

ICALL FOR IMPROVING KOREAN L2 WRITERS' ABILITY TO EDIT GRAMMATICAL ERRORS

Ron Cowan, University of Illinois at Urbana-Champaign

Jinhee Choo, University of Illinois at Urbana-Champaign

Gabseon Sunny Lee, University of Illinois at Urbana-Champaign

This study illustrates how a synergy of two technologies—Intelligent Computer-Assisted Language Learning (ICALL) and corpus linguistic analysis—can produce a lasting improvement in L2 learners' ability to edit persistent grammatical errors from their writing. A large written English corpus produced by Korean undergraduate and graduate students enrolled at an American university was analyzed to determine four persistent grammatical errors in their writing assignments. An ICALL program was then designed to improve these L2 learners' awareness of these errors in texts and provide practice in correcting them. A pre-test/post-test experiment revealed a significant improvement in recognizing and correcting the four errors by Korean L2 learners who had taken the ICALL program over a matched control group that had received standard L2 writing instruction. This improvement held up on a second post-test administered five months later. The implications of the results for L2 instruction and the design of iterative ICALL tutors are discussed.

Keywords: CALL, Corpus Analysis, Persistent Grammatical Errors, L2 Writing

APA Citation: Cowan, R., Choo, J. & Lee, G. S. (2014). ICALL for improving Korean L2 writers' ability to edit grammatical errors. *Language Learning & Technology*, 18(3), 193–207. Retrieved from <http://llt.msu.edu/issues/october2014/cowanchoolee.pdf>

Received: October 10, 2012; **Accepted:** November 26, 2013; **Published:** October 1, 2014

Copyright: © Ron Cowan, Jinhee Choo & Gabseon Sunny Lee

INTRODUCTION

This study has two purposes. The first is to report on the effectiveness of an online Intelligent Computer-Assisted Language Learning (ICALL) program designed to improve Korean second language (L2) learners' grammar awareness and grammatical correction skills. The second is to demonstrate how corpus analysis tools facilitate the selection of L2 writing errors targeted in the design of ICALL grammar correction courseware that addresses persistent/recurring errors in the writing of specific L2 populations.

Background

The program described in this study, *The ESL Writing Tutor*, falls within the domain of ICALL (formerly, Intelligent Language Tutoring Systems [ILTS]) tools, which utilize natural language processing (NLP) to provide individualized feedback to learners about their performance on exercises that support L2 instruction (Meurers, 2012). The exercises may be devoted to a number of language learning activities, such as vocabulary, comprehension, culture, and writing, but a central feature of most ICALL programs is raising L2 grammar awareness. Activities designed for this purpose often take the form of asking students to produce a sentence based on a string of lexical items, (e.g. Heift, 2010) or as a response to a question about a previously presented context, (e.g. Nagata, 1996). The ability of ICALL systems to provide informative and meaningful feedback to the student is crucial to the success of these activities. Based on a number of studies (Bowles, 2005; Heift, 2004; Heift & Rimrott, 2008; Nagata, 1996; Nagata & Swisher, 1995; Pujola, 2001; Rosa & Lowe, 2004), designers of ICALL programs have concluded that explicit feedback, i.e. feedback that immediately identifies a grammatical error and then links the student

to information that can be used to correct it, is the most effective kind for students studying an L2 at all proficiency levels. An example of this is shown in Heift (2010, p. 447), where a student who has made a subject-verb agreement error is first alerted to this and then, as an aid to correction, shown an inflectional paradigm with the verb he or she used in the initial erroneous attempt to form a grammatical sentence. ICALL tutors can store students' attempts at achieving correct grammatical output, and it has been suggested that this database, in addition to being useful for the improving the design of the materials, could have value for investigating interlanguage (IL).

Like the ICALL tools described above, *The ESL Writing Tutor* also uses NLP and is designed for web-based delivery, but it is not intended to support an L2 language course that encompasses multiple levels of proficiency. Instead, it targets a specific population of learners whose L2 proficiency is well above that of students who complete such courses. *The ESL Writing Tutor* was intended as a response to a problem encountered by the consultants at the Writers Workshop, a unit of the library at a large mid-western university in the United States. The consultants had found an inordinate number of grammatical errors in the writing of Korean graduate students, so it was decided to design ICALL courseware capable of improving the ability of these students to detect and correct these persistent grammatical errors. The courseware would, after testing, be improved and made available to all Korean undergraduate and graduate students at the university, and it could also be used in conjunction with the ESL remedial writing courses that most of these students are required to take.¹

Both the structure and feedback used in *The ESL Writing Tutor* also differ from the ICALL programs described above. Since the Korean students had been using English for some time, it seemed likely that many of their grammatical errors would display systematic deviations from specific English structures, such as the order of determiners and quantifiers before head nouns, and the appropriate use of articles. For example, native speakers of Chinese (Yip, 1995), Korean (Ju, 1995), and Japanese (Oshita, 2001) have been shown to make the same type of passivization error with English ergative (unaccusative) verbs. Furthermore, Han (2000) found "pseudo-passive" errors (sentences like **The letter about graphics file is not received*, where the intended form is assumed to be a passive structure in English) to be so persistent in the speech of two subjects he had studied—both Chinese academics who had been living in America for two years but had achieved TOEFL scores of over 600 ten years earlier—that he felt this error was likely to become fossilized. We reasoned that to be effective, the *Tutor* should focus on raising the learners' consciousness of how their productions deviate from syntactic structures that they have been repeatedly exposed to in the past, and that the feedback in our courseware should prompt the learners to think about this difference. Accordingly, when learners made errors, the *Tutor* feedback did not link them to information containing examples of correct forms; rather it prompted them to think about the instruction they had received on what to avoid when using one of these problem structures. Our decision to build this kind of feedback into *The ESL Writing Tutor* was also influenced by a number of studies (Adams, Nuevo, & Egi, 2011; Kim, 2005; Lyster & Ranta, 1997; Sheen, 2004), which indicated that this metalinguistic prompting might enhance the prospect of L2 learners' producing correct responses.

Using a Learner Corpus to Identify Errors

In order to develop a large database containing errors typical of those found for Korean students by the Writers Workshop, we followed the approach advocated by Granger (2003) and applied corpus analysis tools to a large written corpus produced by students who fit the target population: Korean students in the ESL remedial writing courses described above. The corpus, which was compiled over a number of years, comprises over 231, 228 words consisting of text files of essays, writing assignments, and diagnostic tests produced by Korean undergraduate and graduate students. To minimize the possibility of plagiarism, only first drafts of writing assignments were collected. This corpus has the advantage of providing a large sample of written English for Korean native speakers writing academic English in various disciplines. Although similar to corpora like the French Interlanguage Database (FRIDA), the International Corpus of Learner English (ICLE) and the Varieties of English for Specific Purposes dAtabase (VESPA) corpus

compiled at the Center for English Corpus Linguistics (CECL) at the Université catholique de Louvain, it differs in one important way: it is indexed to two tests, the TOEFL and the English as a Second Language Placement Test (EPT), and hence affords the user a more accurate definition of the learners' English proficiency. The fact that the "advanced" level proficiency of the L2 learners is defined in terms of the parameters of a widely used international English proficiency test makes this corpus particularly useful for replications and other applied research on L2 writing. It also facilitates investigation of SLA topics of current interest such as the reality of fossilizations (Long, 2003) and the nature of "hard" rules (DeKeyser, 2003, 2005).

To determine which errors could legitimately be considered persistent for our population of Korean learners of English, we began by manually inspecting our corpus and recording recurrent patterns. From this initial inspection, four errors emerged which displayed consistent patterning and had the highest frequencies in both the undergraduate and graduate writing samples. We therefore decided to make these errors the focus of our investigation and proceeded to carry out an in-depth examination that would reveal their structure in greater detail as well as their relative frequency in much larger sample. To do this, we applied error-tagging software to this larger sample: UCL Error Editor created by CECL. This software allows users to insert error tags that denote the deviation from the correct version of a given structure in the text files. An example of this is the error code UED, which indicates omission of an obligatory definite article before a head noun. The omission is shown in the tagged sequence *0\$ the \$ government*, where *0* represents the omission of the article, and the two dollar signs surround the definite article, *the*, which should have preceded the head noun *government*. Once the text files were tagged, we applied Wordsmith 3 text retrieval software (Scott, 2008), which allowed us to extract every instance of each error type from the texts in our samples. Overall frequencies for individual error types were calculated and converted to percentages (number of errors/number of items in each syntactic category x 100). Inter-rater reliability for this study was the ratio of error corrections agreed upon by two raters to the total number of errors identified. For the purpose of this study, we decided to select only errors for inclusion in *The ESL Writing Tutor* that occurred over ten percent of the time in obligatory contexts. This base level was set to exclude the possibility that we might be selecting errors that were in the final stages of disappearing from the L2 learners' interlanguages.

Four Persistent Grammatical Errors

The first persistent error was the passivization of English ergative (unaccusative) verbs (*fall, die, drown* etc.) and predicates of existing and happening (*exist, happen, occur, last* etc.). This error type has been found for Chinese, Japanese, and Korean learners of English, and has been discussed by a number of researchers (Balcom, 1997; Hirakawa, 1995; Hwang, 1997; Ju, 1999; Oshita, 2001; Yip, 1995). Passivization of these two types of English verbs results in errors like **The lecture was lasted for three hours*, and **Our offspring will be suffered because we neglect the environment*. The corpus analysis revealed extremely high error rates, ranging from 25% to 66.7%. Three "paired" ergative verbs (verbs which have transitive counterparts), *improve, change* and *decrease*, had error rates of 66.7%, 62.9%, and 43.5%, respectively.²

The second type of persistent error made by the Korean learners encompassed aspects of the usage of English articles. Korean does not have an overt article system, and nouns can be unboundedly used as definite or indefinite without explicit grammatical markers. As long as the context allows readers to understand what is being talked about, speakers of Korean tend to omit the closest equivalent to the English definite article. In addition, Korean does have a concept of countability in nouns, but uses a different and optional system to express this. Furthermore, some English non-count nouns, such as *equipment, stuff, vocabulary, homework, education, information*, and so forth are countable in Korean. The result of this combination of factors is that Koreans find English articles extremely difficult to master, and they frequently omit them in running discourse. The corpus analysis revealed a consistent omission of obligatory occurrences of both definite and indefinite articles for both undergraduate and

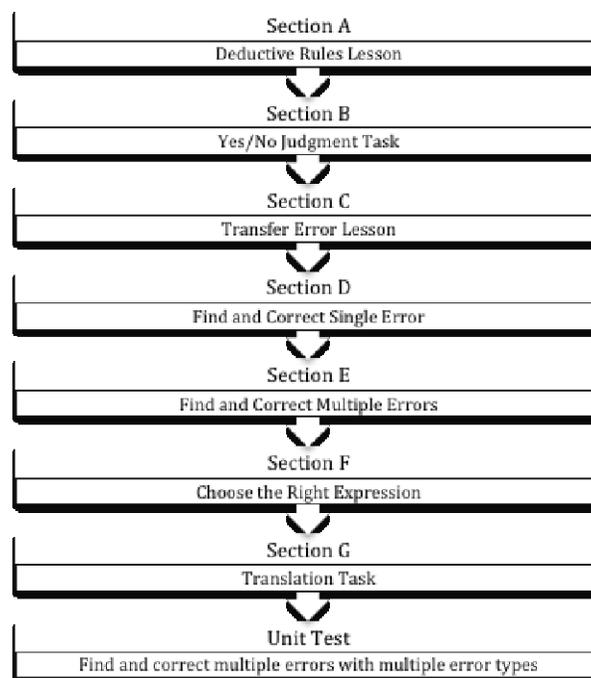
graduate students, with no noticeable differences attributable to course level or essay length. The average number of omissions for graduate essays was 33% of all obligatory occurrences; the range was from 22% to 45%. Overwhelmingly, definite articles constituted the most frequent omissions in obligatory contexts (e.g. **I had \emptyset same experience* for *I had the same experience*), but occasionally they were supplied in contexts where \emptyset article is preferred, for example in referring to generic concepts. So we found examples like **before starting the school* in a context which clearly indicated the broader generic *before starting \emptyset school* was the only appropriate choice. None of the five main discourse contexts where definite articles are obligatory (Poesio & Viera, 1998) predominated. Indefinite articles displayed the same pattern as definite articles.

The form of noun phrases (NPs) with the prenominal quantifiers *most*, *some*, *many* and *one* constituted the third persistent error selected. The Korean students mixed *most of the nouns (Ns)* with *most Ns*, producing errors like **Most of Americans like McDonald hamburgers* and **Some of students didn't go to school today*. Error rates ranged from 17% to 23% for *some* and *most*, and 12% for *one*.

The fourth persistent grammatical error involved the choice of English demonstrative determiners before head nouns. Korean has only one morphemic representation of each set of English demonstrative determiners, hence Koreans tend to produce English noun phrases (NPs) with singular demonstrative determiners before plural head nouns in their writing, for example, **this mistakes*, and the opposite, **those argument*. The rate of occurrence for these types of errors in obligatory contexts was lower than that of the other three rules: 12.7% in undergraduate essays and 11.6% in graduate essays. There were twice as many errors involving plural noun forms than singular noun forms.

The ESL Writing Tutor

All of the units in *The ESL Writing Tutor*, which was developed by Doe-Hyung Kim, have the lesson structure shown in [Figure 1](#).



[Figure 1](#). The ESL Writing Tutor unit design.

The individual sections in each grammar unit utilize a combination of explicit instruction and metalinguistic prompting feedback for error correction. Section A briefly presents a short review of the

salient facts regarding the correct formation of the grammatical construction targeted by each lesson. For example, Section A of the unit that addresses the passivization of ergative verbs presents a brief overview of how active sentences are transformed into the passive voice, part of which is shown in Figure 2.

The screenshot shows the 'ESL Writing Tutor' interface. On the left is a navigation menu with sections A through E, and 'YOU ARE ON...' indicating the current section. The main content area is titled 'How to Make Passive Sentences'. It includes a 'Directions' box with instructions to read carefully and click the 'NEXT' button. Below this, it explains that active sentences have a subject, verb, and object, using the example 'Leonardo da Vinci painted the Mona Lisa.' It then shows how to transform this into a passive sentence: 'The Mona Lisa was painted by Leonardo da Vinci'. A final example shows an active sentence 'Farmers grow a lot of rice in South East Asia' and explains that the words after 'by' can be omitted.

Figure 2. Section A in “Passive Unit”.

This overview is then followed by a grammaticality judgment task with five to ten sample sentences, some of which contain the error that is going to be addressed in the next section. The grammaticality judgment task primes the students for what they will encounter in Section B. When students declare an ungrammatical sentence grammatical, they receive explicit feedback that informs them of this and an explanation of why the sentence is ungrammatical. For example, in the lesson that targeted passivization of ergative verbs, the students, having completed the short review of the passive voice in Section A, encounter a sentence shown in (1) below. If they click a “grammatical” alternative, they receive the message shown under it and an example of the correct sentence.

(1) Most women in developing countries are suffered from extreme poverty.

Student clicks: Grammatical

Feedback: This is not correct, Kim. *Suffer* is a verb that only appears in active sentences.

The task serves as an introduction to Section C, which describes how aspects of the learner’s L1 may lead to grammatical errors in their written English. For example, staying with the same error, passivization of ergative verbs, Section C draws the learner’s attention to the fact that the Korean equivalents of many English verbs like *seem*, *happen*, *occur*, and so forth appear in Korean with passive morphemes or verbs. Next, students are told that these English verbs cannot appear in passive sentences, as shown in Figure 3. Finally, the problem of English paired ergative verbs is presented. Students are shown that English verbs like *boil*, *bounce*, *break*, *change*, and so forth can appear in active and passive sentences, and they are cautioned not to use the passive form of the verb when they want to write a sentence indicating that the

subject experiences the action.

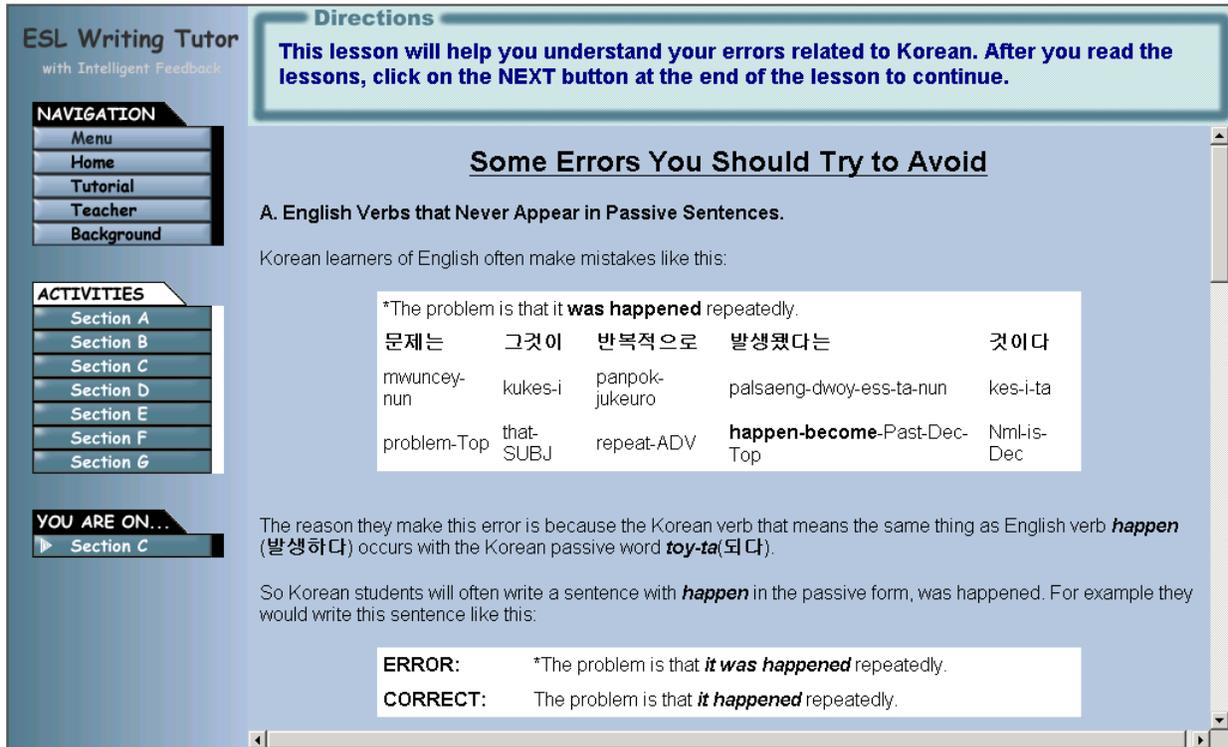


Figure 3. Section C in “Passive Unit”.

After completing Section C, students proceed to Section D, where they read single sentences and then short passages of several sentences and examine them for errors. When they find an error, they highlight it. If the highlighted portion of a sentence contains no error, they receive a message telling them to continue examining the passage in relation to the grammar topic of that lesson. When they highlight an error, they are complimented (*Good choice!*) and told to correct it by typing changes in a box that appears below the sentence. Here the feedback supplied to students differs from the explicit approach used in ICALL courseware described earlier. Because the instruction they have just received in Section C focused on showing them what to pay attention to in order to avoid specific errors, the ICALL feedback in this section prompts the students to think about this, rather than immediately linking them to specific information that would be helpful in correction. After each unsuccessful attempt, the original error reappears in the box, and students are encouraged to consider their answers in relation to the content in Sections A and C. An example of this feedback is shown in (2) below. Although the student has correctly highlighted the error, *have been improved*, he or she initially types in an inaccurate correction.

(2) Sentence:

Over the past three years relations between the United States and China have been improved considerably. This is due partly to an increase in trade between the two countries.

Attempt 1: Student types: had been improved.

Feedback: Try to think of the kinds of problems we saw in “PASSIVE.”

Attempt 2: Student types: were improved

Feedback: Let’s try again. Hint: Is a relationship “improved” or does it improve?

Attempt 3: Student deletes *been*, leaving *have improved*.

Feedback: Good job! Check to see if there are any other errors in the sentences.

Section E has the same format as Section D, but it presents an entire paragraph with multiple examples of the grammar topic in the unit. After completing Section E, students are directed to Section F, where they are presented with sentences that require the choice of the appropriate form of the target grammar point. They then move to Section G, which contains translation tasks on the target grammar topic. Upon completing Sections A through G, students receive a summary of their performance that shows the correct and incorrect choices as well as the percentage of correct choices.

All of the ungrammatical sentences in Sections D and E are taken from the error corpus, and about twenty-five percent of the sentences used in these two sections contain no errors. Each time section D is selected, the sentences are randomly reordered. The beta version of *The ESL Writing Tutor* tested in this study contains a teacher module shown on the navigation box in the top left hand corner of the screenshots shown in Figures 2 and 3. Teachers can enter sentences containing new examples of a particular error type that they find in a student's compositions in the teacher module, which has two purposes: (1) to provide a constantly expanding database of sentences that can be used in Sections D and E, and (2) to afford teachers the possibility of customizing Sections D and E for individual students. As teachers using *The ESL Writing Tutor* become more familiar with the error types, they can tag newly discovered examples and add these to the database with cautionary messages to individual students.

THE STUDY

Method

The effectiveness the ICALL program in raising Korean ESL students' awareness of correct grammatical form and enhancing their ability to correct grammatical errors in English passages was tested in a controlled experiment. The experiment compared the performance of a group of students who received four hours of supplementary instruction with *The ESL Writing Tutor* with another group that did not. The participants, who were randomly assigned to each group, were enrolled in the ESL remedial writing courses, where they continued to receive instruction on how to write academic English. This instruction focuses on the rhetorical organization of academic papers, but teachers may enter grammatical corrections on writing assignments. However, the type of feedback that the students in those classes receive is primarily about rhetorical organization such as topic and transitional sentences, paragraph development etc. The group that did not receive instruction via *The ESL Writing Tutor* did not receive in-class focused grammatical feedback on the four errors described above.

Instrument

The assessment measure was designed to mirror the task for which the ICALL program had been designed: editing grammatical errors from written prose. It consisted of a one and a half page English passage of 615 words. Thirty-two test items (eight items in each of the four grammatical rules in the ICALL program) were embedded along with an equal number of correct examples of each. Since the test replicated a task that the participants perform regularly in their remedial writing classes, that is, editing grammatical mistakes in their compositions before handing in a final version, it can be said to have good face and content validity. The same essay was used for three tests: the pretest, the first posttest, and the second posttest. The topic of the passage was a current event related to environmental pollution. An abridged example of the first two paragraphs of the essay with italicized errors is shown in (3). The errors were not italicized on the test, and the null sign, \emptyset , which indicates a missing indefinite article in this example, also did not appear on the test.

(3) *Most of scientists* would agree that atmospheric pollution *is being increased* throughout the

world. Governments are coming under increasing pressure to take measures to reduce *these problem*. Interestingly, not *all of scientists* agree on how serious atmospheric pollution has become, and this has allowed nations that depend on heavy industry to claim that pollution will decrease without any special efforts by governments. But *all of nations* who signed the Kyoto Protocol in 1997 believe that increased pollution *will be resulted* if industries that send harmful gases into the air are not subject to governmental controls.

New forms of pollution are discovered every year. Recently, using planes, ships and balloons, *ø international team of scientists* tracked a gritty, black cloud that was nearly two miles thick...

Only corrections that demonstrated accuracy on both finding and correcting an error received a score of one. Partial corrections (correct identification only) received a score of zero.

Participants

Forty Korean-speaking English learners volunteered for this experiment and were paid for their participation. Twenty-two of the L2 learners were assigned to the instructional treatment group, the ICALL group, and eighteen to a NO ICALL control group, which received no such instruction. We also included a comparison group of nineteen native speakers of English, the NS group, for two reasons. First, even though the test measure had been pre-tested extensively, variability clearly exists among native speakers on any proofreading task. We felt that the performance of the native speaker comparison group would provide a baseline competency in recognizing errors in running prose and, concomitantly, a measure of the test's reliability. Secondly, we wanted a means of determining whether any improvement found in the performance of the treatment group approached the level of native speaker competence. Because the grammatical rules selected for this study represented persistent errors made by L2 learners with advanced English proficiency, it was reasonable to assume that they might be near fossilizations and hence nearly impossible to eradicate. Comparison of the L2 results with the NS scores would provide an indication of how much farther the ICALL courseware would have to be modified and extended to provide this particular group of L2 learners with the ability to achieve native speaker parity when editing their English writing.

The NS group was made up of native speakers who taught English writing to either native or non-native speakers: three were writing teaching assistants in the English Department, thirteen were instructors from an Intensive English Institute at the university, and three taught in the ESL writing service courses. All of the Korean participants were graduate students and were enrolled in one of the graduate writing courses described earlier. Their English proficiency fell within the TOEFL score range (scores of 550–610 Paper, 213–253 CBT or 79–102 iBT, inclusive) that placed them in one of the three remedial writing courses. Language experience and biographical background were assessed by a brief questionnaire. All of the L2 participants spoke Korean as their first language and had started learning English at the age of 12 in Korea. Their ages at the time of the experiment and their majors varied.

Procedure

The experimental procedure consisted of a pre-test for all groups, computer delivered instruction for the CALL group, writing instruction without attention to grammar or correction for the NO CALL group, followed by two post-tests. The NS group took the editing test twice within a three-week interval. On their first visit, participants from all three groups were given a brief background survey questionnaire, examples of erroneous sentences, and examples of correction procedures prior to the pre-test. After they had completed the questionnaire and the correction procedure section, they were given the pre-test essay which had the following instructions:

A friend of yours has written this passage and asked you to look it over and see that it doesn't

contain any grammatical errors. Help your friend out. Correct any grammatical errors you find. You can insert words, cross them out and change them to make grammatical sentences. Please read the passage very carefully.

Below these instructions were two example sentences corrected with cross-outs and insertions of the type that the students had just practiced. Once all of the participants had completed the pre-test, they were randomly assigned to two groups, the CALL group and the NO CALL group. Only the participants of the CALL group were given brief instructions about the computer administered instruction and correction. They were randomly assigned to orders for taking the four CALL lessons and came to a computer language lab in small groups on different days.

In addition to their regular instruction, the CALL group received the computer delivered instruction once a week for four weeks in the language laboratory. The duration of the sessions averaged about one hour. During the first week, half of the CALL group completed Sections A, B and C of the PASSIVE lesson and the same sections in the QUANTIFIERS lesson. The other half began with the same sections in the QUANTIFIERS lesson and then did the PASSIVE lesson. During the second week, the two groups worked through Sections C, D, E, F and G of these lessons. The same procedure was followed for weeks three and four with the remaining two ICALL lessons. The CALL group also took a unit test that included a large paragraph in which all four grammar rules were embedded. The NO CALL group received no explicit instruction on any of these grammar topics; rather they continued to receive the same instruction that the CALL group was getting in their regular classes. This emphasized rhetorical organization, paragraph coherence, transitions and so forth. Corrections throughout their compositions were by and large related to these rhetorical issues and took the form of comments like: "Nice intro; more detail needed here; what's your thesis?" The participants in the CALL and NO CALL groups returned five weeks after the pre-test to take the first post-test. Five months later they returned and took the test again.

RESULTS

Thirty-two items (eight items in each of the four syntactic error categories) were scored by two graders, who reached 97 percent agreement. All cases of disagreement were then submitted to the judgment of a third rater for final decision. The Cronbach alpha value was .949 with all participants ($n = 59$) and .893 with the L2 learners.

In order to examine the immediate effect of the CALL instruction, the scores of the pre and post-tests were initially submitted to a three-way ($3 \times 2 \times 4$) mixed analysis of variance with test time (pre-test and post-test) and syntactic category (articles, passives, quantifiers, and demonstrative determiners) as within-subject factors, and with three groups (CALL group, NO CALL group, and NS group) as a between-subject factor. However, the assumption of homogeneity of variance was violated with all three groups (Box's $M = 174.37$, $F(72, 8275) = 1.91$, $p = .001$) while the assumption was met without the NS group (Box's $M = 150.85$, $F(78, 4179) = 1.25$, $p = .05$). This means that the variance of the NS group was significantly different from that of the two non-native speaker groups, so ANOVA with three groups as a between groups factor cannot be performed. We therefore undertook a series of two-way analyses between test time and syntactic category per group.

The NO CALL group showed only a significant main effect for syntactic category, $F(3, 168) = 20.91$, $p < .01$.³ Post-hoc comparisons were performed using Fisher Hayter multiple comparison tests. There were significant mean differences between demonstrative determiners and the other categories: articles, $qFH(3, 168) = 2.22$, $p < .01$, passives, $qFH(3, 168) = 2.22$, $p < .01$, and quantifiers, $qFH(3, 168) = 2.36$, $p < .01$. Participants in this group displayed a much better knowledge of the rules for English demonstrative determiners than for the other three categories.

The CALL group showed dramatic improvement between the pre-test and the post-test in all four syntactic categories. Significant main effects for test time, $F(1, 56) = 198.99, p < .01$ and syntactic category, $F(3, 168) = 12.85, p < .01$, were found as well as a significant two-way interaction between test time and syntactic category, $F(3, 168) = 13.39, p < .01$. A series of repeated measures one-way ANOVAs on the CALL group data revealed that all of the categories in the CALL group have significant mean differences between the pre-test and post-test: articles, $F(1, 168) = 23.48, p = .01$, passives, $F(1, 168) = 146.78, p < .01$, quantifiers, $F(1, 168) = 169.62, p < .01$, and demonstrative determiners, $F(1, 168) = 98.90, p < .01$. However, improvement in the articles between the pre-test and the post-test ($M = 1.96, SD = .28, p < .01$) is smaller than that of the other three categories.

The analysis for the NS comparison group showed a significant main effect for syntactic category, $F(3, 168) = 2.89, p = .04$, but there were no significant differences in means between the pre-test and post-test for all the syntactic categories. There were significant differences between articles and passives, $qFH(3, 168) = 1.03, p < .01$, and between passives and quantifiers, $qFH(3, 168) = 1.13, p < .01$. Due to the violation of the homogeneity of variance assumption, a non-parametric ANOVA was performed by transforming the score data to global ranks in order to compare scores of the NS group and the CALL group. Significant differences in means on all syntactic categories were found between NS and the CALL group on the pre- and post-test administrations, $F(1, 39) = 329.49, p < .001$.

To measure long-term retention of the instruction from the administration of post-test 1 to post-test 2, five months later, the mean scores from only the CALL and NO CALL groups were analyzed in a three-way ($2 \times 3 \times 4$) mixed analysis of variance with test time (pre-test, post-test 1 and post-test 2) and syntactic category (articles, passives, quantifiers, and demonstrative determiners) as within-subject factors and with two groups (NO CALL and CALL) as a between-subject factor. The overall means by group are displayed in Figure 4.

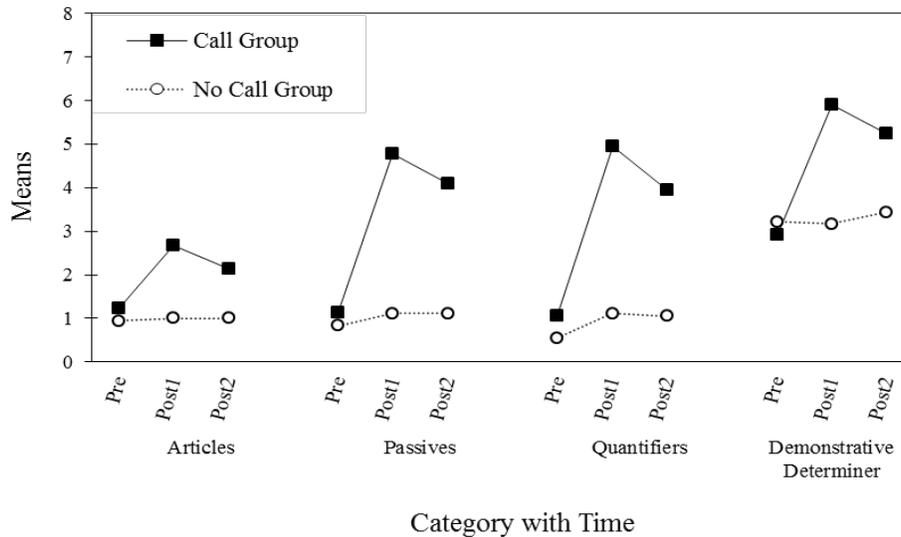


Figure 4. Means per Test Time and Syntactic Category for CALL and NO CALL groups.

Three significant two-way interactions were found: test time x syntactic category, $F(6, 228) = 3.20, p = .005$, test time x group, $F(2,76) = 30.97, p < .001$, and syntactic category x group, $F(3, 114) = 2.97, p = .04$. Significant main effects for test time, $F(2,76) = 42.76, p < .001$, syntactic category, $F(3, 114) = 29.57, p < .001$, and group $F(1, 38) = 50.28, p < .001$ were also found. The significant interaction between test time and syntactic category for the treatment group, $F(6, 228) = 5.11, p < .001$ implies that there were significant means score differences of syntactic categories across all of the rest of the

administrations for the CALL group. In contrast, the NO CALL group showed no improvement over all syntactic categories between test administrations.

There were significant differences between test times for all syntactic categories in the CALL group: articles, $F(2, 228) = 7.69, p < .01$, passives, $F(2, 228) = 53.20, p < .001$, quantifiers, $F(2, 228) = 58.72, p < .001$, and demonstrative determiners, $F(2, 228) = 47.42, p < .001$. Post-hoc comparisons revealed that the CALL group made significant improvement between the pre-test and post-test 1 in all the categories: articles, $qFH(2, 228) = .83, p < .01$, passives, $qFH(2, 228) = 2.13, p < .01$, quantifiers, $qFH(2, 228) = 2.4, p < .01$, and demonstrative determiners, $qFH(2, 228) = 1.63, p < .01$. Moreover, there were still significant differences between pre-test and post-test 2 for three categories: passives $qFH(2, 228) = 1.75, p < .01$, quantifiers, $qFH(2, 228) = 1.82, p < .01$, and demonstrative determiners, $qFH(2, 228) = 1.4, p < .01$. No significant differences between post-test 1 and post-test 2 were found for articles, passives and demonstrative determiners. On the other hand, for quantifiers, there was significant difference between the post-test 1 and the post-test 2, $qFH(2, 228) = .58, p < .05$. This indicates that significant immediate impact of the instruction lasted up to five months after the initial instruction. In addition, there was no significant long-term loss of knowledge gained from the instruction for three of the four categories.

Figure 5 displays the time x group interaction for the two groups over three test administrations at different times. This simply reconfirms the previously reported results that the CALL group retains the effect of the instruction up to the second post-test without significant drop for a semester. In contrast the NO CALL group displays a very minimal increase over time.

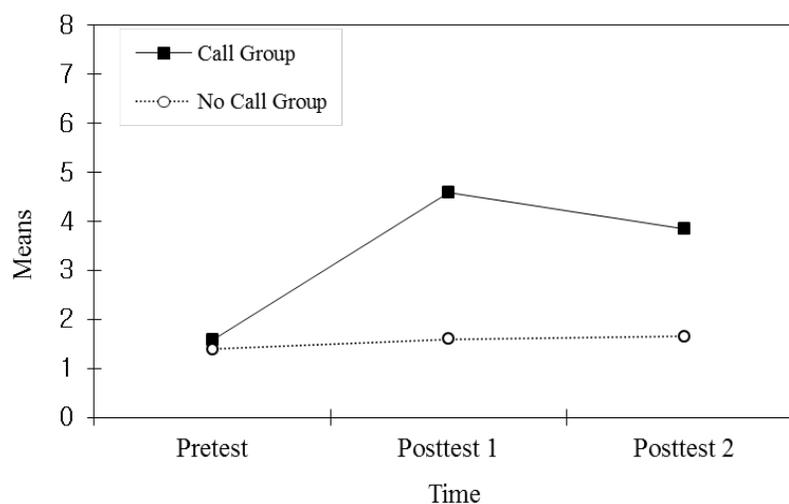


Figure 5. Means per group over test times.

DISCUSSION

The native speakers did not achieve a perfect score on all four rules, but this was due largely to the substitution of a different grammatical alternative that was not allowed under the strict scoring criteria. For example, on demonstrative determiner items a few native speakers substituted *such statements* for *those statements*, and this was counted as an error under the scoring criteria. Similarly, some vacillation was found for definite, indefinite, and zero article before nouns used in the generic sense, as noted by Kachru (2003), for example: *the industrialist/an industrialist is primarily interested in profits* vs. *industrialists are primarily interested in profits*. Overall, however, variation in the performance of the NS group was minor, and performance on all four rules is nearly identical on the pre-test and post-test 1, an indication that the evaluation measure has good internal reliability.

The fact that the CALL group significantly improved their scores on the first post test, while the NO CALL group did not, supports the conclusion that the additional CALL instruction was effective in getting this group of advanced proficiency L2 learners of English to recognize and correct persistent errors in written drafts. Furthermore, this improvement did not diminish substantially over five months. In contrast, as shown in Figure 5, the NO CALL group made, at best, only miniscule progress in recognizing and correcting any of the selected errors over time. These results suggest that an effective pedagogical approach for addressing the problem of persistent grammatical errors found in the writing of advanced proficiency L2 learners may be through supplementing in-class writing instruction with ICALL courseware like *The ESL Writing Tutor*. Although there was still a significant difference between the performance of the CALL group and the NS group after only four hours of the computer delivered instruction, the overall improvement of the former was very encouraging, raising the possibility that a more extended version of the courseware, or repeating portions of it that target particularly recalcitrant errors, could produce gains that approach native speaker performance. The article units would clearly require more refinement and expansion before similar improvement could be expected. The improvement evidenced on overpassivization of ergative verbs was very encouraging, given the overall frequency of these and Han's (2000) conclusions about their prevalence in the writing of L2 academics.

Our results also suggest that the generally assumed superiority of explicit feedback incorporated in ICALL programs, which is supported by the studies cited earlier, should be modified to take into consideration the proficiency level of students and instructional goals. Whereas the conventional approach of identifying grammatical errors and then linking students to information that can be used to correct these errors has been shown to be effective for beginning and intermediate students, ICALL programs that target advanced level students, in particular remedial programs that attempt to eradicate persistent grammatical errors, will probably benefit more from the metalinguistic prompting strategies used in this study. When coupled with instruction that draws advanced level students' attention to the differences between their systematic ungrammatical renditions of certain L2 structures that they are familiar with, metalinguistic prompting appears to be quite effective.

As we have shown in this study, a prerequisite for discovering persistent errors is a very large corpus of written samples produced by L2 students with advanced level proficiency. In addition to providing a basis for ICALL instruction that raises students' consciousness of their persistent grammatical errors, corpus analysis tools provide data that can refine our picture of what kinds of structures are most prone to becoming fossilizations in the interlanguage of L2 learners with different L1s. This issue, which was broached in DeKeyser (2005), presents an interesting new area of exploration for L2 research. Currently interlanguage development is thought to proceed through different stages, but much remains to be learned about how factors like language typology affect the acquisition of different structures. For example, Cowan (2008) catalogues and discusses documented errors made by L2 learners with different L1s attempting to acquire the major syntactic structures of English. The degree of difficulty that some of these errors present to learners with different L1s has been examined in a few SLA studies, but the potential problems presented in mastering others have not been explored. Still, the widespread prevalence of such problems suggests that it might be worthwhile for developers of ICALL courseware to consider addressing them in special units that utilize the design described in this study. The best source of data on errors that might be worth targeting is, in our opinion, not the production tasks currently used in ICALL programs, since these are somewhat artificial and restrict output. Instead, free writing tasks used in conjunction with error identification tasks like the one shown in this study are preferable. Compiling corpora containing written errors made by L2 learners with high intermediate to advanced proficiency can provide ICALL developers with a rich source of L2 learner language that can be used in the courseware, and it might also lead to the discovery of lesser known and/or yet undiscovered persistent errors made by learners with different L1s. Thus the important synergy between corpus linguistics and CALL, first broached by Granger (2003) as a means of improving L2 instruction, also has the possibility of expanding

our understanding of developmental stages and potential limits in second language acquisition.

CONCLUSION

This study has presented evidence that ICALL courseware can raise Korean graduate students' awareness of specific recurrent grammatical errors in their writing, and this heightened awareness resulted in the ability of these L2 learners to detect and correct these errors in unfamiliar English passages of equivalent difficulty. These results suggest that this type of ICALL courseware would be a valuable asset for advanced level L2 writing instruction in learning contexts where the goal is to prepare learners whose professions entail regular writing in the L2. One such context that is widespread in the United States is ESL writing instruction for foreign students admitted to undergraduate and graduate programs. These students have a command of English that enables them to read textbooks and participate in academic discussions with native speaker counterparts, but their written English betrays consistent grammatical errors that appear to be near-fossilizations.

The courseware described in this study differs from that found in other ICALL tutors in that it is not linked to specific language courses, rather it targets specific errors made by a particular L2 learner population. A basic prerequisite for the development of this type of ICALL tutor is therefore a large database that has been produced by L2 learners who fall within the targeted population. In this study we described how corpus analysis tools were applied to a large written corpus produced by L2 learners with the same L1 and a specific, independently determined proficiency level to yield a representative picture of which grammatical errors should be targeted in the ICALL tutor. This picture is greatly enhanced if the learner corpus is as large as the one described in this study and contains samples from a variety of writing assignments. The results of this study therefore represent yet another example of the promising synergy between CALL instruction and corpus linguistic analysis for improving L2 instruction as discussed in Granger (2002).

NOTES

1. All international students admitted to university whose native language is not English are required to submit their Test of English as a Foreign Language (TOEFL) or International English Language Testing System (IELTS) scores. If their TOEFL scores indicate that further English is needed (scores of 550-610 Paper, 213-253 CBT, or 79-102 iBT, inclusive), they are required to take an English as a Second Language Placement Test (EPT) upon arrival at the university. Then, based on the EPT results, they are assigned to one of several remedial ESL writing courses, which are intended to teach them how to write academic English.
2. An example of an error of this type would be: **From last year until now, politics have been changed rapidly.*
3. All p-values were adjusted using the Greenhouse-Geisser (G-G) epsilon correction factor for the violation of the sphericity assumption for the repeated measures ANOVA. The corrected p-values together with uncorrected degrees of freedom are reported.

ACKNOWLEDGEMENTS

The authors would like to thank Laura Collins and Pavel Trofimovich for their comments on an earlier version of this article and two anonymous readers, whose comments led to several improvements in the version that appears here.

ABOUT THE AUTHORS

Ron Cowan is Associate Professor Emeritus of Linguistics at the University of Illinois at Urbana-Champaign. His research interests include Second Language Acquisition, language processing, L2 grammar instruction, and applications of CALL to the teaching L2 writing.

E-mail: jcowan@illinois.edu

Jinhee Choo is a Visiting Lecturer teaching ESL writing at the University of Illinois at Urbana-Champaign. She has a PhD in Educational Psychology specializing in SLA and masters degrees in TESOL and Applied Statistics. Her research interests include ESL writing, CALL, and teacher education for online education and technology integration.

E-mail: jchoo@illinois.edu

Gabseon Sunny Lee has BAs in East Asian Languages and Cultures and English from Indiana University at Bloomington, an MA in TESOL and a Certificate of Advanced Study in Educational Psychology from the University of Illinois at Urbana-Champaign. She has taught English rhetoric and ESL courses and managed interactive web courses.

E-mail: sunnyglee2@gmail.com

REFERENCES

- Adams, R., Nuevo, A. M. & Egi, T. (2011). Explicit and implicit feedback, modified output, and SLA: Does explicit and implicit feedback promote learning and learner-learner interactions? *Modern Language Journal*, 95(1), 42–63.
- Balcom, P. (1997). Why is this happened? Passive morphology and unaccusativity. *Second Language Research*, 13(1), 1–9.
- Bowles, M. (2005). *Effects of verbalization condition and type of feedback on L2 development in a CALL task*. (Unpublished doctoral dissertation). Georgetown University, Washington, D.C.
- Cowan, R. (2008). *The teacher's grammar of English*. Cambridge, UK: Cambridge University Press.
- DeKeyser, R.M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Oxford, UK: Blackwell.
- DeKeyser, R. M. (2005). What makes learning L2 grammar difficult? *Language Learning*, 55, 1–25.
- French Interlanguage Database. (FRIDA). Retrieved from <http://www.uclouvain.be/en-cecl-frida.html>
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20, 465–480.
- Han, Z.-H. (2000). Persistence of the implicit influence of NL: The case of the pseudo-passive. *Applied Linguistics*, 21(1), 78–105.
- Heift, T. (2004). Corrective feedback and learner uptake in CALL. *ReCALL*, 16, 416–431.
- Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. *System*, 36, 196–213.
- Heift, T. (2010). Developing an intelligent language tutor. *CALICO Journal*, 27(3), 443–459.
- Hirakawa, M. (1995). L2 acquisition of English unaccusative constructions. In D. MacLaughlin & S. McEwen (Eds.), *Proceedings of the 19th Boston University conference on language development* (pp.

291–302). Summerville, MA: Cascadilla.

Hwang, J.-B. (1997). *Implicit and explicit instruction for the L2 acquisition of English unaccusative verbs*. (Unpublished manuscript). University of Oregon at Eugene, Eugene, Oregon.

International Corpus of Learner English (ICLE) Retrieved from <http://www.uclouvain.be/en-cecl.html>

Ju, M.-K. (1999). Overpassivization errors by second language learners: The effect of conceptualizable agents in discourse. *Studies in Second Language Acquisition*, 22, 85–111.

Kachru, Y. (2003). On definite reference in world Englishes. *World Englishes*, 22, 497–510.

Kim, D.-H. (2005). *Finding feedback: An effective computer assisted language learning model for second language acquisition*. (Unpublished masters thesis). University of Illinois at Urbana-Champaign. Urbana, Illinois.

Long, M. H. (2003). Stabilization and fossilization in interlanguage development. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 487–536). Oxford, UK: Blackwell.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, 19, 37–66.

Meurers, D. (2012). Natural language processing and language learning. In C. A. Chapelle, (Ed.), *Encyclopedia of applied linguistics*. Oxford, UK: Wiley-Blackwell.

Nagata, N. (1996). Computer vs. workbook instruction in second language acquisition. *CALICO Journal*, 14, 53–75.

Nagata N., & Swisher, M. V. (1995). A study of consciousness-raising by computer: The effect of metalinguistic feedback on second language learning. *Foreign Language Annals*, 28, 337–347.

Oshita, H. (2001). The unaccusative trap in second language acquisition. *Studies in Second Language Acquisition*, 23(2), 279–304.

Poesio, M. & Vieira, R. (1998). A corpus based investigation of definite description use. *Computational Linguistics*, 24(2), 183–216.

Pujola, J.-T. (2001). Did CALL feedback feed back? Researching learners' use of feedback. *ReCALL*, 13, 79–98.

Rosa, H. M., & Lowe, R. P. (2004). Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *Modern Language Journal*, 88, 192–216.

Scott, M. (2008). WordSmith Tools (Version 5) [computer software]. Liverpool: Lexical Analysis Software. Available from <http://www.lexically.net/wordsmith/>

Sheen, Y. (2004). Corrective feedback and learner uptake in communicative classrooms across instructional settings. *Language Teaching Research*, 8, 263–300.

Varieties of English for Specific Purposes dAtabase (VESPA). Retrieved from <http://www.uclouvain.be/en-cecl-vespa.html>

Yip, V. (1995). *Interlanguage and learnability: From Chinese to English*. Amsterdam/Philadelphia: John Benjamins.