

Interactive Phrasebook – Embedding Human Social Dynamics in Language Training

Peggy Wu, Christopher Miller
Smart Information Flow Technologies
211 N 1st St. Suite 300 Minneapolis, MN 55401
{PWu, CMiller} @ siftech.com

Abstract - Today, many individuals must interact with others who not only use different languages, but have fundamentally different belief structures and concepts of trust and collaboration. While proficiency in a language is important, it is only one aspect of effective communication. Knowledge about culturally influenced social norms, or etiquette, directly impacts interaction outcomes. Computer-based training systems are a promising and cost effective solution for providing simulated social interactions for a learner to both practice language use and observe cultural nuances in the dynamics of an interaction. A computerized actor which displays social characteristics consistent with its cultural background can provide low cost training and rehearsal opportunities. Such a system would allow humans to train with realistically-behaving simulations in high fidelity situational contexts, and interactions with virtual agents that reflect the types of people they will be required to work with. The authors have adapted a well grounded socio-linguistic model of human-human interaction, and applied it in a language and culture training application. Interactive Phrasebook leverages work in a scalable, customizable, and computationally tractable model of a universal code of conduct that defines how social dynamics such as power and social distance manifests into the expression and interpretation of interaction behaviors.

I. INTRODUCTION

An undertaking in learning a new language goes hand in hand with learning about a new culture. Language not only conveys semantics, but is also an important channel with which humans detect and express affective states in culturally sensitive ways. In a wide range of training domains, high levels of teacher student interactions have been shown to impact a student's learning experience [1], but this is costly and impractical for most students in the domain of

language training. To reduce the cost of high human teacher-to-student ratios, computerized simulations are being developed to replace or supplement human instructors.

Computerized simulations (such as OneSAF or the Future Immersive Training Environment for military training) may provide simulated environments with 'indigenous' virtual actors, creating rehearsal opportunities for the human player. However, they are extremely complex and expensive to build, maintain, and update. Efforts in advancing serious games [2], such as DARPA's DARWARs effort have resulted in the development of highly regarded games with dramatic effects—but corresponding research [3] shows that games are effective only if they are precisely tuned for their environment and for the material to be taught.

Many of these training environments lack the interpersonal facets of interaction that are critical to understanding a culture. Our work in examining human-machine collaborations suggest that a theoretically based model of human-human interaction can be used to inform virtual agent behavior so they exhibit elements of the socially sophistication shown by the humans they are simulating. A computational model can dynamically generate agent behaviors to react to a human learner based on contextual factors that are socially and culturally biased. Below, we present our work in transferring sound socio-anthropological research on human-human social protocol to the domain of human-NPC interactions for language and training tools.

II. SOCIAL PROTOCOL AND BEHAVIOR

We believe that to by embedding social dynamics of the culture of interest in language training, the learner not only gains knowledge in language, but also the ability to use and interpret it correctly in different social contexts. Observable behaviors (i.e. an actor's

spoken utterances, facial expressions, and gestures) can provide insight into the actor’s cultural biases, motivations and thought processes. By interpreting observable behaviors against a framework of expectations, one can draw conclusions about social norms that are culturally and situationally influenced. One such framework is politeness, a well-studied phenomenon in anthropology, sociology and linguistics having to do with the processes by which we determine and manage the “threat” inherent in communication and interaction between two intentional agents in a social interaction—that is, actors that are presumed to have goals and the potential to take offense at having those goals thwarted in any interaction where those intentional attributes are relevant [4, 5]. In this sense, politeness is the method by which we signal, interpret, maintain and alter power relationships, familiarity relationships and interpretations of the degree of imposition of an act.

A seminal body of work in cross-cultural linguistics is Brown and Levinson’s theory of politeness [6]. In short, they believe any communication exchange implies that the speaker is causing some burden, or *face threat*, upon the hearer. By simply speaking, the speaker has demanded the hearer’s attention, regardless of the message’s content, and has therefore performed a *Face Threatening Act (FTA)*. The fact that the speaker’s message is, for example, a request for fifty dollars, further adds to the threat. The amount of face threat is dependent on situational variables. The same request is more threatening if asked of a stranger rather than a friend, or a superior rather than a peer.

To mitigate the threat the speaker imposes, politeness strategies are used. Over years of cross linguistic and cross cultural studies, Brown and Levinson created an extensive catalogue of politeness strategies used by different cultures, ranging from “being apologetic”, to “using honorifics”. Further, they hypothesize that in nominal cases, the face threat of the interaction should be balanced with an equal amount of politeness

III. A COMPUTATIONAL MODEL OF SOCIAL DYNAMICS

Brown and Levinson themselves do not operationalize these parameters; instead, they are offered as

qualitative constructs. In order to apply this theory to computerized simulations, we have, under various sources of funding, modified the Brown and Levinson’s to create a computationally tractable and scalable. This work has resulted in a content coding scheme, a software architecture, and set of algorithms which we call the *Etiquette Engine™*. We use the term *etiquette* to refer to the set of expectations about observable behaviors that allow interpretations to be made in a cultural context. The Etiquette Engine takes into account influences at the individual and cultural levels to calculate how an action may be interpreted in terms of politeness, and therefore whether it follows the cultural code of conduct. It would be trivial to craft agent behaviors that are consistently rude or overly polite, but most communication strategies occur in the dynamic *believable* middle ground, e.g. being assertive without being rude or being polite without appearing subservient. This is the danger zone where many miscommunications and cultural misunderstandings occur, and is the area of focus for Etiquette Engine.

Brown and Levinson’s expression of the degree of face threat of an action is provided by the function:

$$W_x = D(S,H) + P(H,S) + R_x \quad (1)$$

- W_x is the ‘weightiness’ or severity of the Face Threatening Act (FTA), the degree of threat.
- $D(S,H)$ is the *social distance* between the speaker (S) and the hearer (H). It decreases with contact and interaction, but may also be based on factors such as membership in the same family, clan or organization.
- $P(H,S)$ is the *relative power* that H has over S.
- R_x is the *ranked imposition* of the raw act itself and may be culturally influenced. As an example, the imposition of asking someone for \$5 is less than the imposition of asking someone for \$500.

Based on the severity of face threat, various politeness strategies are selected to mitigate the threat. More precisely, Brown and Levinson claim that the degree of face threat posed by an act must be balanced by the value of the politeness behaviors used if the social status quo is to be maintained.

That is:

$$W_x \approx V(A_x) \quad (2)$$

where $V(A_x)$ is the combined redressive value of the set of politeness behaviors (A_x) used in the interaction.

Expanding on the Brown and Levinson calculation of face threat, we implemented the use of weights for each component to allow the possibility to value D, P, and R differently, and added another component, character (C), to represent the speaker's personal tendencies to be polite or rude. In addition, we created a coding strategy and manual with which independent coders can evaluate and assign numeric scores to P, D, R, C, as well as politeness strategies in order to add content to the simulation. While this mechanism was only tested with three raters, its Robinson's A correlation of .931 was well above traditional thresholds of .7-.8 for multiple-judge rating correlations [7].

In validating our model [8], we found that subjects perceived levels of politeness as designed, even when utterances came from machines with no physical embodiment, suggesting that 1) a computational model can be used to manipulate and predict perceived politeness, and 2) subjects readily anthropomorphize machines, even if interactions are limited to voice or text messages, confirming Nass's findings that humans readily generalize patterns of conduct and expectations for human-human interaction to human-computer interaction—a relationship he calls “the media equation” [9]. The human tendency to anthropomorphize machines and interact with them socially suggests that it may be possible for social skills learned through human-computer interactions to be transferred to human-human interactions.

A “socially intelligent” character would take offense believably if not addressed in a culturally appropriate fashion or may appear recalcitrant or ignorant when it is merely trying to follow its culturally-derived notions of polite turn taking in discourse. Human interactions involve highly subjective variables that are not clearly defined and are difficult to quantify. Interactions also occur in complex contexts, with the variables within each context containing many possible interpretations.

Currently, agent behaviors are frequently emulated in hand-written scripts and simple, locally-relevant rules. Such an approach is labor intensive, extremely brittle, and incapable of conveying the nuances that make it difficult to learn about cultural differences in the first place. While etiquette is not sufficiently rich or precise to explain all aspects of behavior, we believe it is particularly well suited for culture and language learning. Below we describe one implementation of the Etiquette Engine in a language training tool called Interactive Phrasebook.

IV. INTERACTIVE PHRASEBOOK

We have created a demonstration of the Etiquette Engine, called Interactive Phrasebook, a tool for rehearsing conversations with virtual agents that represent indigenous actors. Rather than an elaborate virtual environment with computationally intensive graphics, Interactive Phrasebook focuses on verbal behaviors and gestures displayed in text (with references available in audio and video clips) to reduce processing requirements and provide a highly portable PDA form factor. Similar to a paper phrasebook, a student can select various predetermined scenarios (e.g. going to a restaurant, asking for directions) and is then presented with relevant phrases in a foreign language. However, unlike a paper phrasebook, Interactive Phrasebook offers an interactive experience and a number of features that allow the user to tailor the scenario so it better reflects the real life situation the user wants to rehearse. Below, we step through a use case of Interactive Phrasebook to demonstrate the user interactions and features.

Interactive Phrasebook exposes several social dynamics variables that may affect interactions. The user can customize each virtual agent's ethnic culture, character (i.e. individual politeness tendencies), occupation which has implications on power, and relationships between agents in terms of power and social distance (see Figure 1).



Figure 1 Interactive Phrasebook - Customizing virtual agents

These parameters influence each virtual agent's interpretation of behaviors, and ultimately shape their exhibited behaviors.

The user then selects a scenario and may change the level of imposition associated with the objective the user is trying to achieve. The user interface provides a brief

outline of the scenario. In this use case, the user has selected the scenario "asking for a favor" (see Figure 2). This generic description may be applicable for small favors (with a low imposition) or large favors (with a high imposition). The level of imposition affects the scenario outcome.

The user can then proceed to the Play Mode, where he or she can interact with one or more virtual actors to complete the objective of the chosen scenario (see Figure 3). The user first chooses an actor to speak to from the addressee combo box. A list of possible phrases then populates the list below, sorted by the amount of politeness. Note that the most polite phrase may not always be the most appropriate phrase, and it is the goal of the user to select the most appropriate phrase based on the contextual variables set in the prior screens. The phrases may be displayed in the foreign language or its English translations, with accompanying audio of utterances and video of gestures available for the user's reference. When the user highlights a specific phrase, a gesture check box may appear below, indicating that the displayed non-verbal gesture may be performed with that phrase. The user can then select the phrase and gesture by pressing the select button, similar to sending a text message. The virtual addressee responds with a text

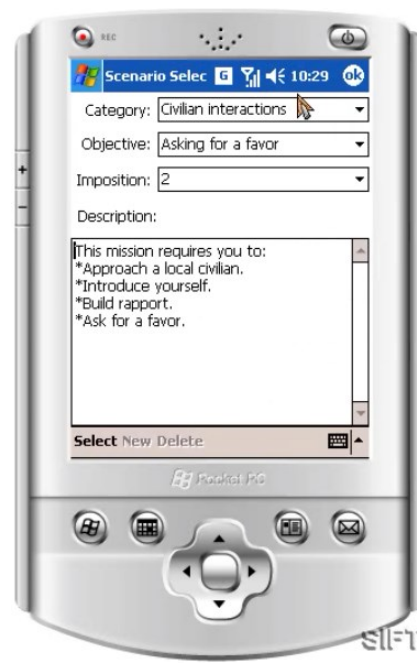


Figure 2 Interactive Phrasebook – Selecting a scenario

phrase below. The address's interpretation of the appropriateness of the user's action is displayed in the lower bar, providing the user with some insight on how his or her phrase and gesture combination was perceived. The etiquette engine calculates this appropriateness as the imbalance

between the expected and actual politeness based on the relationship variables (e.g. power and social distance) between the virtual agent and the speaker. In other words, appropriateness is a measure of politeness relative to expectations, which in turn is dependent on the situational contexts. If, for example, the user is in a heated argument, the most appropriate action may be to use a "rude" phrase. If



Figure 3 Interactive Phrasebook - Play Mode

the human is overly polite instead, the large difference between expected and exhibited behavior would be shown as inappropriately polite. A nominal (most appropriate) action would result in the pale green bar in the middle of the display, extending either to the left (slightly rude) or the right (slightly polite), whereas

overly rude or overly polite actions would be indicated by yellow or red bars extending either to the left or right. This visualization provides instant feedback to the user, akin to reading a facial expression of a smile or frown, but presented in an unambiguous way. The user can immediately refer to the rating to “see” what the NPC was “thinking” in terms of its reaction to the user’s utterances.



Figure 4 Interactive Phrasebook - Review Mode

Figure 4 shows the Review Mode, where the user may review the previous thread of interactions with the addressee. The user can select any phrase and gesture combination and view the expected politeness, the actual politeness, and the imbalance or appropriateness. In addition, the user can choose different

observers from the

observer pull down menu. Since expectations depend on the relationships between actors, and the perceived relationship between actors can vary with different observers, the perceived politeness of an interaction also changes with the observer. For example, an interaction may be seen as appropriate by an observer who is aware that the actors involved are close friends, whereas the same interaction can be seen as inappropriate by an observer unaware of the context. The ability to change observers allows the user to explore how the same utterance may be interpreted differently by individuals of varying ethnicity, power distance, or social distance. By explicitly showing the student when and how social expectations are violated, Interactive Phrasebook can provide insight into the behaviors and underlying motivations of the virtual agents, and therefore the specific actors representative of those the student will encounter.

V. CONCLUSION

The notion of selective-fidelity for simulations [10] places focus on aspects of the real situation that make a functional difference from the human user’s perspective. In the culture and language training domain, we believe social dynamics that affect language use and interpretation is critical to success of the interaction and therefore requires high fidelity in simulation. Many cultural misunderstandings occur because of nuanced differences in the interpretation of actions. While politeness of an utterance by itself does not convey all aspects of social protocols, it is a universal dimension that most people can relate to, and provides a framework for intuitively understanding cultural influences. Politeness conveys interpretations of social relationships that are heavily influenced by culture and has been shown to significantly influence interaction outcomes (e.g. trust, affect, compliance)[7**Error! Bookmark not defined.**]. However, politeness in cultural and language training is often distilled to a minimal set of faux-pas’s as opposed to an overarching theme in interaction. A computational approach with strong theoretical underpinnings can equip the training mechanism with culturally-biased models to provide dynamic user experiences, providing a way to not only learn the most culturally appropriate actions, but also reason about and anticipate reactions and next steps. Training tools like Interaction Phrasebook can allow the learner to practice interactions, as well as provide a structure for the human user to understand how their actions may be viewed as violations of social expectations, the severity of the violation, and how they may singularly or cumulatively alter the scenario outcome.

ACKNOWLEDGMENT

The authors would like to acknowledge the support for Interactive Phrasebook by the U.S. Army Research Laboratories under contract # W911QX-07-C-0039.

REFERENCES

-
- [1] Graesser, A. C., Person, N. K., and Magliano, J. P., (1995). *Collaborative dialogue patterns in naturalistic one-to-one tutoring*. Applied Cognitive Psychology, 9, 1-28. Accessed: http://www.cs.pitt.edu/~chopin/references/tig/grae_sser_person.pdf
- [2] Michael, D. & Chen, S. (2005). *Serious Games: Games that educate, train & inform*. Florence, KY:Course Technology.
- [3] Greitzer, F., Kuchar, O., & Huston, K. (2007) Cognitive science implications for enhancing training effectiveness in a serious gaming context. *J. Educational Resources in Computing., Vol. 7, No. 3, Article 2.*
- [4] Dennet, D., *The Intentional Stance*, MIT Press; Cambridge, MA. 1989.
- [5] Goffman, Erving (1955). "On Face-work: An Analysis of Ritual Elements of Social Interaction." *Psychiatry: Journal for the Study of Interpersonal Processes.* 18(3), 213-231.
- [6] Brown, P. & Levinson, S. *Politeness: Some Universals in Language Usage*. Cambridge Univ. Press; Cambridge, UK. 1987.
- [7] Miller, C., Wu, P., Funk, H., Wilson, P., Johnson, W. L. (2006) A Computational Approach to Etiquette and Politeness: Initial Test Cases. In *Proceedings of 2006 BRIMS Conference*, May 15-18, Baltimore, MD
- [8] Miller, C., Wu, P. and Funk, H. (2007). A Computational Approach to Etiquette and Politeness: Validation Experiments. In *Proceedings of the First International Conference on Computational Cultural Dynamics*, August 27-28; College Park, MD.
- [9] Reeves, B. and Nass, C. 1996. *The Media Equation*. Cambridge University Press, Cambridge, UK.
- [10] Schricker, Bradley C., Schricker, Stephen A., and Franceschini, Robert W. 2001. Considerations for Selective-Fidelity Simulation. In *Proceedings of SPIE – The International Society for Optical Engineering* Volume 4367, p 62-70, Enabling Technology for Simulation Science V, Alex F. Sisti; Dawn A. Trevisani, Eds.