Wong, L.-H. et al. (Eds.) (2013). Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

Automated test assembly tool for Chinese word-segmentation test

Chen-Huei LIAO[a*], Bor-Chen KUO[b], Kai-Chih PAI[b], Chih-Ning WU[b]
[a]*Department of Special Education, National Taichung University of Education, Taiwan*
[b]*Graduate Institute of Educational Measurement and Statistics, National Taichung University of Education, Taiwan*
*chenhueiliao@gmail.com

Abstract: The purpose of present study is to develop an automated test assembly tool for Chinese word-segmentation test. Word familiarity and word similarity were both controlled in the experiments. The study examined the effectiveness of automated test by comparing its performance with the traditional test, which was developed by linguistic experts. The results of the present study indicated that the children's performances on traditional Chinese word-segmentation tests and automated Chinese word-segmentation tests were consistent and comparable. Data collected in different age groups and with larger sample size should be investigated for future research.

Keywords: automated test assembly tool, Chinese word-segmentation test, word familiarity, word similarity.

1. Introduction

In the past, there were few studies investigated the importance of Chinese word-segmentation skill (Hsu & Huang, 2000; Li, Rayner, & Cave, 2009). One major characteristic of Chinese text is that there is no space between words, therefore, for Chinese readers, it's critical to find word boundaries to reach reading comprehension. In other words, word segmentation is a very important skill involved in Chinese reading. Two variables were generally considered by human experts when a Chinese word-segmentation test was developed. Usually, linguistic or language experts spend tremendous amount of time defining word familiarity and word similarity based on their expertise and experience. To solve this problem, the present study aims to develop a test assembly tool to generate Chinese word-segmentation test automatically. The degrees of word familiarity and word frequency were manipulated in the experiments to assess children's word-segmentation skill. The purpose of the present study are:
To define word familiarity and word similarity.
To develop an automated test assembly tool for Chinese word-segmentation test.
To examine the effectiveness of the automated test assembly tool.

2. Literature Review

2.1 Chinese word segmentation

In the past, some studies have examined the effect of Chinese reading when the spaces added between the Chinese words. However, the results found no benefit for Chinese reading (Hsu & Huang, 2000; Inhoff, Liu, Wang, & Fu, 1997). Moreover, some researchers investigated the eye movement control to Chinese readers, and the research results showed that word frequency and word predictability could effected the Chinese readers (Rayner, Li, Juhasz, & Yan, 2005; Yan, Tian, Bai, & Rayner, 2006). In summary, the Chinese word-segmentation is a very important skill for Chinese readers. However, the traditional word-segmentation tests were developed by language or linguistic experts, and they have to spend more time to define the words similarity and familiarity. It's a practical values for developing the automated test assembly tool for word-segmentation tests.

2.2 Latent semantic analysis

Latent semantic analysis (LSA) is a technology for representing the contextual-usage meaning of words by using mathematical method which is called singular value decomposition (SVD) (Landuaer & Dumais, 1997). The advantage of LSA is that it produces the similarity measures of two words (or two sentences, or two texts) relations based on the semantic space. In the semantic space, each word or text unit is represented as a vector. LSA computes their similarity by computing the cosines of their vectors. The more similar the vectors are, the higher the LSA cosine is. Some past studies showed the reliabilities of the LSA-based assessments have been as good as the reliabilities of human raters asked to make the same judgments (Landauer & Dumais, 1997; Landauer, Laham, Rehder, & Schreiner, 1997).

LSA provided the semantic similarity of any two words (texts) in more applications such as AutoTutor, Summary Street, and iSTART (VanLehn et al., 2007; Graesser et al, 2004; Kintsch et al, 2000, Wade-Stein & Kintsch, 2004; MCNamara, Levinstein, & Boonthum, 2004). Past studies have shown that the importance of vocabulary learning and the advantage of computing the words' semantic relation. The presented study defined the words similarity by using LSA and developed deferent types of Chinese word-segmentation tests to examine the effectiveness of automated tests

3. Method

3.1 The definition of word familiarity

Word familiarity was defined by character frequency and word frequency. The original word familiarity was defined as follows:

$$F = w * \frac{wf_a + wf_b + ... + wf_z}{\sum_{i=1}^{n} wf_i} + (1-w) * \frac{tf_k}{\sum_{j=1}^{n} tf_j} \quad (1)$$

Where w is the weight for the character frequency and 1-w is the weight for word frequency. wf_i is the character frequency and tf_j is the word frequency. The purposes of the presented study examined the effectiveness of the weight w to define the optimal weight for the word familiarity by comparing the experts' scoring. The presented study selected 60 terms from child corpus (Liao, 2011) and asked 8 experts in language and literacy education to rate the score of the familiarity of these terms on 7-points scale (1~7). The term has higher score means more familiar.

3.2 The definition of word similarity

The present study used Latent Semantic Space (LSA) to define the similarity of words. LSA is a mathematical and statistical method for extracting and presenting the meaning of words and passages based on a large corpus of texts (Landauer & Dumais, 1997). LSA uses a statistical technique called singular value decomposition (SVD) to condense a very large corpus of texts to 100–500 dimensions spaces (Landauer & Dumais, 1997). In the LSA space, it can provide the semantic similarity between each word and passage by computing the values of cosine. The equation of word similarity as follows:

$$\cos(t_i, t_j) = \frac{t_i t_j^{\,T}}{\|t_i\| \|t_j\|} \quad (2)$$

Where t_i is the vector representation of word i, and t_j is the vector representation of word j. $\|t_i\|$ is the vector length of word i.

3.3 The automated test assembly tool for Chinese word-segmentation

One of the objectives in the present study is to define word familiarity and word similarity for developing the automated test assembly tool for Chinese word-segmentation test. The interface and instruction of the proposed tool is illustrated in Figure 1 and Figure 2.

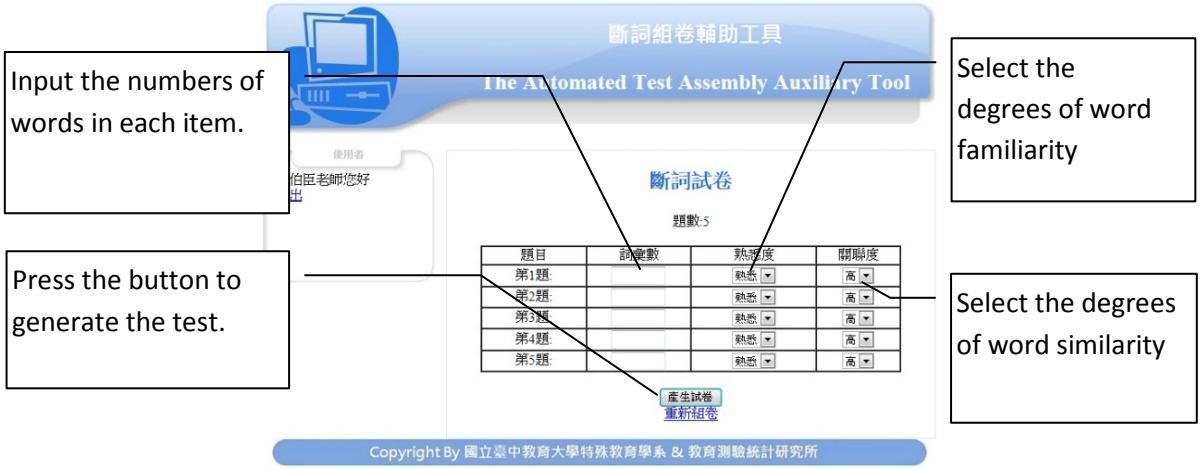

Figure 1. Interface of the automated test assembly auxiliary tool for the Chinese word-segmentation test.

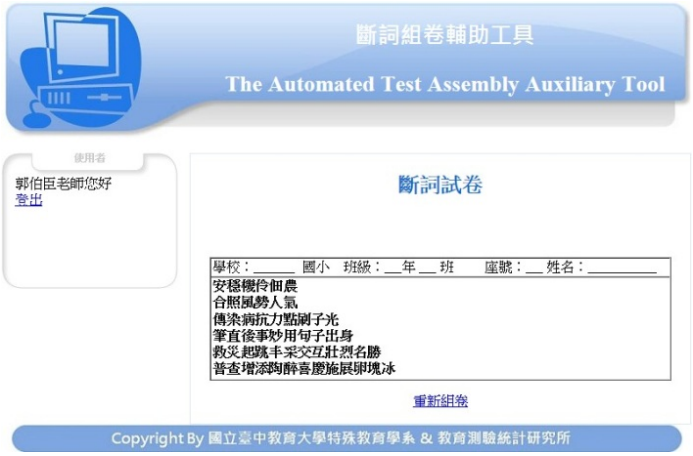

Figure 2. Interface of the Chinese word-segmentation test.

3.4 The effectiveness of the automated Chinese word-segmentation test

The LSA semantic space in the present study was generated from children corpus, which consists of 10568 words and 5219 documents. Three hundred dimensions were used in the present study.

This study examined the effectiveness of different types of Chinese word-segmentation tests by comparing them with the traditional tests, which were developed by linguistic and language experts. Table 1 presents the types of Chinese word-segmentation test used in the study. Three levels of word (semantic) similarity were manipulated: high, medium and low, and the two levels of the word familiarity were manipulated: familiar and unfamiliar. Fifty-two 5th graders were assessed with both traditional word-segmentation tests and automated word-segmentation tests. None of the children was previously diagnosed with any emotional, behavioral or sensory difficulties.

Table 1: The different types of Chinese word-segmentation tests.

	Traditional word-segmentation test	Automated word-segmentation test
Familiarity	Familiar/Unfamiliar	Familiar/Unfamiliar
Similarity	High/Medium/Low	High/Medium/Low

4. Results

This study devised an experiment to evaluate the effectiveness of the familiarity of Chinese terms. Table 2 presents the results of this experiment. It showed that when the function given the weight w=0.1 to the single word frequency, 1-w=0.9 to term frequency, the familiarity of terms and experts' score were the highest correlated (r= 0.511). This result indicated that the effectiveness of the assessment was better than just considered only single word frequency or term frequency.

Table 2: Correlation between experts and the familiarity of terms.

	Human scores
Single word frequency	0.448
Term frequency	0.418
w=0.1	0.511
w=0.2	0.473
w=0.3	0.456
w=0.4	0.445
w=0.5	0.437
w=0.6	0.431
w=0.7	0.426
w=0.8	0.423
w=0.9	0.420

Children's performance on traditional Chinese word-segmentation tests and automated Chinese word-segmentation tests was examined. The correlation between the traditional word-segmentation tests and automated tests showed in Table 3 and Table 4. The results showed that the scores of traditional Chinese word-segmentation tests and the automated Chinese word-segmentation tests were highly correlated (r=0.845, r=0.803, r=0.809) between familiar words and different degrees of word similarity (high, medium, and low). As well, the unfamiliar words and different degrees of word similarity (high, medium, and low) were also strongly correlated (r= 0.674, r=0.803, r=0.721). The strong correlations indicate that traditional Chinese word-segmentation tests and automated Chinese word-segmentation tests are comparable.

Table 3: Correlation between traditional tests and automated tests (familiar words)

		Traditional tests			Automated tests		
		High	Medium	Low	High	Medium	Low
Traditional tests	High	-	.688**	.661**	.845**	.617**	.694**
	Medium		-	.668**	.755**	.803**	.800**
	Low			-	.651**	.647**	.809**
Automated tests	High				-	.748**	.743**
	Medium					-	.792**
	Low						-

**p<.01

Table 4: Correlation between traditional tests and automated tests (unfamiliar words)

		Traditional tests			Automated tests		
		High	Medium	Low	High	Medium	Low
Traditional tests	High	-	.804**	.631**	.674**	.739**	.654**
	Medium		-	.697**	.747**	.803**	.690**
	Low			-	.409**	.557**	.721**
Automated tests	High				-	.725**	.587**
	Medium					-	.697**
	Low						-

**p<.01

5. Conclusion and Future Work

The study aimed to develop an automated test assembly tool to assist teachers to assess children's word segmentation skills effectively and efficiently. Moreover, the study defined word familiarity and word similarity to establish the automated test assembly tool for Chinese word-segmentation test. In summary, the experiment results showed that traditional Chinese word-segmentation tests and automated Chinese word-segmentation tests are comparable; however, further investigation is required. Data collected in different age groups and with larger sample size should be suggested.

Acknowledgements

The authors gratefully acknowledge the support of this study by the National Science Council of Taiwan, under the Grant No. NSC-100-2420-H-142-001-MY3.

References

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 180–193.

Hsu, S. H., & Huang, K. C. (2000). Effects of word spacing on reading Chinese text from a video display terminal. *Perceptual and Motor Skills*, *90*, 81-92.

Inhoff, A. W., Liu, W., Wang, J., & Fu, D. J. (1997). Use of spatial information during the reading of Chinese text. In D. L. Peng, H.Shu, & H. C. Chen (Eds.), *Cognitive research on Chinese language* (pp. 296–329). Beijing: Shan Dong Educational Publishing.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-140.

Landauer, T. K., Laham, D., Rehder, B., & Schreiner, M. E. (1997). How well can passage meaning be derived without using word order? a comparison of latent semantic analysis and humans. In M. G. Shafto and P. Langley (Eds.), *Proceedings of the 19th annual meeting of the Cognitive Science Society* (pp. 412-417). Mawhwah, NJ, Erlbaum.

Li, X. S., Rayner, K., & Cave, K. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology*, *58*, 525-552.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. *Behavioral Research Methods, Instruments, and Computers*, *36*, 222-233.

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review*, *12*, 1089–1093.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *31*, 3-62.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction*, *22*, 333-362.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, *97*, 259–268.