Developing a corpus-based online grammar tutorial prototype

Second language (L2) learners are likely to produce the same kinds of grammatical errors over and over again in their L2 written products despite explicit feedback and review in the classroom by their writing teachers. In an attempt to address this persistent problem, the present study collected and analyzed 61 learner essays to develop a prototype of a corpus-based online grammar tutorial for use outside the classroom. Quantitative analyses of the learner data revealed problematic features common to the learners: the most frequent grammatical errors and poor management of topical progression. These especially prevalent problems with L2 grammatical knowledge were incorporated into the prototype system for special treatment. The usability of the prototype was then evaluated by 26 undergraduate university students. Their evaluation ratings were overall positive. Constructive feedback obtained from the students suggests that this corpus-based online grammar tutorial can be a vital asset for improving the L2 grammatical knowledge of learners.

Masumi Narita
Tokyo International University

The Language Teacher • FEATURE ARTICLE | 23
commercially available, are beneficial in that they can automatically recognize and clean up grammatical errors in writing. Although this corrective feedback is much favored by L2 writers, its pedagogical advantages are seemingly temporary, because the writers are satisfied with their cleaned-up writing and can hardly be empowered to find out relevant things to develop their grammar skills for themselves. Language processing limitations inherent to grammar checkers may also restrain the learner’s learning uptake.

An online writing lab (OWL) is well recognized in western countries to play an essential role in reinforcing tutoring sessions at a writing center. In Japan, however, it is only recently that universities have begun to establish a writing center and/or an OWL. OWLs offer an extensive and useful guide to grammar resources; nevertheless, it is undeniable that they are short, digitized versions of paper-based grammar books and exercises. They tend to be general-purpose learning materials and thus sometimes fail to meet specific EFL learners’ needs. Direct interactivity with users is supported, but not promoted as fully as recent computer-assisted language learning (CALL) programs.

Detailed assessment studies of CALL programs are still sparse, but their pedagogical potential has been positively reported by several researchers (Blake, 2011; Boulton, 2009; Johns et al., 2008; U.S. Department of Education, 2010). Most significantly, Blake (2011) stresses that online learning “stimulates students to spend more time engaged with the second language (L2) materials, which ultimately promotes greater learning” and that this expanded learning time “constitutes the most important value added with respect to online learning” (p. 21). Chapelle (2007) also suggests that CALL tasks could provide learners with informative feedback via beneficial interactions, thereby raising their L2 grammatical knowledge as well as promoting their error awareness.

Rapid advances in computer technology have also enabled a combination of a computer-assisted language learning mechanism and an electronic learner corpus. As pointed out by Römer (2011), pedagogical corpus applications have not been widely implemented, yet the exploratory aspect of corpus-based activities seems to have great potential to improve L2 teaching and learning practices. In addition, more focus may well be placed on “language items that are of central importance and/or troublesome for learners” (Römer, 2011, p. 216). Thus, Römer’s overview of corpus technology and applications signals further possibilities for developing corpus-based online language tutorials.

The present study describes a newly developed prototype of an online grammar tutorial based on frequent grammatical errors that were found in a corpus of student essays in English. The usability of this prototype is also reported through a questionnaire survey of 26 Japanese college students studying English as a foreign language (EFL). This preliminary study aims to explore how learners’ written texts can be effectively used to identify recurrent grammatical errors and how useful web-based interactivity tends to be for promoting L2 learners’ awareness of erroneous forms. In what follows, two recent studies relevant to learner corpus-based CALL programs are reviewed, and then the present research framework is explained in detail.

Recent studies on learner corpus-based CALL programs

Learner corpora become more powerful linguistic resources when tagged with grammatical errors or used for error analysis. In error-tagged corpora, errors are marked using a specific coding scheme and are thus easy to search for on a computer. As mentioned in the previous section, students’ grammatical errors are usually treated in the L2 classroom, but eradicating them remains difficult. Indeed, these errors are repeatedly produced regardless of the degree of grammatical difficulty. For instance, subject-verb disagreement is one of the students’ recurrent errors, even though this type of error seems to be rather easy to avoid. Thus, it could be beneficial to L2 learners if an online grammar tutorial for improving error awareness is developed and made accessible at any time outside the classroom.

Cowan, Choi, and Kim (2003) developed a computer-assisted language learning (CALL) program, the ESL Tutor, to help Korean ESL students. They conducted a study that involved 58 Korean ESL students who were divided into control and experimental groups. The control group was taught using traditional classroom methods, while the experimental group was taught using the ESL Tutor.

The results of the study showed that the experimental group made significantly fewer errors in their writing compared to the control group. The ESL Tutor was found to be effective in improving students’ writing skills and error awareness. The study also highlighted the potential of computer-assisted language learning programs in promoting language learning and enhancing students’ grammatical accuracy.

1 Grammar checking functions have been built into Microsoft Word and are widely used by L1 and L2 writers. Other automatic grammar checkers include the Criterion® Online Writing Evaluation Service and the Grammarly®, both of which are designed to detect typical errors made by native speakers of English. Criterion® Online Writing Evaluation Service: <criterion.ets.org> Grammarly®: <www.grammarly.com>

2 Erroneous sentences produced by language learners are quite hard to analyze structurally even with the current state-of-the-art of parsing technology.
learners recognize and correct their repeated errors. They collected 359 essays from Korean ESL learners enrolled in 4 proficiency-based writing courses at the University of Illinois and identified grammatical errors that were still recurrent among the higher-level learners. Target errors in the CALL program were chosen based on two sources: (1) second language acquisition (SLA) research that reports on frequent errors made by ESL/EFL learners who speak languages typologically similar to Korean and (2) frequent errors that were found in a sampling of essays produced in the four writing courses.

The learner corpus in the ESL Tutor was not error-tagged, so a language processing tool called “concordancer” was used to assist in identifying persistent errors. In contrast, a fully error-tagged learner corpus was used in the online grammar resource Internet Writing Resource for the Innovative Teaching of English (iWRITE) developed by Hegelheimer and Fisher (2006). The learner corpus in iWRITE was a collection of 45 ESL placement essays for Iowa State University. Learner errors were numerically coded. Moreover, error descriptions and corrections were separately prepared and retrieved so that learners could raise their awareness of major grammatical errors and also find possible solutions to these errors.

ESL Tutor and iWRITE differ in the way a learner corpus was implemented into their respective systems. However, both systems were designed to help ESL students recognize their major grammatical errors and learn how to avoid the errors through mini-lessons and practice exercises supported by advanced web interactivity. With the guidance of an L2 teacher, learners are encouraged to use the system, mostly inside the classroom, to improve their grammatical knowledge. The role of the language teacher as a facilitator is significant, and seemingly elaborate data-driven learning procedures may often put high demands on the learners.

The pedagogical effectiveness of both systems is still under review, but their learner corpus-based approach to grammar improvement motivated the present study. Drawing on the advantages of the two systems, a prototype of corpus-based online grammar tutorial was developed as a self-learning tool for use outside the classroom as well as at the writing center of Tokyo International University. In this educational context, students are given explicit feedback on their significant grammatical errors by their writing teachers in the classroom or by tutors at the TIU writing center, and then they are strongly encouraged to use the online self-access grammar tutorial.

Prototype development

Learner data collection and analysis

To identify most frequent problems with sentence grammar and discourse, 61 essays were collected in an experimental session at Tokyo International University. All the participants were Japanese first-year EFL learners. They were given 40 minutes to produce an essay on the topic “Why are cell phones so popular with young teenagers today?” without using any reference tools.

All of the essay data were manually annotated by the present researcher with two kinds of linguistic information: (1) error description and (2) inter-sentential topical progression (for details, see Narita, 2010). Errors were examined by comparing a learner essay with a version of it reconstructed by a native English proofreader on a word-by-word basis and described according to a set of descriptive categories developed by Koizumi (2008). Topical progression was annotated based on the procedures described in previous studies (Knoch, 2007; Lautamatti, 1978; Schneider and Connor, 1990; Witte, 1983). Error descriptions of a sample learner essay and the same sample essay annotated with topical progression are shown in Appendices 1 and 2, respectively.

The annotated information was then quantitatively analyzed to identify major problems with grammar and topical progression. Among others, noun countability caused the most frequent errors, which was possibly interrelated with frequent errors in article use and also in subject-verb agreement. Overuse of conjunctions (such as “and,” “but,” “so”) in the sentence-initial position was also found in the present learner data, and these instances were mostly replaced with semantically similar linking adverbials or logical connectors (such as “moreover,” “however,” and “therefore”) by the native English proofreader.

The analysis of topical progression revealed that the topical subject was often the same in successive sentences in the learner data. This pattern deviates from the regular pattern of information structure in English in which “given information” is placed before “new information” in one sentence and the new information becomes given information in the next sentence. Information structure may vary as L2 learners...
become more skillful writers, but trying to follow the basic sequence of information will most likely help them achieve coherence in English-language writing.

Prototype overview
The prototype system in the present study was designed to treat the most frequent grammatical errors and poor management of topical progression described in Section 3.1. The target system is aimed at offering a web-based online grammar tutorial with a wide coverage of English grammar, but this initial stage of system development focused on grammatical and discoursal needs that were readily identified through examining the present learner data. Thus, five kinds of learning materials were prepared under two categories as shown below.

A. Grammatical Errors Common to Japanese EFL Learners
   (1) Countable versus Uncountable Nouns
   (2) Subject-Verb Disagreement
   (3) Missing Articles

B. Linguistic Devices for Textual Coherence / Cohesion
   (1) Linking Adverbials as Logical Connectors
   (2) Given-New Information Structure

Compared with conventional language learning systems, the present prototype is unique in two ways. First, learners are given grammatical explanations along with error examples found in the present learner corpus. In other words, authentic language errors, not those invented by grammarians, are provided for learners. Second, learners are required to read a given paragraph in order to find an answer to each practice exercise; that is, learners are required to gather and analyze contextual information relevant to their grammatical or textual solution. This scheme engages the learners with a discourse beyond isolated sentences, thus expanding their scope of attention in L2.

When the “Missing Articles” menu is chosen on the top page of the system, for instance, a new window opens as shown in Figure 1. On this new page, learners are provided with grammatical explanations of how to avoid missing articles with reference to authentic learner errors. They are also encouraged to choose one of the practice exercises prepared at the bottom of the screen. Then, a new window opens to offer one of the exercises as shown in Figure 2.
When a wrong answer is given by the student, the warning message “Let’s just try that again!” appears on the screen so that he or she can have another try. With the correct answer, a concise discussion of the exercise is presented on the screen and then followed by another exercise. This online tutorial makes it seem like a game while students tackle their own goal of error treatment. Overseas OWLs are also accessible to the students on the screen.

Usability evaluation

The prototype was experimentally evaluated in terms of usability by 26 undergraduate students at the School of English Communication of Tokyo International University. These trial users accessed all of the grammatical explanations prepared on the web and worked on all of the grammar exercises. A questionnaire was then given to the students to examine whether the grammar tutorial under construction could meet their needs, that is, enhance their grammatical knowledge via self-access learning. This experimental session took about an hour and a half.

The usability questionnaire was designed to include several factors that might affect user satisfaction and operability. The present trial users were requested to rate their satisfaction on a five-point scale (5, 4, 3, 2, 1) ranging from “most positive” to “most negative” and finish the survey with their free descriptions on the utility of the prototype. Figures 3 and 4 illustrate the overall evaluation and the perceived effectiveness of individual resource components, respectively.

As is evident from Figure 3, the factors “Overall Effectiveness,” “Response Time,” “Understandability of Grammatical Explanations,” and “Ease of Use” were evaluated as positive by more than 80% of the students. The “Learnability of Grammar Exercises” was evaluated as positive by about 70% of the students. In contrast, the “Volume of Grammar Exercises” was evaluated as positive by less than 50% of the students. This means that the students in this study were inclined to do more grammar practice activities. This finding is interesting considering that they are not so willing to work on multiple exercises in the classroom.

Of the present resource components, linguistic devices operating on a textual level had a higher percentage of positive responses than those operating on a lexical or syntactic level, which can be interpreted to be that these textual features are of more instructional value. The two components “Subject-Verb Agreement (in number)” and “Article Use” were positively evaluated by half of the students, whereas about 40% of the students evaluated them as neutral (= “3”). This may be because the students are often encouraged by their writing teachers to review these grammatical items. It is also important to note that negative feedback was given by about 20% of the students in their response to “Noun Countability” and “Information Structure.” Possible reasons for this can be found in the following comments and requests that the students freely described in the questionnaire.

- The component “Article Use” is the most helpful because our knowledge of article use is unstable. English articles are the most difficult to use, but for this very reason, we need to learn their usage over time.
• The component “Information Structure” is quite new and takes time to become familiar with; nevertheless, this component is beneficial because writing textbooks do not always include textual features relevant to creating a coherent flow of information.

• “Noun Countability” and “Subject-Verb Agreement” are sometimes difficult to handle because both lexical and syntactic features need to be considered in parallel. We need a direct link to a lexical look-up service on the web.

• It is advantageous to read a given paragraph to find the answer in each grammar exercise. This paragraph can be a writing model for us.

• It could be of great benefit if many more grammar exercises were available to us, preferably according to the level of difficulty. This enables us to find our own challenge(s) to address.

• Grammatical explanations given to learner errors that are familiar to us are fairly understandable. The shorter the better, however.

• The online tutorial is helpful and convenient because we can use it whenever we want. Moreover, we can repeatedly work on the same exercises after a certain period of time to reconfirm our L2 grammatical knowledge.

These statements suggest that the present prototype is very helpful for Japanese EFL learners. To make it more effective, however, more extensive grammar items and practice exercises need to be prepared along with more concise grammatical explanations. Assessment studies are also necessary to examine longitudinal effects of the present online grammar tutorial on the students’ writing performance. This web-based grammar tutorial will be open to the public in the future.

Conclusion
Learner essays were collected and annotated with information on errors and inter-sentential topical progression to identify frequent grammatical errors among university students and subsequently to develop a prototype of a self-access online grammar tutorial. The prototype system was experimentally tested by 26 Japanese undergraduate EFL students. The students evaluated the prototype in terms of usability and provided immediate written feedback.

The students’ positive and constructive feedback seems to acknowledge the potential of the prototype system, although further improvements are necessary. Survey results from the questionnaire also imply that learner errors can be used as a positive resource as long as they are clearly sorted out and incorporated into a language learning system in an effective and efficient way. Yet another important implication is that L2 grammatical knowledge can be reinforced appropriately if learners are naturally led to consider inter-sentential relations to find any grammatical solution, in such a way that paragraph-based but not isolated sentence-based grammar exercises are accessible to L2 learners.

Acknowledgments
I am grateful for the insightful comments of the manuscript reviewers on an earlier version of this manuscript. I am indebted to the Japan Society for the Promotion of Science and the Ministry of Education, Culture, Sports, Science, and Technology for funding this research over the past three years (Grant-in-Aid for Scientific Research, Project No. 20520514, 2008–2010). I also thank Mr. Fumiaki Nishihara, a graduate student at the University of Tokyo, for his continuous practical support.

Masumi Narita is a professor at the School of Language Communication of Tokyo International University, teaching academic reading and writing courses. She received a Ph.D. from the Graduate School of International Development at Nagoya University. Her major research interest is in second language writing. She has been conducting learner corpus-based research in order to explore how Japanese EFL writers can improve their English writing skills not only through face-to-face interactions with their writing teacher in the classroom but also through self-access computer-mediated environments outside the classroom. She can be contacted at <mnarita@tiu.ac.jp>.

References


Appendix 1. A sample essay with information on topical progression (with topical subjects underlined)

<text_number>2006-013</text_number>

Why are cell phones so popular with young teenagers today?

(a) I have three reasons. (b) First, cell phones are very useful for us. (c) We can call and send e-mails wherever you are. (d) Second, cell phones have various kinds of systems. (e) For example, we are able to listen the musics, play games, take pictures and so on. (f) The final reason is, today, almost of the teenagers have cell phone, so they tend to think that it is natural to have cell phone. (g) I think, in fact, teenager parents feel relaxed to give cell phone their children. (h) Because, it is dangerous for children to go out. (i) If children met a dangerous situation, they can contact with someone.

(j) The cell phones are very useful and convenient, so it is popular with young teenagers today. (10 sentences; 9 topical progressions)

******************************************************************************

<topical progression>

Direct sequential progression: (b) => (c)

Indirect progression: (e) => (f), (f) => (g)

Superstructure: (a) => (b)

Progression break: (g) => (h)

Extended progression: (c) => (d), (d) => (e), (h) => (i), (i) => (j)

Appendix 2 is available from the online version of this article at <jalt-publications.org/tlt/archive>.
Appendix 1: A sample essay with information on topical progression (with topical subjects underlined)

Why are cell phones so popular with young teenagers today?

(a) I have three reasons. (b) First, cell phones are very useful for us. (c) We can call and send e-mails wherever you are. (d) Second, cell phones have various kinds of systems. (e) For example, we are able to listen the musics, play games, take pictures and so on. (f) The final reason is, today, almost of the teenagers have cell phone, so they tend to think that it is natural to have cell phone. (g) I think, in fact, teenager parents feel relaxed to give cell phone their children. (h) Because, it is dangerous for children to go out. (i) If children met a dangerous situation, they can contact with someone.

(j) The cell phones are very useful and convenient, so it is popular with young teenagers today. (10 sentences; 9 topical progressions)

*****************************************************************************

<topical progression>

Direct sequential progression: (b) => (c)
Indirect progression: (e) => (f), (f) => (g)
Superstructure: (a) => (b)
Progression break: (g) => (h)
Extended progression: (c) => (d), (d) => (e), (h) => (i), (i) => (j)
Appendix 2: Error descriptions of a sample learner essay

<table>
<thead>
<tr>
<th>&lt;text_number&gt;2006-013&lt;/text_number&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why are cell phones so popular with young teenagers today?</td>
</tr>
<tr>
<td>&lt; an original learner essay &gt;</td>
</tr>
<tr>
<td>(a) <em>I</em> have three reasons. (b) <em>First, cell phones are very useful for us.</em> (c) We can call and send e-mails wherever <em>you</em> are. (d) <em>Second, cell phones have various kinds of systems.</em> (e) For example, we are able to listen the musics, play games, take pictures and so on. (f) The final reason is, today, almost of the teenagers have cell phone, so they tend to think that it is natural to have cell phone. (g) I think, in fact, teenager parents feel relaxed to give cell phone their children. (h) Because, it is dangerous for children to go out. (i) If children met a dangerous situation, they can contact with someone.</td>
</tr>
<tr>
<td>(j) The cell phones are very useful and convenient, so it is popular with young teenagers today.</td>
</tr>
<tr>
<td>&lt; a reconstructed essay by a native English proofreader &gt;</td>
</tr>
<tr>
<td>(a) <em>I</em> have three reasons. (b) <em>First, cell phones are very useful for us.</em> (c) We can call and send e-mails wherever <em>we</em> are. (d) <em>Second, cell phones have various kinds of systems.</em> (e) For example, we are able to listen to music, play games, take pictures and so on. (f) The final reason is because almost all teenagers have a cell phone, the average teenager tends to think that it is natural to have one. (g+h) I think because it is dangerous for young people to go out, parents of teenagers feel more comfortable if their children have a cell phone. (i) If their children face a dangerous situation, they can contact someone.</td>
</tr>
<tr>
<td>(j) Cell phones are very useful and convenient, so they are popular with young teenagers today.</td>
</tr>
<tr>
<td>&lt; error descriptions &gt; <em>COM stands for comments.</em></td>
</tr>
<tr>
<td>COM: 代名詞の語彙選択の問題。代名詞 you を代名詞 we に変更。</td>
</tr>
<tr>
<td>= Lexical choice in pronouns. ‘You’ should be replaced by ‘we.’</td>
</tr>
<tr>
<td>COM: 動詞の自他の問題。自動詞 listen の後に前置詞 to が必要。</td>
</tr>
<tr>
<td>= ‘To’ is necessary because ‘listen’ is an intransitive verb.</td>
</tr>
<tr>
<td>COM: 冠詞の問題。定冠詞 the が不要。</td>
</tr>
<tr>
<td>= The definite article ‘the’ should be omitted.</td>
</tr>
<tr>
<td>COM: 名詞の数の問題。不可算名詞のため musics を単数形 music に変更。</td>
</tr>
<tr>
<td>= Due to its uncountability, ‘musics’ should be replaced by ‘music.’</td>
</tr>
<tr>
<td>(error descriptions continued)</td>
</tr>
</tbody>
</table>