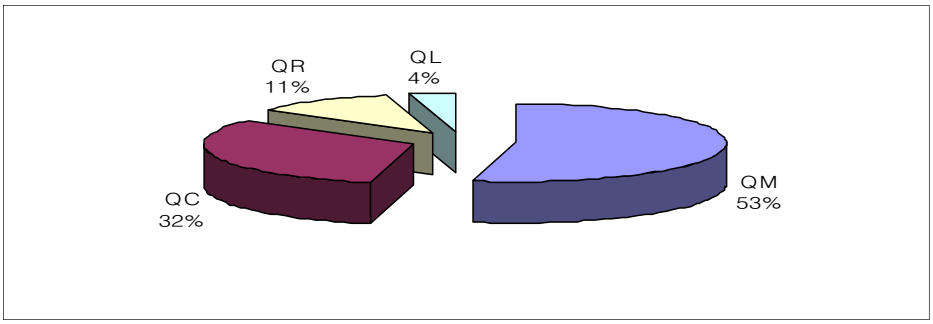

Figure 10. Breakdown of the punctuation error (Q) category.

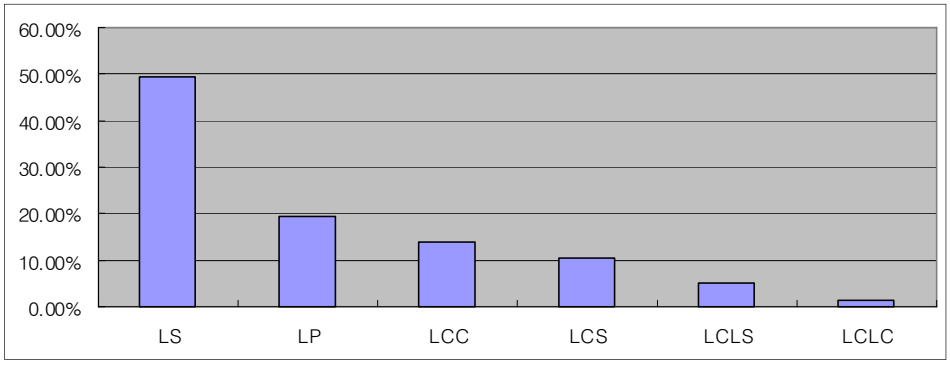

Figure 11. Breakdown of the lexis error (L) category.

$From now on$ I will study hard.) errors (19.6%), LCC (Lexis Conjunctions Coordinating, 14%: *I want to be a teacher (LCC) and or a military cadet.*), LCS (Lexis Conjunctions Subordinating, 10.5%: *(LCS) 0 $When$ I grow up, I will become ...*), LCLS (Lexis Connectors Logical Single, 5.1%: *I'm not a smart student. (LCLS) However So I need to study hard.*) and LCLC (Lexis Connectors Logical Complex, 1.4%: (LCLC) *First of the all $First of all$ ...*).

There were a total of 326 occurrences of the S (Style) error category. Of the two sub-categories, the SI (Style Incomplete: *And one of the most important thing (SI) 0 is studying steadily.*) error type was found to be slightly more frequent (59.5%) than SU (Style Unclear, 40.5%: *I study hard to be a teacher. (SU) Achieve study! $?$*) errors.

As can be seen from Figure 12, the breakdown of the X (Lexico-grammar) category revealed that XVCO (Lexico-grammar Verb Complementation: *... because I (XVCO) like teach $like teaching$.*) was the most frequent error subcategory (62.8% of X errors) followed by XVPR (Lexico-grammar Verb Dependent Preposition, 27%: *My family will (XVPR) move in $move to$ Canada.*), XPRCO (Lexico-grammar Preposition Complementation, 5%: *I am good (XPRCO) at teach $at teaching$ some people.*), XADJPR (Lexico-grammar Adjective Dependent Preposition, 2.5%: *I'm (XADJPR) afraid at $afraid of$ needles.*), XNPR (Lexico-grammar Noun Dependent Preposition, 1.8%: *My (XNPR) enthusiasm in $enthusiasm for$ animals still ...*) and XNUC (Lexico-grammar Noun Uncountable/Countable, 1.1%: *There is much (XNUC) informations $information$ in ...*). There were no errors involving the other three sub-categories (XADJCO, XNCO and XCONJCO).

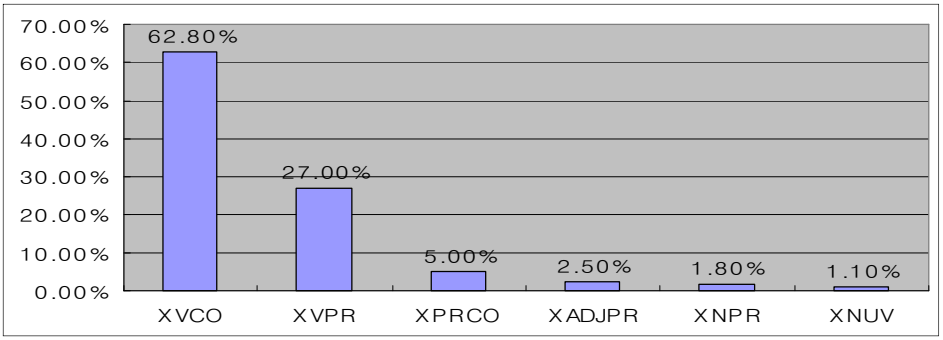

Figure 12. Breakdown of the lexico-grammar error (X) category.

Finally, the breakdown of the W (Word Redundant/Missing/Order) category revealed that WM (Word Missing: *I will give all my love and mind (WM) 0 to my mom* ...) is the most frequently occurring subcategory (39.7% of W errors). The second largest category is WRS (Word Redundant Single, 35.1%: *I have not a dream (WRS) not 0 yet*.) followed by WO (Word Order, 18.0%: *I try to (WO) hard English study $study English hard$*.) and WRM (Word Redundant Multiple, 7.1%: *I can go abroad and (WRM) I can 00 make a lot of money*.). Figure 13 shows the results for W error subcategories.

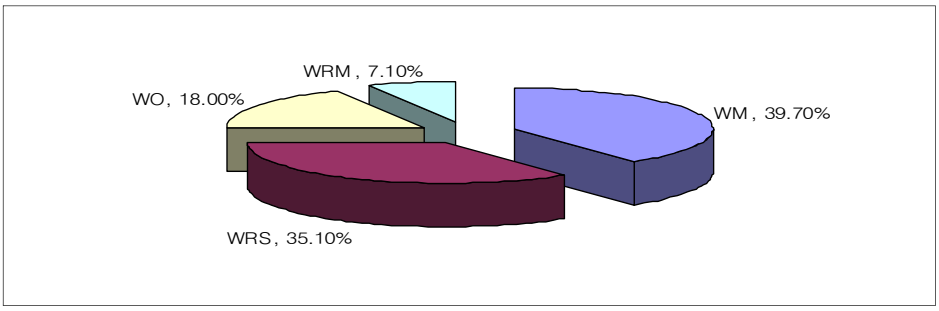

Figure 13. Breakdown of the word redundant/missing/order error (W) category.

The results of this quantitative approach to the fully error-tagged learner corpus are potentially very useful. As discussed above, they show the most common errors which teachers can then focus on when teaching Korean-speaking EFL secondary school grade ten students. Further more detailed analyses of the data can provide more information on the students' problems and should prove useful for materials designers aiming to write appropriate teaching-learning materials best suited to cater for their needs.

III. CONCLUSION

The current study has demonstrated the potential of the error-tagging approach of CEA. It can be used to generate comprehensive lists of specific error types, count and sort them in various ways and view them in context and alongside instances of non-errors. It can also help ELT specialists to identify learners' problems at any given level of proficiency for any given learner group with a much higher degree of accuracy than would be possible simply by using intuition, introspection, or other non-corpus-based grammars and reference works that most studies of traditional error analysis depended on. Studies such as the present one will hopefully thereby lead designers to develop more appropriate ELT materials and methods which take learners' needs more fully into account.

It is hoped that the error-tagged Korean secondary students' corpus will be a very rich source of data for tackling the most common errors this level of students make in their writing. If various learner corpora made up of writing from different school grades were analysed and made into a pedagogical database, and made publicly available, it would enable teachers to raise students' awareness of the items they are likely to find problematic, and teachers would be able to create contextualized exercises using a corpus-based and/or data-driven learning (Johns, 1991a, 1991b; Meunier, 2002) framework.

However, there are some limitations of the current CEA. Firstly, although the error analysis has been conducted in terms of total frequency of errors, it would also be worth carrying out another type of analysis which looks at the proportion of correct and incorrect output of any given item. In other words, it would be interesting to analyse what percentage of each item produced by the student writers is correct in relation to the total number of occurrences of each item. However, mainly due to time limitations, it has not been possible to conduct such an analysis.

Secondly, like all corpus studies, it is evidence of the subjects' performance rather than their competence which has been compiled, and thus it is not possible to judge the latter: just because an item is absent from the writers' repertoire on the basis of corpus evidence, this does not necessarily mean the writers have no knowledge of the said item (cf. Harwood, 2006; Kaltenböck & Mehlmauer-Larcher, 2005). And neither does the corpus evidence guarantee, where there is evidence of correct output, that errors will not be made when the

learner attempts to produce the same language item in a subsequent piece of writing: the corpus is only a snapshot of writers' performance at a given moment of time.

Another limitation is that since the current CEA has primarily investigated erroneous usage of English by a group of students of a similar age and level, the findings will not be equally useful for teachers and students who are teaching/learning English at different school grades or different proficiency levels.

One final issue worthy of comment concerns a practical, rather than a methodological limitation—that of time. Although the error analysis was conducted with the help of a computer and computer software, it was not fully automatic and only certain information (such as error frequencies or tag sequences) could be extracted automatically. A very considerable amount of time and energy was also needed to compile, correct, tag, and analyse the corpus.

On the basis of the findings of the current CEA research, teacher-designed concordance-based materials can be developed as teaching lessons to remedy the students' most frequent error types, designing, adopting and adapting activities similar to those which featured in the studies of Gaskell and Cobb (2004), Granger and Tribble (1998), Sripicharn (2002), Tribble and Jones (1990), etc. These kinds of corpus-based/data-driven learning materials and activities can draw students' attention to problematic language items and to have them identify and classify language patterns and use inductive strategies rather than depending on deductive ones.

REFERENCES

Cho, Byung-Eun. (2004). Issues concerning Korean learners of English: English education in Korea and some common difficulties of Korean students. *The East Asian Learner, 1*(2), 31-36.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System, 26*, 163-174.

Dagneaux, E., Denness, S., Granger, S., & Meunier, F. (1996). *Error tagging manual version 1.1*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

Dagneaux, E., Denness, S., Granger, S., & Meunier, F. (Forthcoming). *Error tagging manual version 1.2*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*, 301-319.

Granger, S. (1993). International corpus of learner English. In J. Aarts, P. de Hann & N. Oostdijk (Eds.), *English language corpora: Design, analysis and exploitation* (pp. 57-71). Amsterdam: Rodopi.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam: John Benjamins.

Granger, S. (2004). Computer learner corpus research: Current status and future prospects. In U. Corner & T. Upton (Eds.), *Applied corpus linguistics: A multidimensional perspective* (pp. 123-145). Amsterdam: Rodopi.

Granger, S., & Meunier. F. (1994). Towards a grammar checker for learners of English. In U. Fries & G. Tottie (Eds.), *Creating and using English language corpora* (pp. 79-91). Amsterdam: Rodopi.

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* (pp. 199-209). London: Longman.

Harwood, N. (2006). (In)appropriate personal pronoun use in political science: A qualitative study and a proposed heuristic for future research. *Written Communication, 23*(4), 424-450.

Hutchinson, J. (1996). *UCL error editor*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

Johns, T. (1991a). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *ELR jourrnal 4: Classroom concordancing* (pp. 1-16). Birmingham: CELS, The University of Birmingham.

Johns, T. (1991b). From printout to handout: Grammar and vocabulary teaching. In T. Johns & P. King (Eds.), *ELR jourrnal 4: Classroom concordancing* (pp. 27-46). Birmingham: CELS, The University of Birmingham.

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

Lee, Dong-Ju. (2007). *Corpora and the classroom: A computer-aided error analysis of Korean students' writing and the design and evaluation of data-driven learning materials*. Unpublished doctoral dissertation. University of Essex, UK.

Leech, G. (1998). Learner corpora: What they are and what can be done with them. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). London: Longman.

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. (1994). A corpus-based online grammar and writing tool for EFL learners: A report on work in progress. In A. Wilson & T. McEnery (Eds.), *Corpora in language education and research* (pp. 65-77). Unit for computer research on the English language: Lancaster University.

Scott, M. (1999). *WordSmith tools*. Oxford: Oxford University Press.

Sripicharn, P. (2002). *Evaluating data-driven learning: The use of classroom concordancing by Thai learners of English*. Unpublished doctoral dissertation. University of Birmingham, UK.

Tribble, C., & Jones, G. (1990). *Concordances in the classroom: A resource book for teacher*. London: Longman.

Tribble, C., & Jones, G. (1997). *Concordances in the classroom: Using corpora in language education*. Houston, TX: Athelstan.

APPENDIX

The 56 Error Categories of the Error Tagging System

Major categories	Sub-categories/ Number of sub-categories
Form (F)	FM (form, morphology), FS (form, spelling), FSR (form, regional, spelling): 3
Grammar (G)	GA (grammar, articles), GDD (grammar, determiner, demonstrative), GDO (grammar, determiner, possessive), GDI (grammar, determiner, indefinite), GDT (grammar, determiner, other), GADJCS (grammar, adjectives, comparative/superlative), GADJN (grammar, adjectives, number), GADJO (grammar, adjectives, order), GADVO (grammar, adverbs, order), GNC (grammar, nouns, case), GNN (grammar, nouns, number), GPD (grammar, pronouns, demonstrative), GPP (grammar, pronouns, personal), GPO (grammar, pronouns, possessive), GPI (grammar, pronouns, indefinite), GPF (grammar, pronouns, reflexive/reciprocal), GPR (grammar, pronouns, relative/interrogative), GPU (grammar, pronouns, unclear reference), GVAUX (grammar, verbs, auxiliaries), GVM (grammar, verbs, morphology), GVN (grammar, verbs, number), GVNF (grammar, verbs, non-finite/finite), GVT (grammar, verbs, tense), GVV (grammar, verbs, voice), GWC (grammar, word class): 25
Lexico-grammar (X)	XADJCO (lexico-grammar, adjectives, complementation), XCONJCO (lexico-grammar, conjunctions, complementation), XNCO (lexico-grammar, nouns, complementation), XPRCO (lexico-grammar, prepositions, complementation), XVCO (lexico-grammar, verbs, complementation), XADJPR (lexico-grammar, adjectives, dependent preposition), XNPR (lexico-grammar, nouns, dependent preposition), XVPR (lexico-grammar, verbs, dependent preposition), XNUC (lexico-grammar, nouns, uncountable/countable): 9
Lexis (L)	LS (lexical, single), LSF (lexical, single, false friends), LP (lexical, phrase), LPF (lexical, phrase, false friends), LCLS (lexis, connectors, logical, single), LCLC (lexis, connectors, logical, complex), LCC (lexis, conjunctions, coordinating), LCS (lexis, conjunctions, subordinating): 8
Punctuation (Q)	QC (punctuation, confusion), QL (punctuation, lexical), QM (punctuation, missing), QR (punctuation, redundant): 4
Word redundant/ missing/order (W)	WM (word missing), WO (word order), WRS (word redundant, single), WRM (word redundant, multiple): 4
Style (S)	SI (style, incomplete), SU (style, unclear): 2
Infelicities (Z)	Z (infelicities): 1

Dong-Ju Lee
Sanggye High School

Geum-ho Apartment 102-808,
Junggyebon-dong, Nowon-gu, Seoul, 139-930, Korea
Tel: (02) 933-2707/ H.P.: 010-2053-2707
Email: maydjlee@hanmail.net

Received in January, 2008
Reviewed in February, 2008
Revised version received in March, 2008