Web-based Dialogue and Translation Games for Spoken Language Learning

Stephanie Seneff

MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar Street, Cambridge, MA 02139
seneff@csail.mit.edu

Abstract

It is widely recognized that one of the best ways to learn a foreign language is through spoken dialogue with a native speaker. However, this is not a practical method in the classroom due to the one-to-one student/teacher ratio it implies. A potential solution to this problem is to rely on computer spoken dialogue systems to role play a conversational partner. This paper describes several multilingual dialogue systems specifically designed to address this need. Students can engage in dialogue with the computer either over the telephone or through audio/typed input at a Web page. Several different domains are being developed, in which a student's conversational interaction is assisted by a software agent functioning as a "tutor" which can provide them with translation assistance at any time. Some of the research issues surrounding high-quality spoken language translation and dialogue modeling for language games are discussed.

1. Introduction

It is widely agreed among educators that the best way to learn to speak a foreign language is to engage in natural conversation with a native speaker of the language. Yet this is also one of the most costly ways to teach a language, due to the inherently one-to-one student-teacher ratio that it implies.

Mandarin Chinese is one of the most difficult languages for a native English speaker to learn. Chinese is substantially more difficult to master than the traditional European languages currently being taught in America – French, Spanish, German, etc., because of the lack of common roots in the vocabulary, the novel tonal and writing systems, and the distinctly different syntactic structure.

With the rapid emergence of China as a major player in the global economy, there is an increased urgency to accelerate the pace at which non-native speakers can acquire proficiency in communicating in Chinese. There is a severe shortage of language educators in Western nations who speak both English and Chinese fluently. Computers can offer a solution to this problem, both by engaging the student in one-on-one spoken conversation, where the computer role plays the conversational partner, and by providing translation assistance when needed to help the student formulate their half of the conversation. Conversations will ultimately support a wide range of topics and will likely be goal directed, to help hold the student's interest and focus their attention. These conversations need not be speech only, but instead could incorporate a display component, ranging from an avatar to embody the voice to an entire video-game-like environment [12, 13].

The explosive expansion of computer usage in households around the world in the last decade is rapidly morphing into the widespread adoption of computers and personal digital assistants (PDA's) as devices for access to remote computational and information resources. Computers, via Voice over IP (VOIP), are also beginning to replace the land line and cellular telephone systems as an alternative way for humans to remotely communicate among one another. *Computer Aided Language Learning* (CALL) systems will be able to take advantage of the widespread availability of high data rate communications networks to support easy accessability to systems operating at remote sites. The student can just enter a Web page, where they would be able to type or speak to the system, with the system responding through displays and synthetic speech, supported by multimodal WIMP-based (Window, Icon, Menu, Pointing) interaction.

Clearly, for this vision to become a reality, a considerable amount of research is necessary. While significant progress has been made on human language technologies, it is not clear that the technology is sufficiently mature to succeed in enticing students of Chinese to play computer conversational games. At issue is the very hard problem of speech recognition not only for a non-native speaker, but also for a hesitant and disfluent speaker. Environmental issues are another risk factor, as students could be using whatever set-up they have at home, and the developer has no control over microphone quality or placement, or over environmental noise. The quality of the provided translations must be essentially perfect, and the dialogue interaction must be able to gracefully recover from digressions and misinformation due to unavoidable recognition errors. Any multimodal interactions need to be intuitive and easily integrated into the conversational thread. Finally, computers should be able to analyze the recorded utterances, and, in a subsequent interaction, critique selected production errors, involving aspects such as phonetic accuracy [5, 20], tone production [21, 15], lexical and grammar usage [16], and fluency [9].

At the Spoken Language Systems group in the Computer Science and Artificial Intelligence Laboratory at MIT, we have been developing multilingual spoken dialogue systems for nearly two decades [35]. A focus of our recent research has been to configure multilingual systems to support language learning applications [23, 26]. Thus far, we have been concentrating on technology goals, but we hope to achieve a milestone of introducing the technology into the classroom within the coming year. Feedback from students and educators will lead to design changes which will eventually converge on a design that works best, given the constraints of the technology and the needs and interests of the students. Most especially, we hope to design application domains that will be entertaining to the students, thus engaging them in the activity and providing a rewarding and non-threatening learning experience.

Figure 1: Screen shot of Web-based drill exercise, in which the student must solve a weather scenario. The student is provided explicit feedback on any tone errors (correcting "jia1," "qi1," and "hui4" in the example).

2. Web-based Exercises and Games

All aspects of learning a new language are difficult to master, and require persistent exposure and concentration. Computers can potentially assist the student in many aspects, such as vocabulary and pronunciation acquisition, as well as reading skills. Computerized flash cards can aid in vocabulary building, and pronunciation assessment technology can provide feedback on the quality of the student's oral productions [5, 20]. A computer can help a student to read by following along as they read out loud a presented passage [4].

Our focus however is on *oral communication skills*. Our intent is to design a variety of different interactive activities that are based on *spoken communication*. Each activity is associated with a particular set of vocabulary items and linguistic constructs. It is important for the game to be *goal-directed*, and it is crucial to provide adequate *scaffolding* such that the student can make progress without getting unduly frustrated. We are also coming up with ways to reward success by *graduating* to higher difficulty levels. By contextualizing their speech, we can greatly improve the recognition accuracy, since the design can support a narrowly defined linguistic space at any time.

To help the student prepare for a given interactive activity, we are developing a set of *translation exercises* which each focus on precisely the subset of the language needed to master the corresponding interactive game. The student would logically first practice the focused translation exercises, where the goal is to learn how to formulate and speak sentences in the domain. They would then be able to *eavesdrop* on computer-computer simulations of dialogues in the domain. Finally, after such extensive preparation, they would be ready to play the interactive dialogue game.

2.1. Translation Exercises

We have developed a set of translation exercises in a number of different domains, allowing a student to practice speaking phrases and sentences in Chinese. The system exploits a spoken language translation system both to provide translation assistance and to assess the student's performance. The initial prototypes of this system were designed for telephone-based interaction. The system allows the student to speak in either English or Chinese, and it replies first with a paraphrase in the language spoken, followed by a translation into the other language. Thus, the student can say a sentence in English, and then attempt to repeat the spoken translation provided by the system. Versions of this prototype were built in three domains: weather [28], flights [30], and a general travel (phrasebook) domain.

This system design was problematic in that it put the burden on the student to think of things to say within the domain, and did not set any explicit goals. In order to convert it into a more task-oriented approach, we needed to design a Web-based version such that the assigned goal(s) could be presented on the screen. Our first attempt along these lines made use of scenarios in the weather domain, randomly generated by the computer and presented to the student as simple English keywords: "rain – Dallas – Friday," as illustrated in Figure 1. The intial interface only supported typed input. The student was tasked with typing in a sentence using pinyin format to query the computer for the constraints specified by English slot fills. A nice feature is that the system can automatically correct any errors in tone, and highlights the corrected tones. Since English speakers have tremendous difficulty learning tone, this is an effective way to help them acquire tonal knowledge.

Once we solved the problem of speech capture at a Web page, it was relatively straightforward to design computer guided translation exercises on the Web, that allowed the student to practice *speaking* in Chinese. An example interface for a flight domain translation game is shown in Figure 2. The system randomly generates a small set of English utterances from templates and displays them on the screen. The student is prompted for each utterance in turn, and they then attempt to speak an ut-

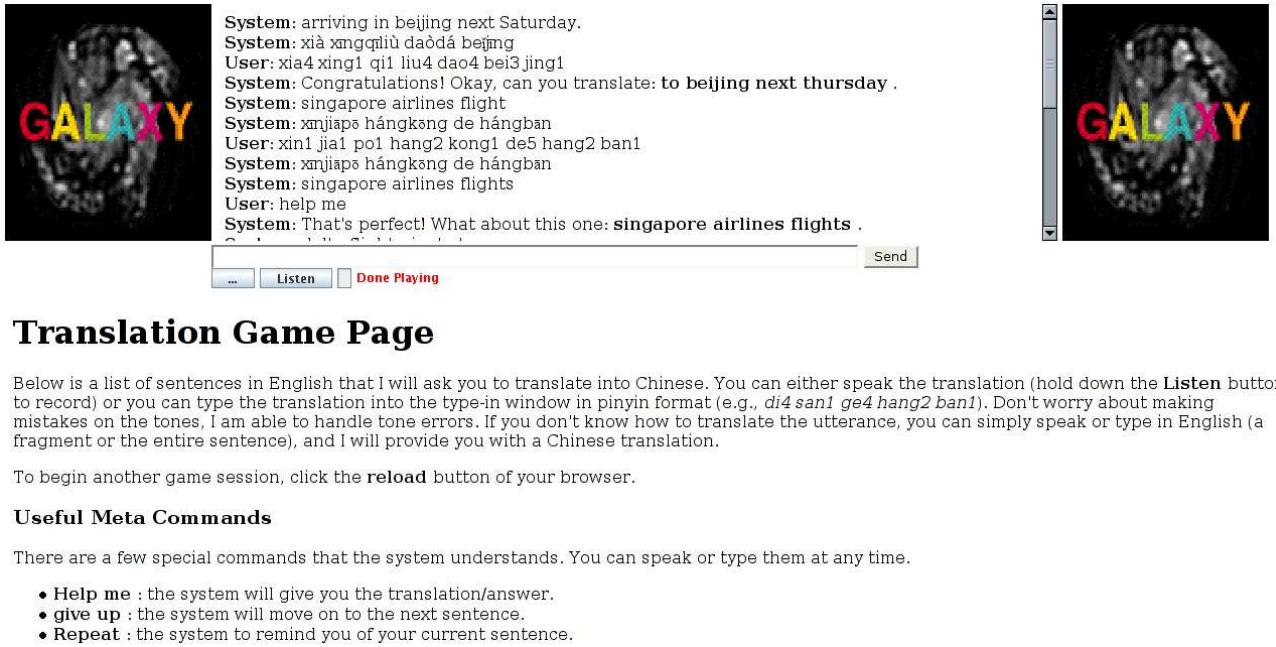

Figure 2: Web interface for a translation game in the flight domain. The system poses English phrases for which the student is tasked with speaking or typing an equivalent sentence in Mandarin. The dialogue unfolds from bottom to top in the upper window. The student can at any time type a word or phrase in English in the type-in window to obtain a spoken translation.

terance of equivalent meaning in Chinese. The system responds with a paraphrase in Chinese, followed by a translation into English, and then congratulates the student if it matches the prompt at the meaning level, and moves on to the next utterance. If the student is stuck, they can ask the system to provide a translation which they can then attempt to repeat. A natural "score" emerges from this process based on the total number of turns taken on average for each utterance.

This system has been configured with ten graded difficulty levels, ranging from isolated words at the lowest level to compound and complex sentences at the highest level. The student is automatically "graduated" to higher levels based on good performance, and can slide to lower levels if they do poorly. Sentences are generated randomly from thousands of templates generated from real user data, such that the game does not appear monotonous or repetitive, even if they stay at the same level. The student's translation is evaluated at an [attribute: value] level rather than as a string match, such that multiple variants of the Chinese translation are accepted. The translation and student evaluation procedures are described more fully in Section 3. Please see [30] for further details.

We have piloted this flight domain translation game in a user study which yielded encouraging results in an exit poll [33]. All of the twelve subjects thought the system was helpful at improving their Chinese, and most would play it again and recommend it to their Chinese-learning friends.

The exercise demands extremely high performance in terms of translation accuracy, and reasonably high speech recognition accuracy, in the face of the heavily accented and disfluent speech of a language learner. Speech recognition accuracy can be significantly improved by exploiting context - the utterance prompt – to influence N-best selection, as described in [32]. Furthermore, the student can exploit the "help" mechanism to hear a translation spoken by a speech synthesizer. They then only need to produce an accurate *imitation* of what they heard.

We now have systems capable of high quality translation in the weather domain [31], the flight domain [30] and the hobbies and schedules domain [3]. Translation reversibly between English and Chinese utilizes an interlingual method. We have configured a slightly different variant for the hobbies and schedules domain, which gives the student a little more control. Instead of the system prompting for each uttterance in turn, it simply presents the entire set for the session on the page, and invites the student to translate any of the utterances. The system performs an NxM match between the N-best list from the recognizer and the M candidates on the screen.

2.2. Interactive Dialogue Games

Telephone-based Systems

When we first began to think of using spoken dialogue interaction for language learning, it was logical to simply adapt our existing multilingual dialogue systems for language learning purposes. Thus, we developed phone-based systems configured to support speech understanding in both English and Chinese. Dialogue interaction involved asking questions about the weather [34, 27] or planning an air travel itinerary [25, 30]. The student was required to communicate with the system in Chinese, but could speak an English sentence at any time, in which case the system would speak a sentence of equivalent meaning in Chinese. The student could also ask for a translation of the system's response at any time.

While these systems were challenging to build and fun to play with, the lack of a structured game made the student less motivated to persist in conversing with the system on random topics. What was needed was a set of explicit goals, along with a schema for advancement to higher difficulty levels. Also, the topic of conversation was not the most appropriate for a beginner student.

The Novice Student

We have recently come up with a design for an interactive

game based on family relationships, which assumes *no* prior knowledge of Chinese. This game is designed to permit the student to deduce the Chinese names for relationships by trial and error. It is configured with multiple difficulty levels, to provide the student with a concrete sense of progress. A novely of the interface is that pictures of family members also operate as record buttons, such that the student can simultaneously select an image for context and initiate recording. I will not say more about this game here, as it is fully described in [18].

A "Video" Game Design

We are exploring a number of different options for configuring games that will achieve the look-and-feel of a video game, while still focusing on teaching a foreign language. Ideally, the game would include some element of competition, either with the clock or with an opponent, in order to increase its intrinsic entertainment value. At the same time, it would be convenient to allow a developer to easily swap in new vocabulary items, to help with vocabulary building. We are currently in a brainstorming phase, considering options for a "card game," similar to "Go Fish" or "Gin Rummy," where the cards would be manipulated through spoken interaction in Chinese, and participants could be either virtual or human, making use of a collaborative game interface on the Web.

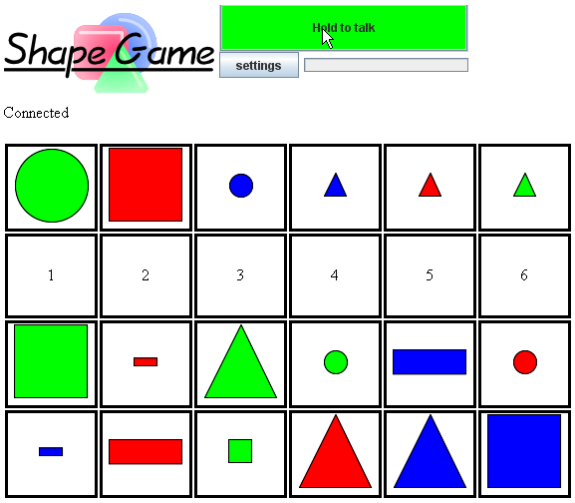

Figure 3: Screen shot of the *Shapegame* Web interface.

A game that meets many of these design goals was recently developed by a student in our group, Alex Gruenstein [7]. The screen is populated with displayed objects that represent large/small colored shapes – triangles, rectangles, squares, circles, as illustrated in Figure 3. The student is tasked with first selecting (verbally) one of the shapes in the last two rows, and then moving the selected object to one of the empty slots in row 2. They are then required to alter the shape (verbally) to make it look identical to the one displayed just above it.

This game now exists in both English and Chinese. Typical commands in the game are, "Pick up the large red triangle and place it in the fifth slot." "Make it small." Interestingly, the system communicates its understanding of the user's request entirely by its actions on the screen. It is able to understand the student's recorded speech *incrementally*, such that it can give feedback on partial understanding without having to interrupt them. This allows the student to get immediate feedback on whether they have been understood correctly and to make corrections midstream. It will also alert them to ambiguities due to inadequate specification by highlighting multiple items.

Student:	What are you doing tomorrow afternoon?
System:	I'm playing basketball. Would you like to join me?
Student:	No, I don't like to play basketball.
System:	What do you like to do?
Student:	I like playing video games and watching movies.
System:	Would you like to get together to watch a movie this evening?
Student:	No, I'm going to study this evening.
System:	Are you free tomorrow evening?
Student:	Yes! Let's watch a movie tomorrow evening.

Figure 4: Example English dialogue in the Hobbies and Schedules domain.

We have an opportunity with this game design to explore two-party human-human computer games, where two players compete at separate computers on filling the slots in the second row in a race condition. We have built a prototype of this design and members of our laboratory have played the two-party game in Chinese. We are encouraged by the potential of this game to achieve an element of fun and excitement rarely found in language drill exercises.

Furthermore, this game represents a prototype for a number of variants that will allow us to substitute other objects in place of the basic shapes, which will then make it more generally useful for vocabulary building.

Spoken Dialogue Interaction Game

Our most sophisticated dialogue interaction game was developed with the goal of supporting the curriculum of a typical first-year student. We have provided a significant amount of scaffolding to assist the student with the task. However, the scaffolding is incrementally peeled away as the student advances to higher difficulty levels. We consulted a high school text book [29] for guidance, and settled on a "Hobbies and Schedules" game that would exercise vocabulary and grammar constructs introduced in the first year textbook. The game is based on controlled scenarios where the student is assigned a specified "persona" and a "future schedule," and is tasked with finding a time in the future to jointly participate with the simulated dialogue partner in an activity that both parties "like."

Access to the dialogue game is obtained by simply entering a Web page, following a log-in step. The student's assigned preferences, as well as a schedule of events in the next few days, are displayed on the screen. The student engages in interactive spoken dialogue with the computer in Chinese, to solve the scenario. A typical dialogue interaction (in English) is shown in Figure 4.

A screen shot of the dialogue game is shown in Figure 5. Here, the student has just completed a successful dialogue with the system to arrange to meet to watch television together tomorrow evening. The system misunderstood the user when they said they liked to watch television, thinking they had said "watch a movie" instead. But the user was able to correct this misconception in subsequent dialogue. Once the user says "good bye" the system will launch an entirely new game.

The system supports five distinct "difficulty" levels. At the lowest level, the conversational partner's answer is displayed

Figure 5: A screen shot of a dialogue interaction within the Hobbies and Schedules dialogue game. In spite of a recognition error misunderstanding "dian4 ying3" when the student said, "dian4 shi4," the student was able to correctly achieve the dialogue goal.

in English, tone-marked pinyin[1], and characters in a dialogue box. The system also speaks the answer in Chinese. More significantly, in addition to the conversational partner, there is a "robotic tutor" who assists the student in composing their next turn in the dialogue. This tutor "listens in" on the conversation and plans the next dialogue turn alongside the student. It can then propose a candidate utterance that, according to its "judgment," would be appropriate to say next. At the lowest level, the student can simply read off (in pinyin) the proposed prompts. At higher levels, the prompt is provided in English or in Chinese characters, and is thus more challenging to the student.

At any time, the student can, with the click of a mouse, "submit" the tutor's prompt to the system. The robotic tutor (a different voice from the dialogue partner) will then speak the appropriate Chinese utterance corresponding to the prompt. The student then only needs to parrot what they just heard to push the dialogue forward. At the very lowest level, the mouse click invokes not only the *tutor's* utterance but also the *dialogue partner's* response. Thus the student can experience the game in a totally passive mode if they choose to, by listening to a completely simulated dialogue unfold turn by turn. At the highest level, the tutor disappears altogether (the prompt space is left blank), so the student must solve the scenario on their own.

There is another dimension in which difficulty level can be manipulated, which concerns the amount of overlap between the student's persona and the system's persona. If the two conversational partners have very little in common and a great deal of scheduling conflict, it requires statistically much longer to reach a common goal. We have designed the interface such that a student graduates in difficulty level with respect to tutor assistance only after completing five episodes, arranged from easy to hard in terms of the difficulty of solving the task.

[1] A Roman encoding of word pronunciations, including tone.

3. Underlying Technologies

In this section, we briefly describe the underlying technologies that support the language learning systems we are developing, highlighting three aspects in particular: spoken language translation, assessment of student translations, and symmetrical dialogue interaction.

Our systems are all configured as a set of technology and interface servers that communicate among one another via a programmable central hub using the Galaxy Communicator architecture [22]. A Java audio program is automatically downloaded to support audio input at the computer. For speech recognition we use the SUMMIT landmark based system [6]. The natural language understanding component, TINA [24], processes an N-best list of utterance hypotheses from the recognizer. It produces a *semantic frame*, encoding the meaning, which is then translated (paraphrased into Chinese by the language generation server [2]), or answered (dispatched to the dialogue manager), depending on the application. The dialogue manager interprets the sentence in context, assisted by the context resolution server, and retrieves appropriate information pertinent to the question from the database (flights, weather, etc.). The dialogue manager prepares a *reply frame* which is passed on to the language generation server to produce a string response in Chinese. Each translation or response string is directed to the appropriate synthesizer (English or Chinese) by the hub pro-

a.	where did you put the book I bought yesterday?
b.	did you put I bought the book yesterday where?
c.	you put I buy book yesterday where?
d.	you ba5 I yesterday buy de5 book put where?
e.	ni3 ba5 wo3 zuo2 tian1 mai3 de5 shu1 fang4 zai4 na2 li3?

Figure 6: Illustration of steps for converting an English generation system into a Chinese generation system, based on an underlying interlingua. (a) fully English, (b) omit movement, (c) omit inflectional endings and function words, (d) insert Chinese ordering rules, (e) map to Chinese lexicon.

gram. When relevant, a separate HTML response is displayed as a table of appropriate information returned from the database. The system response is also displayed in a dialogue box that shows a sequence of all preceding user-system turns.

3.1. Speech Translation

Both the dialogue interaction games and, obviously, the translation exercises, depend heavily on a high-quality speech-to-speech translation system. Although statistical methods dominate the current literature for translation [1, 14], we have decided instead to use linguistic analysis/synthesis as our central process, with statistics playing a supporting role in training the grammar and in *n*-gram selection on the final hypothesis [30]. We argue that, because we require essentially perfect translation quality, we can not rely on statistical methods alone, because they do not explicitly take into account linguistic well-formedness.

In parsing, we have taken care to explicitly account for movement phenomena, which are prevalent, particularly for English *wh*-marked questions where the who/what/where/when/why units are obligatorily preposed to the front of the sentence. We use a technique that assigns certain parse categories special privileges to *generate, activate,* or *absorb* a moved constitutent, as described fully in [24].

Another important point to make is that our gramars are trained on a large corpus of example sentences in the domain. For the flight domain, we have available a corpus of over 20,000 utterances spoken by English users and subjects interacting with a flight travel planning system [25]. We used this corpus to generate random templates in English, which can in turn be translated into Chinese to yield a corpus to train the Chinese grammar (and recognizer). Our grammars are predominantly syntax-based, and the probability model provides important information to help resolve modifier attachment ambiguity.

Figure 6 illustrates schematically by an example how an English-to-English paraphrase system can be converted into an English-to-Chinese translation system by modifying (mainly simplifying) the English generation rules. We argue that the most difficult aspect of English-to-Chinese reordering involves the English movement phenomena mentioned earlier, which apply both in wh-questions and in relative clauses, both of which are illustrated in this example. The most challenging aspect for *English* generation is the restoration of the moved constituent to its surface form position. By simply omitting this step, the generated string appears as shown in Figure 6b. The second change is a further simplification to omit many of the function words and all of the inflectional endings (Figure 6c). Chinese-specific alterations involve reordering the temporal modifier, invoking a *ba-construction* rule, and inserting a special marker (de5) for the relative clause. Finally, an English-to-Chinese lexicon converts the individual words to Chinese (Figure 6e).

3.2. Assessment of Translation Accuracy

The process of evaluating whether the student has correctly translated the system's prompt in the translation game takes into account the possibility that the student may not say exactly the same utterance as the system produces in its automatic translation of the prompt. Care must be taken however to assure, as much as possible, that the student's utterance is in fact a well-formed Chinese utterance and contains the same semantic information as the prompt. The automatically translated English prompt and the hypothesized N-best list produced by the recognizer are processed through the same procedure to extract semantic content, as illustrated in Figure 7. This process first assures that the hypothesis is parsable (a good test of well-formedness), and then uses a generation mechanism to reduce the meaning to a simplified [key: value] (KV) representation. A perfect match is achieved if the KV pairs of the N-best hypothesis and reference are identical. However, partial matches are common, especially when substitution errors on dates and times occur as a result of misrecognition. In such situations, it is natural for the user to just repeat the "incorrect" piece, especially when the sentence is long. A partial match mode in the comparison algorithm allows the student to complete the translation in multiple turns. Further details for the assessment algorithm and evaluation results are described in [32].

3.3. Symmetrical Dialogue Interaction

In this section we briefly describe the dialogue management component of the Hobbies and Schedules domain, which represents a significant departure from the traditional information-access domains we have developed in the past. The inherent symmetry in this kind of dialogue allows the computer to role play both sides of the conversation, thus enabling an effective resource for system development and refinement, as well as providing tutorial support for the student. The dialogue unfolds differently with each episode, due in part to the randomly generated personas specifying preferences and schedules. Thus, thousands of dialogue interactions can be automatically generated, yielding a large corpus of simulated utterances to train the initial statistics of the grammar and recognizer language model. The initial question is randomly generated from among a number of different possibilities, e.g., "When are you free <day>?" What are you doing <day> <time>?" "Do you like <hobby>?" "What do you like to do?" etc. Whenever the system is unable to come up with a logical follow-up question, it invokes a random question to encourage further dialogue.

The system maintains both a *self model* and a *user model*, where the self model initially defines its own personal likes, dislikes, and schedule, but is frequently updated over the course of the dialogue to record which information has already been communicated to the user. The user model is gradually expanded to record any information obtained through dialogue with the user, including user preferences and previous commitments. An additional *joint model* maintains proposed future events under consideration. These events are adjusted over time as different times and/or activities are entertained. The status of a proposed future event reflects whether individual facts have been confirmed with the user, in order to deal with miscommunications due to both recognition errors and the student's inadequate comprehension of the language.

Similar to our strategies in the past, we have separated

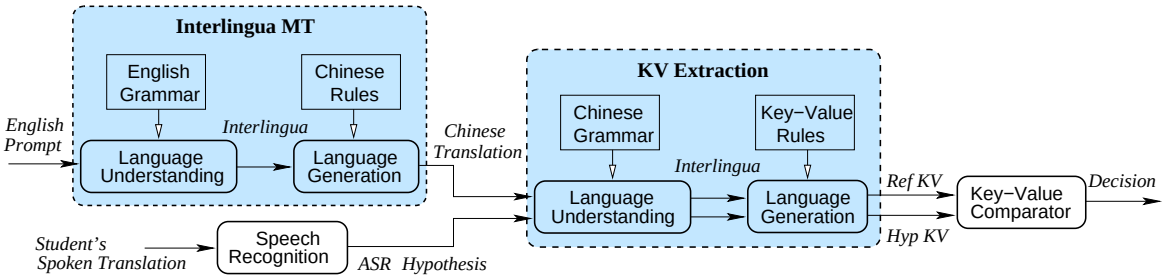

Figure 7: Flow chart of the user utterance verification process.

out aspects particular to the scenario into an external domain-specific configuration file. This will make it relatively straightforward to redesign dialogue interaction scenarios around other topics that share a common abstraction with our original scenario, and thus lead to a substantially accelerated development cycle for an expanded curriculum. For example, a later scenario might involve making plans to dine together at a mutually agreeable restaurant. The dialogue manager adopts a simple E-form representation of its linguistic messages, which are converted into well-formed English and/or Chinese sentences using formal generation rules.

4. Related Research

In a 1998 review paper assessing the state of the art in computer aids for language teaching, Ehsani and Knodt [8] wrote: "Students' ability to engage in meaningful conversational interaction in the target language is considered an important, if not the most important, goal of second language education. This shift of emphasis has generated a growing need for instructional materials that provide an opportunity for controlled interactive speaking practice outside the classroom." However, perhaps because of the complex requirements associated with human-computer dialogue interaction, there has been surprisingly little research in spoken dialogue systems aimed towards this goal up to the present time.

There are a couple of promising ongoing initiatives, one in the U.S. and one in China, which are rapidly changing this picture. The U.S. initiative is the Tactical Language and Culture Training System (TLCTS) [12, 13], which began as part of the DARPA Training Superiority program. This ambitious program has focused mainly on Arabic languages. The idea is to embed language learning into a video-game-like environment, where the student assumes the role of a character in the video game, and interacts with other characters they encounter as they explore a virtual space, for example, role playing a soldier in an Iraqi village. The student communicates with the other characters through speech and mouse-based gestures, and the options available at any point are based on the situational setting.

One of the presumably many ongoing efforts in China for learning English is the CSIEC Project [10, 11], which is similar to ours in that the main delivery model is interactive dialogue at a Web page. Similar to the TLCTS project, the student interacts with embodied characters. No attempt is made to situate them in a complex scene, but rather each character simply role plays a conversational partner, mainly using a chatbot concept. The student can choose from among six different "virtual chatting partners," each of which has a distinct style of conversational interaction. For example, one personality will simply rephrase a user's statement into a question: "Why do you like to play baseball?" Another character will tell jokes or stories, or sing a song, upon request. Thus far interaction has been restricted to typed input, with the target language being English, although their intention is to eventually support spoken inputs. No translation assistance is offered.

Any *multilingual* spoken dialogue system could be relatively easily reconfigured as a language learning activity. For example, the ISIS system [19] is an impressive trilingual spoken dialogue system, supporting English, Mandarin, and Cantonese, which involves topics related to the stock domain and simulated personal portfolios.

A research topic that has some synergy with dialogue systems for language learning is the more general area of educational tutoring scenarios. An example involving spoken dialogue interaction to help a student solve simple physics problems can be found in [17].

5. Summary and Future Work

This paper summarizes the current status of our research over the past several years, which is aimed towards providing an enriching, entertaining, and effective environment for practicing a foreign language by interacting with a computer. We are now at the threshold of a new phase of our research, in which we will introduce our technology into the classroom setting, and evaluate its effectiveness in teaching Mandarin to native speakers of English. We have barely begun research on the assessment phase, which will involve post-processing the student's recorded utterances and providing focused corrective feedback on errors in prosodics, pronunciation, lexical, and grammar usage.

While our research has predominantly involved the paradigm of a native English speaker learning Mandarin, it would be quite straightforward to reverse the roles of the two languages to support a native Mandarin speaker learning English. Our attempts to support portability issues allow the techniques to generalize to other language pairs as well, but of course this needs to be demonstrated in future research.

Our symmetrical dialogue interaction paradigm could support the intriguing possibility of *humans* role playing both sides of the conversation, via their respective Web-based interfaces. Two students could interact with each other to solve the scenario, with the computer playing a tutorial role for both students, providing them with translation assistance when needed and filtering their utterances such that the other student only receives sentences spoken in Chinese.

In future research, we plan to greatly enrich the graphics component of our systems, ultimately supporting an interactive immersive video contextualization, thus blurring the boundary between educational exercises and video games.

6. Acknowledgements

Many former and current students and researchers in the Spoken Language Systems group at MIT have contributed to this research. They include Chih-Yu Chao, Alex Gruenstein, John Lee, Ian McGraw, Mitch Peabody, Chao Wang, and YuShi Xu.

Julian Wheatley has been very supportive of our work and has helped integrate it into MIT's language laboratory. We are indebted to both students and educators at the Defense Language Institute, especially Mike Emonts, for their participation in pilot experiments introducing our software into the classroom.

This research was funded by grants from the Cambridge MIT Initiative, the Industrial Technology Research Institute, and the U.S. Department of Defense.

7. References

[1] P. F. Brown, J. Cocke, S. A. Della Pietra, V. J. Della Pietra, F. Jelinek, R. L. Mercer, and P. S. Roossin, "A Statistical Approach to Machine Translation," *Computational Linguistics*,, 16:2, 79–85, 1990.

[2] L. Baptist and S. Seneff, "Genesis-II: A Versatile System for Language Generation in Conversational System Applications," *Proc. ICSLP '00*, V. III, 271–274, Beijing, China, Oct. 2000.

[3] C-Y Chao, S. Seneff, and C. Wang, "An Interactive Interpretation Game for Learning Chinese," *Proc. SIGSLaTe These Proceedings*, 2007.

[4] R. Cole, S. van Vuuren, B. Pellom, K. Hacioglu, J. Ma, J. Movellan, S. Schwartz, D. Wade-Stein, W. Ward, J. Yan "Perceptive Animated Interfaces: First Steps Toward a New Paradigm for Human Computer Interaction," *Proc. IEEE Special Issue on Multimodal Human Computer Interface*, 2003.

[5] B. Dong, Q. Zhao, J. Zhang, and Y. Yan, "Automatic Assessment of Pronunciation Quality," ISCSLP '04, (2004) 137–140, Hong Kong.

[6] J. Glass, "A Probabilistic Framework for Segment-Based Speech Recognition," *Computer, Speech, and Language*, 17, 137-152, 2003.

[7] A. Gruenstein, "Shape Game: A Multimodal Game Featuring Incremental Understanding," Term Project, 6.870 Intelligent Multimodal Interfaces, May, 2007. http://people.csail.mit.edu/alexgru/shapegame/

[8] F. Ehsani and E. Knodt, "Speech Technology in Computer-aided Language Learning: Strengths and Limitations of a new CALL Paradigm Language Learning & Technology," V. 2, No. 1, 45-60, 1998.

[9] M. Eskenazi, "Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype," Language Learning and Technology, V. 2, No. 2, 62-76, 1999.

[10] J. Jia, "The study of the application of a web-based chatbot system on the teaching of foreign languages," *Proceedings of SITE 04*, AACE Press, USA. 2004. 1201-1207.

[11] J. Jia, "CSIEC (Computer Simulator in Educational Communication): A Virtual Context-Adaptive Chatting Partner for Foreign Language Learners," *Proceedings of ICALT 04*, IEEE Computer Society Press, USA. 2004. 690-692.

[12] W. L. Johnson, S Marsella, N. Mote, M Si, H. Vihjalmsson, S. Wu, Balanced Perception and Action in the Tactical Language Training System, *Proc. International Conference on Autonomous and Multiagent Systems*, 2004.

[13] W. L. Johnson, "Serious Use of a Serious Game for Language Learning," *Proc. AIED*, 67–74, 2007.

[14] P. Koehn and F. J. Och and D. Marcu, "Statistical Phrase-Based Translation," *Proc. HLT-NAACL*, 2003.

[15] J. Leather, "Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers." In J. Leather and A. James, eds., *New Sounds 90*, University of Amsterdam, 72–97, 1990.

[16] J.S.Y. Lee and S. Seneff, "Automatic Grammar Correction for Second-Language Learners," *to appear*, *Proc. INTERSPEECH*, 2006.

[17] D. Litman, "Spoken Dialogue for Intelligent Tutoring Systems; Opportunities and Challenges," Keynote speech, *Proc. HLT-NAACL*, 2006.

[18] I. McGraw and S. Seneff, "Immersive Second Language Acquisition in Narrow Domains: A Prototype ISLAND Dialogue System," *Proc. SIGSLaTe, These Proceedings*, 2007.

[19] H. Meng, P.C. Ching, S.F. Chan, Y.F. Wong, and C.C. Chan, "ISIS: An Adaptive Trilingual Conversational System with Interleaving, Interaction, and Delegation Dialogues," *ACM Transactions on Computer-Human Interaction (TOCHI)*, V. 11, No. 3, pp 268–299, 2004.

[20] A. Neri, C. Cucchiarini, and H. Strik, "Feedback in computer assisted pronunciation training: technology push or demand pull?" *Proceedings of ICSLP*, 1209–1212, 2002.

[21] M. Peabody, S. Seneff, and C. Wang, "Mandarin Tone Acquisition through Typed Dialogues," 173–176, *InSTIL Symposium on Computer Assisted Language Learning*, Venice, Italy, 2004.

[22] S. Seneff, E. Hurley, R. Lau, C. Pao, P. Schmid, and V. Zue, "Galaxy-II: A Reference Architecture for Conversational System Development," *ICSLP '98*, 931–934, Sydney, Australia, December, 1998.

[23] S. Seneff, C. Wang, M. Peabody, and V. Zue, "Second Language Acquisition through Human Computer Dialogue," *Proc. ISCSLP '04*, Hong Kong, 2004.

[24] S. Seneff, "TINA: A Natural Language System for Spoken Language Applications," *Computational Linguistics*, V. 18, No. 1, 61–86, 1992.

[25] S. Seneff, "Response Planning and Generation in the MERCURY Flight Reservation System," *Computer Speech and Language*, V. 16, 283–312, 2002.

[26] S. Seneff. "Interactive Computer Aids for Acquiring Proficiency in Mandarin," Keynote Speech, pp. 1–11, *Proc. ISCSLP*, Singapore, 2006.

[27] C. Wang, D. S. Cyphers, X. Mou, J. Polifroni, S. Seneff, J. Yi, and V. Zue, "MUXING: A Telephone-access Mandarin Conversational System," *Proc. ICSLP '00*, V. II, 715–718, Beijing, China, Oct. 2000.

[28] C. Wang and S.Seneff, "High-quality Speech Translation for Language Learning," 99–102, *InSTIL Symposium on Computer Assisted Language Learning*, Venice, Italy, 2004.

[29] T.C. Yao and Y. Liu Yao, *Integrated Chinese, 2nd Edition, Cheng and Tsui Company*, Boston, MA, 2005.

[30] S. Seneff, C. Wang, and J.S.Y. Lee, "Combining Linguistic and Statistical Methods
for Bi-directional English Chinese Translation in the Flight Domain," *Proc. AMTA '06*, 2006.

[31] C. Wang and S. Seneff, "High-quality Speech Translation for Language Learning," pp. 99–102, *InSTIL Symposium on Computer Assisted Language Learning*, Venice, Italy, 2004.

[32] C. Wang and S. Seneff, "Automatic Assessment of Student Translations for Foreign Language Tutoring," *Proc. NAACL-HLT*, Rochester, NY, 2007.

[33] C. Wang and S. Seneff, "A Spoken Translation Game for Second Language Learning," *Proc. AIED*, Los Angeles, California, 2007.

[34] V. Zue, S. Seneff, J. Glass, J. Polifroni, C. Pao, T.J. Hazen, and L. Hetherington, "Jupiter: A Telephone-Based Conversational Interface for Weather Information," *IEEE Trans. Speech and Audio Proc.*, 8(1), 85–96, 2000.

[35] V. Zue and J. Glass, "Conversational Interfaces: Advances and Challenges," *Proc. IEEE*, V. 88, No. 8, 1166–1180, 2000.