ANALYSIS OF THE ASSOCIATION BETWEEN READING PROFICIENCY AND FREQUENCY OF WRITING ERRORS USING AN ERROR-TAGGED LEARNER CORPUS FOR MULTIPLE LINGUISTIC SKILLS

Katsunori Kotani[1] and Takehiko Yoshimi[2]

[1]College of Foreign Studies
Kansai Gaidai University
16-1 Nakamiyahigashino-cho, Hirakata, Osaka 573-1001, Japan
kkotani@kansaigaidai.ac.jp

[2]Faculty of Science and Technology
Ryukoku University
1-5 Yokotani, Seta, Otsu, Shiga 520-2194, Japan
yoshimi@rins.ryukoku.ac.jp

Abstract. *In order to help teachers of English as a foreign language develop more effective writing exercises, it is first necessary to clarify whether linguistic knowledge regarding how to correct writing errors should be taught not only in writing exercises, but also in other skills such as reading. This study addresses the following research questions: 1) "Is there an association between reading proficiency and frequency of writing errors among learners of English as a foreign language?" and 2) "If such an association exists, does the extent of the association differ depending on the type of error?" To develop a learner corpus capable of answering these questions, we annotated writing error-tags on our learner corpus data. Based on the results of our experiment using this error-tagged learner corpus, an association was found between reading proficiency and frequency of writing errors. Our results also identified typical errors made by learners at each proficiency level.*
Keywords: Error tag, Learner corpus, Multiple linguistic skills, Reading, Writing

1. **Introduction.** This study investigates the effectiveness of using the frequency of writing errors made by learners of English as a foreign language (EFL) – in consideration of their information about the information regarding their proficiency of skills other than writing, such as reading, listening, and speaking – for analysis in an error-tagged learner corpus. This type of analysis should provide insight into the types and frequencies of writing errors EFL learners make at different levels of reading, listening, and speaking proficiency. If an association is found between a writing error and skills other than writing, linguistic knowledge to correct this error should be both learned and taught not only in writing exercises, but also in each relevant skill. The use of similar learning and teaching methods is known as integrative communicative language teaching [1,2]. In this approach, EFL learners practice multiple as opposed to single skills, which strengthens the establishment of their linguistic knowledge. Another advantage of this approach is that it achieves a learning process that follows a conceptually desirable direction, from input (reading and listening) to output (writing and speaking), which enables EFL learners to familiarize themselves with words and grammar through reading and listening [3,4].

In an error-tagged learner corpus for multiple linguistic skills, the association between writing errors and proficiency of reading, listening, and speaking is testable. However, in our previous survey of learner corpora [5], we found that previous learner corpus research

had focused on speaking or writing skills [6-11]. Therefore, we developed a learner corpus that included writing, reading, listening, and speaking (pronunciation) data.

In the context of developing a learner corpus for multiple linguistic skills, this study developed an error-tagged learner corpus in order to investigate the association of writing errors with reading proficiency since writing and reading skills are essential for university EFL learners to achieve academic literacy.

Our research questions were as follows: 1) "Is there an association between frequency of writing errors and reading proficiency?" and 2) "If such an association exists, does it differ depending on the type of error?" Although an association between reading and writing test scores has been reported [12,13], our first research question addresses whether this empirical observation holds true in terms of the frequency of writing errors, while our second question further examines the association between writing and reading based on the type of writing error. This is because although some writing errors may be associated with reading proficiency, others may not.

The organization of the present paper is as follows: Section 2 describes our learner corpus that demonstrated EFL learners' reading and writing capabilities; Section 3 explains the tags for writing errors that were annotated with the learner corpus data; Section 4 reports the experimental results for the above research questions; and Section 5 concludes this paper.

2. **Learner Corpus for Annotation.** Our previously developed learner corpus [5] consisted of data for analyzing EFL learners' reading, writing, listening, and pronunciation skills. This corpus was compiled using data from 90 university EFL learners (mean age, 21.5 years; S.D. 2.6). From among the four types of data in that corpus, the reading and writing data were used in this study.

The reading data were compiled using a reading task in which each EFL learner read four news articles. After reading each article, they answered five comprehension questions. Their comprehension rate was determined by calculating the percent of correct answers to multiple-choice questions (mean rate, 0.6; S.D. 0.1). The reading rate, which refers to the number of words read in one minute (words per minute: WPM) (mean reading rate, 104.8 WPM; S.D. 38.1), and difficulty judgment score, which refers to the comprehension difficulty as judged by an EFL learner on a five-point Likert scale (1: difficult, 2: somewhat difficult, 3: average, 4: somewhat easy, or 5: easy) (mean score, 3.4; S.D. 0.7), were measured for each sentence ($n = 80$).

The writing data were compiled using a picture description task in which each EFL learner described four pictures, and a question answering task in which each EFL learner answered 20 questions about their English learning background and computer literacy. The writing data comprised a total of 33,385 words.

3. **Annotation of Error-Tags.** This study annotated error-tags as previously described [7]. Although their error-tags were developed for annotating speech, as opposed to writing errors, the target EFL learners in that study were similar to those in this study (EFL learners whose first language is Japanese). Therefore, the previously proposed error-tags [7] cover writing errors made by EFL learners whose first language is Japanese. The annotation task was performed by a native English speaker. The error-tags used in this study were classified into 12 types based on part of speech. An error tag contains three pieces of information: the part of speech, a grammatical and lexical rule, and a corrected form.

The annotation results of the writing data had 6,128 errors. Errors irrelevant to specific categories ('Other' in Table 3) were observed most frequently (2218 errors). The second and third most frequently observed error types were verb-type error (1183 errors) and noun-type error (939 errors), respectively.

4. **Association between Writing Errors and Reading Proficiency.** To answer our research questions regarding the association between frequency of writing errors and reading proficiency, we conducted experiments in which an EFL learner's reading proficiency was determined using the effective reading rate (eWPM: the product of comprehension rate and reading rate) [14]. EFL learners with high reading proficiency are supposed to have high effective reading rates, and vice versa. Effective reading rate data were derived from the comprehension and reading rates stored in the reading data of our learner corpus.

EFL learners were then classified into three groups of equal number ($n = 30$) based on reading proficiency (low group: mean, 32.6 eWPM, S.D. 14.3; intermediate group: mean, 53.4 eWPM, S.D. 14.7; high group: mean, 91.7 eWPM, S.D. 33.8).

The frequency of errors in each group is shown in Table 1. The chi-square test of independence was used to examine the association between the frequency of writing errors and each reading proficiency group; a significant association was found ($\chi^2(2) = 460.22$, $p < 0.01$). We then examined this association using a residual test, which statistically examines the difference between expected and observed values (residuals). The results of this test are shown in Table 2.

TABLE 1. Contingency table showing the number of erroneous and correctly used words by each reading proficiency group

	Reading proficiency group		
	Low	Intermediate	High
Number of erroneous words	2319	2265	1544
Number of correctly used words	7806	8667	10784

TABLE 2. Adjusted residual values for the frequencies of writing errors among the reading proficiency groups

	Reading proficiency group		
	Low	Intermediate	High
Residuals of erroneous words	14.16**	7.78**	−21.06**
Residuals of correctly used words	−14.16**	−7.78**	21.06**

$^{**}p < 0.01$

Regarding residuals, a significantly different residual indicates that an error occurred either more or less frequently than expected. A significantly positive difference indicates that a relevant error occurred more frequently than expected, and vice versa. However, both positive and negative differences show EFL learners' unfamiliarity of words and grammar in writing. Regarding familiarity from a reading proficiency viewpoint, a positive difference suggests that EFL learners might also be unfamiliar with words and grammar in reading, while a negative difference suggests that they were already familiar.

A significant difference ($p < 0.01$) was observed in all of the reading proficiency groups (Table 2); therefore, in response to our first research question, a positive association appears to exist between frequency of writing errors and reading proficiency, which suggests that learning and teaching linguistic knowledge relevant to writing errors is effective in both writing and reading exercises.

Regarding familiarity, EFL learners with low reading proficiency need to practice reading exercises in order to familiarize themselves with relevant words and grammar; however, EFL learners with high reading proficiency do not.

Next, to address our second research question, using the chi-square test of independence, we investigated whether the association between reading proficiency and the frequency of writing errors held true for all types of errors. The frequency of writing errors based on

TABLE 3. Contingency table showing the number and type of errors made by each reading proficiency group

Type of errors (part of speech)	Reading proficiency group		
	Low	Intermediate	High
Verb	436	421	326
Noun	392	331	216
Article	288	227	124
Preposition	190	165	137
Adjective	66	63	52
Pronoun	53	57	44
Adverb	44	49	49
Conjunction	35	50	23
Modal verb	12	17	12
Relative pronoun	11	11	8
Interrogative	0	0	1
Other	792	874	552

TABLE 4. Adjusted residual values for the frequencies of writing errors among the reading proficiency groups

Type of errors (part of speech)	Reading proficiency group		
	Low	Intermediate	High
Verb	−0.78	−1.09	2.08*
Noun	2.68**	−1.18	−1.68
Article	3.98**	−0.80	−3.56**
Preposition	0.37	−1.64	1.41
Adjective	−0.39	−0.61	1.11
Pronoun	−0.89	0.01	0.98
Adverb	−1.71	−0.61	2.59**
Conjunction	−1.18	2.03*	−0.94
Modal verb	−1.14	0.60	0.60
Relative pronoun	−0.13	−0.03	0.19
Interrogative	−0.78	−0.77	1.72
Other	−2.60**	2.98**	−0.42

$^*p < 0.05$, $^{**}p < 0.01$

TABLE 5. Error type with positive difference by the reading proficiency group

Group	Low	Intermediate	High
Frequent error type	noun, article	conjunction	verb, adverb

type for each reading proficiency group is shown in Table 3; a significant difference was evident ($\chi^2(22) = 54.48$, $p < 0.01$). Next, we examined this association using a residual test. The results are shown in Table 4.

A significant association was observed in the error types of verb, noun, article, adverb, conjunction, and other, which suggests that the association between frequency of writing errors and reading proficiency holds true for these types of writing errors.

Interestingly, as shown in Table 5, we found markedly positive differences in each of the reading proficiency groups, which suggests that each level of reading proficiency has its own typical error type. For example, EFL learners in the low reading proficiency group tended to make errors of noun and article types.

Contrary to our expectations, the low reading proficiency group showed a markedly negative difference in the error type of other, which suggests that they were already familiar with relevant words and grammar. Therefore, we consider that a negative difference reveals another aspect of language use: the less frequent use of words and grammar relevant to writing errors. EFL learners cannot make errors with some words and grammar unless they actually use them. From the perspective of less frequent word and grammar use, the significantly negative difference, at least in the low reading proficiency group, highlights the necessity of writing and reading exercises for EFL learners.

5. **Conclusion.** This paper described an error-tagged learner corpus for multiple linguistic skills. An association was found between the frequency of writing errors and reading proficiency. We also found that each level of reading proficiency appears to have its own typical error type. These results are expected to help EFL researchers and teachers develop more effective learning and teaching methods.

Future research of this topic includes analyzing the association of writing errors with reading proficiency in more detail. Although this paper analyzed writing errors in terms of the part of speech, future research needs to explicate which kinds of errors, such as spelling, inflection, and co-occurrence, occur (in)frequently.

REFERENCES

[1] I. S. P. Nation, *Teaching ESL/EFL Reading and Writing*, Routledge, New York, 2009.
[2] I. S. P. Nation and J. Newton, *Teaching ESL/EFFL Listening and Speaking*, Routledge, New York, 2009.
[3] S. Krashen, We acquire vocabulary and spelling by reading: Additional evidence for the Input hypothesis, *The Modern Language Journal*, vol.73, pp.440-464, 1989.
[4] S. Krashen, Reading and vocabulary acquisition: Supporting evidence and some objections, *Iranian Journal of Language Teaching Research*, vol.1, no.1, pp.27-43, 2013.
[5] K. Kotani, T. Yoshimi, H. Nanjo and H. Isahara, Compiling learner corpus data of linguistic output and language processing in speaking, listening, writing, and reading, *Proc. of the 5th International Joint Conference on Natural Language Processing*, pp.1418-1422, 2011.
[6] M. Narita, C. Sato and M. Sugiura, Connector usage in the English essay writing of Japanese EFL learners, *Proc. of the 4th International Conference on Language Resources and Evaluation*, pp.1171-1174, 2004.
[7] E. Izumi, K. Uchimoto and H. Isahara, Error annotation for corpus of Japanese learner English, *Proc. of the 6th International Workshop on Linguistically Interpreted Corpora*, pp.71-80, 2005.
[8] D. H. Dahlmeier, T. Ng and S. M. Wu, Building a large annotated corpus of learner English: The NUS corpus of learner English, *Proc. of the 8th Workshop on Innovative Use of NLP for Building Educational Applications*, pp.22-31, 2013.
[9] P. MacDonald, A. Garca-Carbonell and J. M. Carot-Sierra, Computer learner corpora: Analysing interlanguage errors in synchronous and asynchronous communication, *Language Learning & Technology*, vol.17, no.2, pp.36-56, 2013.
[10] D. Nicholls, The Cambridge learner corpus-error coding and analysis for lexicography and ELT, *Proc. of the Corpus Linguistics 2003 Conference*, pp.572-581, 2013.
[11] B. Seidlhofer, A. Breiteneder, T. Klimpfinger, S. Majewski, R. Osimk-Teasdale, M.-L. Pitzl and M. Radeka, *The Vienna-Oxford International Corpus of English (version 2.0 XML)*, https://www.univie.ac.at/voice/page/index.php, 2013.
[12] M. Hirai, Correlations between active skill and passive skill test scores, *Shiken: JALT Testing & Evaluation SIG Newsletter*, vol.6, no.3, pp.2-8, 2002.
[13] C. W. Liao, Y. Qu and R. Morgan, The relationships of test scores measured by the TOEIC listening and reading test and TOEIC speaking and writing tests, *TOEIC Compendium 13*, 2010.
[14] M. D. Jackson and J. L. McClelland, Processing determinants of reading speed, *Journal of Verbal Learning and Verbal Behavior*, vol.108, pp.151-181, 1979.