Learner Centered Language Programs: Integrating Disparate Resources for Task-based Interaction

Deryle Lonsdale
Department of Linguistics
Brigham Young University
4039 JFSB
Provo, UT 84602
(801) 422-4067
lonz@byu.edu

C. Ray Graham
Department of Linguistics
Brigham Young University
4059 JFSB
Provo, UT 84602
(801) 422-2208
ray_graham@byu.edu

Rebecca Madsen
Department of Computer Science
Brigham Young University
4064 JFSB
Provo, UT 84602
(801) 422-2937
rmadsen@byu.net

Learner-centered Language Programs: Integrating Disparate Resources for Task-based Interaction

KEYWORDS: animated agent, computer-assisted language learning, dialogue move engine, interaction, language learning, pronunciation, scenario, speech

ABSTRACT

In this paper we first discuss three factors that are believed to be important for success in second language learning: comprehensible input, comprehensible output, and noticing discrepancies. We then discuss our current research work in integrating various components of human language technology to address these three language acquisition factors. Our efforts involve creating a wide spectrum of interesting language learning applications including question answering and pronunciation tutoring. These applications show the potential of combining speech processing with other important natural-language tools, such as external knowledge sources and dialogue move engines. The applications which we have developed not only show that this integration can be successful in creating non-trivial applications, but that there is much work which can be done to build on what we have been able to accomplish thus far.

INTRODUCTION

Recent reviews of computer-assisted language learning (CALL) (Pennington,1996; Beatty, 2003; Zhao, 2003) suggest that while developers and learners are often enthusiastic about CALL programs, there is little empirical evidence which shows that they are effective in helping

learners to develop oral communication skills, particularly speaking skills. This lack of evidence is in part due to the fact that researchers have simply failed to conduct the needed research with the myriad of software programs available for teaching such skills. A great deal of the fault, however, must be placed at the feet of program developers who have failed to create programs which enable learners to practice oral interaction skills as a part of their learning activities.

When one examines the nature of software programs reviewed in such journals as *CALICO* and *Language Learning & Technology* or online at the CTI Centre for Modern Language Resources and the CALL@ChorusSoftware Reviews, one is struck by the fact that, with few exceptions, software programs provide learners with very little opportunity to go beyond a rather mechanical reproduction of oral language. While many programs may provide learners with some high quality listening opportunities and lots of graphic support and enhanced input, opportunities for oral interaction are very limited. In fact, most of these programs require that students activate the learning materials by clicking on buttons to watch actors participate in oral interactions which serve as models for them to repeat and memorize. Thus, the learner becomes a third person participant in the interaction. Instructions are most often presented in written form and often in the native language and learners interface with the program via the keyboard and/or buttons with clicks of the mouse. In some programs learners are encouraged to insert their names into structured dialogues that they have seen enacted by others and to participate in presenting alternate lines following previously given models. In some, they are encouraged to record their own voices performing the dialogs and then compare them with the native speaker model. These exercises are much more akin to dialog memorization than they are to communicative interaction.

The major goal of our work is to show how non-communicative learning activities in conventional CALL programs can become a part of a communicative learning environment where the student becomes a first person participant in language interaction through the use of animated agents and speech technology. We show that animated conversational humanoid agents can direct the language experience of the learner through verbal communication and that the learner can interact both verbally and non-verbally with the agent. Our strategy has been to develop complex, multi-component applications—selecting appropriate toolkits supporting existing technologies and integrating them together as seamlessly as possible.

In this paper will discusses: (i) a theory of language learning which supports the interactive nature of our learning activities, (ii) core technologies that serve as the foundation for our work, and (iii) novel applications that we have developed to engage the learner in simulated communication with agents... Finally, we will conclude with a discussion of possible further work.

BACKGROUND: A THEORY OF SECOND LANGUAGE ACQUISITION

It has long been recognized that input plays a crucial role in language acquisition (Krashen, 1985; Gass, 1997). Comprehensible input can be presented to learners in a variety of ways, including an agent speaking directly to the learner, video segments of native speakers performing real world tasks, video and audio segments of presentations and monologues, etc. The comprehensibility of the input can be enhanced by using visual aids and pre-listening schema building exercises, by presenting crucial vocabulary items prior to the listening exercise, by providing the written form of the text to accompany the listening exercise, and by providing glossing of difficult words, to mention only a few methods.

However, for input to get incorporated into the learner's interlanguage, it must not only be comprehensible but it must be comprehended (Gass, 1988, 1997, 2003). Verification of comprehension requires interaction on the part of the learner. Long's (1996) Interaction Hypothesis claims that attention, achieved through interaction, is a crucial part of the mechanism of acquisition. Thus one of the contributors to the effectiveness of second language instruction is the degree to which it provides the learner with input that is comprehended by the learner. We believe that this can be done most effectively in CALL by having an animated agent speak directly to the learner, giving instructions and interacting with the learner regarding the instructional contents and by engaging the learner in verbal exchanges.

A second factor which has been shown to contribute to the development of language is that of comprehensible output (Swain, 1985). Not only does interaction increase the comprehensibility of input for a given learner, but attempts to produce language that is comprehensible to interlocutors contributes to the acquisition process in other ways. Recently (Swain, 1995) has claimed that the mechanism behind the influence of output on language acquisition is that it moves the learner from a general nondeterministic semantic processing mode for comprehension to a complete grammatical processing mode necessary for production. Thus as learners are required to formulate novel utterances in a communicative task they are forced to focus on the more temporal and structural aspects of the forms and process them in much greater detail. Hence, task-based approaches to second language instruction in which learners are required to participate in communicative interactions are widely considered to be the most efficient way to develop interlanguage skills (Bygate, Skehan, and Swain, 2001; Lightbrown and Spada, 1999; Swain, Brooks, and Tocalli-Beller, 2002; Hall and Walsh, 2002). We attempt to create this interaction via animated agents using speech recognition and synthesis.

A third factor which is hypothesized to contribute to acquisition is that of "noticing." Many researchers in second language acquisition believe that in order for the learners' interlanguage system to evolve toward more native like forms, the learner must notice, either through positive or negative evidence, that their system is at variance with the native speaker forms. This is believed to be accomplished through various mechanisms, including those already discussed. As mentioned above, Long's (1996) Interaction Hypothesis claims that attention to form is accomplished through negotiated input. As interlocutors use strategies such as confirmation checks, comprehension checks, clarifications requests, recasts, and overt corrections, learners are made aware of the fact that their interlanguage needs modifying. This, along with positive modeling enables the learner to make corrections in their interlanguage system.

An example of how noticing with negative evidence is accomplished in our program is illustrated by one of our pronunciation activities discussed in detail below. In the activity, the animated agent describes a situation involving a character performing a certain action (for example, a blacksmith heating metal with a fire or hitting the metal with a hammer). If the agent tells the learner that the blacksmith is hitting the metal, the learner must respond that it is with a hammer. If the learner misperceives what the agent has said and chooses the fire instead of the hammer, the agent gives corrective feedback, but always in the context of communicating the correct meaning. When it becomes the learners turn to tell the agent what the blacksmith is doing, the learners must pronounce "hitting" and "heating" well enough for the agent to correctly identify the object... If the agent chooses the wrong object the learner knows that he has not pronounced it correctly and must try again. It is expected that by reducing the redundancy of language to a minimum so that miscommunication results from a mispronounced sound, the

learner will be induced to notice and begin the process of modifying the interlanguage phonological system, especially if the experience is repeated with several different situations involving the same sound.

CORE TECHNOLOGIES: ISSUES AND SOLUTIONS

Speech recognition and synthesis

Our work incorporates two fundamental technologies: speech recognition and speech synthesis. Although both are familiar, we deem it necessary to mention some of the criteria and features that make these technologies desirable for language learning, as well as some of the difficulties that preclude their widespread use in current educational applications.

Speech recognition is a complex problem that relies on various architectures that have been instantiated in several developer toolkits and end user programs. While work has been progressing on increasing the versatility of such technology, its performance is still far short of the widely sought-for ideal: large-vocabulary, high-accuracy, speaker-independent continuous speech recognition for any language. Still, speech recognition technology can be a viable option for educational applications when pragmatically implemented and integrated with other components, particularly when developers are able to adapt system components to the particular issues being addressed. In order for an application to be successful, the questions the developers must ask themselves are "What do we want to use it for?" and "How do we get it to perform the task?" (Ehsani and Knodt, 1998). Pursuing technological tradeoffs via available toolkits becomes the central question for the use of speech recognition.

A speech synthesis component is equally important for highly interactive speech processing systems. Most widely-used speech synthesis systems use text-to-speech (TTS)

processing. Although the quality of TTS spoken output is improving, all existing systems produce unnatural-sounding voices to varying degrees. The use of diphone concatenation techniques seems the most promising solution to phonetic and phonological variation. However, significant progress remains to be made in the area of suprasegmental properties: intonational contours, stress, rhythm, and so on. It has been observed that the "speech synthesis component is the one that often leaves the most lasting impression on users" (Glass, 1999).

In the rest of this section we indicate how our work has involved selecting and developing interactive speech-based tasks that build upon the strengths of existing speech technologies.

Pronunciation modeling

For adults, proper pronunciation is one of the most difficult areas to achieve in learning another language (Ellis, 1994; Gass and Selinker, 1994; Lightbrown and Spada, 1999). To this end, various pronunciation tutors have been developed to assist language learners in their pronunciation (Bernstein et al., 1999). One system (Knoerr, 1994) even allows students to view and compare waveforms from their own utterances with those of an idealized native speaker (for example, the teacher).

On the other hand, the use of animated humanoid agents as pronunciation tutors has increased lately. Whereas traditional animated agents were not designed to show fine-grained articulatory movements (Ladefoged, 1993), newer articulatorily-correct animated agents have been developed specifically for visually modeling correct pronunciation in three-dimensional space. Some agents even allow for control of such properties as intonational patterns, speech rate, and pitch levels.

Our pronunciation tutor relies on such an agent to model pronunciation in a communicative environment in which the negotiation of meaning is at the center of all practice.

Conversation agents

Beyond pronunciation tutors, language-based agents are becoming more useful in carrying out linguistic interactions with human users (Mostow and Aist, 1999; Hatless et al., 1999). Some interact with users in a virtual reality environment for specific tasks. Though the earliest conversational agents were purely textual, more recent ones interact via speech with users.

When conversational interactions take place, considerably more attention must be paid to pragmatic factors: discourse participants, context, previous utterances, participants' goals and assumptions, etc. This nontrivial aspect of interaction is often implemented via a dialogue manager.

A dialogue manager is the component of conversational agents that controls the flow of the dialogue, the higher-level decisions of how the agent should proceed in the conversation—what questions to ask or statements to make, and when to ask or make them (Rees, 2002). A dialogue agent is one that can interact and communicate with other agents in a coherent way, with coordinated utterances serving to accomplish the same end goal or to collaborate on the same topic.

The previously discussed core technologies have all been developed in different ways by various research groups. Their products span a wide range of components and associated functionality. In this section we discuss which specific components we have chosen to use in our work, along with the rationale behind their selection.

The speech toolkit

We use the CSLU/OGI Speech Toolkit (Cole, 1999) as our speech processing platform of choice for several reasons. First, this toolkit supports both speech recognition and text-to-speech synthesis. It has also included Baldi, an articulatorily-correct animated humanoid agent (a "talking head") whose movements can be closely controlled.[1] We have found Baldi's articulation modeling capabilities to be useful in modeling pronunciation in applications to be described shortly. Another benefit of using the OGI toolkit is that the programming environment consists of a user friendly Rapid Application Developer (RAD) component. This is an object-oriented graphical interface consisting of different widgets that can be placed on a canvas to create various interactive speech-based dialogue scenarios. RAD also allows adjustment of numerous low-level aspects of the speech recognizer and synthesizer performance via a menu-driven environment. With RAD the developer can specify the linguistic properties of an interactive scenario: context-free phrase-structure grammars for recognizing utterances, lexicons and vocabularies for word spotting, and user-specifiable phonemicizations for any desired words. Finally, the toolkit is freely available, widely used, and actively supported.[2]

It should be mentioned, though, that while the OGI toolkit has been used extensively for our research and pedagogical work, our work involves the development of additional resources that are as general as possible so that our results could be implemented on other platforms if necessary.

The dialogue move engine

Dialogue move engines (DME's) are increasingly popular in designing and implementing conversational scenarios. One approach, Trindi, addresses task-oriented instructional dialogue (Larsson, Ljunglöf, Cooper, Engdahl, and Ericsson, 2000). Its associated toolkit, TrindiKit, enables developers to build and experiment with dialogue move engines and information states.

It supports the design of a general dialogue system architecture: information state formats, update rules, algorithms, and dialogue moves. The system developer must define task-specific update rules, discourse moves, and utterance structure. The common ground between discourse participants is tracked by the system as much as possible, including agendas, shared assumptions, and shared referents.

Typical approaches to dialogue managers include finite state models, form fillers, and belief-desire-intention models. The finite state approach uses a different node to represent each possible state in the conversation. Each node then precisely dictates the system output used at that point in the conversation–the system's response. The finite state model utilizes user-input to determine which transition to follow, from the current state to some new state in the system. This produces "canned" dialogues, in effect, because the programmer must predetermine acceptable input and what output will be generated. This type of approach makes the dialogue managers grow exponentially large as the desired complexity for conversations grows. It does have the advantage of being a quick system to build, though it does not allow for much human user control over the flow of the conversation.

The use of forms extends the finite state model to allow mixed-initiative dialogues (both the human user and the system could help decide the next state in the dialogue). Instead of specifying all the states in the system, developers specify a set of inputs desired from the user. For example, if the user is asking about airplane flight information, the set of inputs might include: destination, departure city, dates, whether there will be a return flight, and the class. A form-based approach to dialogue management would accept as much information from the user at once as the user desired (user control), but would also generate questions (system control) based on the next empty element in the set of desired inputs until all the slots were filled. This

allows for a more robust system, than the finite state model does, but does not explain the motivation for each step in the dialogue.

Belief-desire-intention models were developed to give context to a dialogue and to provide an explanation of the human user's goals in communicating. Trindi (Larsson, 2004) is an example of this type of dialogue manager. Based on the observed exchanges in the conversation, dialogue move engines update the current information state in the dialogue manager and select the next appropriate move. The information state is the dialogue manager's method of modeling its perception of both the system and user goals in conversing. It can model information as it becomes apparent what both participants understand in the course of a dialogue. This gives a richer capability to modeling context and motivation in dialogue management techniques. Other approaches to dialogue management include discourse plans and recipes (Green, 2002).

With DME's it is possible to implement various kinds of conversations. System-initiated conversations put the system in charge, and the human participant is relegated to simple answers to questions. Human-initiated conversations put the system in the role of question-answerer or respondent to actions requested of the user. In mixed-initiative interactions, both participants (i.e. the system and the human) share the initiative as the conversation unfolds. Clearly mixed-initiative discourses are the most engaging to a human to the degree that they can be coherent and sufficiently constrained. On the other hand, they are the most difficult to implement with a high degree of success. We discuss in this paper systems that we have developed with various types of initiative.

Prolog

The Prolog programming language was instrumental in providing to some of our applications the ability to do high-level knowledge-rich processing. Some advantages to using

Prolog are that it can be used to do forward inferencing; it can efficiently encode relationships and rules between sets of data; and it can efficiently match queries to those sets of data regarding relationships which might exist. Also, Prolog components of a multi-modal application can be tested prior to integration, thus aiding in the discovery of programming errors. Certain Prolog systems, such as SICStus, also come packaged with added functionality, including more built in predicates and a library of modules for interfacing with other programming languages such as Java or Tcl/Tk. This greatly increases the power of Prolog to be used for a myriad of different types of applications.

External Knowledge Sources

In order to support a dynamic, realistic interactive environment, real-world knowledge about such topics as language, geography, and events of interest to a user is necessary. Extensive hand-coding of such resources for one-time applications is prohibitive and tedious. On the other hand, appropriate resources are becoming increasingly available for public use. The applications that we have developed make use of some well-structured databases and other knowledge sources.

- The UCI zoo database contains information for about a hundred animal instances, with features for salient properties like number of legs, fur or feathers, etc. Though used primarily for machine learning applications, this database is useful (as are others in this repository)[3] for dialogue purposes in our applications.
- The freely available CIA World Factbook[4] is a rich source of information about different countries of the world: each country's bordering countries, major industries, and climate.

● Genealogical information is rich with low-level data such as dates, names, locations, family relationships, and documentary references. A GEDCOM file is a standardized format for encoding and exchanging such information[5].

● WordNet (Fellbaum, 1998) is a freely available lexical database[6]... It provides large-scale coverage of lexical relationships such as synonymy, homonymy, hyponymy, and meronymy for words and their various senses.

● Our university's online events calendar is an example of web-based information we have used in our applications. The calendar lists such events as music concerts, sporting matches, and theater productions. The information is semi-structured and hence can be fairly easily queried in order to access the data. Its domain-specific, closed-world nature makes it ideal for the applications discussed below.

● The internet itself is a large-scale repository of usable information for conversational tasks. Content concerning almost any topic can be freely accessed, manipulated, and organized in a form that will allow for easy integration into conversational applications.

APPLICATIONS: INTEGRATED SOLUTIONS

In this section we survey applications that reflect our integration of the core technologies, components, and knowledge sources mentioned above. As with any software integration effort, issues of modularity, interface mechanisms, and data structures have been paramount. Fortunately, our calculated choice on which toolkits and knowledge sources to use have minimized the amount of integration work necessary.

For example, the Toolkit Command language (Tcl) is widely used to integrate various computer applications and toolkits, acting as "glue" between the various software components

and knowledge sources. Tcl is used in the OGI toolkit, and is supported by the dialogue move engine we used. Given our use of various engines and web-based knowledge sources, the use of sockets is also critical for establishing interprocess communications.

Pronunciation Tutor

Our pronunciation tutor is an application that combines the OGI Toolkit with multimedia images in interactive practice in which the learner must be able to hear and produce certain sound distinctions in a second language in order to perform a communicative task. It is based on a technique for teaching second language pronunciation, developed originally by Bowen (1972) and expanded by Henrichsen, Green, Nishitani, and Bagley (1999). The learner is introduced to a sound distinction through a brief story presented in narrative form with pictures. Through the story a plausible but ambiguous sentence (e.g. “The blacksmith hits/heats the metal.”) is introduced in which two meanings are possible based on the sound distinction in question.

(Place figure 1 about here.)

After telling the story, the agent articulates one member of the sentence pair while displaying two pictures and the learner must choose the appropriate picture to represent the meaning of the sentence. Then the learner must produce (i.e. speak) the appropriate sentence as the agent displays one of the two pictures... The system recognizes the user's answer and the agent comments on its (in)correctness. The system then starts an interactive activity where images are presented to the user who must describe them to the system. Correct responses are met with congratulations and positive feedback, while incorrect responses are met with hedging, requests for clarification, reformulations, etc. Clearly this approach assigns all discourse initiative to the

system. Figure 1 shows a dialogue structure automaton for a task which helps learners distinguish between the vowels in "hit" and "heat". We have created dozens of such pronunciation stories teaching a variety of sound contrasts in English and Spanish. In our experience, current speech technology is capable of discriminating between most but not all of the sound contrasts of interest.

Empirical research in pronunciation instruction, whether with live teachers or with CALL tutors suggest that there is a vast difference between developing mechanical articulation skills in a drill environment, and producing correct pronunciation in a communicative context... What is needed, then, is not more sophisticated ways of drilling students in the articulation of particular segmental or even suprasegmental features in isolation. Such practice must be integrated into a communicative context in order for it to get instantiated in productive spontaneous speech. (Morley, 1987; Celce-Murcia, Brinton and Goodwin, 1996) This is what we attempt to do with our pronunciation tutor. As was mentioned earlier, comprehensible input is provided through the narrative by the agent, accompanied by visuals. The learner interacts with the system both verbally and non-verbally and receives both positive and negative evidence to help in noticing and thus in reformulating the interlanguage system.

Language pedagogy task scenarios

Besides the pronunciation tutor, we have developed a number of scenarios for teaching English using an animated agent and speech processing, as well as some external sources such as WordNet. Following is a brief summary of some scenarios of various types:

- One such activity involves practicing buying tickets for various modes of transportation. The learner is guided through an interactive task-oriented scenario in which s/he learns how to

accomplish different subtasks relevant to buying a ticket such as differentiating amounts of money, telling time, and recognizing different places and types of transportation. An animated agent guides the learner through the entire activity, giving feedback regarding the interaction of the user. The activity culminates in an information gap activity in which the learner negotiates the ordering of a ticket from the agent. Again, we have attempted to bridge the gap between conventional CALL activities in which learners practice form in a rather mechanical way and real communication in which they apply those forms in a simulated communicative context with an agent. Again all the conditions for language acquisition are met in that learners receive comprehensible input specific to the task being learned. They then have the opportunity to practices producing the forms, first in a mechanical way with feedback... Then they are given the opportunity to employ those form in a communicative activity in which the agent, which controls discourse initiative, recognizes their negotiations and responds in an appropriate way.

● Another activity is a twenty-questions game. For this we use the hierarchical relations from WordNet to pursue increasingly specific semantic/hierarchic goals. An agent interacts with the learner regarding certain predefined topics (e.g. languages of the world, animals, plants, etc.) and tries to guess which item a user has chosen. Meanwhile, the user answers 'yes' or 'no' to increasingly specific questions posed by the agent. Or, if the user fails to understand a term used by the agent, s/he can ask for clarification. The game ends when the agent guesses the pre-selected object that the user has chosen. For example, the agent might say, 'Think of a language and I will guess what language you are thinking of.' The participant may choose any artificial or natural language and the agent will ask yes/no questions until he is able to identify the language. This activity is designed primarily to provide comprehensible input and the opportunity to

manifest comprehension by responding to yes/no questions. We have developed a limited number of applications of this game in which the roles are reversed so that the learner is asking the yes-no questions and the agent is responding. Initiative, then, can be implemented as either user-driven or system-driven.

● Another task-oriented scenario involves learning how to give directions. In the directions scenario, users are shown a map consisting of various streets and corresponding city blocks. The goal is to have a user guide a friend from a starting point on the map to the user's house located at a different spot on the map. As correct directions are given, progress is shown on the map (see Figure 2). After a number of refinements were made to the system, an overall interaction success rate of 93% was achieved. In this scenario, we have found the need to develop a hedging function to avoid giving false feedback to learners as they improve in their ability to give directions...

(Place Figure 2 about here)

Figure 2: Direction-giving exercise display at start (left) and after one utterance (right).

● Another task involves interacting with an agent in a lost-and-found-booth scenario. Here the user plays the role of a person who works in such a booth. The agent approaches and asks a series of questions to ascertain whether a given item (e.g. a large purple backpack) that he has lost has been turned in. The user must decide whether, among the items in the booth, one matching the agent's description is present. Each time the scenario is run, the booth is populated with a random set of items of various descriptions, assuring a novel situation each time. Sometimes the sought item is present in the booth, and sometimes it is not. In this scenario the

system controls the initiative, though we intend to implement the inverse scenario where the learner must seek an item from the booth and thus control the initiative.

Country Talk

● This application has an agent ask questions and make comments to a user concerning a specific country. The application makes use of the OGI toolkit, the CIA World Factbook, and WordNet. The system asks users where they are from; after they respond with a specific country, the system retrieves appropriate information enabling it either to ask a question concerning a specific fact about that country or to make comments about the same. Users then may answer any questions asked of them and the system may respond with either another question or a comment. The mixed-initiative nature of this task allows for interesting, realistic conversation, albeit on a narrow topic. An interesting challenge in this task was that some country names overlap with common nouns (e.g. Turkey). WordNet was employed to help distinguish the correct sense in such cases. Another issue was that because of the amount of data listed in the Factbook, a comprehensive vocabulary would have been prohibitively large. Accordingly, we implemented a dynamic recognizer in the system. The dynamic recognizer allows the system to anticipate what kind of vocabulary would be needed by the speech recognizer at a given stage in the conversation (i.e. when a given country was chosen). This implementation has increased the accuracy of the system significantly. A similar discussion engine on the topic of animals leverages the zoo database mentioned earlier.

GEDspeak and GEDquiz

● GEDspeak is a speech-based application designed to enable a user to query information contained in GEDCOM files. The OGI toolkit along with its RAD canvas was used to structure

the dialogue and specify requisite vocabulary and grammar for spoken interactions. Middle-level functionality in system was supplied by data-specific interface routines... They allow the system to receive and formulate queries from the user, and then send those queries to a library of Tcl routines that directly access the GEDCOM file for relevant information. The correct answer is then sent back to the dialogue level for response generation. In GEDspeak scenarios the initiative belonges to the user of the system. A related system called GEDquiz focuses on how GEDCOM data can be used to drive an interactive natural-language game where the system has the initiative. The goal was to develop a system that allows a user to assimilate, in a fun way, a global impression from the myriad of low-level facts contained in a typical GEDCOM file.

(Place figure 3 about here)

Figure 3: GEDquiz system architecture: the engine mediates between genealogical and real-world knowledge sources, the conversational agent, and a dialogue move engine.

A GEDCOM file is supplied to the engine, which then parses out the file's contents and stores it as a database of Prolog assertions. Then a set of pre-specified inferential relationships is automatically generated by the system. Questions might involve specific data items about an individual (e.g. "Where was your paternal grandfather born?") or might be of a very global nature (e.g. "Name two of your ancestors who immigrated to America.").

At run-time a minimal amount of information about the user was also supplied to situate the user with respect to information in the file, and to ascertain the level of expertise of the user with respect to the data in question (e.g. minimal=very little, average, expert=very knowledgeable). This helps the system set an appropriate level of specificity and difficulty for the interaction.

Once the system has been initialized and the data compiled, the engine enters into an interactive, goal-directed dialogue with the user. The system presents to the user a series of family history questions for which one or more alternative answers have been determined from the fact base. The system employs the GoDIS/TrindiKit dialogue move engine to track the multi-participant goal-directed discourse. Questions are generated from propositional content of the knowledge base via a phrase-structure grammar designed specifically for the task. The system gauges the correctness of the user's response(s) for each question and responds accordingly.

Set-a-date program

- The set-a-date program combines our university's events calendar, a Prolog database of closed-world knowledge, and a dialogue move engine. With this application a user can query a database of campus events using speech. The dialogue move engine keeps track of specific event times, types, and locations. The user answers queries by the system about suggested possible events based on cost, preferences, time constraints, etc.

For this system-initiated application we coded up specific domain-dependant information into a Prolog database which provided a core set of commonsense knowledge available to the system. This included such facts as where particular rooms and buildings were located on campus, time scheduling conventions, and different categories of events (e.g. sports, lectures, musical events).

- Another application allows students to listen to weather reports and to practice saying numbers and times of the day. It begins by asking the learner for a U.S. zip code. The system then accesses a Web site for weather information for that zip code, parses out the information, and returns to the user. The user is then asked for a time frame (e.g. Friday afternoon), and the

system reports the weather forecast for that time period. Again, the learner is subject to system initiative in this type of interaction.

FUTURE WORK

The applications discussed in this paper show how integrating speech, dialogue, and knowledge representation technologies can result in highly interactive, dynamic, knowledge-rich, and realistic scenarios. These scenarios can be used in task-oriented applications for question answering and language instruction, as well as many other possible uses. Our work has sought a pragmatic balance between the current limited state of the art in these technologies on the one hand, and the unlimited possibilities for instruction and data access that spoken-language dialogue can provide.

We expect to pursue current directions in future work such as the following: (i) developing increasingly complex task scenarios; (ii) developing foreign-accented English acoustic models; (iii) integrating learning tasks and speech engines for other languages (including less-commonly taught ones) (iv) integrating speech components with richer off-the-shelf language tutoring environments (v) developing and supporting more complicated dialogue structure: multi-person conversations, more interaction error recovery, and other text-, corpus-, and web-driven interactions. All of these technologies are becoming viable as separate language processing paradigms, and we believe that our novel implementations will leverage the strengths of each approach. Each type of processing resource mentioned in this paper either models or directly interacts with a human user, and the integrated solution we propose and are developing addresses many of the pitfalls of current CALL approaches.

Our work in the immediate future involves going beyond simply bundling these components together to create a new systems architecture. We are also integrating these

resources into a fully functional large-scale CALL system (Elzinga, 2000) that, while innovative in its user feedback mechanisms (Parry, 2000), only uses the traditional, limited user interaction modalities (mouse clicks, icons, and keyboard input). We expect to see substantial gains in student learning due to this synergistic combination of proven and innovative practices and technologies.

CONCLUSIONS

In this paper we have outlined a theory of second language acquisition consistent with learner-centered task-based instruction. We have argued that there are three essential features of current SLA theory which must be incorporated into such a program: (i) it must provide ample opportunity for the learner to receive comprehensible input, (ii) it must provide opportunities to interact both at the comprehension level and at the production level, and (iii) through modeling and feedback, it must provide both positive and negative evidence which will enable learners to incorporate new patterns into developing interlanguage.

We have attempted to show how current technologies lend themselves to the construction of activities which meet the demands of such a theory. In particular we have shown that by having a humanoid agent speak directly to learners and guide them through activities rather than having them activate multi-media material through clicking on a series of buttons, the learner becomes a first/second-person participant in the communication process. This enhances opportunities to demonstrate that input is not only comprehensible but comprehended. In addition we have shown that current technologies can provide interesting and motivating opportunities for learners to participate in the negotiation of meaning with humanoid agents through the use of speech recognition, speech synthesis, dialogue move engines and external

knowledge sources. We have shown how such interaction can provide learners not only with large amounts of comprehended input and appropriately modeled speech, but also that it can give them significant opportunities to interact verbally and to receive feedback regarding the correctness of their attempts to communicate.

While most of our learning activities are currently in prototype form, we anticipate that in the near future we will incorporate them into a full-blown instructional system which will enable us to examine empirically the extent to which the activities achieve the intended goals.

REFERENCES

Beatty, K. (2003). Teaching and researching computer-assisted language learning. Longman: London.

Bernstein, J., Najmi, A. & Ehsani, F. (1999). Subarashii: Encounters in Japanese spoken language education. *CALICO Journal,* 16:361-384.

Bowen, J. D. (1972)... Contextualizing pronunciation practice in the ESOL classroom. *TESOL Quarterly, 6 (1): 83-94.*

Bygate. M., Skehan, P. & Swain, M. (2001). Researching Pedagogic Tasks: Second Language Learning, Teaching, and Testing. Applied Linguistics and Language Study. New York: Longman.

Celce-Murcia, M. Brinton, D. & Goodwin, J. (1996)... *Teaching pronunciation: A Reference for teachiers of English to speakers of other languages.* New York: Cambridge University Press.

Cole, R. (1999)... Tools for research and education in speech science. In *Proceedings of the International Conference of Phonetic Sciences*, San Francisco, CA, August.

Ehsani, F. & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning and Technology,* 2(1):45-60, July.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press

Elzinga, C. Bret (2000). Technology assisted language learning. Brigham Young University Department of Instructional Psychology and Technology: MS Project...

Fellbaum, Christiane. (1998). *WordNet: An electronic lexical database.* MIT Press, Cambridge, MA.

Gass, S. (1988). Integrating research areas: A framework for second language studies. *Applied Linguistics*, 9, 198-217.

Gass, S. (1997). *Input, Interaction, and the Second Language Learner.* Mahwah, N.J.: Lawrence Earlbaum Associates, Publishers.

Gass, S. (2003). Chapter 9: Input and interaction. In C. Doughty & M. Long (Eds) *The Handbook of Second Language Acquisition.* Madsen, MA: Blackwell Publishing, Ltd., 224-255.

Gass, S. & Selinker, L. (1994). *Second Language Acquisition: An lntroductory Course.* Hillside, New Jersey: Erlbaum,

Glass, J. (1999). Challenges for spoken dialogue systems. In *Proceedings of the 1999 IEEE ASRU Workshop,* Keystone, CO, December.

Green, N. & Lehman, J. F. (2002). An Integrated Discourse Recipe-Based Model for Task-Oriented Dialogue. *Discourse Processes*, 33(2).

Hall, J. K. & Walsh, M. (2002). Teacher-student interaction and learning. *Anual Review of Applied Linguistics*, 22, 186-203.

Henrichsen, L. E., Green, B. A., Nishitani, A., & Bagley, C.L. (1999). *Pronunciation matters: Communicative, story-based activities for mastering the sounds of North American English.* Ann Arbor: University of Michigan Press.

Harless, W. C., Zier, M. A.& Duncan, R.C. (1999). Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal,* 16(3):313-337.

Krashen. S. (1985). *The input hypothesis: Issues and implications.* New York: Longman.

Knoerr, H. (1994). *Elaboration d'un didacticiel pour l'enseignement de l'intonation en Français Langue Seconde.* CIRAL: Quebec.

Ladefoged, P. (1993). *A Course in Phonetics*. Harcourt, Brace and Jovanovich, New York, 2 edition. .

Larsson, S., Ljunglöf, P. Cooper, R., Engdahl, E., & Ericsson, S. (2000)... *Trindikit 2.0 manual. Technical report*... http://www.ling.gu.se/research/projects/trindi.

Lightbrown, P. M. & Spada, N. (1999). *How languages are learned.* Oxford: Oxford University Press.

Long, M. (1996)... The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (eds), *Handbook of Second Language Acqisition.* San Diego: Academic Press, 413-68.

Morley, J. (Ed...). (1987). *Current Perspectives on Pronunciation: Practices Anchored in Theory*. Washington, D.C.: TESOL.

Mostow, J. & Aist, G. (1999). Giving help and praise in a reading tutor with imperfect listening because automatic speech recognition means never being able to say you're certain. CALICO Journal, 16(3):407- 424.

Parry, Kent. (2000). Improving implementation fidelity of large scale, distributed, computer assisted learning systems using scalable automated feedback... Brigham Young University Department of Instructional Psychology and Technology: PhD Dissertation.

Pennington, M.C. (Ed.)... (1996). The power of CALL. Athelstan:Houston, TX.

Rees, R. D. (2002). Investigating dialogue managers: building and comparing FSA models to BDI architectures, and the advantages to modeling human cognition in dialogue. Brigham Young University Department of Physics: Honors Thesis.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition.* Rowley, MA: Newbury House, 235-253.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds...). *Principle and practice in applied linguistics: Studies in Honour of H. G. Widdowson*... Oxford : Oxford University Press, 125-144.

Swain, M., Brooks, L. & Tocalli-Beller, A. (2002)... Peer-peer dialogue as a means of second language learning. *Anual Review of Applied Linguistics*, 22, 171-185.

Zhao, Y. (2003). Recent Developments in Technology and Language Learning: A Literature Review and Meta-analysis. CALICO Journal, 21 (1). 7-27.

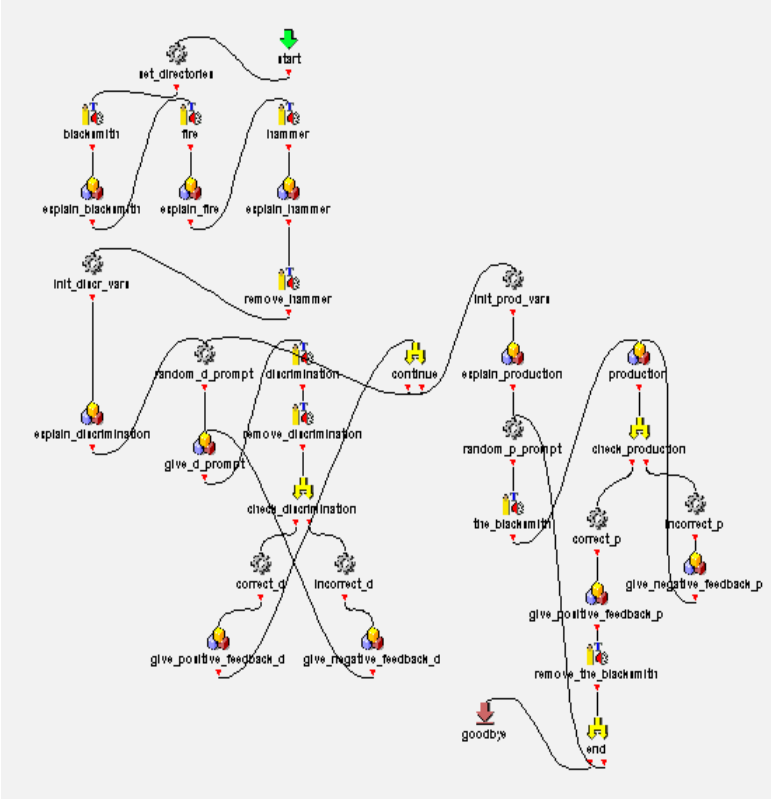

Figure 1: RAD canvas showing automaton for minimal pair pronunciation drill.

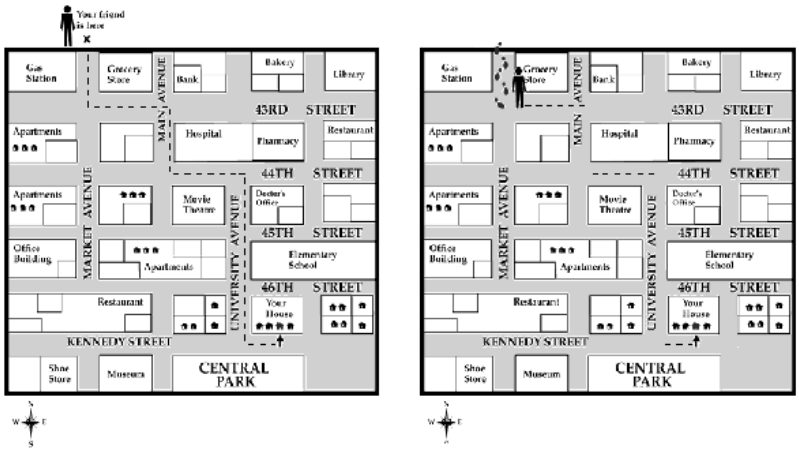

Figure 2: Direction-giving exercise display at start (left) and after one utterance (right).

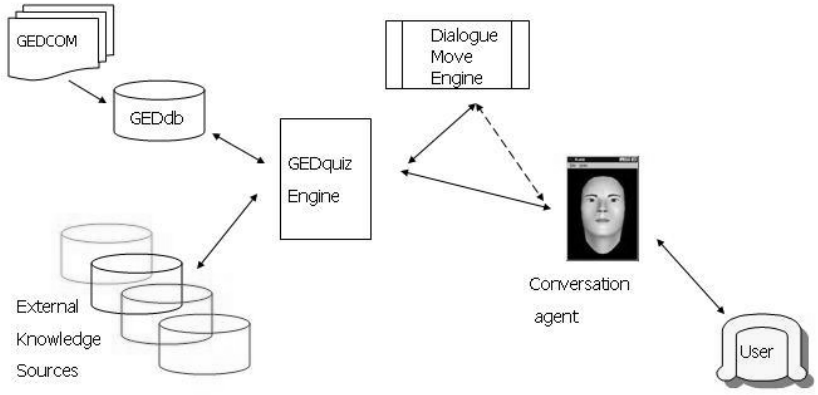

Figure 3: GEDquiz system architecture: the engine mediates between genealogical and real-world knowledge sources, the conversational agent, and a dialogue move engine.

[1] Recently Baldi has been replaced by a collection of other animated agents.
[2] See http://cslu.cse.ogi.edu/toolkit/.
[3] See http://www.ics.uci.edu/~mlearn/MLRepository.html.
[4] See http://www.cia.gov/cia/publications/factbook.
[5] See http://homepages.rootsweb.com/~pmcbride/gedcom/55gctoc.htm.
[6] See www.cogsci.princeton.edu/~wn.