Automatically Generated Lessons and Practices of the Knowledge of Chinese Characters and Words

Hsueh-Cheng Wang
Graduate Institute of Learning and Instruction
National Central University, Taiwan
hcwang@cl.ncu.edu.tw

Hwawei Ko
Graduate Institute of Learning and Instruction
National Central University, Taiwan
hwawei@cc.ncu.edu.tw

Abstract: To assist students to learn semantic and phonetic components of Chinese characters, valid Chinese words, and their usages in context, this study provides automatically generated practices and lessons with difficulty levels based on linguistic data sources such as corpus and frequency. To enhance students' interests, the web-based practices provide multimedia and interactive interfaces, and to achieve individualized teaching purpose, the authoring tools helps teachers to create own contents with the data-driven interfaces. All lessons and practices are developed as learning activity modules in Moodle, an open source course management system. Both PC and mobile device can access the system with proper layout. Moreover, personalized portfolio reports accuracy, reaction time, and feedback for teachers to monitor students' learning progress and errors. The assessments are proved valid in an elementary school try out. It is considered to be used for Chinese as 2nd language learners.

1. Objective

Most of the Chinese characters are constructed with a semantic and a phonetic component. Considerable evidences indicate that the knowledge of Chinese character radicals is significant in recognizing Chinese characters (Shu & Anderson, 1997; Ko & Wu, 2003). Moreover, most Chinese words consist of one, two, or three individual characters, and words are essential to read Chinese. Since there is no obvious rules to learn either the structure of characters or the rule of constructing words (Ko & Tzeng, 2000), practice and experience is a must to master Chinese. This study aims at beginning and poor readers and provides them self-generated contents for learning and practicing Chinese characters and words. The characteristics of the design are as following: 1) self-generated assessments and lessons from linguistic data for practice and learning, 2) different difficulty level for different reading ability students, 3) multimedia, interactive, and data-driven interfaces which embed strategies, 4) portfolios and performances of students, and 5) automatic error analysis and feedback for self-learning of students.

2. System Framework

Server Side. There are four learning activity modules, Character Categorization, Word Identification, Daily Chinese, and Reading Comprehension, on Moodle course management system using PHP, and all system and linguistic data are stored in MySQL database. In Character Categorization module, the semantic and phonetic component knowledge and consonant and vowel of character pronunciation are diagnosed. In Word Identification module, whether two individual characters can compose a valid word is tested. Moreover, the usages of Chinese characters and words are given in Daily Chinese Lesson module to help students learn form texts retrieved from Academic Sinica Balance Corpus (ASBC). Reading Comprehension module transferred the traditional reading comprehension assessment into a preliminary eye tracking equipment in moving window. Moodle is an open source course management system and is used because it's free and provides many useful learning activity modules. Moodle also support many languages and in our website, the UTF-8 code of English and Chinese are used because the UTF-8 code is more compatible. Last but not least, it is broadly used and developed all over the world.

Client Side. Both PC and PDA interfaces are developed for ubiquitous learning. The data-driven flash applications provide interactive and multimedia environment and the content can be updated from server side.

2.1 Character Categorization

2.1.1 Test Item Editor and User Interface

Administrators or teachers can setup the test item by character frequency, number of distracter, correct group number, and task type. As shown in Figure 1, the test item is randomly generated, and the character frequency range is the top 1~500 frequent characters. There are 2 characters as distracter and 3 characters with the same semantic component. Students are instructed to drag and drop the characters with the same semantic component into the circle in user interface. The reaction time (RT) and accuracy rate are recorded during the assessment. After completing the task, students can access dictionary by clicking the characters. There are 4 task types, semantic component, phonetic component, consonant, and vowel, and students are instructed to categorize the characters with the same component feature designated in task type.

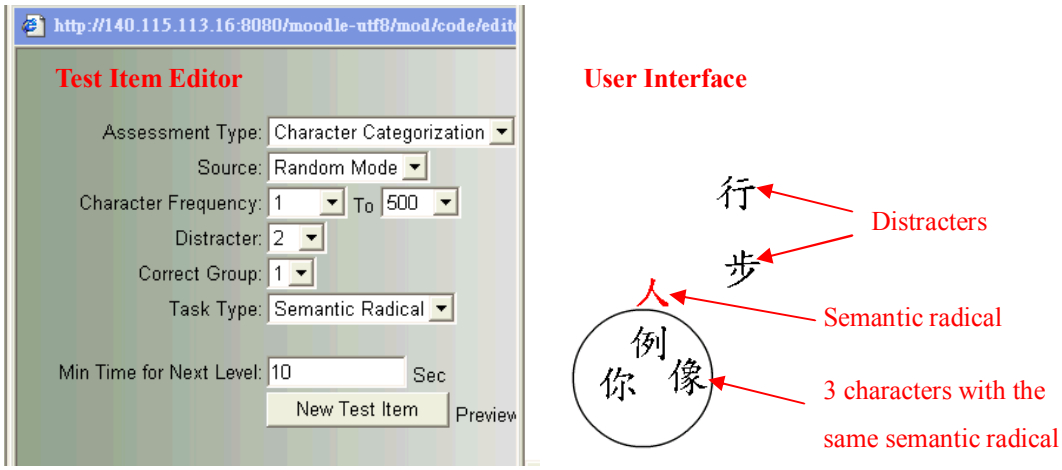

Figure 1: Test Item Editor and User Interface of Character Categorization Module

2.1.2 Quiz Editor

The quiz editor allows teachers to organize the test items, instructions, or feedbacks in a quiz. As shown in Figure 2, the random mode add a randomly generated test item in quiz while the fixed mode select a test item from test item bank. The instruction mode provides HTML editor to edit static instruction and hyperlink to include dynamic resources such as PHP script of feedback.

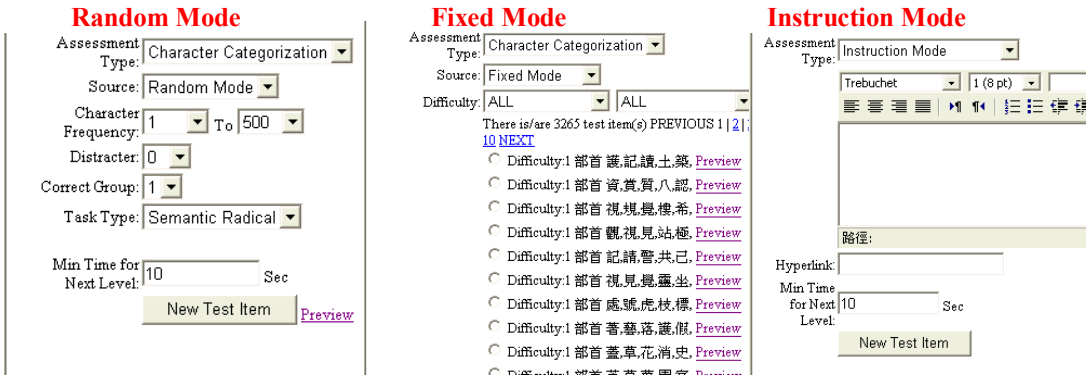

Figure 2: Random, Fixed, and Instruction Mode of Quiz Editor

2.1.3 Personal Score, Error Report, and Feedback

As shown in Figure 3, the personal score shows the RT and accuracy. Error Report presents the characters of wrongly categorized group, and each character can be clicked to look up the on-line dictionary. The section of feedback evaluates the performance of the students based on the standard line which is established according a try-out in an elementary school. Participated students are from Grades 2 to 6. Each grade has about 30 students. History in RT and accuracy present the diagrams that indicate the improvement if RT is decrease or accuracy is increase.

Figure 3: Personal Score, Error Report, and Feedback in Character Categorization Task

2.2 Word Identification

2.2.1 Test Item Editor and User Interface

Figure 4 demonstrates the randomly generated test item of 5x5 in the top 1~3000 frequent words. User interface shows 25 combinations of 10 characters which arranging the 1st character of word in row and 2nd character in column. Students are instructed to click the valid two-character words. The question mark will become smile if the clicked word is a valid one but bomb otherwise. Valid words in this study are defined according to dictionary.

RT is recorded from the nine question marks appear on screen to students complete the task, that is, to find all required valid words. Accuracy is calculated according to the ratio of bomb number to total combinations and errors are also collected. The quiz editor, personal score, error report, feedback, and history are also available in this module.

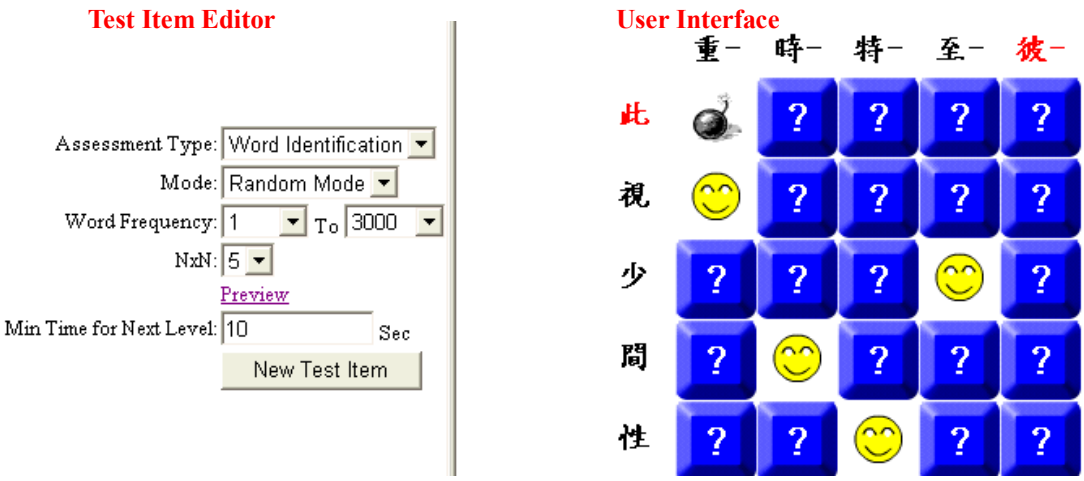

Figure 4: Test Item Editor and User Interface in Word Identification Module

2.2.2 Automatic Error Pattern Analysis

To let students learn from their own errors, the automatic error pattern analysis is performed to provide feedback of the possible correct words. A database containing phonetic information, including phonologically inconsistent characters, and component information, including semantic and phonetic component, is created. Table 1 shows the error patterns of orthographically similar, phonologically similar, and misplaced character. The examples of wrongly selected words are retrieved from a try-out in elementary and the suggested correct words are given programmatically. The automatic error pattern analysis serves as a tutoring tool for word identification. Mental lexicon

Wrongly Selected Word	Error Patterns and Suggested Correct Words		
	Orthographically Similar	Phonologically Similar	Misplaced Character
打紛 (Da Fen)	打扮 (Da Ban)		
倒歉 (Dau Chian)		道歉 (Dau Chian)	
具面 (Jiu Mian)			面具 (Mian Jiu)

Table 1: Automatic Error Patterns Analysis and Suggestions

2.3 Daily Chinese Module

2.3.1 Content Editor and User Interface

The contents of courses often rely on human maintenance and therefore costly and time consuming. As shown in Figure 5, the Daily Chinese Module provide automatically generated course by selecting a character and the related words from required character and word frequency. Example sentences are given by retrieving the sentences containing related words from corpus.

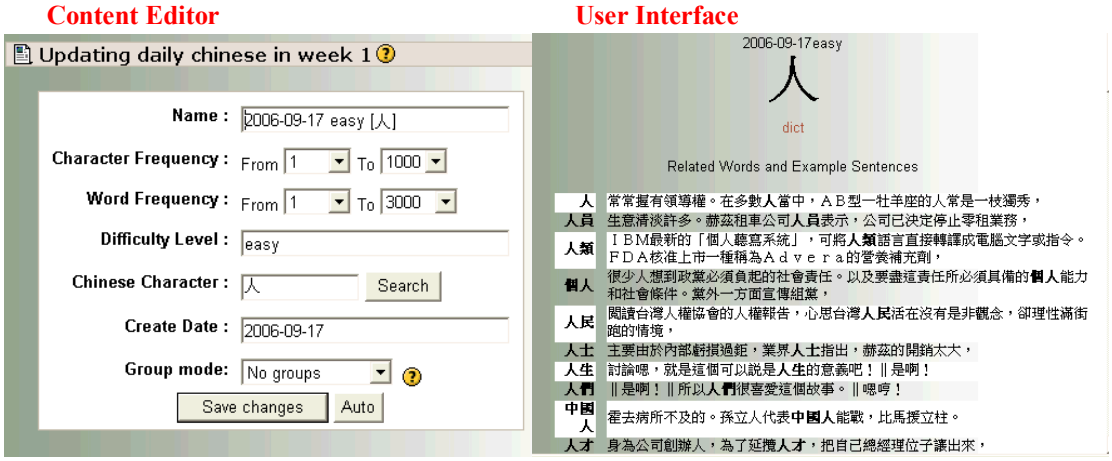

Figure 5: Content Editor and User Interface in Daily Chinese Module

2.3.2 Related Assessment of Chinese Categorization and Word Identification

Taking advantage of the data-driven interface, system can generate Character Categorization and Word Identification assessments based on the given course of Daily Chinese to evaluate the learning outcome of students.

2.4 Reading Comprehension Module

The module is design according to eye tracker measurements and put the traditional reading comprehension assessment in the moving window task (e.g. Tsai, Lee, Tzeng, Hung, &Yen, 2004). Based on the gaze contingent experiment of eye movement, the size of the opened window is set as 3 characters on the left of cursor and 1 character on the right. The average fixation duration is 250 ms (Ko, Chen & Liao, 2006) and the characters in this module become burned after that, as shown in Figure 6. Students who have difficulty in reading often fixate longer and this module serves as an assisted tool for students to know that they have already fixated too long. Students are asked to move alone the mouse while they read. This module records the move and fixation duration on each character for further analysis.

Figure 6: Reading Comprehension Module

3. Validation

The performance of different grade students shows the language development in a try-out. The try-out is performed in elementary school including 1^{st} ~ 6^{th} grade students on character categorization and word identification module and there are over 50 students in each grade. The collecting data of assessments shows that 1) higher grade students often complete the tasks faster and more accuracy than lower grade students, 2) the design of difficulty level is qualified because students have to process longer in hard ones, 3) different tasks reflect different responses, that is, the more characters or combination of words, the more processing time is needed. Besides, the consonant / vowel tasks are harder than semantic / phonetic component tasks in character categorization module because there are always visual clues in the latter tasks while the pronunciations of characters are not always regular from orthographic forms. The scores and reaction times of each grade in this try-out can also serve as standard line of different difficulty levels for further studies. In addition, students were having fun and also learned strategies of reading in the automatic mechanism such as the semantic and phonetic component strategies which is critical in Chinese reading.

4. Educational Importance of the Study

In our design, students with different reading ability can acquire suitable learning of Chinese characters and words according to their difficulty levels. The proposed computer-assisted language learning tools work to enhance students' interests of learning because they were having fun. It is evidenced that during the try-out, after practicing and assessment at school, students revisited our website at homes. Teachers and parents can learn the performance and problem of each student from the portfolios which present the learning progress and error patterns. Students can learn Chinese characters, words, and their usages by using the presented system. It is

proved that students learned radical knowledge and gained speed of word identification by practicing in the self-generated mechanism. Since the modules generate new items every time, it serves as a good self-paced practice and learning platform for Chinese language learners. The standard line of each task can be established from our try-out data in each grade and we consider performing adaptive learning to provide suitable difficulty materials. We also consider putting the program for those who would like to take Chinese as 2nd language in the near future.

Reference

Academia Sinica. (1998). *Academia Sinica Balanced Corpus* (version 3) [Electronic database]. Taipei, Taiwan.

Academia Sinica. (1998). *Word List with Accumulated Word Frequency in Sinica Corpus 3.0* Academic Sinica. Taipei, Taiwan.

Huang, L. Y. (2006). *Fluency Assessments of Common Characters*, Division of Special Education, Minister of Education, Taiwan.

Juel, C. (1998). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology,* 80 (4), 437-447.

Just, M. A. , & Carpenter, P. A. (1987). *The Psychology of Reading and Language Comprehension*, Ma: Allyn & Bacon..

Ko, Hwawei &Wu, Chia Feng (2003). The role of character components in reading Chinese. In C. McBride-Chang & H.-C. Chen (eds.) *Reading Development in Chinese Children.* NY: Praeger.

Ko, H. W., & Tzeng, O.J.L. (2000). The role of phonological awareness in a phonetically opaque script. *Studies in the linguistic Sciences,* 30, 119-132.

KO, Hwawei, Chan, Ming-Le, & Liao, Chia Ning (2006)

Shu, H., & Anderson, R. (1997). Role of radical awareness in the character and word acquisition of Chinese children. *Reading Research Quarterly*, 32, 78-89.

Tsai, J.-L., Lee, C.-Y., Tzeng, O.J.L., Hung, D., & N.-S. Yen (2004). Use of phonological codes for Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain & Language,* 91:235-244.

Tsao, F. F. (2002). *Common Words for Elementary Students*. National Language Committee, Ministry of Education, Taiwan, 2002. (In Chinese)