

Intelligent Tutoring for Interpersonal and Intercultural Skills

**H. Chad Lane, Mark G. Core,
David Gomboc, Ashish Karnavat, & Milton Rosenberg**

**Institute for Creative Technologies
University of Southern California**

Marina del Rey, CA USA

**lane@ict.usc.edu, core@ict.usc.edu,
gomboc@ict.usc.edu, karnavat@ict.usc.edu, rosenberg@ict.usc.edu**

ABSTRACT

We describe some key issues involved in building an intelligent tutoring system for the ill-defined domain of interpersonal and intercultural skill acquisition. We discuss the consideration of mixed-result actions (actions with pros and cons), categories of actions (e.g., required steps vs. rules of thumb), the role of narrative, and reflective tutoring, among other topics. We present these ideas in the context our work on an intelligent tutor for ELECT BiLAT, a game-based system to teach cultural awareness and negotiation skills for bilateral engagements. The tutor provides guidance in two forms: (1) as a coach that gives hints and feedback during an engagement with a virtual character, and (2) during an after-action review to help the learner reflect on their choices. Learner activities are mapped to learning objectives, which include whether the actions represent positive or negative evidence of learning. These underlie an expert model, student model, and models of coaching and reflective tutoring that support the learner. We describe several other cultural and interpersonal training systems that situate learners in goal-based social contexts that include interaction with virtual characters and automated guidance. Finally, our future work includes evaluations of learning, expansion of the coach and reflective tutoring strategies, and integration of deeper knowledge-based resources that capture more nuanced cultural aspects of interaction.

ABOUT THE AUTHORS

H. Chad Lane is a Research Scientist at the Institute for Creative Technologies (ICT) at the University of Southern California (USC). His research is in the areas of intelligent tutoring systems and serious games with a particular focus on ill-defined domains. He holds a Ph.D. in Computer Science from the University of Pittsburgh where he worked in the Learning Research and Development Center (LRDC) on dialogue-based tutoring systems.

Mark G. Core is a Research Scientist at ICT/USC. He applies research in computational linguistics to the task of building intelligent tutoring systems that interact with students in natural language. His Ph.D. in Computer Science from the University of Rochester focused on natural language understanding and dialogue analysis.

David Gomboc is a Research Engineer at ICT/USC. He holds an M.Sc. in Computing Science from the University of Alberta-Edmonton where he worked on the tuning of evaluation functions for heuristic search.

Ashish Karnavat is a Research Programmer at ICT/USC. He holds an M.S. in Computer Science from the University of Memphis where he worked with latent semantic analysis and dialogue-based tutoring systems.

Milton Rosenberg is a Special Project Manager at ICT/USC. He holds a BS in Business Administration with over 20 years of experience managing multiple projects. Milton supervised the construction and global rollout of the Whirlpool manufacturing system and holds a patent in methods for maintaining computer software.

Intelligent Tutoring for Interpersonal and Intercultural Skills

H. Chad Lane, Mark G. Core,
David Gomboc, Ashish Karnavat, & Milton Rosenberg

Institute for Creative Technologies
University of Southern California

Marina del Rey, CA USA

lane@ict.usc.edu, core@ict.usc.edu,
gomboc@ict.usc.edu, karnavat@ict.usc.edu, rosenberg@ict.usc.edu

INTRODUCTION

“Ill-defined” domains, in contrast to those that are well-defined, are characterized by problems that tend to lack consistent, unambiguous, and generalizable solutions. One of the earliest characterizations of ill-defined problems came from Rittel & Webber (1973) in their research on social policy. They describe characteristics of “wicked problems” in the planning domain and contrast them with problems from more “benign”, well-defined domains such as mathematics, chemistry, and chess, all of which provide clear goal structure and conditions that indicate success (p.160). They suggest that wicked problems have no “definitive formulation” (p. 161), no “stopping rule” (p.162), no “true-or-false” but only “good-or-bad” solutions (p.162), no “ultimate test” (p.163), no “well-described set of permissible operations” (p. 164), among other undesirable traits.

Thankfully, domains extreme as social policy and planning are rare. It is debatable that *every* problem in the social policy space is as mysterious as Rittel & Webber claim. More likely is that domains fall within a spectrum between well-defined and ill-defined, consisting of aspects that are widely agreed upon and other “hot spots” representing areas of expert disagreement and of underspecified or tacit problem solving procedures. It is these amorphous sub-areas that present unique challenges to educators, educational technologists, and learning science researchers who seek to teach and understand how learning occurs in ill-defined domains. Here, proficiency tends to be less about applying formulaic patterns and more about observing, interpreting, hypothesizing, and adapting.

We have been engaged in research that seeks to apply intelligent tutoring system (ITS) techniques to guide learning in the domain of interpersonal and intercultural competence. While it may be possible to

divorce the two, our target has been on a learner from one culture who is interacting with a virtual human character from a different culture. Thus, it is important to repeatedly consider seemingly universal principles (e.g., “Listen and respond to what your counterpart says.”) as well as seemingly individual cultural factors (e.g., “Avoid sitting such that the soles of your feet face your counterpart.”; Nydell, 2006, p.63). Although cultural rules such as these often seem unambiguous, it is *potential disagreement between experts* that makes it difficult to encode formally. The approach we have taken is to keep this knowledge authorable so that the system can be configured to allow variance on cultural rules depending on an instructor’s or learner’s specific goals and particular geographic regions.

Not surprisingly, we have found that many traditional ITS components are applicable (e.g., Shute & Psozka, 1996) and provide an appropriate framework for delivering automated assessment and feedback on learner performance. It does however become necessary to augment some modules with more robust functionality. The rest of the paper explores the pedagogical and technical issues involved with building ITSs for ill-defined domains with the greatest emphasis on interpersonal and intercultural competence. We describe the Enhanced Learning Environments with Creative Technologies Bi-Lateral negotiation serious game (ELECT BiLAT), which is the target system for our ITS. We then describe the Intelligent Guided Experiential Learning (IGEL) ITS that provides guidance both during BiLAT meetings and after them via a reflective tutor that supports review and post-practice reflection. We conclude with a summary and discussions of related and future work.

TUTORING IN ILL-DEFINED DOMAINS

The ironic truth is that it is difficult to define ill-defined domains. The wicked problems of Rittel &

Webber (1973) situate them almost exclusively in contrast with well-defined problem spaces, and thus constitute a primarily negative definition (listing everything ill-defined domains do *not* consist of). Lynch et. al. (2006) provide an excellent summary of several other conceptions of what ill-defined domains are and are not. Although they do include some negative aspects like Rittel & Webber, they ultimately adopt a definition with three key characteristics of problems in ill-defined domains (p. 2):

1. they lack a definitive answer,
2. the answer is heavily dependent upon the problem's conception, and
3. problem solving requires both retrieving concepts and mapping them to the task.

Lynch et. al. also conclude that many existing ITS techniques are still appropriate, but that many lack sufficient flexibility to account for these additional characteristics. Assessment of student work is especially challenging given that classroom approaches tend to use open-ended questions and support solution variability (Ogan et. al., 2006). Lynch et. al. suggest that tutoring systems can still provide useful guidance even when it is not possible to maintain a deep model of the domain or perform detailed plan recognition of learner activities. In the case of ill-defined domains, it may even sometimes be impossible to track and assess actions with precision (else the domain would be well-defined), and so *knowledge-lean* methods may be appropriate for guidance. This is the approach taken by ELECT BiLAT.

ELECT BiLAT

The role of narrative and story is believed to have increased importance in ill-defined domains such as leadership (Sternberg et. al., 2000). Given that *experience* is the best way to acquire tacit knowledge, stories represent one way to describe experience, and are often used as a teaching tool – *case method* teaching approaches fit squarely into this category (e.g., Kim et. al., 2007). Interactive storytelling systems (e.g., Riedl & Stern, 2006 and Magerko et al. 2006) attempt to go one step further by placing the learner in the middle of a story with the ability to affect how it plays out through dynamic manipulations of the environment and events that take place. ELECT BiLAT (Hill et al. 2006) is also situates the learner in the middle of a story. It represents a dovetailing of several research efforts at the USC Institute for

Creative Technologies, including the intelligent tutoring research described in this paper, interactive narrative, games design, cognitive modeling, and virtual human research (Swartout et. al., 2006).

Interactive narrative

One approach to interactive storytelling is to situate the learner in an *emergent narrative* in which interactions and story unfold as a natural consequence of the learner's interactions with complex agents that reason about their goals, beliefs, and possible behaviors (Si et. al., 2005; Swartout et. al., 2006). With this rich artificial intelligence (AI) approach, there is the potential to build an infrastructure that automatically assesses student actions based on causal connections between these actions and reactions of the artificial agents. In an educational context, the key drawback of emergent narrative is that the experiences the learner ends up creating may not contribute to the intended or desired learning goals. Another hurdle is that it typically requires advanced AI skills to create the interactive agents.

To support ease-of-authoring of narrative structures and to achieve greater control of learner experiences, a more scripted approach was adopted for ELECT BiLAT. Interactive stories are conveyed through the dialogue utterances of virtual characters. Agent reactions can be augmented with scripted events to achieve specific goals in the interactive narrative. With this approach, knowledge-based assessment is not possible because of the reliance on raw English text (which is hard for an AI system to understand or explain) and parameter tweaking to produce realistic agent behavior. We discuss our intelligent tutoring work with the ELECT BiLAT simulation as a case study in how a scripted approach to interactive narrative must be augmented to allow for learner assessment and automated guidance.

Intercultural competence and negotiation

Achieving intercultural competence is a long and difficult process (Bennett, 1993) that requires advanced self-monitoring and self-assessment skills (Lane, 2007). The IGEL ITS supports the development of these *metacognitive* and intercultural skills, but currently provides limited support for negotiation guidance such as evaluating the non-cultural aspects of specific offers, agreements, and so on.



Figure 1. Screenshots of the research & preparation (left) and meeting (right) screens of ELECT BiLAT.

The authorable nature of BiLAT means that the general framework can be used for simulated meetings and negotiations with characters from any culture. The system models the distinct elements of bilateral negotiation as described by Wunderle (2005): *meeting preparation, dialogue between the learner and counterpart, actual negotiation, and a review of the meeting with your supervisor*. The last element typically involves the learner reporting back what happened during the meeting; however, we use this review instead as an opportunity for our reflective tutor to discuss with the student the cultural aspects and lessons learned from the meeting. Below we describe these aspects of bilateral negotiation and their instantiation in BiLAT in more detail.

Preparing for a meeting

BiLAT's simulation of meeting preparation was heavily influenced by the approach of Wunderle (2005). Wunderle's leader preparation sheet is a form designed to be filled out before a negotiator meets with a foreign counterpart. An electronic version of the form (figure 1, left) unifies the aspects of meeting preparation, including selecting intended outcomes, researching aspects of your counterpart such as possible impasses, and brainstorming about strategies for addressing these potential stumbling blocks. The preparation process can be thought of as a situation awareness building activity: information from the available resources must be combined and synthesized in a way that supports decision making and problem solving.

As is the case for other parts of the simulation, meeting preparation relies heavily on authored English text and scripted events. Researching a counterpart consists of selecting information resources represented with pictures (e.g., laptop computer in figure 1) and English names (e.g., SIPRNet, Battle Update Briefing, newspaper). Within each of these resources the learner selects a particular source (e.g., you can use SIPRNet to get an update from a fictional S2 officer) represented with a name and picture. The simulation gives the learner the resulting information in the form of text, then the learner "commits" this to the leader preparation sheet if they agree with it. Instances of resources are associated with costs (in units of time) and learners are given a fixed amount of time to prepare for the meeting. Meeting preparation is linked to the dialogue and negotiation through the process of "unlocking" dialogue and negotiation actions. Learners are given some default actions to perform during dialogue and negotiation, but typically to successfully conduct a meeting they must earn actions by recording crucial pieces of information into the leader preparation sheet.

In future work, we hope to enhance the knowledge representation in meeting preparation to enable assessment and feedback for the student in this phase of the simulation. This effort will involve developing a model of building situation awareness and answering questions such as how should learners prioritize their choices given limited time, and in the case of conflicting information what criteria impact what to believe.

BiLAT Meetings

A BiLAT meeting consists of two modes: dialogue and negotiation. The dialogue window is shown on the right side of figure 1 and consists of a menu of conversational actions (e.g., questions and statements – lower left in the screenshot) as well as physical actions (e.g., removing sunglasses, giving a gift). The character responds to the learner with a synthesized voice and physical gestures. Although there are dozens of variables governing the actions of the character, the variable of primary importance is *trust* (displayed in the upper left corner of the screen). Although characters may say nice things or display anger in their responses, *trust is the persistent record of how well you have used your interpersonal and intercultural skills to win the character to your side*. We will not cover the negotiation aspect of the game as our focus currently is interpersonal and intercultural skills. The two aspects of the game are related in that trust is a major factor in whether the character will agree to negotiate and what deals they will accept (a mistrusting character will demand more).

Because our ITS focuses on interpersonal and intercultural skills, the mechanics of the dialogue phase are crucially important. For the most part, the character is reactive and chooses its response from a set of hand authored alternatives based on a virtual dice roll. The dice roll simulates the uncertainty of human behavior which requires more advanced AI cognitive and emotional modeling to simulate in a principled way. The responses for a learner action include the text spoken by the character, appropriate gestural feedback, and possible changes to the trust variable. Because the game actions presented to the user are textual strings, it is the game author that assigns meaning to them in crafting the responses: a positive response means no loss of trust is possible, a negative response means no gain in trust is possible, and a mixed response can either increase or decrease trust depending on the dice roll. A well known weakness of this kind of approach is that it leaves out why actions are good, bad, or mixed, and it may be the case that a character's response is part of the story or character personality. For example some characters may appreciate gifts and flattery more than others who may even be insulted. These decisions facilitate rapid scenario development, but limit the depth and explanation power behind intelligent assessment.

A second aspect of the dialogue phase is context-dependent reactions. Because of the importance of time in cultural modeling, there are distinct time spans corresponding to a business period and social periods before and after business. During authoring, actions are annotated with the times for which they are appropriate. There is also a more general scripting mechanism where authors can specify a set of conditions where the character will take the initiative rather than simply responding to learner actions. The conditions make use of internal simulation variables tracking what actions the user has selected in the past and trigger the character to reward, punish or challenge the learner. In challenges, the learner must select a response to the character's initiative from a small set of options. These scripted encounters are easily authorable, but give the ITS little to work with since it is not obvious what set of actions triggered the script and why, and how to assess learner's choices from the presented responses. In the next section, we discuss the learning objectives for ELECT BiLAT and the process we developed for establishing connections between these game elements both in terms of positive and negative support that the student has met the objective and how the game element and learning objective are related.

PEDAGOGICAL CONTENT

Typically when developing a virtual practice environment, a small number of training objectives guide the process. These training objectives may come from an existing course, practice environment, or from instructional designers who are creating new course materials. As discussed in the previous section, the type of simulation built determines how the learner can be assessed. In ELECT BiLAT, the artificial agents are able to behave appropriately, but do not have the capability to introspect on their behavior, explain how it relates to the learner's actions, or trace back how a course of events relates to training objectives. Instead these relationships are delineated explicitly during pedagogical authoring, which in turn drive the actions of the ITS.

Linking training objectives to game content

As a starting point, we were fortunate to receive a hierarchy of training objectives based on a cognitive task analysis of U.S. Army officers negotiating with Arab counterparts. We used a modified version of the

hierarchy as a way of grouping game actions based on how they support the top-level training objectives. To further support the ill-defined aspects of the domain, we defined a typing system for the lowest-level elements in the hierarchy to facilitate the ITS' understanding of the different categories of activity. These include:

1. *required steps*: agreed upon standards and sequences of actions that are best
2. *usual steps*: actions that are generally good to take; do them when time permits
3. *rules of thumb*: conditions that should one should strive to meet; general guidelines
4. *avoids*: actions or gestures that will have negative consequences

These represent consensus results of a cognitive task analysis of subject-matter experts from real-world bilateral meetings. A few, such as *avoid* and *usual* steps are specifically designed to handle ill-defined issues. For example, a learner may succeed in achieving an objective, but if a few *avoids* were violated, the overall solution is not considered optimal. Similarly, steps in the *usual* category are things some experts do while others do not – but that all agree is something that at least won't hurt one's cause. The difference between *required* steps and those that are less agreed upon helps the ITS assess the learner and give feedback geared toward more hardened aspects of the domain.

Once the learning objectives are defined the next step is linking this hierarchy of learning objectives to game elements. For example, one action allows the learner to talk about historical events. At the lowest level, this game action addresses a topic that Arabs appreciate given the elevated importance of history (Nydall, 2006, p. 35). At a middle level of abstraction, this is a type of small talk (e.g., other possibilities could include sports, weather, etc.) and the top level training objective is learning to develop relationships through socialization. Thus, we link the game action, "talk about Iraqi history" as positive support of this training objective. The action is positive evidence that the student understands socialization but also that small talk regarding history is a specific way to achieve this goal. To support a meeting, all player dialogue actions must be linked with learning objectives in this way. Scripted counterpart actions and player responses to these actions must also be considered in this linking process.

There is no limit in theory to how many links can be created between a game action and the learning objective hierarchy. It was often necessary to assign two links to some actions that had both positive and

negative elements. For example, if a learner promises to give a character what he wants, this will help build your relationship but it could lead to problems down the road (e.g., his neighbors may grow jealous and demand the same favors from you). An abstracted view of these linked structures is shown in figure 2. When action is tagged as both positive and negative, it usually represents an ill-defined aspect of the domain, as discussed earlier. There are usually reasons to take the action (or respond) and reasons not to do it – which is best and what the payoff is depends both on the specific problem being solved, and how the learner has conceptualized the problem (Lynch et. al., 2006).

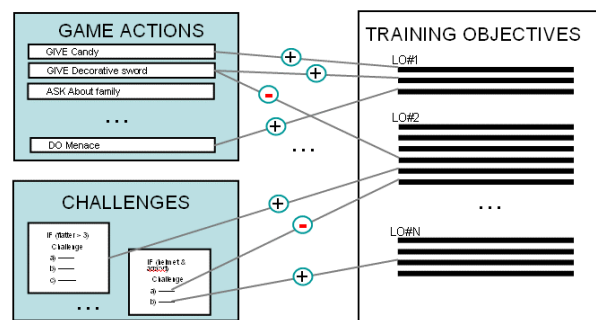


Figure 2. Linking learning objectives to game content.

COACHING

As discussed, during the meeting phase of BiLAT, the learner must repeatedly select conversational actions with the goal of achieving game objectives (see figure 1, right). These actions must be selected according to cultural norms and at appropriate times. The coach is an observer of these learner activities and is given the opportunity to interject messages before and after each action. Currently, these messages show up as dialogue utterances that the student can read.

Assessing meeting actions

Each time an action is taken in a meeting or the learner responds to a character challenge, the expert model is called to judge the quality of the action. It labels each action as *correct*, *incorrect*, or *mixed*. This assessment is done in two stages:

1. The action is checked to see if it is phase appropriate.
2. (If no phase mismatch is found) The linked learning objectives are determined along with the associated polarities.

As discussed, meeting phases are windows of time during a meeting that define when certain categories of actions are appropriate or not. They are culture dependent and the expert module dynamically assigns negative assessments when a wrong action is chosen. This is implemented via a link to the appropriate training objective. For example, a learning objective that says that one should learn not to rush into business (Nydall, 2006, p.58) implies conversational actions are needed early in meetings. When a business-phase action is taken in an opening phase, then, the expert model will record an “incorrect” and save negative evidence toward that learning objective.

When an action is phase-appropriate, it is assessed based on the authored polarity of the learning objective links as shown in figure 2. If all links are of positive polarity, then the action is classified as correct. Similarly, if all are negative, the action is incorrect. Any combination that includes links of both polarities, the action is classified as mixed. Each of these assessments is recorded in a rudimentary *student model*, which consists of a mirror image of the full set of training objectives with counters attached that keep track of positive and negative instances of each.

Feedback content

As part of the pedagogical authoring process, coach feedback and hinting content needs to be created. We use two levels of feedback: training objective and action level. There is one universal set of training objectives that cover all characters in a scenario. Each character, on the other hand, has its own set of available actions and challenges (although some are shared) and many that must be “unlocked” during the preparation phase. This is also displayed in figure 2. Authors must “attach” feedback messages to training objectives, actions, and challenges. Since all actions link back to training objectives, which are general across all characters, the coach can use this difference to deliver vague feedback, or jump to the action level to deliver more specific, directive feedback.

Four categories of coach messages exist to be authored (an author enters only what he or she desires):

- **hints:** forward-looking suggestion on what is an appropriate next action
- **warnings:** forward-looking suggestion to avoid a certain action or class of actions

- **negative feedback:** backward-referring comment describing what was wrong or the problem with a certain action.
- **positive feedback:** backward-referring praise for a good action taken.

Some combinations turn out to be incoherent – if there is no situation when an action can be incorrect, for example, negative feedback associated with it will be unreachable. Mixed actions are ideal for using both negative and positive feedback messages, however, to help the student understand the reasons for and against taking certain actions.

Hints are generated by consulting the expert model. The expert model runs a search algorithm that (1) identifies all available actions (some may have remained “locked” if preparation was inadequate), (2) filters out actions that are not phase-appropriate, (3) filters out previously performed actions, (4) identifies those actions that are positively linked to learning objectives, (5) sorts the remaining actions according to the *type*. Actions tagged as *required* are given precedence for hinting over those that are *usually* good, and so on. This step either returns a list of required actions or a list of actions associated with other types. The current hinting strategy is to first give help at the learning objective level (i.e., abstract level) and in step (6) the learning objectives for each action output by 5 are retrieved. The coach will randomly hint on one of these learning objectives that has not been mentioned previously. If no novel hint is available, the coach will hint randomly on one of the actions returned by step 5. Hints constitute one of the four potential messages the coach can deliver – two before the student acts (hints and warnings) and two after (positive and negative feedback as applicable). Clearly, a coach that intervenes this much will be a distraction to the learner, which brings us to timing.

The timing of feedback

There are seemingly limitless combinations of feedback timing (and content) rules a coach could follow. Our current implementation supports a variety of options including very simple patterns, such as:

- Give negative feedback every third error.
- Give feedback (either) every second turn.
- Give feedback only on certain training objectives (i.e., categories in the domain).



Figure 3. Screenshots of the scoreboard (left) and reflective tutoring (right) screens of ELECT BiLAT.

These simple mechanisms – counting and filtering – use the rudimentary student model to keep track of the learner’s progress and when the test is passed (e.g., a third error was committed), generation is as simple as looking up the message in a table.

The default is to give action-level messages only, but if a *progressive strategy* is selected, the coach will begin with the more vague, training objective-level messages, then transition to the more specific action-level messages if another error is made later in the same category. The most advanced strategy the coach is able to follow is a *model-scaffold-fade* algorithm, inspired by cognitive apprenticeship (Collins et. al., 1989). In this model, the coach provides forward looking guidance and feedback very frequently at first (the modeling and scaffolding) then pulls the support away as the student demonstrates more successful interactions until, finally, the student is successful on his or her own with no help. This algorithm is configurable in terms of how quickly fading occurs and to the level of abstraction in the content.

REFLECTIVE TUTORING

After a meeting ends, the reflective tutor receives a history of the actions that occurred including learner actions, character actions (including changes to trust), and coach actions. In addition, each of the learner actions will have been assessed and assigned a series of positive/negative links to learning objectives. The role of the reflective tutor is to engage learners in reviewing their performance.

Scoreboard

The first element of the review is the overview shown on the left side of figure 3 that we refer to as the scoreboard. The color coded boxes in the middle of the screen display the support shown by the student for the top-level general training objectives and just above these is a short summary of the overall performance (e.g., “you did a good job building a relationship in general but you should avoid tactics such as insults and threats”). The final element of the overview is a history of the learner actions labeled as correct, incorrect, or mixed as determined by the coach and expert model during the meeting.

The second part of the review is interactive as shown on the right of figure 3. The avatar for the reflective tutor stands on the left of the screen in back of a dialogue window where he communicates with the student through text. The current configuration reviews the meeting in roughly chronological order and as the tutor mentions part of the dialogue, it is replayed in the right window (both in video form and as a scrolling transcript). The video replay is designed to resemble current video player software and has a progress bar across the top with hash marks corresponding to student actions. The hash marks are color coded (green=correct, yellow=mixed, and red=incorrect) and light up when the tutor is discussing them.

The review is planned in advance although in future work we aim for the flexibility to change this plan (called the agenda of the reflective tutor) during the review. The planning process is primarily a grouping task and starts with the game action having the most

negative links to learning objectives. This game action forms the “seed” of an agenda item. The agenda item’s nucleus is formed from the seed and actions adjacent in time to the seed that share links to the same learning objectives. The next step is searching the history for actions sharing links to the same learning objectives that are not temporally adjacent. These actions form the satellite of the agenda (note, these terms come from Rhetorical Structure Theory <http://www.sfu.ca/rst/>). The idea is that although the agenda will be traversed chronologically there is no need for the tutor to repeat itself each time it encounters a link to a learning objective that it has mentioned before.

To conduct the interactive review, the reflective tutor uses a rule-based system (a derivative of our earlier work, Core et. al., 2006). The primary elements on the left-hand side of these rules are patterns of learning objectives matching against the learning objectives associated with the current agenda item. The right-hand side of the rules contain templates that the system’s natural language generator translates into English text. There are two major types of templates:

1. **“say” templates:** produces utterances that are simply meant to be read by learners, and
2. **multiple-choice-questions:** requires the learner to answer a deep reasoning question from a short list of possible answers.

We currently use a simple approach to deal with cases where learners select an incorrect answer; the learner gets a second try and if they fail a second time they are told the correct answer. The natural language generator has a number of notable features. We construct templates for all game actions allowing the tutor to discuss the actions in the review in different tenses and forms (e.g., good job talking about Iraqi history, Farid appreciates Iraqi history, and Farid appreciated the small talk). The templates are generic so the name of the counterpart (e.g., Farid) is inserted automatically and not hard coded.

RELATED WORK

Very few tutoring systems for interpersonal and cultural skills have been built. Four examples of significant efforts are the Tactical Language and Culture Training System, TLCTS (Johnston et. al., 2005), the VECTOR system (McCollum et. al., 2004), the Virtual Objective Structured Clinical Exam (VOSCE) system (Johnsen et. al., 2006), the

responsive virtual human technology (RVHT) from Hubal & Frank (2001).

In the TLCTS's mission practice environment, learners explore a virtual town speaking to locals in Arabic via speech recognition technology, seeking to accomplish goals such as getting the names of contacts and directions to help find them. TLCTS' mission practice environment includes a coach in the form of a helpful aide who accompanies the learner and gives suggestions during the game. Conversational hints are generated in a way similar to ELECT BiLAT by associating them with scripted dialogue interactions and are given only when requested by the learner.

In the VECTOR system, learners also explore a virtual foreign town; this time speaking to locals by selecting English utterances from a menu with the goal of finding a bomber and stopping him from attacking his next target. VECTOR's tutor is described as monitoring the game not giving suggestions or conducting an AAR.

In the VOSCE system, learners diagnose pain in a virtual patient through a standardized series of questions and observations. VOSCE's tutor conveys system messages (e.g., introduction and closing messages) and reports questions the learner should have asked (but did not) and the correct diagnosis. VOSCE (as well as the other systems described) has addressed issues of evaluating learner actions and giving short feedback but not the problem of planning and conducting reviews.

RVHT provides practice environments for interaction skill training. It has been applied to security domains such as interviewing suspects, hostage negotiations, and military checkpoint conversations. RVHT shares many similarities to the work of Swartout et. al. (2006) through the use of behavior modeling and natural language dialogue. RVHT has also been applied to create virtual tutors that interact with students to answer questions and requests for help. They also maintain a student model of the learner that tracks progress through a mission and maintains an estimation of student understanding.

CONCLUSIONS & FUTURE WORK

Intelligent tutoring systems (ITS) are beginning to emerge that operate in ill-defined domains. We have

described an ITS built for the domain of interpersonal and intercultural competence that consists of traditional ITS components (action assessor, student model, expert model, etc), but with several extensions to handle actions with mixed results and easily changeable representations for efficient updates to expert knowledge.

Two experiments to evaluate learning in ELECT BiLAT and of the ITS are currently underway. The first experiment is ongoing with learners using the system as part of a larger curriculum. This experiment uses a situational judgment test (SJT) that presents the learner with near-transfer situations, similar to those from the game, and asks them to rate a list of responses on a Likert scale as ideal or poor choices (Legree & Psotka, 2006). An answer key was created by using the average responses of three subject-matter experts. In a second experiment intended to identify optimal patterns of pedagogical interventions, we are comparing different coach configurations and inspecting (1) subsequent game performance, as well as (2) results on culture quizzes and a subset of the full SJT. We also hope this experiment will shed light on the question of distraction; i.e., how much coaching help is too much in ELECT BiLAT?

Development-wise, we are working in three primary directions. First, the current reflective tutoring model does not include tactics that consider coaching actions. For example, if a learner does not successfully follow up on a hint, it may be evidence that the hint was not understood or that the learner did not believe it was good advice. In either case, this is likely good evidence that discussing the issues and related learning objectives would be helpful. Second, to support more effective coach and reflective tutor feedback decision making, we are investigating more advanced forms of student modeling that persist over time and include estimations of competence in domain knowledge. This work is being performed in collaboration with the Interactive Storytelling Architecture for Training (ISAT) project (Magerko et al. 2006). Finally, and in collaboration with the BiLAT team, we are taking steps to integrate the deeper cognitive models of domain knowledge, emotions, and culture from virtual humans (Swartout et. al., 2006) and using them to provide improved feedback and explanation (extending our earlier work with virtual humans described in Core et. al., 2006).

The goal of this work has been to create an authorable practice environment for interpersonal and cultural skills, using storytelling techniques to situate and motivate practice and applying the principles of guided learning via an ITS. The flexibility resulting from mixed-result assessment and easily changed content has helped us create a system that can handle a variety of ill-defined aspects of the domain. We are now seeking to move to a stronger knowledge-based approach that will facilitate deeper feedback and tutorial explanations, but without losing the properties of authorability and robustness towards mixed-assessment.

ACKNOWLEDGEMENTS

We thank other members of the ELECT team at ICT as well as the broader Learning with Adaptive Simulations for Training (LAST) team, including members from the U.S. Army Research Institute (ARI), U.S. Army Research Lab – Human Research & Engineering Directorate (ARL HRED), U.S. Army Training and Doctrine Command (TRADOC), and U.S. Army Research, Development, & Engineering Command (RDECOM). We specifically thank Dr. Jim Belanich of ARI for his insights on the design of the IGEL intelligent tutor and scoreboard, Eric Forbell, lead engineer on the ELECT team, and Julia Kim, project director, for their support in integrating the IGEL ITS with BiLAT.

The project described here has been sponsored by RDECOM. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

REFERENCES

- Bennett, M.J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the Intercultural Experience* (21-71). Yarmouth: Intercultural Press.
- Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, 453-493.
- Core, M.G., Traum, D., Lane, H.C., Swartout, W., Gratch, J., van Lent, M., & Marsella, S. (2006).

- Teaching negotiation skills through practice and reflection with virtual humans. *SIMULATION*, 82, 685-701.
- Hill, R.W., Belanich, J., Lane, H.C., Core, M.G., Dixon, M., Forbell, E., Kim, J., & Hart, J. (2006). Pedagogically Structured Game-based Training: Development of the ELECT BiLAT Simulation. Proceedings of ASC06: *The 25th Army Science Conference*. Dept. of the Army.
- Hubal, R.C. & Frank, G.A. (2001). Interactive training applications using responsive virtual human technology. Proceedings of I/ITSEC: *Interservice/Industry Training, Simulation, and Education Conference*.
- Johnsen, K., Dickerson, R., Raji, A., Harrison, C., Lok, B., & Stevens, A. (2006). Evolving an immersive medical communication skills trainer. *Presence: Teleoperators and Virtual Environments*, 15, 1.
- Johnson, W.L., Vilhjalmsson, H., Marsella, S. (2005). Serious games for language learning: How much game, how much AI? Proceedings of AIED 2005: *The 12th International Conference on Artificial Intelligence in Education*, 306–313. Amsterdam: IOS Press.
- Kim, J.M., Hill, R.W., Zbylut, M.L. (2007). Film-based cases in Interactive Learning Environments for Leaders. In *Proceedings of the Workshop on Narrative Learning Environments* at the 13th International Conference on Artificial Intelligence in Education (AIED). Marina del Rey, CA.
- Lane, H.C. (2007). Metacognition and the development of intercultural competence. In *Proceedings of the Workshop on Metacognition and Self-Regulated Learning in Intelligent Tutoring Systems* at the 13th International Conference on Artificial Intelligence in Education (AIED), 23-32. Marina del Rey, CA.
- Legree, P. & Psotka, J. (2006). Refining situational judgment test methods. Proceedings of ASC06: *The 25th Army Science Conference*. Dept. of the Army.
- Lynch, C.F., Ashley, K., Alevan, V., & Pinkwart, N. (2006). Defining “ill-defined” domains; A literature survey. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at Intelligent Tutoring Systems* at the 8th International Conference on Intelligent Tutoring Systems, 1-10.
- Magerko, B., Stensrud, B. and Holt, L. S. (2006). Bringing the schoolhouse inside the box - A tool for engaging, individualized training. Proceedings of the 25th Army Science Conference. Orlando, FL. November 27-30, 2006.
- McCullum, C., Deaton, J., Barba, C., Santerelli, T., Singer, M.J., & Kerr, B.W. (2004). Developing an immersive, cultural training system. Proceedings of I/ITSEC: *Interservice/Industry Training, Simulation, and Education Conference*.
- Nydall, M.K. (2006). *Understanding Arabs: A guide for modern times* (4th ed.). Boston: Intercultural Press.
- Ogan, A., Alevan, V., & Jones, C. (2006). Culture in the classroom: Challenges for assessment in ill-defined domains. In *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at Intelligent Tutoring Systems* at the 8th International Conference on Intelligent Tutoring Systems, 92-100.
- Riedl, M.O. & Stern, A. (2006). Believable Agents and Intelligent Story Adaptation for Interactive Storytelling. Proceedings of *The 3rd International Conference on Technologies for Interactive Digital Storytelling and Entertainment*.
- Si, M., Marsella, S.C., Pynadath, D.V. (2005). Thespian: Using multi-agent fitting to craft interactive drama. Proceedings of AAMAS: *International Conference on Autonomous Agents and Multi-Agent Systems*, 21-28.
- Shute, V.J. & Psotka, J. (1996). Intelligent tutoring systems: Past, present, and future. In D. Jonassen (Ed.), *Handbook of research on educational communication and technology*. New York: Simon & Schuster Macmillan.
- Sternberg, R.J., Forsythe, G.B., Hedlund, J., Horvath, J.A., Wagner, R.K., Williams, W.M., Snook, S., & Grigorenko, E.L. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.
- Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., & Rickel, J. (2006). Toward virtual humans. *AI Magazine*, 27, 2, 96–108.
- Wunderle, W. (2005). *Engagement: A preparation guide*. Presentation at the USC Institute for Creative Technologies, 6 July 2005.