

Paper:

Supporting the Translation and Authoring of Test Items with Techniques of Natural Language Processing

Ming-Shin Lu*, Yu-Chun Wang**, Jen-Hsiang Lin*,
Chao-Lin Liu*, Zhao-Ming Gao**, and Chun-Yen Chang***

*National Chengchi University

E-mail: {g9407, g9429, chaolin}@cs.nccu.edu.tw

**National Taiwan University

E-mail: albyu35@gmail.com, zmgao@ntu.edu.tw

***National Taiwan Normal University

E-mail: changcy@ntnu.edu.tw

[Received April 21, 2007; accepted September 15, 2007]

Using techniques of natural language processing to assist the preparation of educational resources for language learning has become an important field. We report two software systems that are designed for assisting the tasks of test item translation and test item authoring. We built a software environment to help experts translate the test items for the Trends in International Mathematics and Science Study (TIMSS). Test items of TIMSS are prepared in American English and will be translated to traditional Chinese. We also built a software environment for composing test items for introductory Chinese courses. The system currently aids the preparation of four important categories of test items, and the resulting test items can be administrated on the Internet.

Keywords: natural language processing, computer assisted education, controlled languages, test item translation, test item writing

1. Introduction

Using techniques of natural language processing (NLP) to assist the preparation of educational resources for language learning has become an important field in real world applications [11, 20]. In this paper, we report two applications of NLP techniques to the preparation of education resources. The first application is to assist the translation of TIMSS [3] test items from English to Chinese. The second application is to assist teachers to prepare test items for introductory Chinese courses.

TIMSS aims at studying the achievements (in mathematics and science) of the fourth and the eighth grade students in participating countries with a common set of test items. The original test items are prepared in American English, and the test items are translated into local languages that are used in the participating countries. Although researchers may have different opinions about the validity of using translated items in tests [14, 27], trans-



Fig. 1. Main user interface for translating TIMSS items.

lated items offer a common base upon which we can compare students' performance. To ensure the quality of the translated test items, the translation of TIMSS test items is governed by a set of guidelines [19]. While the translation must consider the cultural differences of the participating countries [23], all translators must try to make the translation as close to the original items as possible. The bottom line is that the translation must maintain the challenge levels of test items so that the test results remain reliable for further studies. Hence, the translation must be reviewed carefully by the local TIMSS organisation and an international review committee.

Our system aims at helping translators abide by the translation guidelines as much as possible, while making the translation process more efficient. Fig. 1 shows the main user interface of the system. After the translators choose a test item in the upper left corner, the system shows the selected item in the upper right area, and recommends translations for the English words in the middle of the window. Translators can select and modify the recommended words, and change the orders of the selected words to make a complete test item at the bottom of the interface.

We also introduce an environment for preparing test items for students who are learning introductory Chinese. Chinese characters are hard to learn and remember. A typ-

ical way to test students' vocabulary is asking students to identify and correct wrong characters in a sentence. In addition, asking students to find a grammatical ordering of a set of shuffled words is a good way to practice Chinese grammars. We discuss two functions of our system: authoring of test items for character correction and sentence reconstruction.

We present the system for translating TIMSS items and its evaluation in Section 2, overview the design of the environment for preparing test items for introductory Chinese courses in Section 3, and make a brief conclusion in Section 4.

2. Translating TIMSS Items

In the current system, it takes three major steps to translate a test item, after we convert the TIMSS files from the Microsoft WORD format to pure texts with a JACOB service [10]. The translator first chooses a test item in English from an item set. Our system will look up the lexicon and provide a list of candidate Chinese translations for words and phrases in the selected item. The translator will then choose the best candidate translation for each word and phrase, and edit the selected sequence of translation into a Chinese test item. During this post-editing phase, the translator can add supplementary Chinese words that do not directly correspond to any English words or phrases in the original item. The translator may need to change the word orders to make the translation grammatically correct in Chinese, and the translator may also remove and/or modify the words that were chosen from the list of candidate translations.

It is undeniable that our system should attempt to recommend a Chinese translation which considers the change of word orders, and allows the translator to improve the recommended Chinese sequence. In order to offer this function, we need to have a sufficient number of translated TIMSS items to learn the correspondence between syntactic structures of English and Chinese items. However, we have only test times for TIMSS 1999 and TIMSS 2003, which include only hundreds of test items. A possible substitute is that we try to learn the syntactic correspondence between English and Chinese from other text. This is a more feasible approach and we are working in this direction.

2.1. Consistency in Translation

To achieve a high quality of the translated items, it is important to translate specific terms and phrases in a consistent way. These terms and phrases include "as shown below", "explain why", "one has been done for you" and many others. Every translator must use the same Chinese patterns for these specific phrases, according to the guidelines for all translators. Translations of units for weights and length as well as localisation of English names are taken care of too. Hence, our system must identify these special phrases for recommending appropriate translations.

In addition, there are occasions when translators will want to find how a term or a pattern of terms were previously translated in the TIMSS item bank. Knowing how the patterns were translated in the past years helps the translators maintain the consistency in the test items.

Hence our system will help translators find previous test items that contain specific word patterns. We achieve this by implementing a component that can recognise regular expressions, and apply a concordancer (cf. [18]), which aligns the queried terms, to present the previous test items to the translators.

2.2. Ordering Candidate Translations

Except the special patterns that we just discussed in Section 2.1, our system finds all candidate translations for the English words in the test item from the Concise Oxford English-Chinese Dictionary (OED) [4]. We employ MINIPAR [17] to locate some special patterns, and MXPOST [26], the Porter algorithm [25], and WordNet [4] to determine the part of speech of words and their root forms. Hence, each of the special patterns and individual words has a list of candidate translations.

Let E_1, E_2, \dots , and E_n represent the units, i.e., individual words or idioms, in an English sentence S_e . Let C_i denote the set of possible Chinese translations of E_i , and $C_i = \{C_{i,1}, C_{i,2}, \dots, C_{i,q(i)}\}$, where each $C_{i,j}$ represents a candidate translation of E_i , and E_i has $q(i)$ candidate translations. If E_i is a special term, that we explained in Section 2.1, we use the standardised translations for E_i . If not, we set up C_i for E_i with the OED. Let $C_{i,t(i)}$ denote a word that is selected from C_i . In the following subsections, we use S_c to represent a sequence of $C_{i,t(i)}$, $i = 1, 2, \dots, n$.

Given the lists of candidate translations, a translator can choose the best candidate for each word in a pull-down menu. Hence, placing more promising candidates at the tops of the menus facilitates the translator to find the best candidates easier. We would like to offer better orderings of the candidate words, but we will not try to solve this word sense disambiguation [18] problem in the system.

We consider four possible factors for ordering the candidate translations. We may record the frequency of a candidate Chinese translation being chosen for an English term, and prefer the one that has the highest frequency. We may look into relevant publications from which we obtain the relative frequency of a word being used. In addition to collecting word frequencies with monolingual corpora, more advanced NLP techniques [18] can be helpful. By aligning English words and their Chinese translations in parallel corpora, we can estimate the probability, $\Pr(C|E)$, of an English word, E , being translated into a particular Chinese word, C . We can also learn the n-gram statistics from the corpora with public domain software. To explore these possibilities, we employed GIZA++ [22] to learn $\Pr(C|E)$ and SRILM [28] to learn the bi-gram statistics with the Chinese-English bilingual version of Scientific American [8].

We discuss these factors next. Following the convention used in [18], we use superscripts and subscripts to