Applications of NLP techniques to computer-assisted authoring of test items for elementary Chinese*

LIU Chao-Lin, LIN Jen-Hsiang, WANG Yu-Chun
(Department of Computer Science, National Chengchi University, Taipei 11605, Taiwan)

Abstract: We report an implemented environment for computer-assisted authoring of test items and provide a brief discussion about the applications of NLP techniques for computer assisted language learning. Test items can serve as a tool for language learners to examine their competence in the target language. We apply techniques for natural language processing to help teachers prepare test items for elementary Chinese. Teachers can then compile the test items to form test sheets for formal examinations on the Internet. Our system can record the results of the tests for post-test analysis of students' achievements so that teachers can gain insight into the distributions over students' competence levels and can adjust the teaching activities for the students who may need individualized care. At this moment, our system offers assistance for authoring test items for basic listening comprehension, cloze tests, incorrect-character identification, sentence reconstruction, and usage of measure words.

Key words: computer assisted test item authoring; natural language processing; mandarin grammar; sentence reconstruction; Chinese-character formation

1. Introduction

The history of applying computers in assisting language learning and teaching can be dated back as least 40 years ago, when the Programmed Logic for Automatic Teaching Operations, which is referred as PLATO usually, was initiated in 1960 (Heift & Schulze, 2007, p. 70). The computing powers of modern computers and the accessibility to information supported by the Internet offer a very good environment for language learning that has never been seen before.

Computers and the Internet help language learning in a wide variety of ways. Applying multimedia-related technologies, teachers can present teaching material in attractive and lifelike ways. Students can practice conversation with their peers via the Internet. The exploding growth in accessibility of the Internet in the past decade not only has made English the de facto official language for international communication, but also has offered students free access to a humongous amount of materials prepared in English. This applies to other

* Work reported in this paper was supported in part by the research plans NSC-95-2221-E-004-013-MY2 and NSC-97-2221-E-004-007-MY2 from the National Science Council and in part by the funding ATU-NCCU-96H061 and ATU-NCCU-98 from the Ministry of Education of Taiwan.

LIU Chao-Lin, Ph.D., professor, Department of Computer Science, National Chengchi University; research fields: natural language processing, artificial intelligence.

LIN Jen-Hsiang, M.S., Department of Computer Science, National Chengchi University; research fields: computer-assisted language learning, natural language processing.

WANG Yu-Chun, M.S., Department of Electrical Engineering, National Taiwan University; research fields: natural language processing, computational linguistics, information retrieval.

languages as well, thought the amount of information prepared in other languages may not be comparable. Although language learning is not just about obtaining the basic material, the ability to access the information prepared in a particular language does offer a chance to enhance the environment for learning that language.

The techniques for natural language processing (NLP) (Manning & Schütze, 1999) are useful for designing systems for information retrieval, knowledge management, and language learning, teaching, and testing. In recent years, the applications of NLP techniques have received attention of researchers in the Computer Assisted Language Instruction Consortium (often referred as CALICO, http://calico.org/, instituted in 1983) and the researchers in the computational linguistics, e.g., in United States of America (Burstein & Leacock, 2003; Burstein & Leacock, 2005; Tetreault et al., 2008) and in Europe (Metcalf & Meurers, 2006; Ezeiza et al., 2007). Heift and Schulze (2007) report that there are over 100 documented projects that employed NLP techniques for assisting language learning. In a recent issue of this journal, SUN (2008) discussed the availability of online resources for learning English as well.

The applications of computing technologies to the learning of Chinese language can be traced back as far as 40 years ago, when researchers applied computers to collate and present Chinese text for educational purposes (WANG, 1966). The superior computing powers of modern computers offer researchers and practitioners to invent more complicated tools for language learning. As a result, these so-called Computer-Assisted Language Learning (CALL) applications are no longer limited to academic laboratories, and have expanded their existence into real-world classrooms, e.g., in Taiwan (LIN et al., 2004) and in USA (ZHANG, 2007). ZHANG (1998) and Bourgerie (2003) have provided good reviews of the past achievements in CALL for Chinese.

In this paper, we focus on how computers may help teachers assess students' competence in Chinese, and present the platform that we have built for assisting teachers to prepare test items for elementary Chinese. There are five categories of test items in the current system, including basic listening comprehension, cloze tests, incorrect-character identification, sentence reconstruction, and usage of measure words. In addition to allowing the teachers to build databases of test items, we offer methods for teachers to compile test sheets by selecting different test items from the test-item databases. The students can then take the tests on the Internet, and, if desirable, the test results can be reported to the students right away. In addition, teachers can analyze the test results so that the teachers can find the deficiency of individual students, and offer individualized help for students.

We explain the main operations and working procedure of our system in section 2. We elaborate the types of test items a teacher may prepare with our system in section 3, and overview how our system assists the tasks of test sheet compilation and post-test analysis in section 4. Finally, we introduce the NLP techniques that we have employed to build our system in section 5 before reviewing more related work in section 6 and making concluding remarks in section 7.

2. Main operations of the proposed system

Figure 1 shows the main operations and flow of our system. From left to right, we extract raw material from appropriate sources and help the teachers produce test sheets with processed data. In this figure, we use squares with rounded corners to represent data, ovals to represent processes. We employ a different font, arial, to refer to the objects in the figure within our text, and links between objects denote the flow of information.

We obtain appropriate material from different sources with the material collector which saves the obtained material in the raw corpora. The annotator processes the raw material in the raw corpora, annotates the preprocessed material with linguistic information, and saves the results in the tagged corpora. The preprocessing tasks include some low-level operations, including the removal of HTML tags and the extraction of text from the raw data. Teachers will author the test items through the test-item authoring interface which will find appropriate material with the help of material extractor. The annotator and the material extractor employ NLP techniques to enrich and search the material that are useful for authoring test items. The teachers save the test items in the test item databases, and can compile test sheets with the test items from the test item databases in the future. The test items can be reused directly, or can be revised for a new test. Finally, students can take the tests online with any test sheets that locate in the test sheet databases.

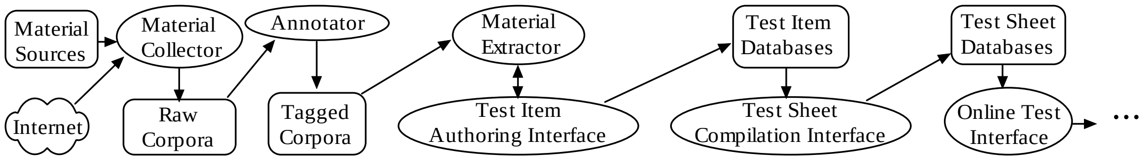

Figure 1 Main flow for producing test items from raw material

Currently, we mainly make use of the course material for elementary schools in Taiwan to demonstrate the functions of our system. Some language materials were extracted from the Internet. We also included contents of books that are available in the bookstores for illustration, but we have not solved the copyright problem yet. As a consequence, we cannot publicize the current system, and demonstration can be conducted in private only.

The current aim of our system is to provide a platform based on which teachers can prepare and share test-related materials. By offering a public platform, we attempt to attract volunteer teachers who are willingly to share the test items that they author. We hope to make the test items publically available (CHU et al., 2001) to help the financially disadvantaged students reduce the burden of buying many exercise books. It is clear that a platform itself cannot accomplish the goal of assisting language learning and testing, and a reliable and steady source of teaching materials is required for facilitating the students to learn languages.

3. Types of test items

We describe the types of test items and explain the assistance that our system offers in separate subsections.

3.1 Basic listening comprehension

There are five tones (including the neutral tone) in Mandarin Chinese, and it is a fundamental ability for students to learn to differentiate sounds pronounced in different tones. For instance, guang(1) bo(1)[1] and guang(1) bo(2) are different. The former means "broadcasting" (广播) in English, and the latter is close to "omniscient" (广博). These two words are different in how the second characters in them are pronounced.

In our system, we can offer sounds that were recorded by human experts, use these recorded sounds in a multiple-choice test item, and ask students to select the correct sound that is specified in the test item. In Mandarin Chinese, there are only a limited number of actual combinations of onsets and rhymes. Hence it is possible to record the pronunciations of all characters by human at a relatively low cost than recording the pronunciations for all English words. To obtain more accurate recordings, we need to consider the tone sandhi problems (CHEN, 2000), which we discuss later in Section 3.4. A typical test item looks like the one shown below, where *va*, *vb*, *vc*,

[1] We use Hanyu Pinyin(漢語拼音) in this paper. Retrieved from http://en.wikipedia.org/wiki/Pinyin.

and *vd* denote clickable links.

(1) Which of the following sound is “guang(1) bo(2)”?
(a) va (b) vb (c) vc (d) vd

By clicking on those links, students can hear the recorded sounds. In a typical test, one and only one of these choices is the correct answer. The other choices can be chosen by the teacher who prepares the test item to control the item difficulty.

3.2 Cloze tests

A cloze test is a multiple-choice test, in which one and only one of the candidate words is correct. The examinee has to find the correct answer that fits the blank position in the sentence. A typical item looks like the following.

(2) 州长于昨日正式_____中文教师学会，与会长深入讨论加州的中文教育问题。(The governor officially visited the Society of Chinese Teachers and discussed with the president about the education of Chinese in California yesterday.)
(a) 见面 (meet) (b) 走访 (go to) (c) 拜访 (visit) (d) 访视 (oversee)

Cloze tests are quite common in English tests, such as GRE verbal tests and TOEFL. We have built a working system that can help us find English words with a specific meaning with a reasonable accuracy (LIU et al., 2005). The current system offers a similar service for Chinese, but we have not worked on the problem of word sense disambiguation in Chinese yet.

To create a cloze test item, a teacher determines which Chinese word that will be the answer, and our system will choose (from the tagged corpora) and present the sample sentences that contain the answer word to the teacher. The teacher will choose one of the sample sentences for the test item. Our system will then show an interface where the answer will be removed from the chosen sentence (usually called *stem* in computer assisted item generation).

A cloze item needs distracters, in addition to the stem and the correct answer in the choices. To assist the teachers prepare the distracters, we rely on a Web-based service at the Institute of Linguistics at the Academia Sinica to find Chinese words of similar meanings (CHENG, 2004), and present these words to the teachers as candidates for the distracters. For instance, “造访” (zao(4) fang(3), visit), “拜会” (bai(4) hui(4), visit politely), “走访” (zou(3) fang(3), go to), and “见面” (jian(4) mian(4), meet) carry a related meaning with “拜访” (bai(4) fang(3), visit). The teachers can either choose or avoid those semantically similar, yet possibly different in ordinary usage, words for the test item.

Our system can also present words that contain the same characters with the answer as possible distracters. For instance, both “喝酒” (he(1) jiu(3), drink wine) and “奉茶” (feng(4) cha(2), provide tea for drinking) can serve as a distracter for “喝茶” (he(1) cha(2), drink tea) because they share one character at exactly the same position in the words. We employ HowNet (http://www.keenage.com/) to find candidate words of this category.

3.3 Incorrect-character identification

Just like that there are English words that are spelled similarly, there are Chinese characters that are pronounced or written in very similar ways. For instance, the sentence “今天上午我们来试场买菜” (We came to the test site to buy vegetables.) contains an incorrect character. We should replace “试场” (a place for taking examinations) with “市场” (a market), whilethese two words have the same pronunciation: shi(4) chang(3). The sentence “经理要我构买一部计算器” (The manager asked me to buy a calculator.) also contains an error, and we need to replace “构买” (This is an incorrect Chinese word, so no appropriate English translation can be provided.) with “购买” (buy).

Figure 2 shows the interface through which teachers can create test items for incorrect-character identification. A teacher provides a Chinese sentence which contains the character that will be replaced by an incorrect character. The teacher has to choose the answer character which s/he wants to have a test item, and our system will provide four types of incorrect characters. The first type (“建议字”: recommended selections in Figure 2) is a list of candidate characters that are maintained by experts; the second (相同音, same sound) and the third (相似音, similar sound) include characters that have, respectively, the same and similar pronunciation with the answer character; and the last (相似字, visually similar words) includes characters that look similar to the answer character. The teacher can choose a character from either of these lists, and our system will replace the answer character with the chosen (incorrect) character to form a test item.

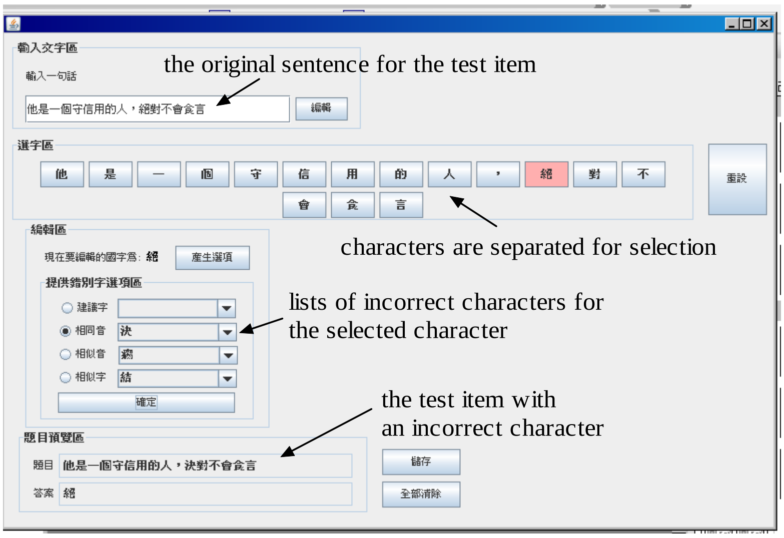

Figure 2 The interface for authoring test items for identifying incorrect characters

Given a machine readable lexicon, it is not difficult to list all those characters that have the same or similar pronunciation, if we ignore special rules, e.g., tone sandhi (CHEN, 2000), for pronouncing some word sequences. (A Chinese character is the most basic unit in written Chinese, while one or more Chinese characters form a Chinese word. For instance, “喝” (drink) and “茶” (tea) are two Chinese characters, and “喝茶” (drink tea) is a Chinese word.) It is relatively difficult to find words that are written in similar ways, e.g., “构” with “购”, in an efficient way. These two characters share the same component, but that particular component is not the radical of these two characters. We have invented an efficient algorithm (LIU & LIN, 2008; LIU et al., 2009) that allows us to find visually similar Chinese characters without forcing ourselves to apply image processing techniques that are typically more computationally more costly.

3.4 Sentence reconstruction

Grammar is important for students who learn a second language. Ultimately, students should become capable of using the language without consciously thinking about the grammar—just like ordinary native speakers. However, during the learning stage, learning about the grammar of the target language is crucial.

There are multiple forms of drills that require the knowledge of grammar (LI & SONG, 2007). Writing short sentences and oral conversation are very common. For students of the beginning level, one possible and simpler way is to reconstruct some original sentences with shuffled segments of the sentences. This strategy should apply to languages in which semantics depend reasonably strongly on the word orders, and the ability of reconstructing shuffled segments shows the competence in grammars. For instance, we can segment and shuffle the sentence "中文是一门有趣的课程" (Chinese is an interesting course.) into five parts: "一门" (an), "中文" (Chinese), "有趣的" (interesting), "是" (is), and "课程" (course), and ask students to rebuild the original sentence.

Figure 3 shows the interface where students can respond to test items for sentence reconstruction. The upper left corner shows the context of the conversation between two persons, where we see three short sentences in this particular example. The first person begins the conversation with the first sentence (shown in red in the colored print of this paper), and responds to the second sentence (shown in blue) with the third sentence (shown in green). We show only the leading part of the second sentence, which the second person uses to respond to the first sentence. The rest part of the second sentence is separated into a few words, which are shown in the middle of Figure 3. A student can try different orders of the segments and finalizes his/her answer at the bottom of the interface.

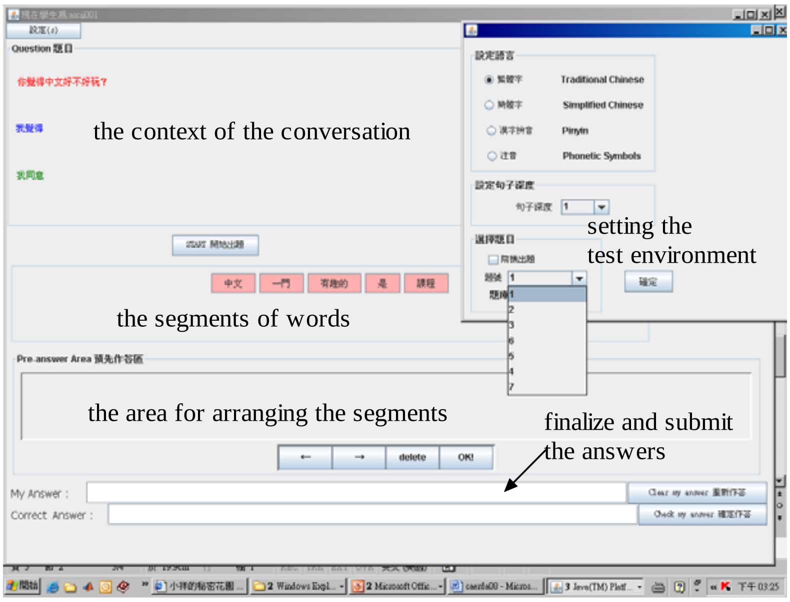

Figure 3 The interface for students to answer test items for sentence reconstruction

We ought to expect that we should not segment the sentences in arbitrary ways. In the previous example, even native speakers of Chinese will find it unacceptable if we segmented the sentence into "中", "文是", "一门有", "趣的课", and "程". On the other hand, it would not be a very interesting task if we ask teachers to segment the sentences manually.

Hence, the NLP techniques are instrumental for generating test items of this category. Applying grammar rules that are provided by experts, the parsers segment sentences into meaningful parts. We can apply parsers for natural languages to generate the syntactic structures for the sentences, e.g., the *parse tree* shown in Figure 4, and employ the structures to create test items for sentence reconstruction.

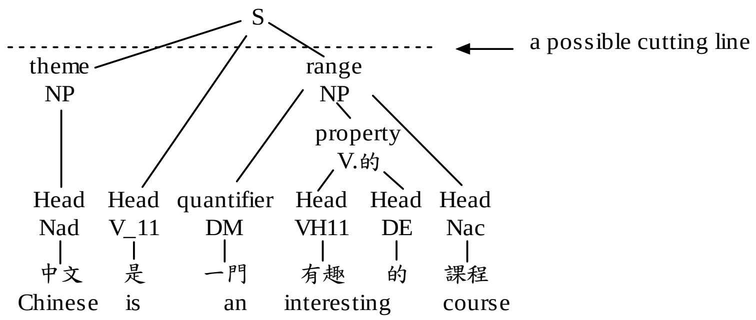

Figure 4 A sample parse tree (generated by the parser of Academia Sinica, http://parser.iis.sninca.edu.tw, 2008/09/30)

Replying on the information provided by a parse tree allows us to segment the original sentences at different levels. This becomes an essential tool for us to design tests that adapt to students' levels of competence in Chinese grammar. If we treat all words under the interior node "range NP" in Figure 4 as a unit, we will segment the sentence into "中文", "是", and "一门有趣的课程". This is the case when we adopt the cutting line shown in Figure 4. If we split the original sentence into the lowest word level, we will segment the sentence into "中文", "是", "一门", "有趣" , "的", and "课程". Intuitively, it is harder to rebuild the original sentence if the sentence is split into more units. Hence, in an adaptive test (TIAN et al., 2007), we may want to split the test sentences into more units for more advanced students. We set the depth of cutting lines ("句子深度") in Figure 3 to a larger quantity, if we want to convert the original sentence into more number of words or phrases.

A major problem that we can encounter when we put the items for sentence reconstruction in real world tests is the existence of multiple answers. This can also occur in English. For instance, both "I joined the CAERDA last year." and "Last year I joined the CAERDA." are correct sentences. There are other possible situations when a natural language allows some flexibility in the word orders. It is thus necessary for our system to generate multiple legal parses for a given sentence whenever possible. This can be a difficult challenge because not all syntactic ambiguities can be captured perfectly in a parser easily (Manning & Schültz, 1999).

We employ two strategies to make up for this problem. The first and easy, but costly, method is to expect the teachers to add more answers to the database manually whenever possible. The second method is to add conversational context for the sentence that is being reconstructed, shown in the upper left corner in Figure 3. Using the context, we may reduce the number of permissible sentences given the segments, though this may not solve all possible problems.

Currently, we imagine that test items of this category should be useful for students who learn Chinese as a second language. To meet the needs for this application, we must prepare to present the test items in different forms because student may learn Chinese in different countries in the world, and the Chinese text may have been presented in different forms in students' home countries. There are two different forms of written Chinese, namely, the traditional Chinese and the simplified Chinese. There are also different ways to Romanize Chinese, namely, Zhuyin, Hanyu Pinyin, and possibly others. With the help of a small translator, our system can present Chinese in traditional Chinese, simplified Chinese, Zhuyin, and Hanyu Pinyin. A student or a teacher can choose his/her preference by clicking on appropriate buttons shown on the right hand side of Figure 3.

When presenting the test items in the Romanized form and the five tones in Mandarin, we have to deal with the tone sandhi problems (CHEN, 2000). Like other natural languages, people may pronounce a character differently when the character is in a special context. For instance, when a character of the third tone follows another character of the third tone in mandarin Chinese, people pronounce the first character in the second tone.

3.5 Usage of Measure Words

Unlike English, Chinese includes many special words for describing the units of objects. These so-called measure words are important for producing Chinese text and utterance (ZHANG et al., 2008). For instance, the translation for "an apple", "a movie", and "a letter" are "一颗苹果", "一场电影", and "一封信", respectively. There are no strict rules for governing which measure word goes with a special class of objects, so it is important to learn the appropriate measure words whenever possible.

Given a parser for Chinese, it is not difficult for us to find sentences that contain a particular measure word and to present those sentences to a teacher who wants to prepare test items for measure words. The online parser that is maintained by the Academia Sinica (http://parser.iis.sninca.edu.tw) assigns a special part-of-speech tag, i.e., "M" to measure words in Chinese. Therefore, we can submit a sentence to the parser and check the returned results to find the measure words in the sentences.

4. More supported functions

The test items that are prepared with the functions described in the previous section can be used separately for practice. They can also be used for formal examinations. Hence, it is natural for our system to assist the task of test sheet preparation, test administration, and post-test analysis.

Our system stores the test items in databases, and our system provides methods with which teachers can search for language materials and test items to compile test sheets. For instance, if a teacher needs a Chinese cloze test item for "美好" (good and beautiful), the teacher can ask our system to search for sentences that contain the target word from our corpora and databases. As an additional feature, we can assign unique identification numbers to test sheets so that different groups of students can take different tests.

We have built several functions that are forming a full-scale computer-assisted testing and evaluation system on the Internet (CHOU, 2000). Students need to have their own account. When administrating an examination, teachers can tell students the identification number of the test sheets, and students need to respond to test items on those test sheets. We did not show these services for post-test analyses in Figure 1 to simplify the figure, but the test results can be graded on the spot, and the results are stored in another database so that the teachers can analyze students' performance. In addition, teachers can look into the performance of individual students or a group of students. We hope and believe that such post-test analyses makes individualized teaching more realizable than the old days.

5. NLP techniques

We have applied NLP techniques, including phonological, lexical, syntactical and semantic considerations, to assist the preparation of the test items and test sheets that we explained in the previous sections.

We are relying on human experts to record the sound for basic pronunciations in Chinese in this system. If we will expand the scale of this type of tests, we may want to synthesize recordings with computer software. We obtained recordings for English text with the AT&T Natural Voice (http://www.naturalvoices.att.com) in a previous work (HUANG et al., 2005). Research projects exist for text-to-speech synthesis for Chinese, e.g., IBM Text-to-Speech Research (http://www.research.ibm.com/tts/).

Word sense disambiguation (WSD) is a research topic that is closely related to how we provide assistance for the authoring of cloze test items. Just like what we have done for the authoring of cloze test items for English (LIU et al., 2005), we will have to tackle the problem of WSD in Chinese, c.f. (HSIAO et al., 2007), if it is necessary to tell the difference in lexical semantics for one word in two different contexts. Doing so will help us to extract sample sentences of higher quality than just extracting sentences that contain the desired answer words, which perhaps carry different meanings.

田=[田]		購=[月金廿廿月]
由=[中田]	間=[日弓日]	構=[木廿廿月]
甲=[田中]	閒=[日弓月]	刑=[一十一十弓]
申=[中田中]	閃=[日弓人]	型=[一弓土]

Figure 5 Cangjie codes for some characters

The capability of finding semantically similar words (CHENG, 2004) allows our system to provide useful recommendations for the teachers to prepare test items for cloze tests. In fact, we have designed a method that applies the semantic annotation for words in HowNet to find semantically related words in Chinese (LU et al., 2008). Two words can be considered semantically related if they share the same defining semantic units (義原) in HowNet. Using semantically related words that are contextually inappropriate for the distracters, e.g., the sample test item shown in section 3.2, makes the resulting test items more challenging for cloze tests.

The parser provided by the Academia Sinica is crucial for serving the needs to prepare the test items for sentence reconstruction and the usage of measure words. In both applications, the knowledge at the syntactic level matters. In addition to segmenting the Chinese sentences in meaningful ways, a parse tree that is similar to the one shown in Figure 4 will allow us to segment Chinese sentences into different numbers of units. For instance, we could have segmented the sentence into "中文", "是", and "一门有趣的课程", thus making a very simple test item. The ability to create test items of varying levels of difficulties is the starting point of adaptive testing.

We have invented the method to find similar Chinese characters ourselves (LIU & LIN, 2008; LIU et al., 2009), by splitting Chinese characters into components. This is inspired and adapted from the design of the Cangjie input method for Chinese (仓颉输入法) (http://en.wikipedia.org/wiki/Cangjie_method) and a related work reported in (JUANG et al., 2005). We created a database that stored the results of decomposing Chinese characters into parts. When we need to find Chinese characters that have similar shapes, we simply compare the list of subparts of each character, and we can find the characters that have similar shapes.

Figure 5 shows some interesting examples. "购" and "构" share "廿", "廿", and "月", which result from the decomposition of "冓"; all three characters in the middle column share "日" and "弓", which result from the decomposition of "门". Hence, it is possible to identify visually similar characters from their decompositions. The original Cangjie codes for Chinese characters were designed for an input method, so are not perfect for telling the similarity among characters. We have actually refined this intuitive method in (LIU & LIN, 2008) and conducted a more comprehensive evaluation of our methods in (LIU et al., 2009), so we do not provide all details for the proposed methods in this paper.

6. Related work and advanced applications

NLP techniques can facilitate CALL for Chinese in various ways. We have presented a few examples of assisting the authoring of test items, and have briefly mentioned the applications of assisting the administration of online tests and post-test analyses. "Basic listening comprehension" is necessary for beginning learners, "incorrect-character identification" aims at the learning of individual characters within specific contexts, "cloze tests" and "usage of measure words" focus more on the level of Chinese words, and "sentence reconstruction" demands students to practice their knowledge in Chinese grammar.

NLP techniques that can be applied to CALL for any languages should be able to be used for CALL for Chinese. An ideal review ought to cover issues for listening, speaking, reading, and writing at the character, word, phrase, sentence, conversation and essay writing levels. For instance, it is worthwhile to mention that writing and remembering Chinese characters are challenging for everyone, including native speakers of Chinese, and Professor XIE has collected a list of software for learning Chinese characters at http://www.csulb.edu/~txie/character.htm. We have applied information about collocation and selection for recommending distracters for test items for English cloze tests in (LIU et al., 2005), and we have explained how we employ semantically similar words for recommending distracters for test items for Chinese cloze tests in this paper. It is also possible to rely on heuristics for recommending distracters (Aldabe et al., 2007). In addition to cloze tests, there are different forms of tests for assessing students' vocabulary. Sumita, et al. (2005) study ways to generate fill-in-the-blank test items.

However, it should not be surprising to find that there are a lot of new developments of CALL for Chinese and other languages, given the explosive interests in applying new technologies to assist language learning. Hence, a comprehensive review requires a much longer essay than the current writing, and we choose to list few items that we thought would be practically useful for future students.

Becoming competent in using Chinese in everyday conversation should be one of the main goals for many learners. Network-based applications, such as MSN and Skype, offer students around the globe the platforms to communicate conveniently by typing or talking via the Internet. Can we apply NLP techniques to support this type of language learning activities? Is it possible to monitor the contents of the conversation, and offer hints about sample sentences or useful words, without compromising privacy? If students would like to discuss about some specific topics, how may the computers assist the students to find appropriate conversation partners? Moreover, could a computer chat with a student (JIA, 2004), and provide supports for learning Chinese?

Reading texts that are written in the targeted language is another important goal for most learners. Can the computers recommend language materials that are appropriate for a learner to read, where appropriateness may depend on the learners' competence and interests? NLP researchers have looked into this direction for a long time (Chall & Dale, 1995), and are still making progress (Heilman et al., 2008).

Writing in the targeted language is yet another important goal for learners. It is now possible to grade English essays with NLP techniques (Attali & Burstein, 2006), and researchers have started to look for algorithms for grading Chinese essays (CHANG et al., 2007). If we can develop algorithms for grading written essays, can we find ways to identify bugs in those essays and provide suggestions for improving the graded essays?

7. Concluding remarks

We report the applications of NLP techniques at the phonological, lexical, syntactic, and semantic levels to the design of a computer-assisted item authoring environment, and discuss a list of interesting research topics in this paper. Our experience indicates that the environment can improve the efficiency for item authoring. However, the quality of the test items depends on the quality of the underlying corpora, based on which our system searches for recommendations. Obviously, the pedagogical expertise of those who actually prepare the test items influences the quality of the test items and test sheets (XU & LIU, 2008). Since NLP techniques are good at manipulating linguistic information, they can serve as an important basis for computer-assisted language learning, and we anticipate that more NLP-based CALL systems to come in the future. Linguistic information alone may not provide us a comprehensive view of how students learn languages. Integrating components that consider related aspects, such as the cognitive processes for interpreting linguistic information and students' cultural background (Bailin & Grafstein, 2001), will help us invent better CALL systems.

References:

Aldabe, I., Maritxalar, M. & Martinez, E. (2007). *Evaluating and improving the distractor-generating heuristics*. In: Ezeiza et al. (Eds.), 7-14.

Attali, Y. & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *JTLA*, 4(3). Retrieved from http://www.jtla.org

Bailin, A. & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language and Communication*, *21*(2): 285-301.

Bourgerie, D. S. (2003). Computer assisted language learning for Chinese: A survey and annotated bibliography. *JCLTA*, *38*(2), 17-48.

Burstein, J. & Leacock, C. (Eds.). (2003). *Proc. of 1st Workshop on Building Educational Applications Using NLP*. 41st ACL.

Burstein, J. & Leacock, C. (Eds.) (2005). *Proc. of 2nd Workshop on Building Educational Applications Using NLP*. 43rd ACL.

Chall, J. & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Brookline Books.

CHANG, T. H., LEE, C. H. & TAM, H. P. (2007). On issues of feature extraction in Chinese automatic essay scoring system. *Proc. of the 13th Int'l Conf. on Artificial Intelligence in Education*, 545–547.

CHEN, M. Y. (2000). *Tone Sandhi: Patterns across Chinese dialects*. Cambridge University Press.

CHENG, C.C. (2004). Word-focused extensive reading with guidance. *The 13th Int'l Symposium on English Teaching*, 24–32, Crane Publishing. Retrieved from http://elearning.ling.sinica.edu.tw/

CHOU, C. (2000). Constructing a computer-assisted testing evaluation system on the World Wide Web–The CATES experience. *IEEE Trans. on Education*, *43*(3), 266–272.

CHU, C. L., HWANG, R. H. & HSU, J. M. (2001). 全民题库系统：理念与建置，*2001 年台湾因特网研讨会论文集* (A proposition for national item bank: concepts and implementation, *Proc. of the 2001 TANET Conference*)，243–247。

Ezeiza, N., Maritxalar, M. & Schulze, M. (Eds.) (2007). Proc. of the Workshop on Natural Language Processing for Educational Resources. *Int'l Conf. on Recent Advances in Natural Language Processing.*

GAO, C.-M. (2007). Influence of contexts on vocabulary testing, *UCER*, *4*(7):42–46.

Heift, T. & Schulze, M. (2007). *Errors and Intelligence in Computer-Assisted Language Learning*. Routledge.

Heilman, M., Collins-Thompson, K., & Eskenazi, M. (2008). An analysis of statistical models and features for reading difficulty prediction, in (Tetreault et al., 2008), 71–79.

HSIAO, M. C., YANG, K. J., & CHANG, J. S. (2007). Word translation disambiguation via dependency. *Proc. of the 19th Conf. on Computational Linguistics and Speech Processing*, 145–159.

HUANG, S. M., LIU, C. L., & GAO, Z. M. (2005). Computer-assisted generation of listening cloze items and dictation practice in English. *LNCS*, (3583), 197–208.

JIA, J. (2004). The study of the application of a web-based chatbot system on the teaching of foreign languages. *Proc. of Society for Information Technology and Teacher Education Int'l Conf. 2004*, 1201–1207.

JUANG, D., WANG, J. H., LAI, C. Y., HSIEH, C. C., CHIEN, L. F., & HO, J. M. (2005). Resolving the unencoded character problem for Chinese digital libraries. *Proc. of the 5th ACM/IEEE Joint Conf. on Digital Libraries*, 311–319.

LI, Z. & SONG, M. (2007). The relationship between traditional English grammar teaching and communicative language teaching, *UCER*, *4*(1), 62–65

LIN, J. M. C., LEE, G. C. & CHEN, H. Y. (2004). Exploring potential uses of ICT in Chinese language arts instruction: Eight teachers' perspectives. *Computers & Education*, *42*(2), 133–148.

LIU, C. L., TIEN, K. W., LAI, M. H., CHUANG, Y. H. & WU, S. H. (2009). Capturing errors in written Chinese words. *47th ACL*, 25–28.

LIU, C. L. & LIN, J. H. (2008). Using structural information for identifying similar Chinese characters. *46th ACL*, 93–96.

LIU, C. L., WANG, C. H., & GAO, Z. M. (2005). Using lexical constraints for enhancing computer-generated multiple-choice cloze items. *IJCLCLP*, *10*(3), 303–328.

LU, M. S., WANG, Y. C., LIN, J. H., LIU, C. L., GAO, Z. M. & CHANG, C. Y. (2008). Supporting the translation and authoring of test items with techniques of natural language processing. *JACIII, 12*(3), 234–242.

Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. The MIT Press.

Metcalf, V. & Meurers, D. (2006). Generating web-based English preposition exercises from real-world texts. Presented in *EUROCALL*.

Sumita, E., Sugaya, F. & Yamamoto, S. (2005). *Measuring non-native speakers proficiency of English by using a test with automatically-generated fill-in-the-blank questions*. In: Burstein & Leacock, 61–68.

SUN, S. Z. (2008). Application of computer technology and web resources to assisting English learning. *UCER, 5*(4), 24–27. (in Chinese)

TIAN, J. Q., MIAO, D. M., ZHU, X. & GONH, J. J. (2007). An introduction to the computerized adaptive testing. *UCER, 4*(1), 72–81.

Tetreault, J., Burstein, J. & De Felice, R. (Eds.) (2008). Proc. of the Third Workshop on Innovative Use of NLP for Building Educational Applications. *The Forty Sixth Annual Meeting of the Association for Computational Linguistics*.

WANG, F. F. Y. (1966). Report on Chinese language concordances made by computer. *JCLTA, 1*(2), 73–76.

XU, J. & LIU, X. H. (2008). Application of study theories on digital Chinese teaching and learning to foreigners. *UCER, 5*(4), 9–15. (in Chinese)

ZHANG, D., LI, M., DUAN, N., LI, C. H. & ZHOU, M. (2008). Measure word generation for English-Chinese SMT Systems. *46th ACL*, 89–96.

ZHANG, H. Y. (2007). Computer-assisted elementary Chinese learning for American students. *UCER, 4*(5), 55–60.

ZHANG, Z. S. (1998). CALL for Chinese–Issues and practice. *JCLTA, 33*(1), 51–82.

(Edited by Nicole and Lily)