An Interactive Interpretation Game for Learning Chinese

Chih-yu Chao, Stephanie Seneff, and Chao Wang

Spoken Language Systems Group, Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology, Cambridge, MA, USA
{chihyu, seneff, wangc}@csail.mit.edu

Abstract

In this paper, we present an interactive interpretation game for learning Chinese. We extend our previous work on a flight domain translation game by introducing a new topic that is more appropriate for language learners. We discuss new features that have been added to the existing translation game system. We also report results from a pilot study to evaluate if the game helps learners improve their ability to speak the target language.

Index Terms: dialogue system, second language acquisition, computer-assisted language learning, translation

1. Introduction

It is widely agreed that the best way to learn to speak a language is to engage in natural conversation with native speakers [1]. However, this is very costly. Although various off-the-shelf applications/tools assist language learners, very few of them provide the learners with interactive speaking exercises.

In a traditional self-taught setting, learners may listen to the audio, watch the video, or simply read the text material. To practice speaking, they typically compare their own speech with the native speech they hear. In a classroom setting, the instructor will give feedback and correct the students' mistakes; however, given the number of students in a typical class, it is impossible for any instructor to pay full attention to every student.

Table 1 compares a computer tutor with a human tutor. Most notably, a computer system is affordable, available and accessible any time, infinitely patient, and it pays full attention to each user and provides immediate feedback.

Our language learning game system focuses on the scenario of a native speaker of English learning Mandarin Chinese, extending previous work – a web-based spoken translation game serving as a preparation for dialogue interactions of booking flights [2]. The new content of activity scheduling was designed based on a lesson in a beginning-level Chinese textbook [3] used in high schools. A high quality domain-specific speech synthesizer was employed for system output. A pre- and post-test approach was incorporated to evaluate the helpfulness of the game system.

2. The system

The domain of the system is related to activity scheduling. We envision that after users master the vocabulary and sentence structures by playing the game, they will be able to schedule an activity with a human interlocutor, or with a computer system designed to simulate a human conversational partner [4]. Example sentences involved in a relevant dialogue may include: "I am playing basketball Tuesday evening; would you like to join me?", "Playing basketball is not bad, but I prefer dancing.", and so on.

2.1. Overview of the game

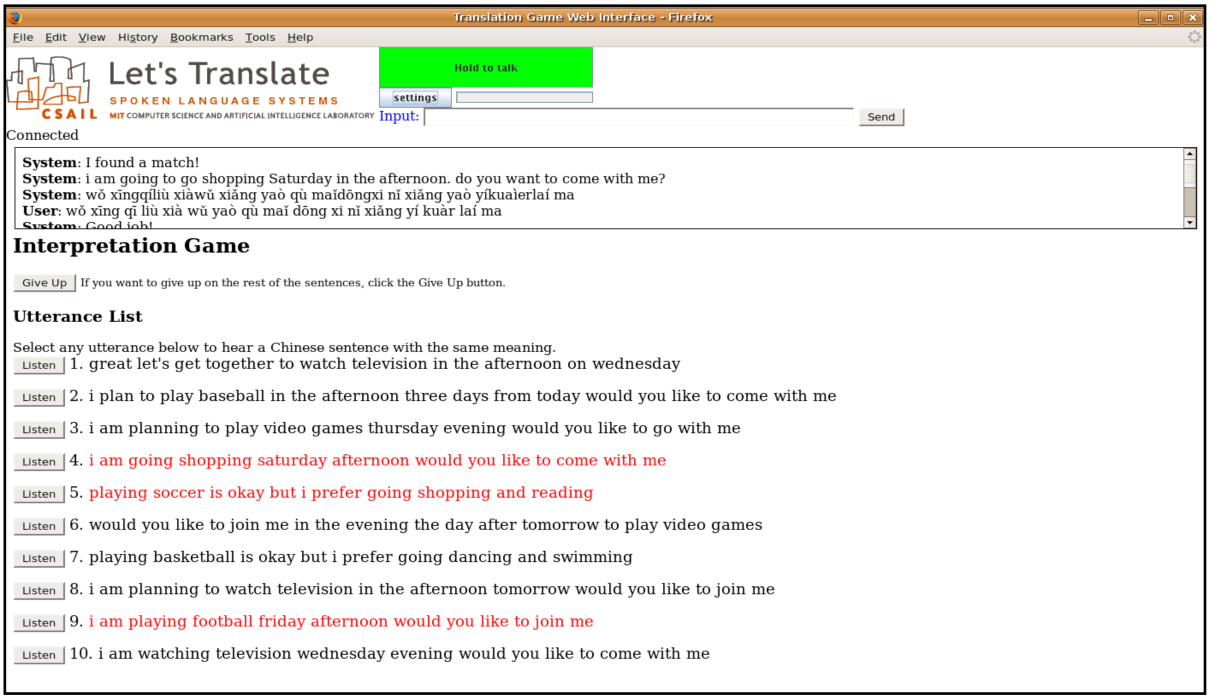
The game design was motivated by the learning approach advocated by Pimsleur [5]. By practicing translation repeatedly, language learners are able to internalize the structures of the target language, and thus the vocabulary, grammar rules, and pronunciation are practiced concurrently. In our game there are five difficulty levels. The user begins at Level 1 by translating isolated vocabulary items, advancing to phrases and short sentences at higher levels. The most difficult level, Level 5, involves relatively long and complicated sentences.

Table 1. *Comparison between a human tutor and a computer-based tutoring system on a number of dimensions.*

	Computer	**Human Tutor**
Availability	24/7	Limited
Attention to learner	100%	Quite low, except for 1-on-1 tutoring
Cost	A computer with a browser and an Internet connection	1-on-1: very high
Stress level	Difficulties with adjusting the microphone gain; recognition errors	Embarrassment from speaking in front of the class
Pronunciation / Intonation	Concatenative speech synthesis, can be quite natural	Good, natural

When the users complete a session, a score is calculated based on the average number of turns they took to successfully translate each utterance, and how frequently they asked for help. The difficulty level is automatically adjusted up or down depending on the monitored score.

Figure 1 shows a screenshot of the game interface. The system presents a list of randomly generated utterances in English, and the user is tasked with translating them into Chinese. The user can speak a translation of one of the sentences (in any order), while holding down the "Hold to Talk" button. If they need help, they may click the "Listen" button to hear a system-provided translation of any utterance. Alternatively, they may type English in the text area, and the system will attempt to automatically translate the typed utterance and speak the translation.


Translation Game Web Interface - Firefox
File Edit View History Bookmarks Tools Help
Let's Translate
SPOKEN LANGUAGE SYSTEMS
CSAIL
MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY
Hold to talk
settings
Input:
Send
Give Up
Listen

Connected

System: I found a match!
System: i am going to go shopping Saturday in the afternoon. do you want to come with me?
System: wǒ xīngqīliù xiàwǔ xiǎng yaò qù maǐdōngxi nǐ xiǎng yaò yíkuaìerlaí ma
User: wǒ xīng qī liù xià wǔ yaò qù maǐ dōng xi nǐ xiǎng yí kuàr laí ma
System: Good job!

Interpretation Game

If you want to give up on the rest of the sentences, click the Give Up button.

Utterance List

Select any utterance below to hear a Chinese sentence with the same meaning.

1. great let's get together to watch television in the afternoon on wednesday
2. i plan to play baseball in the afternoon three days from today would you like to come with me
3. i am planning to play video games thursday evening would you like to go with me
4. i am going shopping saturday afternoon would you like to come with me
5. playing soccer is okay but i prefer going shopping and reading
6. would you like to join me in the evening the day after tomorrow to play video games
7. playing basketball is okay but i prefer going dancing and swimming
8. i am planning to watch television in the afternoon tomorrow would you like to join me
9. i am playing football friday afternoon would you like to join me
10. i am watching television wednesday evening would you like to come with me

Figure 1 *The interpretation game interface. A list of (10) translation tasks is presented to the user. Each "Listen" button on the left produces a suggested Chinese translation of the task on the right. The feedback from the system and the paraphrased translation utterances are displayed on the top. The user may directly speak to the system by holding down the green "Hold to Talk" button. Once an utterance is translated correctly, the text turns red. The user may also type English in the text area beneath the talk button to get help. If the user is not confident enough on translating the current tasks, clicking the "Give Up" button resets the game with a new set of utterances.*

2.2. Technical components

Each user utterance is processed through several steps in order to determine whether it is a translation of one of the utterances in the list. The system also paraphrases each sentence into both Chinese and English and speaks both paraphrases via a speech synthesizer to provide feedback to the user.

Each user utterance is first processed through a segment-based speech recognition system [7], where the acoustic models were trained on Mandarin Chinese speech data from native speakers [8][9]. In the acoustic models, the tone features were ignored, but some tone constraints were captured implicitly in the language model. The language model was trained on Chinese translations of the English sentences used in the game, augmented with additional sentence patterns derived from a pilot data collection effort. The recognizer produces an N-best list of hypotheses for further processing.

To perform the translation evaluation, the N-best list is converted into N-best key-value representations of the meaning via a parsing [10] and generation [11] paradigm. These are then compared against the set of key-value representations previously computed for each of the utterances in the task list. If any item in the N-best list matches any candidate task item, the system chooses that candidate and considers it a successful match. It congratulates the user and changes the color of the matching utterance. Because the match is performed at the key-value level, multiple variants of the Chinese translation are accepted. If none matches, it chooses the top hypothesis that parses. The generation system is also used to generate Chinese and English paraphrases of the selected hypothesis.

It is important to produce high quality speech synthesis for language learners. Anecdotally, users of the flight domain game system [2] did not like the quality of speech produced by the general-purpose synthesizer. To improve the synthesis quality, in our current system we utilize the ENVOICE synthesizer [12], a concatenative text-to-speech engine with a scalable finite-state transducer implementation for unit selection. The costs of concatenation and substitution are calculated based on local phonetic context. Phones that exhibit similar concatenation behavior are grouped into classes and share the same concatenation and substitution costs. When combined with a domain-specific corpus, as is the case in our system, the synthesizer produces very natural-sounding speech.

3. Experimental design

It is difficult to quantify how much users would learn from interacting with the system. Therefore, a pilot study was conducted with a balanced pre-test/post-test design, to quantify user behavior and understand the utility of the game.

3.1. Subjects

The pilot study was conducted with 7 volunteer subjects, all of whom are students from a local university. Four subjects have taken formal Chinese lessons at school for 1 to 3 years; one subject was self-taught for 9 months and studied formally with a tutor while staying in China; two subjects have family members speaking Chinese, as well as one year of college-level Chinese lesson. Table 2 shows the self-assessed Chinese proficiency of the subject pool. All subjects were asked to fill out a survey [13] after they completed the experiment.

3.2. Experimental procedures

Before each subject started playing the game, a pre-test was presented (disguised as a warm-up exercise), displaying five translation tasks manually selected from the most difficult level. The interface was exactly the same as in the actual game. The subject could make use of all the help features in the system (*i.e.* the "Listen" button and the type-to-translate text area), as well as pencil and paper, to complete these tasks. Then the actual game started at Level 1 with 5 English sentences to translate for each session. The game proceeded as described in the previous section.

After Level 5 was completed, the subject was asked to fill out a survey about the game, as well as to provide suggestions for improvement. After the survey, there was a post-test (disguised as another session that helped the experimenter collect more data) with the same interface design. The post-test consisted of a different set of five translation tasks manually selected from the most difficult level, and the subject could still use all forms of help if necessary.

Table 2 *Self-rated Chinese proficiency of the subject pool, in terms of reading, writing, speaking, and listening. The proficiency levels range from none, poor, fair, fluent, to native. Subject 3 and subject 7 have family members speaking Chinese.*

	Reading	**Writing**	**Speaking**	**Listening**
subj1	poor	fair	fair	fair
subj2	fair	poor	fair	fair
subj3	poor	poor	fair	native
subj4	poor	poor	poor	fair
subj5	poor	poor	poor	fair
subj6	poor	Poor	poor	poor
subj7	none	Poor	poor	fluent

4. Results and discussions

The language model was initially trained on sentences automatically generated from a template, and thus the coverage on the vocabulary entries and sentence structures was inadequate. After the first few pilot study sessions, the templates were augmented with patterns derived from user input to improve system coverage.

It was our intention to motivate users by showing a score after the completion of each session. It was observed that users would accidentally click the "Hold to Talk" button, which would be counted as an unsuccessful turn. Some subjects figured out that clicking the "Listen" button would deduct points from the score, but that they could instead type the sentence into the text area and avoid deductions. These user behaviors provided guidelines for future system improvement.

From the pilot study, we would like to see how each subject interacted with the system, how much the subjects liked the game, and whether the system helped subjects learn spoken Chinese.

4.1. Observation from the experiment

Figure 2 shows the amount of interaction each subject had with the system when completing all 5 levels of the game. The interaction was measured by the number of spoken turns. Given that the experiment, excluding pre-test and post-test, lasted for 45-60 minutes, the interaction between the subjects and the system was extensive.

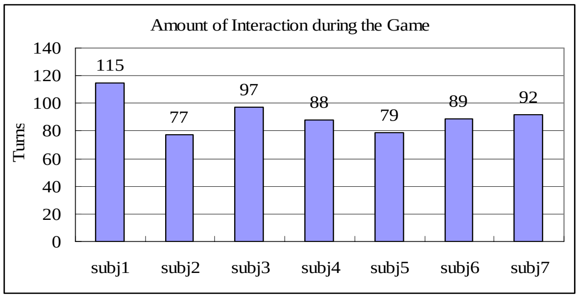


Figure 2 *The amount of interaction (i.e. number of spoken turns) with the system over a 45- to 60-minute game session.*

Figure 3 shows the average time-on-task for each subject during pre-test and post-test. During pre-test, subjects spent 2 to 6 minutes to correctly translate each utterance, while during post-test, it only took them 1/3 to 1/4 of the time to complete a task of the same difficulty.

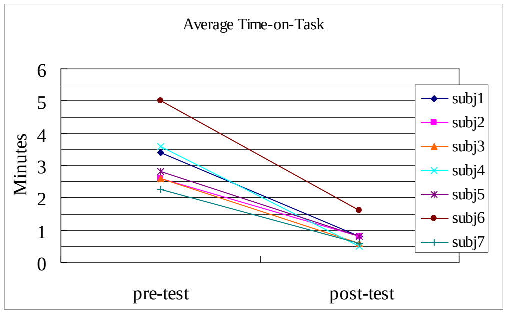


Figure 3 *Comparison of pre-test and post-test time-on-task.*

One may argue that the decrease in time was due to the familiarity with the game interface. However, it was observed that during pre-test, subjects spent a lot of time constructing translated utterances with various levels of assistance, while during post-test, all subjects started speaking to the system immediately after reading the English sentences to be translated, nearly without any form of help. Figure 4 shows the number of times the subjects asked for help during the two tests. Also, we observed that there was no "parroting" (*i.e.* speaking right after clicking the "Listen" button) during the post-test.

4.2. Survey responses

Subjects were asked how they thought of the difficulty of the game. On a Likert scale of 1 to 5 (1: very easy; 5: very difficult), the average perceived difficulty was 3.29 (std. dev.= 0.76). A subject responded with "it was just above my level and it was good to stretch my skills". As for how fun it was to play the game (1: not at all; 5: absolutely enjoyable), the average rating was 3.43 (std. dev.= 0.98). As one subject pointed out, "I really enjoyed getting phrases right. I like how it built my confidence translating sentences".

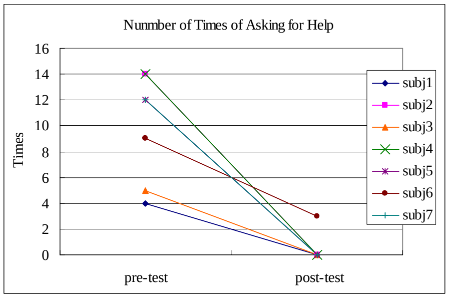


Figure 4 *The frequency of each subject clicking the "Listen" button or typing English text during pre-test and post-test.*

Six out of seven subjects found the game helpful and would like to play it again in the future, because "(the game) forced me to use correct pronunciation; my Chinese teacher usually allows small mispronunciations to slip by", "hearing the thing I said spoken back properly makes me improve my pronunciation", and "I usually get more practice translating English to written Mandarin instead of spoken Mandarin". Though one subject felt that the content was repetitive, others thought that "vocabulary is good to remember; repeating similar sentences reinforces common grammatical patterns".

Six out of seven subjects would recommend the game system to friends, because "books don't suffice!", "it's a good way to get speaking practice; otherwise you need a fluent friend who is patient", and "it was fun and interactive, and better than anything else". Meanwhile, the subjects also pointed out it would be better if one did not have to correctly translate every part of an utterance all at once. In fact, we are planning to adopt the "partial match" scheme from [6], so that future users will be able to complete a complex translation task cumulatively. All subjects were satisfied with the synthesized speech, but some did notice the recognition issue. We have so far collected 854 non-native utterances from the pilot study; by collecting more non-native speech data in the future, we will be able to improve the acoustic model with the approach used in [14] and [15].

5. Conclusion and future work

We have introduced an interactive interpretation game based on a topic domain more suitable for beginning students and with a higher quality speech synthesizer than our previous flight domain translation game. The pilot study showed that the system provided users with opportunities to practice spoken Chinese and with appropriate feedback. It also built up users' confidence in speaking Chinese. We plan to follow up with a later post-test to assess long-term retention.

Observing user behavior has helped us identify future directions. In addition to expanding the language model and acoustic model with more non-native speech data, we will also redesign the scoring algorithm to prevent users from gaming the system.

We are currently developing a dialogue interaction game [4], in which the user negotiates with a "virtual buddy" to find a time to meet in the near future to jointly participate in a task that both parties find enjoyable. We hope to conduct experiments shortly with the same subjects, to allow them to practice communication with this system.

6. Acknowledgements

This research is supported in part by ITRI and the Cambridge MIT Initiative. The authors would like to thank Scott Cyphers and Alexander Gruenstein for rectifying many system and server issues. We would also like to thank Anna Goldie, Ian McGraw, and Michael Cohen for preliminary system testing and helpful suggestions.

7. References

[1] Ehsani, F. and Knodt, E., "Speech Technology in Computer-Aided Learning: Strengths and Limitations of a New CALL Paradigm", Language Learning and Technology, 2(1): 45-60, 1998.

[2] Wang, C. and Seneff, S., "A Spoken Translation Game for Second Language Learning", to appear in Proc. Artificial Intelligence in Education, 2007.

[3] Yao, T., Liu, Y., Ge, L., Chen, Y., Bi, N., Wang, X., and Shi, Y., Integrated Chinese, Level 1, Part 1, 2nd Ed., Cheng & Tsui Company, Boston, 2005.

[4] Seneff, S., "Interactive Computer Aids for Acquiring Proficiency in Mandarin", keynote speech in Proc. of ISCSLP, pp. 1-11, Singapore, 2006.

[5] Pimsleur, P., "A Memory Schedule", Modern Language Journal, 51(2): 73-75, 1967.

[6] Wang, C. and Seneff, S., "Automatic Assessment of Student Translations for Foreign Language Tutoring", Proc. HLT-NAACL 2007, pp. 468-475, Rochester, NY, USA.

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

[8] Wang, C., Cyphers, D. S., Mou, X., Polifroni, J., Seneff, S., Yi, J., and Zue, V., "Muxing: A telephone-access mandarin conversational system", Proc. ICSLP, Vol. II, pp. 715-718, Beijing, China, 2000.

[9] Wang, H. C., Seide, F., Tseng, C. Y., Lee, L.S., "MAT2000-Design, collection, and validation of a Mandarin 2000-speaker telephone speech database", Proc. ICSLP, pp. 460-463, Beijing, China, 2000.

[10] S. Seneff, "TINA: A natural language system for spoken language applications", Computational Linguistics, 18(1): 61-86, 1992.

[11] Baptist, L. and Seneff, S., "Genesis-II: A versatile system for language generation in conversational system applications", Proc. ICSLP, Vol. III, pp. 271-274, Beijing, China, 2000.

[12] Yi, J., Glass, J., and Hetherington, L., "A flexible scalable finite-state transducer architecture for corpus-based concatenative speech synthesis", Proc. ICSLP, Vol. III, pp. 322-325, 2000.

[13] Located at http://people.csail.mit.edu/chihyu/survey.cgi

[14] Raux, A., Langner, B., Black, A., Eskenazi, M., "LET'S GO: Improving Spoken Dialog Systems for the Elderly and Non-natives", Eurospeech 2003, Geneva, Switzerland.

[15] Raux, A. and Eskenazi, M., "Non-Native Users in the Let's Go!! Spoken Dialogue System: Dealing with Linguistic Mismatch", HLT-NAACL 2004, pp. 217-224, Boston, MA, USA.