

Automated Essay Scoring Systems: Can We Use them in Japanese Language Learning Environment?

Jun Imaki and Shunichi Ishihara

Faculty of Asian Studies, College of Asia and the Pacific, ANU

Abstract:

This study aims to empirically test whether the automated essay scoring systems—which were designed and developed for the Japanese essays written by native speakers—can be used for the Japanese compositions written by non-native speakers. According to Lonsdale (2003), automated scoring is desirable for the following three reasons: (1) practicality – essay grading is time-consuming; (2) consistency – essay grading is subjective in nature and consistency may sometimes suffer; and (3) feedback – providing feedback to a student is important, and automated scoring can provide specific suggestions. This is true for Japanese compositions written in Japanese classes as a foreign language. For example, in class placement tests where Japanese teachers are required to evaluate their students' compositions and provide their holistic scores promptly despite the fact that they only have a limited time to mark numerous essays. For regular assessments, composition marking is time exhaustive, resulting in failing to provide analytical feedback promptly despite that it is a truism that prompt feedback is important in language learning. One way of rectifying these situations is the automation of the assessment process.

Much research exists regarding the automatic evaluation of English essays written by both native and non-native speakers. Some automatic essay scoring systems, such as e-rater and criterion, have been used in real situations (Bacha, 2001). For Japanese essays written by Japanese native speakers, the automated essay scoring systems, such as Jess and Moririn, are readily available online. Previous analyses have shown that reliable results were gained from both analytic and holistic scorings with these systems (Ishioka, 2004). However, presently little research exists regarding these automated scoring systems for Japanese as a foreign language, and to our knowledge, nothing has been specifically developed to automatically assess the compositions written by non-native speakers of Japanese. With this sort of rating system yet to be developed for Japanese language learning and teaching, a sensible starting point is to see how well the systems which were developed for the essays written by Japanese native speakers work for the compositions written by non-native speakers.

In this study, 50 Japanese compositions written by Thai and Korean students which were drawn from Taiyaku database (The National Institute for Japanese Language, 2009) will be assessed by 10 Japanese teachers including native and non-native speakers. These compositions will be also put in the Japanese automated scoring system, Jess and Moririn, and

the scores from each rater will be statistically compared with those from the automated scoring systems. This paper discusses the correlations in scores between the human raters and the Japanese essay scoring systems when evaluating Japanese compositions written by non-native students of the Japanese language. If the scores from the automated scoring system are within the variables among the human raters, we consider that these systems have potential to be used in the environment of Japanese language teaching and learning.

Reference:

- Bacha,N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell us? *System* 29. pp.371-383. Retrieved December 2008 from www.elsevier.com/locate/system.
- 石岡恒憲(2004)「記述式テストにおける自動採点システムの最新動向」『行動計量学』Vol. 31. No.pp67-87
- Losndayle,D. (2003). Automated Rating of ESL Essays. *BYU Linguistics and English Language*.
- The National Institute of Japanese Language (2009). 対訳作文データベース. Retrieved January 2009 from <http://www2.kokken.go.jp/eag/wiki.cgi?page=taiyakuDB%2Ftop>

Keywords:

Evaluation, Automated scoring system, Japanese essay, rater difference, taiyaku database