MANDARIN LEARNING USING SPEECH AND LANGUAGE TECHNOLOGIES: A TRANSLATION GAME IN THE TRAVEL DOMAIN

Yushi Xu, Stephanie Seneff

Spoken Language Systems Group, MIT Computer Science and Artificial Intelligence Laboratory

ABSTRACT

This paper describes a new Web-based translation game we have designed to help a student learn spoken Chinese. The student talks to the system in Chinese and the system compares the recognized sentence against a set of English prompts to judge whether it is a suitable translation of any one of them. The game can also provide translation assistance upon request. The game was developed using the IWSLT corpus of utterances in the tourist domain, and is oriented towards helping the student communicate effectively during foreign travel. In a preliminary evaluation, the system performed correctly on over 90% of test utterances. The system received positive feedback from the subjects.

Index Terms— Language Learning, Machine Translation, Language Understanding, Language Generation

1. INTRODUCTION

Second language learning is becoming a compelling activity to enhance people's skill set. While the demand for learning foreign languages is increasing, teaching methods for second language acquisition have changed little over decades. Learning mainly takes place in the classroom, where the teacher explains difficult grammar points and assigns written homework. Notably, the student is usually given little chance to practice speaking in the language. The severe shortage of tutors or partners to listen to students provides an opportunity for computer technology to help with this situation. A computer tutor will be affordable any time and anywhere. It is infinitely patient and never judgmental.

In this paper, we present a Web-based spoken translation game for beginning Mandarin learners. The game integrates various computer speech and language processing technologies to generate random English sentences, provide reference Chinese translations and automatically assess students' performance. It covers a relatively broad range of topics such as flight reservations, shopping, dining, city navigation, tourist information, weather, etc. The game curriculum is configured to gradually introduce new vocabulary and language constructs.

2. PREVIOUS RESEARCH

Two similar URL-accessible Web-based games have been developed previously, in the domains of flights [1] and hobbies [2]. The student records speech while holding down the record button. The sentences to be translated are generated from a set of templates. A parse-and-generate procedure produces the reference translations, which are compared with an N-best list of recognizer hypotheses generated from the student's speech input based on a key-value abstraction. Because both games were designed in very limited domains, the parsing, language generation and key-value extraction can be handled by relatively straightforward hand-coded rules. However, the limited domain is also a defect in terms of the scope of the language content.

Our aim is to help students learn to express themselves in Mandarin in a variety of realistic situations for a traveler. We are utilizing predominantly a linguistic approach, because it is fast, compact in memory usage, and easily controllable. But, with a broader domain, linguistic rules are less tractable to write manually. Thus, we exploit the power of statistical methods to help in rule construction, and to select among alternative strings produced by the generator.

3. THE GAME

The game starts with a login page, in which the student can choose the difficulty level and the number of sentences in each round. After log-in, the main game interface is presented, as shown in Figure 1. The student can hold down the green button on the top to speak the translation of any sentence in the presented task list. The box below the green button displays the recognized utterance, as well as a Chinese paraphrase and an English translation. If the translation is judged to match an item in the list, the system marks that sentence in red. After all the sentences are cleared in this manner, the system assesses the student's overall performance, and decides whether to advance by one lesson, drop by one lesson, or remain at the same difficulty level.

At any time, the student can click the button to the left of any sentence to access help. A reference translation is displayed on the Webpage, along with some hints on relevant aspects of Chinese syntax, and a robotic tutor speaks the correct translation for the student to imitate.

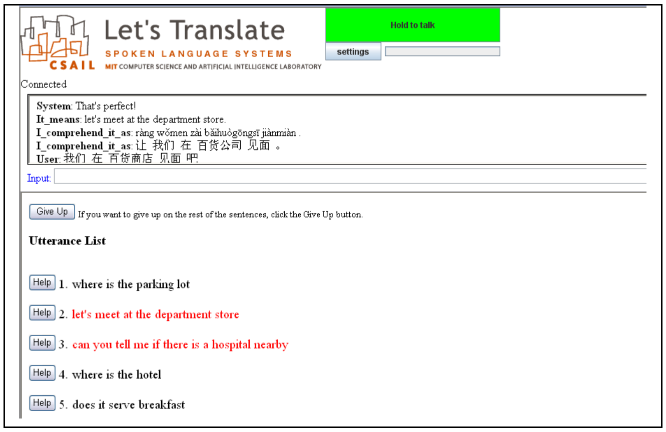

Figure 1. Game interface after login

- **Reading Exercises**

While the game was originally intended for the task of translation from English to Chinese, it can easily be reconfigured as a reading exercise to test the student's knowledge of Chinese characters or simply to test their fluency in speaking out loud a sentence presented in Pinyin format.

We thus designed three distinct prompt modes for students to choose from: Chinese Pinyin, Chinese characters, and English. In the first two modes, the student is expected to read out loud the sentences displayed on the screen. These two modes test whether the student knows the characters, and how well they can pronounce the sentences. Since the system speaks an English translation of the sentence, the student also gets feedback on the meaning of the sentence.

- **Typed Inputs**

Whenever the environment is too noisy for the speech recognizer to work well, or if the recognizer is having difficulty due to the student's poor pronunciations, the student can use the input box to *type in* a translation using Pinyin. Since Chinese tone is difficult for English speakers, the system provides automatic tone correction for the typed Pinyin inputs. The student can see the corrected Pinyin string, as well as an automatically proposed character string in the dialogue box. If the student types English into the input box, the system translates their English string into Chinese. This is especially helpful when the student has forgotten a particular vocabulary word or a short phrase.

4. THE TECHNOLOGY COMPONENTS

4.1. Overview

The game system is composed of several components. A central hub manages communication among the servers, based on a hub program script. Services include speech recognition, language understanding and generation, meaning comparison, speech synthesis, and the graphical user interface, which also handles audio capture. Figure 2 provides a flow chart of system processing steps.

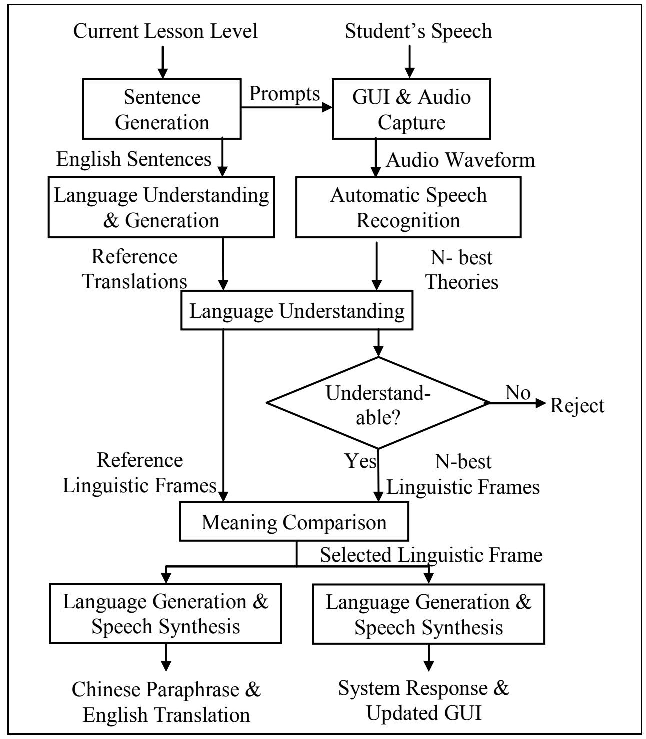

Figure 2. Flow chart illustrating system processing steps.

4.2. Generating Game Sentences

Each game sentence is randomly generated from templates. Each template encodes the sentence pattern recursively, using context-free rewrite rules. The patterns are organized according to topics to form individual lessons, following the conventions of mainstream textbooks. The lessons are blended such that each game episode has on average 50% new material and 50% review material, drawn from all the preceding lessons using a fading algorithm. Thus, the student is gradually exposed to new material while continually reviewing previous material for reinforcement.

The manually composed templates are written in English in a simple format, and can be augmented by anyone with only a little training. This would allow language teachers, or even students, to design their own personalized games.

4.3. Speech Recognition and Synthesis

For recognition we used the SUMMIT [3] landmark-based recognizer, trained on native speech data. The recognizer language model was trained on a mixture corpus of the template-generated sentences and the pre-tokenized IWSLT[1] 2006 data, so that the recognizer is biased towards what the student is expected to learn, but still covers a lot of other common vocabulary and sentence patterns. A Dectalk synthesizer is used to output English speech, and a synthesizer developed by ITRI in Taiwan handles the Chinese speech.

4.4. Language Understanding and Generation

Language understanding and generation are the most important components in the system. In order to deal with the relatively broad travel domain, they are designed to be general-purpose. The main approach is linguistic and rule-based, with statistical methods used selectively to provide more robustness. We have emphasized efficiency such that the system can produce high-quality results at very high speed and very low memory cost, with a future goal in mind of running the entire system on a handheld device.

As in the previous game systems, we use TINA [4] for language understanding and GENESIS [5, 6] for language generation. TINA produces a parse tree and successively a linguistic frame from a string, and GENESIS takes in a linguistic frame and generates a string. Both work from a set of linguistic rules. Since our group had already developed fairly high coverage generic English grammar and generation rules, we concentrated on developing generic Chinese understanding and generation. Since Chinese is a left-recursive language, many modifying phrases and clauses precede the constituent they modify. TINA's top-down parsing mechanism is thus forced to hypothesize several redundant theories which are later rejected. We initially tried to adapt TINA's trace mechanism to delay construction of higher levels of the parse tree until after a subsequent word validated them. However, this led to inefficiencies due to the relatively costly trace computation. After comparing parse trees before and after movement of the traced constituent, we realized that simple surgery could be performed on the parse tree post-hoc to achieve the same result. This led to a significant increase in parse speed and a more transparent grammar formalism. We also created new "parasite" and "enemy" features to deal with cross-level long-distance constraints that occur frequently in Chinese [7, 8].

We utilize statistical methods to aid in the language generation process in two distinct ways, one in a lexicon-development phase for Chinese generation, and the other during the English generation task of game play. To develop the Chinese generation rules, we began by mutating the constituent ordering rules of the English generator, to produce a string in an intermediate language we call "Zhonglish" (English words, Chinese word order). We then applied GIZA alignments to a Zhonglish/Chinese parallel corpus from the IWSLT domain to produce an initial version of an English-to-Chinese mapping lexicon, which was then manually edited. By instantiating this lexicon, we produced an initial version of the Chinese string generator. Semantic features have subsequently been added, based on the linguistic frame context, to help disambiguate among multiple word senses. This is an ongoing process.

The Chinese grammar generates a linguistic frame that is largely parallel to the one generated by an equivalent English string. However, there are many ambiguities and missing elements due to the differences between the two languages. We developed a powerful formalism for expressing frame restructuring rules that can augment and transform the linguistic frame for English specific aspects. For example, one rule specifies the addition of the missing auxiliary "do" for wh-questions containing a verbal predicate. We also added an extremely useful mechanism for expressing alternatives under conditions of uncertainty, to later be resolved using an *n*-gram language model. For example, in the Chinese to English translation situation, "多少 钱" will be translated into "(how many|how much) money". An *n*-gram language model is then applied to select the most confident path from the word graph.

Another new feature in this game system is the hints. We try to provide some grammar or vocabulary points for the student that are related to the specified sentence. The process of generating accurate Chinese necessarily involves several flags to help GENESIS choose the correct rules and correct lexicon mappings. These flags encode useful information to help the students as well, and therefore we added rules to produce contextualized hints triggered by these flags, in parallel with the generation of the translated string.

4.5. Meaning Comparison

The meaning comparison is done using a key-value approach. From the linguistic frame, we extract key semantic information such as agent, patient, temporal, locative, etc., and use GENESIS to generate key-value pairs. Then we compare the two lists of key-value pairs, one from the reference translation, and the other from the student's translation. In order to make the judgment more flexible, we collapse some synonyms into classes, e.g. adjectives "great", "excellent", "wonderful" are all mapped to "good". Also, since Chinese has a rather flexible grammar, we devised a scheme to handle alternatives. For example, the topic of some sentences can be omitted, as in "它 多少 钱？". In this case, we use the schema "it|*NONE*" to express the two variants.

4.6. Tone Correction

For typed Pinyin inputs, the system offers automatic tone correction. This is treated as a recognition problem. We first generalize the student's input to all the possible tones, and then build a Finite State Transducer (FST) to represent all the alternatives in the graph. This FST is composed with two FSTs, representing Pinyin-to-character mappings and the Chinese trigram language model in characters. The best scoring path contains the proposed Chinese characters. The

[1] International Workshop for Spoken Language Translation

Pinyin mapping of these characters is compared to the student's input, and the ones that differ are marked in red, to indicate likely mistakes.

5. SYSTEM REFINEMENT AND EVALUATION

5.1. System Refinement

The system produces 100% accurate reference translations for the template sentences. However, in order to make sure the system accepts/rejects other possible inputs from student, a refinement process was performed prior to releasing the system for data collection. We asked a native speaker of Chinese to generate a set of utterance templates reflecting several different ways to translate the English templates in our system. From these, we automatically generated 2,300 unique sentences, which were then used to guide refinement of system performance. We processed the sentences through the following three steps: (1) Parse and translate into English, (2) Parse the English sentences obtained in (1), and translate back into Chinese, and (3) Compare the result from (2) and the original Chinese input in terms of a [key: value] matching criterion.

This process can test the accuracy of both Chinese-English and English-Chinese translation, as well as the meaning comparison part. The outcome of Step (2) should be equivalent in meaning to the original input, and thus Step (3) should yield all perfect matches. In the first run, the Chinese parser failed on 7.7% sentences. However, nearly 1/3 of the failures were a consequence of ill-formed templates, so failure is a correct outcome. English translations produced from 87.8% of the parsable sentences were parsable.

Refinements and modifications were then applied to the grammar and generation rules to finally obtain 100% parse coverage for Step (1), 99% parse coverage for Step (2) , and a perfect match rate of 90% in Step (3). Subjectless Chinese sentences like "喜欢 足球" are the main source of error. When these are translated into English, a subject "I" is added automatically, causing an agent key insertion.

5.2. Preliminary Evaluation

We did a preliminary evaluation on 5 subjects whose Chinese proficiencies vary from novice to native. We asked them to talk to the system, and monitored how well the system worked. We collected 615 utterances from these subjects. Among these utterances, 8.6% were falsely rejected. Most of these rejections could be attributed to poorly pronounced words. The false acceptance rate was less than 1%. We noticed that the falsely-accepted utterances only contained very small grammatical problems, e.g. a wrong measure word, which are typically corrected in the Chinese paraphrase spoken out by the system. Thus, the system offers implicit error correction.

All the subjects gave positive opinions on the system. On average, the subjects played 18 rounds, with 5 sentences per round. They asked for help about 1.4 times per round, and took about 2 rounds on average to advance to the next level. They felt that the system is encouraging, and the gradual introduction of new vocabulary and sentence patterns is a great feature of the system. They felt they indeed learned something after using the system.

6. CONCLUSIONS AND FUTURE WORK

We have designed a translation game system for Mandarin learning. The system utilizes speech recognition and language processing technologies to automatically give feedback to the student and provide help. We use language understanding and generation methods, supported by statistical methods to enhance performance and robustness. The preliminary evaluation showed that the system is reliable and useful. For future work, we plan to make the game system available to the public through the Internet and to collect more data. We also plan to port the system to other domains of knowledge.

7. ACKNOWLEDGEMENT

This research has been supported by the Delta Environmental and Educational Foundation. We would like to thank the Industrial Technology Research Institute for supplying the Chinese synthesizer. We would also like to thank the subjects who helped us conduct the preliminary evaluation.

8. REFERENCES

[1] C. Wang and S. Seneff. A Spoken Translation Game for Second Language Learning. In *Proc. of AIED*, Marina del Rey, California, 2007

[2] C. Chao, S. Seneff, and C. Wang. An Interactive Interpretation Game for Learning Chinese. In *Proc. of the SLaTE Workshop*, Farmington, Pennsylvania, 2007.

[3] J. R. Glass. A probabilistic framework for segment-based speech recognition. *Computer Speech and Language*, Vol. 17, No. 2-3, pp. 137-152, April/July 2003.

[4] S. Seneff. TINA: A Natural Language System for Spoken Language Applications. *Computational Linguistics*, Vol. 18, No. 1, 1992.

[5] L. Baptist and S. Seneff. Genesis-II: A Versatile System for Language Generation in Conversational System Applications. In *Proc. of ICSLP*, Beijing, China, 2000.

[6] B. A. Cowan. PLUTO: A Preprocessor for Multilingual Spoken Language Generation. Master's Thesis, MIT, 2004.

[7] Y. Xu and S. Seneff. Combining Linguistics and Statistics for High-Quality Limited Domain English-Chinese Machine Translation. Master's Thesis, MIT, 2008.

[8] Y. Xu, J. Liu and S. Seneff. Mandarin Language Understanding in Dialogue Context. To Appear in *ISCSLP* 2008.