

# A Computer-assisted Evaluation System for On-line Testing of English-Chinese Translation

**Yan Tian**

**School of Foreign Languages,  
Department of Computer Science  
Shanghai Jiao Tong University  
Shanghai, 200240, P.R. China**

**Ruzhan Lu, Jianyong Duan**

**Department of Computer Science  
Shanghai Jiao Tong University  
Shanghai, 200240, P. R. China**

***Abstract:** It is a challenging task to evaluate English-Chinese Translating abilities of non-English majors on-line because it involves not only the processing of the source language—English, but also the target language—Chinese. The most difficult job is to teach the computer to “understand” the different kinds of translated versions and to judge how correct the students’ translated versions are, according to the standard versions. The on-line platform for this research is a self-developed on-line English-Chinese Translation Course. The materials used in this research are the pretest materials for this course. The method adopted is called “concept-equivalence-based keyword matching” through which we roughly achieved the goal of scoring the students’ translated versions on-line.*

**Keywords:** computer-assisted, evaluation system, on-line testing, English-Chinese translation

## 1.0 Introduction

Translating ability is regarded as one of the required foreign language skills in China. All non-English majors are required to master this skill before they graduate from college. Besides, this skill, mainly English-Chinese Translating is tested in the National College English Examination (both in Band IV and Band VI) which is held twice each year. What’s more, it is also a required examination item in the National English Examination for Admission to Graduate School. Since translation test is not an objective one, computer was not used to score the translation test papers. However, because of the time limitation imposed on the test paper scoring and the intensive manual labor for the scorers, the validity cannot be ideally guaranteed.

Another reason for conducting this research is that there is a growing demand for qualified translators in China, especially after the successful application for holding the 2008 Olympic Games in Beijing and 2010 Expos in Shanghai. Naturally, people turn to Web for help. And it is well known that a successful on-line course requires an instant feedback system to support. It proves easy to provide such systems to objective tests, but it would be extremely difficult to provide such system to the subjective tests because it involves the technology of natural language processing. Since translation involves two languages, this work turned out to be even more difficult.

The third reason is that it can help foreign language learners to learn Chinese on-line. Since more

and more foreigners want to learn Chinese, translating from English to Chinese is an effective way to test their Chinese proficiency. And on-line evaluation system for English-Chinese Translation can help them efficiently since Web is so convenient nowadays.

Unfortunately, up to now, we have not found any similar researches which can evaluate on-line English-Chinese translation in China, but we did find some relative reports on on-line assessment systems for subjective tests which only require using Chinese to answer questions.<sup>[3] [4]</sup> In these researches, the researchers calculated the similarities between the students' answers and the standard answers with the so called "single similarity degree", a concept from fuzzy mathematics. They regarded both the students' answers and the standard answers as word strings. Then they counted the words appeared in student's answers and divided them by the number of the words in standard answers.

## 2.0 Materials

To guarantee the validity and the reliability of our translation test papers, we adopted the authentic English-Chinese translation test papers of the National English Examination for Admission to Graduate School (from 1991 to 2005) and some model tests from books. The total number is 105 short English passages with an average length of about 400 words. Each test has 5 underlined sentences with about 150 words. The sentences are usually difficult for Chinese learners both in terms of structure and in terms of vocabulary. The topics of the passages range from natural sciences to social science. At present, our testing set is 525 English sentences, most of which are compound sentences.

To make our system practical, we developed an on-line English-Chinese Translation Course and linked it to College English Web Site (<http://english.sjtu.edu.cn>) of Shanghai Jiao Tong University. Any student of the university can register in this on-line course. The student ID is the registration number. The first step of taking this course is to take a pretest for their current English-Chinese translating abilities. And the pretest database is the above mentioned 105 short passages with 525 underlined English sentences.

## 3.0 Methods

Since we could not find any ideal parsers to analyze the English sentences to meet our needs, we tagged the 525 English sentences manually as follows:

For example: "Pearson has (1){*pieced together*} the work of hundreds of researchers around the world to produce (2)[AN][0]【 *a unique [NN][0] (millennium technology calendar)*】 that gives (3) [AN][0] (*the latest dates*) when we can expect hundreds of (4)[AN][0] (*key breakthroughs and discoveries*) to take place."

The 4 bolded chunks with numbers and brackets are the parts to be scored by the system. In front of each part are the types of the chunks. If the tagged are verbal phrases or prepositional phrases, no types are marked.

Standard version: "皮尔森汇集世界各地数百位研究人员的成果，编制了一个独特的新技术千年历，它列出了人们有望看到数百项重大突破和发现的最迟日期。"

The acceptable versions for each part:

“(1) 汇集；结合；整合；综合；(2) 一个独特的新技术千年历； 一个独一无二的科技千年日历；  
(3) 最迟日期； 最早日期；(4) 重大突破和发现,； 关键性突破和发现”

As shown, the number of the acceptable versions for each part is: (1): 4; (2): 2; (3): 2; (4): 2.

By concept equivalence we mean that the concepts revealed by English phrases can be expressed by many different Chinese equivalents. In this research, we focused mainly on noun phrases which we believe concept equivalence can be achieved most easily. Apart from noun phrases, we also tagged verb phrases and some prepositional phrases which we believe are the connectors of the noun phrases. To make it easy to assign scores to the tagged chunks in a sentence, we only tagged 4 chunks in a sentence and assigned each chunk to 5 scores, similar to the scoring practice adopted in the National English Examination for Admission to Graduate School. Therefore, the score for each translated sentence is 20 and the total score for a test paper is 100.

Then we designed 4 databases. The first one is the 105 passages for translation, with “serial number of a passage”(the sign to link the four databases), “topic field of the passage”(to make it easy for students to choose their familiar passages), “five sentences to be translated”, “5 standard versions”. The second is for each sentence, with “the serial number of a passage”(same as in the above database), “types of the tagged part of a sentence”( such as NN, AN, VN, etc.), “the standard versions of the four tagged chunks”, “the number of the acceptable versions for each part” (in principle, from 1 to numerous). The next database is for the 4 tagged chunks, with “serial number of a passage” (same as in the above database), “sentence number”, “ the number of the tagged chunks”( from 1 to 4); “the number of the acceptable versions”(at present only 4 at maximum), “the specific version of an acceptable version”. The last database is for students’ versions, with “a student’s identification number”, “serial number of a passage”(same as in the above database), “the serial number of a passage”, “5 standard versions of a sentence”.

Our on-line English-Chinese Translation course pretest system can generate a test paper automatically with the set time (elementary level: 20 minutes; intermediate level: 15 minutes; advanced level: 10 minutes, set according to the College English Curriculum Requirements) according to the level chosen by the student. Students must submit their versions; otherwise, the system will do so when time is up.

After submission, the students’ versions will be sent to the fourth database. Then the system will do its job. First it sets the student’s initial score to zero. Then it will match the student’s versions with the acceptable versions. The system will first look for the serial number of the passage, then locate the 5 sentences and the 4 chunks, then match them one part after another. The times of matching depend on the number of the acceptable versions. If the student’s version is matched with the first standard version, then 5 scores will be assigned and the system will go to the next part. If none is matched, instead of assigning zero to this part, the system will split the student’s versions into single characters. Then it will count the emerged characters according to the standard version of the part. If it finds some characters, it will divide the number of the characters by the total number of the standard versions. By doing this, the system avoids giving zero score to a certain part of the student’s version at maximum. This procedure will continue until all the sentences are done.

As soon as the evaluation is done (in only about 1 second), the student’s scores are presented on the Web. If the student wants to know the standard versions, he can click the button and view the 5 standard versions.

## 4.0 Results and Discussion

To test our system, we randomly chose two test papers as the testing samples, and invited 15 non-English major volunteers who have passed the National College English Examination (Band IV) to try our system. All of them finished the first test on time, but only 13 of them finished the second test on time. Their chosen level is intermediate. The results are as follows:

Number of Subjects	Average scores	Below 60 (%)	60-70marks (%)	70-80marks (%)	80-90marks (%)	90-100marks (%)
15	81.14	None	35.7	28.6	42.8	7.1
13	87.67	None	8.3	33.3	58.3	8.3

The above table shows that the average scores are not low; none of the students failed; the majority achieved quite high scores. These results indicate that concept-equivalence-based key matching is applicable. Although these scores are higher compared with the experts' evaluation, yet we believe they are acceptable because favorable scores will encourage on-line learners, if not too exaggerated.

## 5.0 Conclusion and Future Work

To investigate the possibility of evaluating English-Chinese Translation on-line is a pilot research. Although semantic approach would be the final key to Chinese language processing, a practical system has not gone beyond statistical approach. In nature, this system belongs to statistical category, yet it tries to calculate concept equivalents and has achieved favorable results which are quite exciting.

Our future work would be focused on the deep processing of the concept expressions of both English and Chinese. This includes shortening the tagged length of concept units because we found that our algorithm could deal very well with single concept units, but not do so with compound concept units, especially when several compound concepts coexisted in an expression. Simultaneously, we would try semantic approach by drawing NN concept diagrams from English and applying them to Chinese concept equivalents in standard versions.

## 6.0 Acknowledgement

This research has been sponsored by National Natural Science Foundation of China (NSFC: No.: 60496326) and 863 Project of China (No.: 2001 AA 114210-11).

## 7.0 References

- [1] Allen J. "Natural Language Understanding" (2<sup>nd</sup> ed.) The Benjamin/Cummings Publishing Company, 1995.
- [2] Sidan Gao, et al. "The Application of Sentence Similarity Measurement in Automated Assessment Technology of Subjective Tests" Computer Engineering and Application, 2004, (14): 132-135
- [3] Aiguo Meng, et al. "Design and Implementation of an Automated Assessment Algorithm for Subjective Tests in Network Examination System" Computer and Digital Engineering, 2005, 33 (7): 147-150