POLLy: A Conversational System that uses a Shared Representation to Generate Action and Social Language

Swati Gupta
Department of Computer
Science, Regent Court
211 Portobello Street
University of Sheffield
Sheffield, UK
s.gupta@dcs.shef.ac.uk

Marilyn A. Walker
Department of Computer
Science, Regent Court
211 Portobello Street
University of Sheffield
Sheffield, UK
m.walker@dcs.shef.ac.uk

Daniela M. Romano
Department of Computer
Science, Regent Court
211 Portobello Street
University of Sheffield
Sheffield, UK
d.romano@dcs.shef.ac.uk

Abstract

We present a demo of our conversational system POLLy (POliteness in Language Learning) which uses a common planning representation to generate actions to be performed by embodied agents in a virtual environment and to generate spoken utterances for dialogues about the steps involved in completing the task. In order to generate socially appropriate dialogue, Brown and Levinson's theory of politeness is used to constrain the dialogue generation process.

1 Introduction

Research in Embodied Conversational Agents (ECAs) has explored embedding ECAs in domain-specific Virtual Environments (VE) where users interact with them using different modalities, including Spoken Language. However, in order to support dialogic interaction in such environments, an important technical challenge is the synchronization of the ECA Spoken Interaction module with the ECA non-verbal actions in the VE. We propose an approach that uses a common high level representation which is broken down to simpler levels to generate the agents' verbal interaction and the agents' non-verbal actions synchronously for task-oriented applications that involve performing some actions to achieve a goal, while talking about the actions using natural language.

In previous work, Bersot et al (1998) present a conversational agent called Ulysses embedded in a collaborative VE which accepts spoken input from the user and enables him or her to navigate within the VE. They use a 'reference resolver' which maps the entities mentioned in utterances to geometric objects in the VE and to actions.

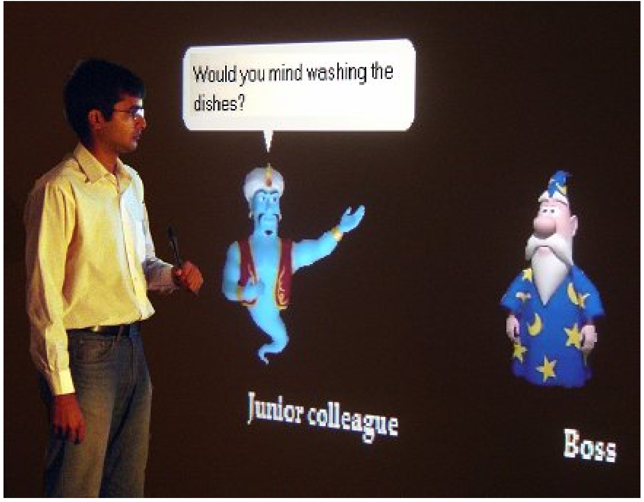

Figure 1. A user interacting with the Agents

Max, a VR based conversational agent by Kopp et al (2003) allows multimodal conversational abilities for task-oriented dialogues in virtual construction tasks. It builds on a database of utterance templates which contain the verbal part, augmented with accompanying gestures and their cross-modal affiliation. In order to deal with the vagueness of language in specifying spatial relations in virtual space, the K_2 system (Takenobu et al 2003) proposed a bilateral symbolic and numeric representation of locations, to bridge the gap between language processing (a symbolic system), and animation generation (a continuous system). K_2 extracts a user's goal from the utterance and translates it into animation data. The FearNot! demonstrator by Paiva et al (2005) provides training to kids against bullying via virtual drama in which one virtual character plays the role of a bully and the other plays the role of victim, who asks the child for advice. FearNot!'s spoken interaction is template-based where the incoming text from the child is matched against a set of language templates. The information about the character's action is defined in a collection which contains the utterance to be spoken as well as the animation. Eichner et al (2007) describe an application in which life-like characters present MP3 players in a virtual showroom. An XML scripting language is used to define the content of the presentation as well as the animations of the agents. A more expressive agent, Greta, developed by Pelachaud et al (Poggi et al, 2005) is capable of producing socially appropriate gestures and facial expressions, and used is in an evaluation of gesture and politeness as reported in Rehm and André (2007).

Since these ECAs function in scenarios where they interact with the world, other agents, and the user, they must be 'socially intelligent' (Dautenhahn, 2000) and exhibit social skills. Our work is based on the hypothesis that the relevant social skills include the ability to communicate appropriately, according to the social situation, by building on theories about the norms of human social behaviour. We believe that an integral part of such skills is the correct use of politeness (Brown & Levinson, 1987; Walker et al 1997). For instance, note the difference in the effect of requesting the hearer to clean the floor by saying 'You must clean the spill on the floor now!' and 'I know I'm asking you for a big favour but could you kindly clean the spill on the floor?'

According to Brown and Levinson (1987) (henceforth B&L), choices of these different forms are driven by sociological norms among human speakers. Walker et al (1997) were the first to propose and implement B&L's theory in ECAs to provide interesting variations of character and personality in an interactive narrative application. Since then B&L's theory has been used in many conversational applications e.g. animated presentation teams (André et al 2000; Rehm & André, 2007), real estate sales (Cassell & Bickmore, 2003), and tutorials (Johnson et al, 2004; Johnson et al, 2005; Porayska-Pomsta 2003; Wang et al 2003). Rehm & André (2007) show that gestures are used consistently with verbal politeness strategies and specific gestures can be used to mitigate face threats. Work in literary analysis has also argued for the utility of B&L's theory, e.g. Culpeper (1996) argues that a notion of 'impoliteness' in dramatic narratives creates conflict by portraying verbal events that are inappropriate in real life. Thus impoliteness often serves as a key to move the plot forward in terms of its consequences.

This demo presents our Conversational System POLLy which produces utterances with a socially appropriate level of politeness as per the theory of Brown and Levinson. We have implemented POLLy in a VE for the domain of teaching English as a second language (ESL). It is rendered in our VE RAVE at Sheffield University as well as on a normal computer screen, as explained in section 3. Figure 1 shows a user interacting with POLLy in RAVE. Since RAVE is not portable, we will demonstrate POLLy on the computer screen where the user will be able to verbally communicate with the agents and the agents will respond with computationally generated utterances with an appropriate level of politeness as per a given situation.

2 POLLy's Architecture

POLLy uses a shared representation for generating actions to be performed by the ECAs in the virtual domain on one hand, and on the other, for generating dialogues to communicate about the actions to be performed. It consists of three components: A Virtual Environment (VE), a Spoken Language Generation (SLG) system and a Shared AI Planning Representation for VE and SLG as illustrated in Figure 2.

A classic STRIPS-style planner called GraphPlan (Blum & Furst, 1997) produces, given a goal e.g. cook pasta, a plan of the steps involved in doing so (Gupta et al., 2007). POLLy then allocates this plan to the Embodied Conversational Agents (ECA) in the VE as a shared collaborative plan to achieve the cooking task with goals to communicate about the plan via speech acts (SAs), needed to accomplish the plan collaboratively, such as Re quests, Offers, Informs, Acceptances and rejections (Grosz, 1990; Sidner, 1994; Walker, 1996). It also allocates this plan to the SLG which generates variations of the dialogue based on B&L's theory of politeness that realizes this collaborative plan as in (André et al,2000;Walker et al, 1997).

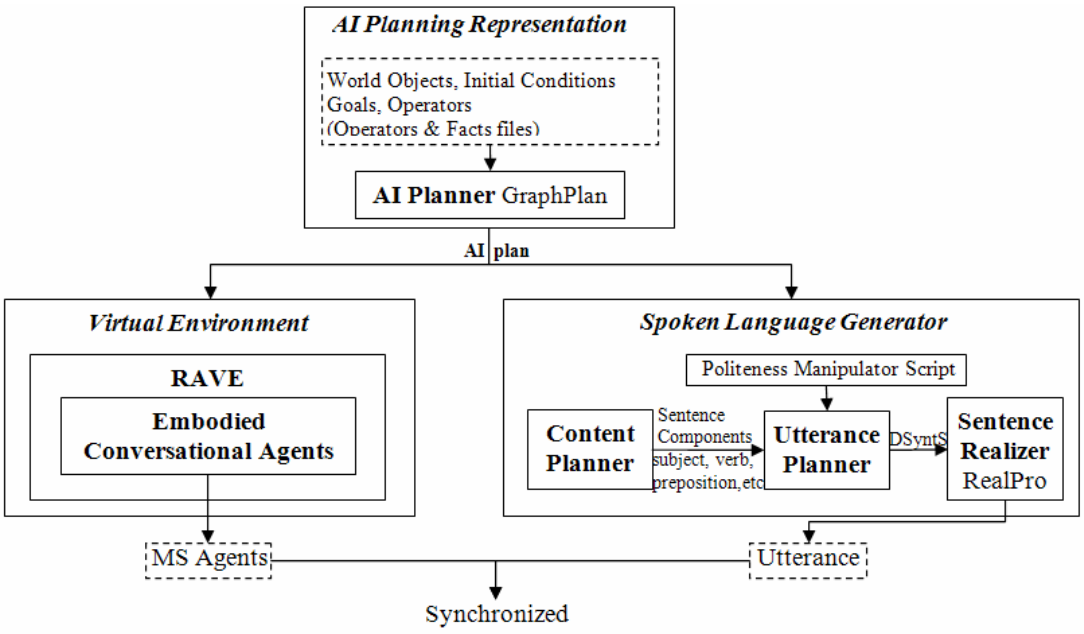
AI Planning Representation
World Objects, Initial Conditions
Goals, Operators
(Operators & Facts files)
AI Planner GraphPlan
AI plan
Virtual Environment
RAVE
Embodied Conversational Agents
Spoken Language Generator
Politeness Manipulator Script
Content Planner
Sentence Components subject, verb, preposition,etc
Utterance Planner
DSyntS
Sentence Realizer RealPro
MS Agents
Utterance
Synchronized

Figure 2: POLLY's Architecture

The SLG (Gupta et al., 2007) is based on a standard architecture (Dale & Reiter, 1995) with three components: Content planning, utterance planning and surface realization. See Figure 2. The politeness strategies are implemented through a combination of content selection and utterance planning. The linguistic realizer RealPro is used for realization of the resulting utterance plan (Lavoie & Rambow, 1997), which takes a dependency structure called the Deep-Syntactic Structure (DSyntS) as input and realizes it as a sentence string. The Content Planner interfaces to the AI Planner, selecting content from the preconditions, steps and effects of the plan. According to B&L, direct strategies are selected from the steps of the plan, while realizations of preconditions and negating the effects of actions are techniques for implementing indirect strategies. The content planner extracts the components of the utterances to be created, from the plan and assigns them their respective categories, for example, lexeme get/add under category verb, knife/oil under direct object etc and sends them as input to the Utterance Planner. The Utterance Planner then converts the utterance components to the lexemes of DSyntS nodes to create basic DsyntS for simple sentences, which are then transformed to create variations as per B&L's politeness strategies, with the 'politeness manipulator script'. For realizing these B&L strategies, transformations to add lexical items such as 'please', 'if you don't mind', and 'mate' were added to the DSyntS to make a sentence less or more polite.

Some example dialogues are shown in section 3. In the VE, the human English language learner is able to interact with the Embodied Conversational Agent and plays the part of one of the agents in order to practice politeness real-time.

2.1 Brown and Levinson's theory

B&L's theory states that speakers in conversation attempt to realize their speech acts (SAs) to avoid threats to one another's face, which consists of two components. Positive face is the desire that at least some of the speaker's and hearer's goals and desires are shared by other speakers. Negative face is the want of a person that his action be unimpeded by others. Utterances that threaten the conversants' face are called Face Threatening Acts (FTAs). B&L predict a universal of language usage that the choice of linguistic form can be determined by the predicted Threat θ as a sum of 3 variables: P: power that the hearer has over the speaker; D: social distance between speaker & hearer; and R: a ranking of imposition of the speech act. Linguistic strategy choice is made according to the value of the Threat θ. We follow Walker et al.'s (1997) four part classification of strategy choice.

The Direct strategy is used when θ is low and executes the SA in the most direct, clear and unambiguous way. It is usually carried out either in urgent situations (Please Help!), or where the face threat is small as in "I have chopped the vegetables" or if the speaker has power over the hearer, "Did you finish your homework today?"

The Approval strategy (Positive Politeness) is used for the next level of threat θ - this strategy is oriented towards the need for the hearer to maintain a positive self-image. Positive politeness is primarily based on how the speaker approaches the hearer, by treating him as a friend, a person whose wants and personality traits are liked, for ex. by using friendly markers "Friend, would you close the door?"

The Autonomy Strategy (Negative Politeness) is used for high face threats, when the speaker may be imposing on the hearer, intruding on their space or violating their freedom of action. These face threats can be mitigated by using hedges, "I wonder if you would mind closing the door for me," or by minimizing imposition, "I just want to ask you if you could close the door."

The Indirect Strategy (Off Record) is the politest strategy and is used when θ is greatest. It usually has more than one attributable intention so that the speaker removes himself from any imposition. For ex. using metaphor and irony, rhetorical questions, understatement, hints etc. For instance, "Its cold in here," which implies a request to close the door.

2.2 Example Dialogues

Here are some example dialogues that illustrate the difference in the politeness strategies used in discourse contexts of varying power. Two Microsoft Agents, Merlin and Genie are involved in a conversation while cooking and cleaning in a kitchen together. Consider the difference in the degree of politeness in each situation.

Conversation 1: This is an example conversation in which Merlin is Genie's boss.

Agent (Speech act: Politeness strategy): Utterance

***Merlin**: Would you mind washing the dishes? (Approval: RequestAct)*

Genie: Sure, I'll wash the dishes. (Direct:AcceptRequest)

Genie: I'm wondering whether I should boil the pasta.(Autonomy:Offer)

***Merlin**: Yeah you can. (Direct: AcceptOffer)*

***Merlin**: You've burnt the vegetables. (Direct: Inform)*

Genie: Yeah. (Direct: AcceptInform)

Genie: The oven is dirty. (Indirect: RequestAct)

***Merlin**: I'm sorry I can't clean the oven. (Approval: RejectRequest)*

Genie: Ok. (Direct: AcceptReject)

Genie: If you don't mind, can I clean the burner? (Autonomy: Offer)

***Merlin**: No, forget it. (Direct: RejectOffer)*

***Merlin**: I must wipe the counter now. (Direct: Offer)*

Genie: Alright, if you insist. (Autonomy: AcceptOffer)

Genie: Do you know that you've broken the baking dish mate? (Approval: Inform)

***Merlin**: Yes. (Direct: AcceptInform)*

***Merlin**: I'll cook the vegetables. (Direct: Offer)*

Genie: No please don't bother yourself. (Autonomy: RejectOffer)

Genie: The vegetables are healthy. (Indirect: RequestAct)

***Merlin**: Alright, I'll add the vegetables to the pasta. (Direct: AcceptRequest)*

***Merlin**: Could you please add the cheese to the pasta for me? (Approval: RequestAct)*

Genie: Please don't mind but I can't do that. (Autonomy: RejectRequest)

***Merlin**: Alright. (Direct: AcceptReject)*

Conversation 2: In this case Genie and Merlin are colleagues at an equal level in hierarchy.

Agent (Speech act: Politeness strategy): Utterance

***Merlin**: Could you possibly clean the oven for me?(Approval:RequestAct)*

Genie: Sure. (Direct:AcceptRequest)

Genie: I'll clean the burner. (Direct:Offer)

***Merlin**: Ok. (Direct:AcceptOffer)*

***Merlin**: You've burnt the vegetables. (Direct:Inform)*

Genie: Yeah. (Direct:AcceptInform)

Genie: Would you mind washing the dishes? (Approval:RequestAct)

***Merlin**: I'm sorry but I can't wash the dishes. (Approval:RejectRequest)*

Genie: Alright. (Direct:AcceptReject)

Genie: I must boil the pasta. (Direct:Offer)

***Merlin**: No, thanks. (Direct:RejectOffer)*

***Merlin**: I can wipe the counter. (Direct:Offer)*

Genie: Yeah you can. (Direct:AcceptOffer)

Genie: You've broken the baking dish. (Direct:Inform)

***Merlin**: Yes. (Direct:AcceptInform)*

***Merlin**: I'll cook the vegetables. (Direct:Offer)*

Genie: No, forget it. (Direct:RejectOffer)

***Merlin**: Could you please add the vegetables to the pasta? (Approval:RequestAct)*

Genie: Please don't mind but I can't do that. (Approval:RejectRequest)

***Merlin**: Ok. (Direct:AcceptReject)*

Genie: Will you please wipe the table mate? (Approval:RequestAct)

***Merlin**: Sure. (Direct:AcceptRequest)*

3 Virtual Environment

We rendered POLLy with Microsoft Agent Characters (Microsoft, 1998) in our Virtual Environment RAVE at Sheffield University as well as on a desktop computer screen. RAVE consists of a 3-dimensional visualisation of computer-generated scenes onto a 10ft x 8ft screen and a complete 3D surround sound system driven by a dedicated computer. Since Microsoft Agents are 2D, they are not rendered 3D, but a life size image of the characters is visible to the users on the screen to make them appear believable. Figure 1 showed a user interacting with POLLy in RAVE. The MS Agent package provides libraries to program control using various developing environments like the .NET framework and visual studio and includes a voice recognizer and a text-to-speech engine. It also provides controls to embed predefined animations which make the characters' behaviour look more interesting and believable (Cassell & Thórisson, 1999). We have programmed MS agent in Visual C++ and have embedded these animations like gesturing in a direction, looking towards the other agents, blinking, tilting the head, extending arms to the side, raising eyebrows, looking up and down etc while the agents speak and listen to the utterances and holding the hand to the ear, extending the ear, turning the head left or right etc when the agents don't understand what the user says or the user doesn't speak anything.

The Agents share the AI plan to collaborate on it together to achieve the cooking task. Goals to communicate about the plan are also allocated to the agents as speech acts (SAs) such as Requests, Offers, Informs, Acceptances and Rejections, needed to accomplish the plan collaboratively. While interacting with the system using a high quality microphone, the user sees one or two agents on the screen and plays the part of the second or the third agent, as per the role given to him/her.

When we extend this to a real-time immersive Virtual Reality environment, a Virtual Kitchen in this case, the ECAs will actually perform the task of cooking a recipe together in the virtual kitchen while conversing about the steps involved in doing so, as laid out by the AI plan.

This setup makes it possible to design a 2x2x2 experiment to test three conditions: *Interactivity*, i.e. whether the user only sees the agents interacting on the screen vs. the user interacts with the agents by playing a role; *immersiveness of the environment*, i.e. rendering in RAVE vs. rendering on a desktop computer; and *culture*, i.e. the difference between the perception of politeness by people from different cultures as in (Gupta et al., 2007). We are now in the process of completing the design of this experiment and running it.

4 Conclusion

We presents a demo of our conversational system POLLy which implements MS Agent characters in a VE and uses an AI Planning based shared representation for generating actions to be performed by the agents and utterances to communicate about the steps involved in performing the action. The utterances generated by POLLy are socially appropriate in terms of their politeness level. The user will be given a role play situation and he/she will be able to have a conversation with the agents on a desktop computer, where some dialogic utterances would be allocated to the user. An evaluation of POLLy (Gupta et al, 2007; Gupta et al, 2008) showed that (1) politeness perceptions of POLLy's output are generally consistent with B&L's predictions for choice of form for discourse situation, i.e. utterances to strangers or a superior person need to be very polite, preferably autonomy oriented (2) our indirect strategies which should be the politest forms, are the rudest (3) English and Indian speakers of English have different perceptions of politeness (4) B&L implicitly state the equality of the P & D variables in their equation ($\theta = P + D + R$), whereas we observe that not only their weights are different as they appear to be subjectively determined, but they are also not independent.

References

André, E., Rist, T., Mulken, S.v., Klesen, M., & Baldes, S. 2000. *The automated design of believable dialogues for animated presentation teams.* In Embodied Conversational Agents (pp. 220–255). MIT Press.

Bersot, O., El-Guedj, P.O., God´ereaux, C. and Nugues. P. 1998. *A conversational agent to help navigation & collaboration in virtual worlds.* Virtual Reality,3(1):71–82.

Blum, A., Furst, M. 1997. *Fast Planning Through Planning Graph Analysis.* Artificial Intelligence 90.

Cassell, J. and Thórisson, K.R. 1999. *The Power of a Nod and a Glance: Envelope vs. Emotional Feedback* *in Animated Conversational Agents.* Applied Artificial Intelligence 13: 519-538.

Cassell, J. and Bickmore, Timothy W. Negotiated Collusion. 2003. *Modeling Social Language and its Relationship Effects in Intelligent Agents.* User Model. User-Adapt.Interact. 13(1-2):89-132.

Culpeper, J. 1996. *(Im)politeness in dramatic dialogue.* Exploring the Language of Drama: From text to context. Routledge, London.

Dale, R. and Reiter, E. 1995. *Building Natural Language Generation Systems.* Studies in Natural Language Processing. Cambridge University Press.

Dautenhahn, K. 2000. *Socially Intelligent Agents: The Human in the Loop* (Papers from the 2000 AAAI Fall Symposium). The AAAI Press, Technical Report.

Eichner, T., Prendinger, H., André, E. and Ishizuka, M. 2007. *Attentive presentation agents.* Proc. 7th International Conference on Intelligent Virtual Agents (IVA-07), Springer LNCS 4722. pp 283-295.

Grosz, B.J., Sidner, C.L. 1990. *Plans for discourse.* In: Cohen, P.R., Morgan, J.L., Pollack, M.E. (eds.) Intentions in Communication, MIT Press, Cambridge.

Gupta, S., Walker, M.A., Romano, D.M. 2007. *How Rude are You?: Evaluating Politeness and Affect in Interaction.* Affective Computing & Intelligent Interaction (ACII-2007).

Gupta , S., Walker, M.A., Romano, D.M. 2008 (to be published). *Using a Shared Representation to Generate Action and Social Language for a Virtual Dialogue Environment.* AAAI Spring Symposium on Emotion, Personality and Social Behavior.

Johnson, L.W. and Rizzo, P. and Bosma, W.E. and Ghijsen, M. and van Welbergen, H. 2004. *Generating socially appropriate tutorial dialog.* In: ISCA Workshop on Affective Dialogue Systems. pp. 254-264.

Johnson, L., Mayer, R., André, E., & Rehm, M. 2005. *Cross-cultural evaluation of politeness in tactics for pedagogical agents.* Proc. of the 12th Int. Conf. on Artificial Intelligence in Education.

Kopp, S., Jung, B., Lessmann, N. and Wachsmuth, I. 2003. *Max – A multimodal assistant in virtual reality construction.* KI Zeitschift (German Magazine of Artificial Intelligence), Special Issue on Embodied Conversational Agents, vol.4, pp.11–17.

Lavoie, B., and Rambow, O. 1997. RealPro – a fast, portable sentence realizer. In Proc. Conference on Applied Natural Language Processing (ANLP'97).

Microsoft. 1998. *Developing for Microsoft Agent*. Microsoft Press.

op den Akker, H.J.A. and Nijholt, A. 2000. *Dialogues for Embodied Agents in Virtual Environments.* In: Natural Language Processing - NLP 2000, 2nd Int. Conf. pp. 358-369. LNAI 1835.

Paiva, A., Dias, J., & Aylett, R.S. 2005. *Learning by feeling: evoking empathy with synthetic characters.* Applied Artificial Intelligence: 19 (3-4), 235-266.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., De Carolis, B. 2005. *GRETA. A Believable Embodied Conversational Agent*. in O. Stock and M. Zancarano, eds, Multimodal Intelligent Information Presentation, Kluwer.

Prendinger, Helmut and Ishizuka, Mitsuru. 2001. *Let's talk! Socially intelligent agents for language conversation training.* IEEE Transactions on xSystems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 31, No. 5, pp 465-471.

Porayska-Pomsta, K. 2003. *Influence of Situational Context on Language Production: Modelling Teachers' Corrective Responses.* PhD Thesis. School of Informatics, University of Edinburgh.

Rehm, M. and André, E. 2007. *Informing the Design of Agents by Corpus Analysis.* Conversational Informatics, Edited by T. Nishida.

Sidner, C.L. 1994. *An artificial discourse language for collaborative negotiation.* In: Proc. 12th National Conf. on AI, pp. 814–819.

Takenobu, T., Tomofumi, K., Suguru, S., Manabu, O. 2003. *Bridging the Gap between Language and Action.* IVA 2003, LNAI 2792, pp. 127-135.

Traum, D., Rickel, J., Gratch, J., Marsella, S. 2003. *Negotiation over Tasks in Hybrid Human-Agent Teams for Simulation-Based Training.* Proceedings of the 2nd Int. Joint Conf. on Autonomous Agents and Multiagent Systems.

Walker, M.A. 1996. *The effect of resource limits and task complexity on collaborative planning in dialogue.* Artificial Intelligence Journal 85, 1–2.

Walker, M., Cahn, J. and Whittaker, S. J. 1997. *Improving linguistic style: Social and affective bases for agent personality.* In Proc. Autonomous Agents'97. 96–105. ACM Press.

Wang, N., Johnson, W.L., Rizzo, P., Shaw,E., & Mayer, R. 2005. *Experimental evaluation of polite interaction tactics for pedagogical agents.* Proceedings of IUI '05. ACM Press.

Watts, Richard J. Ide, S. and Ehlich, K. 1992. Introduction, in Watts, R, Ide, S. and Ehlich, K. (eds.), *Politeness in Language: Studies in History, Theory and Practice.* Berlin: Mouton de Gruyter, pp.1-17.