Identification and Automatic Correction of Common Article Errors in Asian Learner Writing

Laurence Anthony

Center for English Language Education (CELESE)
Faculty of Science and Engineering, Waseda University, Tokyo, Japan

Introduction

Asian learners of English are known to struggle with the English article system (a, an, the), particularly in their writing (Master 1987, 1995; Celce-Murcia & Larsen-Freeman 1999; Han et al. 2006). Previous research has shown that article errors can account for over a quarter of all errors in a given learner corpus (Gamon et al., 2008), and thus, it is no surprise that many books (e.g., Claire & Greenwood 1988; Brender, 1997; Cole 2000) and research papers (e.g., Master 1995; Bitchener, Young & Cameron 2005; Farrow 2008; Chuang 2009) focus on identifying article errors and proposing effective ways to correct them in the classroom. Despite this interest, researchers have shown disagreement about which article errors cause learners the most difficulty. Han et al. (2006: 115), for example, report that the use of the Ø (zero or no) article can be one of the most complex problems faced by non-native speakers of English. On the other hand, Farrow (2008: 30) treats this as less challenging than many other article rules, such as the use of 'the' before superlatives (e.g., 'the nicest guy I ever met') and universally unique items (e.g., 'the sun').

Compounding the problem of article usage is the fact that there are few, if any, mainstream electronic grammar checkers that are able to automatically identify and correct article errors with any degree of accuracy. As part of this study, an investigation of the grammar checker included in Microsoft Word revealed that less than 1% of articles could be identified when they appeared in naturally occurring Japanese learner writing. Others (e.g., Chen, 2009) have investigated the accuracy of automatic grammar checkers designed specifically for second language learners, but again, the results are not encouraging.

This paper addresses two research questions: 1) What are the most common article errors that Asian students commit in their writing? and 2) How can an automatic grammar checking system be developed to automatically identify and correct article errors? To address the first question, I conduct an analysis of article errors found in a corpus of first- and second-year Japanese university student essay reports. In many previous studies on writing errors, highly detailed error tagging systems have been employed, but little attention has been given to the difficult task of identifying what actual errors were committed by the learner. For example, if an article is missing, it is often impossible to identify if the error is an article error or simply an error resulting from the target noun being written in the singular form instead of the correct plural form. To deal with this problem, the error tagging here was carried out by an in-class teacher as part of his regular teaching duties. As a result, all error tags (and subsequent corrections) were returned to the learner for confirmation. In cases when the articles were mistagged, these were corrected prior to the analysis. Results of this study support those of Han et al. (2006) with omission errors accounting for 79% of all article errors, and deletion errors and incorrect article errors accounting for the remaining 11% and 10% of errors, respectively (see Figure 1).

To address the second question, I explain the design and development of a novel automated article error correction tool. To explain the design of the system, I will first review two common approaches to automatic error analysis systems; 1) data-driven approaches (e.g., Turner & Charniak 2007), and 2) rule-based approaches (e.g., Bond, Ogura & Ikehara 1994; Heine 1998). I will then explain their advantages and disadvantages in terms of their flexibility to be adapted to specialized contexts and
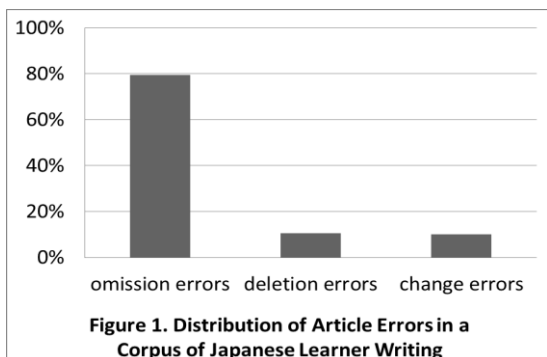
Figure 1. Distribution of Article Errors in a Corpus of Japanese Learner Writing

Table I. Performance measures of proposed system based on a test sample of 50 sentences each containing one error

Precision$_{omission}$	27/27	100%	Recall$_{omission}$	27/38	71%
Precision$_{deletion}$	5/5	100%	Recall$_{deletion}$	5/8	62%
Precision$_{change}$	0/0	-	Recall$_{change}$	0/4	0%
Accuracy = 33/50 = 64%					

domains, their ability to be understood by human users, their scalability, and their speed of processing. In this research, a novel rule-based approach to automatic article error analysis is adopted. Using real-world test data, I will show how the system is simple in design, fast, and accurate, being able to identify and correct the most common learner article errors with 100% precision (see Figure 2). I will also show how the rules adopted by the system can be made accessible to teachers providing them with a useful resource to rank and then teach the most effective article rules to learners.

References

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191-205.

Bond, F., Ogura, K., & Ikehara, S. (1994). Countability and number in Japanese to English machine translation. In D. Coleman (Ed.), *Proceedings of the 15th Conference on Computational Linguistics* (pp. 32-38). Kyoto: Association for Computational Linguistics.

Brender, A. (1989). *Three little words a, an, the*. Japan: McGraw-Hill.

Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The Grammar Book*. Boston: Heinle & Heinle.

Chen, H. (2009). Evaluating Two Web-based Grammar Checkers -Microsoft ESL Assistant and NTNU Statistical Grammar Checker. *Computational Linguistics and Chinese Language Processing*. Vol. 14, No. 2, pp. 161-180.

Chuang, W. (2009). The Effects of Four Different Types of Corrective Feedback on EFL Students' Writing in Taiwan. *Dayeh University Bulletin,* No. 4, pp. 123-128.

Claire, E. & Greenwood, R. (1988). *Three little words: a, an, and the*. Dundee, IL: Delta Systems Co.

Cole, T. (2000). *The article book: Practice towards mastery of a, an, and the*. Ann Arbor, Michigan: University of Michigan Press.

Farrow, N. K. (2008). Learning to Use the Articles, A and The, in One Lesson. *Asian EFL Journal*, Vol. 26, pp. 26-50.

Gamon, M., Gao, J., Brockett, C., Klementiev, A., Dolan, W. B., Belenko, D., & et al. (2008). Using contextual speller techniques and language modeling for ESL error correction. In *Proceedings of the Third International Joint Conference on Natural Language Processing* (pp. 449-455). Hyderabad, India: Asian Federation of Natural Language Processing.

Han, N.-R., Chodorow, M., & Leacock, C. (2006). Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12(2), 115-129.

Heine, J. E. (1998). Definiteness predictions for Japanese noun phrases. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics* (pp. 519-525). Montreal: Association for Computational Linguistics.

Master, P. (1987). A cross-linguistic interlanguage analysis of the acquisition of the English article system. *Unpublished doctoral dissertation*, University of California, Los Angeles.

Master, P. (1995). Consciousness raising and article pedagogy. In D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 183-204). Norwood, NJ: Ablex Publishing Corporation.

Turner, J., & Charniak, E. (2007). Language modeling for determiner selection. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers* (pp. 177-180). Rochester, NY: Association for Computational Linguistics.