Modeling Leadership Behavior of Players in Virtual Worlds

Samira Shaikh[1], Tomek Strzalkowski[1], Jennifer Stromer-Galley[2], George Aaron Broadwell[1], Ting Liu[1] and Rosa Mikeal Martey[3]

[1]State University of New York – University at Albany, [2]Syracuse University, [3]Colorado State University
sshaikh@albany.edu

Abstract

In this article, we describe our method of modeling sociolinguistic behaviors of players in massively multi-player online games. The focus of this paper is leadership, as it is manifested by the participants engaged in discussion, and the automated modeling of this complex behavior in virtual worlds. We first approach the research question of modeling from a social science perspective, and ground our models in theories from human communication literature. We then adapt a two-tiered algorithmic model that derives certain mid-level sociolinguistic behaviors--such as Task Control, Topic Control and Disagreement from discourse linguistic indicators--and combines these in a weighted model to reveal the complex role of Leadership. The algorithm is evaluated by comparing its prediction of leaders against ground truth – the participants' own ratings of leadership of themselves and their conversation peers. We find the algorithm performance to be considerably better than baseline.

Introduction

The popularity of virtual worlds and Massively Multiplayer Online Games (MMOGs) has led to a need for research to understand which norms from the real world are transferred to, and practiced in, virtual environments, and conversely which behaviors manifested in virtual environments reflect or are predictive of real world characteristics. An important aspect of virtual worlds is the ability to craft a persona to navigate the world—an avatar that can persist over time. That persona could represent one aspect of a person's identity, be a faithful reproduction, or be an alternate self (Banks & Bowman, 2014), and that persona could also build a reputation that could be altered over the avatar's history. The emergence of such personas led to research concerning them, and that work has often been tied to matters of identity, including how individuals related their real life identities to their created identities. A growing body of work finds that users' activities online are not so easily divorced from real world characteristics and personal practices and social norms from the real world transfer to virtual worlds (Stromer-Galley and Martey; 2009; Messinger et al., 2008; Yee et al., 2007; Yee et al., 2011).

This research is part of a larger project where we explored behavioral indicators across different virtual worlds - *Second Life* and *World of Warcraft*; as well as between cultures – as characterized by English-speaking and Spanish-speaking players. We designed and conducted a multiple-factor research study to understand the relationships among virtual world behaviors and seven real world characteristics: Gender, Education, Age, Leadership, Social Conformity, Gamer Culture, and Digital Nativity.

In this paper, we focus on the behavioral indicators that are significantly correlated with leadership in the virtual world environments. These indicators are then used to inform an algorithmic model to predict leadership roles in multi-party virtual world environments. The virtual worlds we studied were *Second Life* and *World of Warcraft*, and included players across different cultures – English and Spanish. The focus of this article is on our findings related to our *Second Life* English language study, and the leadership behaviors manifested in the discourse that participants engage in. Other analyses, including those related with World of Warcraft and comparisons across cultures, are the basis of separate publications.

Related Research

We draw upon two main bodies of literature for relevant research. The first is the automated modeling of discourse roles and the other is the determination of leadership in virtual worlds.

Much prior work has been done in communication that focuses on the communicative dimension of discourse. For example, Speech Act theory (Austin, 1962; Searle 1969) provides a generalized framework of multiple levels of discourse analysis; work on dialogue analysis (Blaylock, 2002; Carberry and Lambert, 1999; Stolcke et al., 2000) focuses on information content and structure of dialogues. Automatic modeling of sequences of dialogue acts (Bunt, 1994), in order to predict the next dialogue act (Samuel et al. 1998; Stolcke, et al., 2000; Ji & Bilmes, 2006, inter alia) or to map them onto subsequences or "dialogue games" (Carlson 1983; Levin et al., 1998), from which participants' functional roles in conversation (though not social roles) may be extrapolated (e.g., Linell, 1990; Poesio and Mikheev, 1998) is a relevant line of research. But, there are few systematic studies in the current literature that explore the way language may be used to make predictions of social roles in groups where (a) these roles are not known a priori, or (b) these roles do not exist prior to the beginning of the discourse and only emerge through interaction. Notable among these is work done on modeling complex social phenomena in multi-party discourse environments by Strzalkowski et al. (2010) and Broadwell et al. (2012). The use of language by participants as a feature to determine interpersonal relations has been studied in multi-party discussion contexts. Bracewell et al. (2011) developed a multi-tier learning framework to determine collegiality between discourse participants. Their approach, however, looks at singular instances of linguistic markers or single utterances rather than a sustained demonstration of sociolinguistic behavior over the course of entire discourse. Freedman et al. (2011) designed an approach that looks at the entire discourse to detect behaviors such as persuasion; however, their analysis is conducted on online discussion threads where the proportion of social phenomena of interest may be scarcer than required to obtain adequate untainted data to build initial models.

A number of studies have examined leadership in virtual spaces, looking at the relationships between offline and online characterizations of leadership (Jang & Ryu, 2011; Xanthopoulou & Papagiannidis, 2012; Yee et al., 2007). However, our approach is the first attempt at the automated modeling of leadership in these environments, using the empirical evidence gathered from our regression analyses.

In the next section, we provide details about our experimental protocol, data annotation and analyses.

Data Collection and Analysis

In this research, we evaluated the relationships between real-world characteristics and behaviors in *Second Life*. It is a three-dimensional virtual environment where players create and customize avatars to interact with one another and with the virtual world through movement, chat, gesture, sound, and object-clicks. The chat can be public, meaning players located nearby can see the chat on their screen. It can also be private, which requires players to select a specific recipient for the message.

Second Life's many islands, or "sims," encompass a range of communities. For our research, we created a custom steampunk-themed island, Adamourne on Wells, and populated it with buildings, objects and non-player characters (NPCs). Participants were recruited using in-world advertising, as well as Facebook, Twitter, and *Second Life* forums. Two-hundred and ten qualified participants were each assigned into 48 groups of 3 to 5 people. An experimenter assisted each group, but only minimally, following strict protocols.

Prior to the session, participants took an online survey assessing game experience, demographics, and other characteristics including leadership measures. After the session, participants filled out a 15-minute post-session survey about their experiences in the session, which also included leadership measures. Custom logging software recorded participants' movement, chat, and clicks on interactive objects. Participants who completed all stages of the study received compensation of 5,000 Linden dollars (~$19 USD). They were tasked with solving a mystery centered on the theft of a precious diamond by investigating clues and suspects across five areas. To solve the mystery, players explored buildings and towns, interacted with NPCs, clicked objects, received items, and fought with ghosts. The challenges in the game required players to coordinate actions and share information, engendering their communication via chat.

Measures of Leadership

At the end of each session, participants were asked to rank themselves and their group members by who they thought performed as top leader, second leader, etc. of their group. This self and peer ranking data establishes the ground-truth against which we match algorithm leadership predictions.

Additional measures of leadership included whether participants perceived themselves to be leader-like in the game. The pre-session survey also included at total of 21 survey items that assessed leadership qualities adopted from Bass and Avilio (1995) and Posner and Kouzes (1988).

These measures of self-reported leadership and peer-reported leadership were used to run standard binary logistic regression with a variety of control variables – such as demographics, choice of avatar etc. The resulting analyses allow us to determine which behavioral indicators correlate highly with leadership. A detailed analytical report of these findings is the focus of a separate publication.

In this article, we focus on the sociolinguistic behaviors indicated in the chat discourse the participants engage in, the correlation of these sociolinguistic behavioral indicators with ground-truth measured against surveys and how these correlations were used to inform our algorithmic models of leadership. To analyze the discourse data from a sociolinguistic perspective, we performed human annotation along several dimensions as discussed in the following section.

Data Annotation

A multi-stage annotation process was adopted for our study. The categories of communication were based on several prior discourse content analysis projects including, most closely, that of Shaikh et al. (2010), as well as Allen and Core (1997) and Jurafsky, Shriberg and Biasca (1997). Intercoder reliability was established by three trained coders, and was at least a Krippendorff's alpha (2003) of .76 for all measures on a 10% sample of chat, which is within acceptable parameters of established validity and reliability. The data were annotated along the following categories, notably:

1. *Communicative Links*: The annotation of multi-player human interaction requires tracking who is speaking to whom in any (chat) message that is written to the group. It is necessary to track such information in order to determine the appropriate dialogue act (described next) to annotate, and eventually to discern factors such as power dynamics and leadership characteristics of individuals within group interaction. Our annotation includes tracking who is addressing whom, who is responding to whom, and who is continuing to speak—behaviors that are fairly common in online chat environments because multiple speakers and multiple interwoven threads of conversation make it challenging to ascertain the various, discrete communicative acts.
2. *Dialogue Acts*: Dialogue acts classify the function of each statement (posted chat message) within a conversation among the players. Our annotation guide identifies 16 dialogue acts including assertions, information requests, answers to information requests, agreements, disagreements, directives, greetings and other conventional phrases, and emotional phrases.

Other categories of annotation included Social Acts (interactions that are indicative of social relations in the group, for instance flirtation, humor and swearing); Movement Data (which included annotation of the order in which participants moved from one area of the game into another); Game Behaviors (such as use of pets and use of gestures) and Avatar Appearance and Name. Each of these categories or a combination of them then represents a variable against which to test the correlations of our leadership measures.

Grounding Models with Correlation Analyses

Social science theory indicates that leadership may be manifested in various ways (Bradford, 1978, Huffaker, 2010). We define leadership in the following terms: A leader is someone who guides the group toward an outcome, controls group discussion, manages actions of the group, and whom members recognize as the task leader. Such a leader is a skilled *task* leader, which corresponds to the social science theory put forth in Beebe and Masterson (2006). Consequently, one of the hypotheses we tested was that those who are perceived as leaders by other members of the group communicate more forcefully, which is to say – they produce more directives, control the topic more, produce more words, take more turns and disagree to a higher extent than those who are not leaders. Table 1 indicates peer-reported leaders produced more lines of chat (i.e. were more involved in the discussions), produced more action directives (directing others to perform some action), and exercised greater topic control (count of topics introduced and subsequently discussed by the group). Action directives and topic control were determined to be statistically significant when taking the number of tests into account.

Table 1. Means Tests by Peer-Reported Group Leadership in *Second Life*

	Peer-Reported Leader	
Test	/*t*/	*p*
Chat Lines (Involvement)	2.49	.014
Action Directives	4.10 ǂ	.000* ϕ
Disagreements	1.38	.168
Topic Control	3.88	.000* ϕ

ǂ *Equal variances not assumed*
* *p < .05*
ϕ *Statistically significant using Bonferroni correction for 15 tests (adjust to .0034)*

In the next section, we explain how we used the findings from our correlation analyses to inform the automated leadership models.

Leadership Models

Our research extends the work of Strzalkowski et al. (2010) and Broadwell et al. (2012), who first proposed the two-tiered approach to sociolinguistic modeling and have demonstrated that a subset of mid-level sociolinguistic behaviors may be accurately inferred by a combination of low-level language features. We have adopted their approach and extended it to modeling leadership in the virtual world context. Furthermore, we enhanced their method by adding the evidence learned from correlations to compute weights through which sociolinguistic behaviors may be combined appropriately to infer higher-level social phenomena such as leadership. They have shown this approach to work well on chatroom discussions, face-to-face discussions, forum discussions and other forms of multi-party chat discussions. This approach has not been applied towards predicting leadership in the context of virtual worlds, which is what we sought to do.

In this two-tier approach we use linguistic elements of discourse to first unravel sociolinguistic behaviors and then use the behaviors in turn to determine social roles, as shown in Figure 1. We underscore here that at both levels our analyses are solidly grounded on sociolinguistic theory. Mid-level behaviors that we discuss in this article are Topic Control, Task Control, Disagreement and Involvement that are computed using *indices*. These indices are directly obtained from linguistic elements of discourse, which are described below. For each participant in the game, we compute the degree to which they engage in sociolinguistic behaviors, using *measures,* which are a linear combination of indices. We describe relevant behaviors, component indices and corresponding measures in this section.

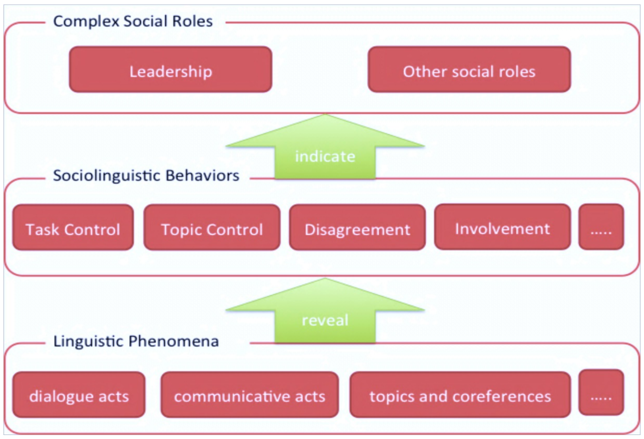

Figure 1. Two-tier approach applied to model social roles in discourse.

Topic Control Measure (TCM)

Topic Control is defined as an attempt by a participant to impose a topic of conversation. This sociolinguistic behavior is found to be consistent with Leadership (Table 1). In any conversation, whether it is focused on a particular issue or task or is just a social conversation, the participants continuously introduce multiple topics and subtopics. These are called *local topics*. These, following the notion put forth by Givon (1983), may be equated with any substantive noun phrases introduced into discourse that are subsequently mentioned again via repetitions, synonyms, or pronouns. Who introduces local topics, who continues to talk about them, and for how long are some of the indicators of topic control in dialogue. We use four indices for Topic Control. Participants who introduce more local topics exert more topic control in dialogue. The first index, called the **Local Topic Introductions Index (LTI)** calculates the proportion of local topics introduced by each participant by counting the number of first mentions of local topics as a percentage of all local topics in a discourse. The **Subsequent Mentions of Local Topics (SMT)** index calculates the percentage of discourse utterances where the local topics introduced are being mentioned (by themselves or others) through repetition, synonym, or pronoun. The **Cite Score (CS)** index calculates the percentage of subsequent mentions of local topics first introduced by each participant, but excluding the self-mentions by this participant. The final measure of topic control is the average **Turn Length (TL)** per participant. This index calculates the average utterance length (words) for each participant, relative to other participants.

Once we computed the scores for each participant on each index, we combine them to compute a single score on the corresponding measure. In this case, the LTI, SMT, CS and TL indices are combined to get a Topic Control Measure (TCM) for each participant. In our current system prototype, TCM score is computed as the mean of component index scores.

Task (or Skilled) Control Measure (SCM)

Task Control is an effort by one or more members of a group to define the group's project or goal and/or steer the group towards it. Task Control is gained by telling others to perform certain tasks, or subtasks, or to accept certain decisions about the task. It can also be gained by the speaker offering to perform a task. This sociolinguistic behavior is primarily consistent with Leadership (Table 1). One index of Task Control is the number of directives (done as statements or questions) made by each participant as a percentage of all directives in discourse, known as the **Directive Index (DI)**. In other words, a participant who tells others what to do (whether overtly or more subtly) is attempting to control the task that the group is performing.

Cumulative Disagreement Measure (CDM)

Disagreement has a role to play with regard to leadership and influence in that it is possible that a person in a small group engages in disagreements with others in order to control the topic by way of identifying or correcting what they see as a problem (Ellis and Fisher, 1994; Sanders, Pomerantz and Stromer-Galley, 2010). We are interested in a sustained phenomenon where participants repeatedly disagree with each other, thus revealing a social relationship between them. One of the indices we have developed to measure disagreement is the proportion of disagree and/or reject turns produced by a participant that are directed at any other participants in the discourse. This index is called the **Disagree-Reject Index (DRI)**.

Involvement Measure (INVX)

Involvement is defined as the degree of in the discussion of a group. This behavior is consistent with Leadership (Table 1). A degree of involvement may be estimated by how much a speaker contributes to the discussion in terms of substantive content. This includes introduction of new local topics, taking up the topics introduced by others, as well as taking sides on the topics being discussed. By topics here, we mean the local topics described previously. We have defined five indices in support of Involvement; we shall expand on three of them here. The **Noun Phrase**

Index (NPI) is the amount of information that each speaker contributes to discourse. The NPI measure is calculated by counting the number of content words (e.g., all occurrences of nouns and pronouns referring to people, objects, etc.) in each speaker's utterances as a percentage of all content words in discourse. The **Turn Index (TI)** is the frequency of turns that different speakers take during a conversation. The **Topic Chain Index (TCI)** is computed by identifying the most frequently mentioned topics in a discourse, i*.e.,* topics chains (i.e., with gaps no longer than 10 turns) and then by computing the percentages of mentions of these persistent topics by each participant.

The component indices of each of the four measures explicated above are automatically computed from text using standard natural language processing tools such as part-of-speech taggers and dialogue act taggers. For instance, the Directive Index for Skilled Control Measures and the Disagree-Reject Index for the Disagreement Measure are automatically computed using an automated dialogue act tagger. We used supervised learning using a Naïve Bayes classifier to learn the utterance level cues and classify the utterances into dialogue act categories. The four measures so computed form the first tier of the algorithm.

At the second tier, we wish to combine the measures in a linear combination to give us a ranked list of participants on their leadership score. Using a weighting scheme learned from the empirical evidence found from our analysis of correlations against survey ratings, we adjust the weights given to the component indices. The Task Control Measure (SCM) has the highest contribution, following by Topic Control (TCM) and then Involvement (INVX) and then Cumulative Disagreement Measure (CDM), as learned from correlations in Table 1.

In this weighting scheme:

$$\text{Leadership score} = (\alpha_{SCM} * SCM) + (\alpha_{CDM} * CDM) + (\alpha_{TCM} * TCM) + (\alpha_{INVX} * INVX)$$

Where $\alpha_{SCM} > \alpha_{TCM} > \alpha_{INVX} > \alpha_{CDM}$

We illustrate the combination of measures using Figure 2. There are seven participants in the chosen session, denoted by their initials (SS, EV and so on). We show scores of participants on three measures discussed above and their combined leadership scores. We note that although participant EV has a higher score for the Cumulative Disagreement Measure, participant SS has higher Topic Control and Task Control, and this is reflected in the overall leadership scores. Participant SS is chosen to be the leader by the algorithm. We next evaluate the ranked list of participants from the post-survey measures and the ranking produced by our algorithm.

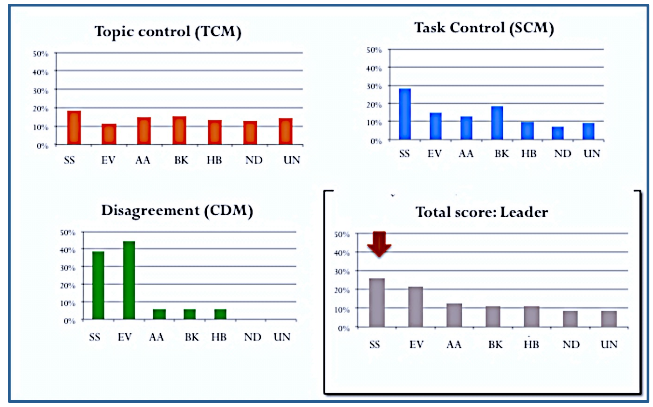

Figure 2. Combining measures to get leadership scores for players.

Evaluation and Results

Although we have a full ranking of participants, both from survey ratings as well as algorithm output, we are only interested in participants who have the highest Leadership score. This means in order to evaluate system performance the top-ranking participant on both rankings should match. In cases where the top two individuals are quite close in the survey scores, we may consider top two participants. In Table 2, we show the algorithm accuracy at predicting the top ranking leaders on our *Second Life* sessions. We compare these results against a random baseline algorithm, one that would pick a leader from the list of participants of a given group at random. Our algorithm proves to be quite accurate in the predictions and has a significant improvement over random baseline.

Table 2. Accuracy of predicting leaders compared against random baseline

	Random Baseline	Our Algorithm
Accuracy in predicting leaders	*30.77%*	*83.33%*

In cases where the algorithm did not pick the correct participant as leader, we noted that the correct choice was either the third or fourth ranked participant and the difference in scores between them was marginal.

Summary and Discussion

In this paper, we describe our approach towards the automated modeling of complex social roles in virtual worlds. We describe the carefully constructed data collection experiments and the resulting regression analyses to determine the appropriate weights of various sociolinguistic models. We discuss mid-level sociolinguistic behaviors such as Task Control, Topic Control, Disagreement and

Involvement and how these behaviors are used in an empirically derived two-tier model to compute a ranked list of participants on their leadership scores. We compare the algorithm ranking to the ground truth collected from participant's exit surveys and find that the algorithm accuracy is quite reassuring. From the literature review section, we posit that ours is the first attempt at modeling such behaviors of players in the context of virtual games.

This model could be beneficial for researchers aiming to understand leadership dynamics in virtual worlds, and could be used to process large volumes of discussion online to identify groups with leaders as compared with groups without a clear leader. Such identification could then be used to predict group successes in the context of online games and virtual worlds. It could also be used to identify leaders in virtual worlds who might then be selected for further study over time to learn how leadership evolves in groups. Most leadership research surveys people in order to ascertain leadership characteristics, whereas our research focuses on leadership that is effectively communicated and can be automatically modeled. This algorithm could be especially useful in driving scholarship on the actual behavior of leaders.

In future work, we aim to conduct the same analyses for the data we collected from *World of Warcraft* sessions. In addition, we aim to analyze and compare our models across different languages – English and Spanish.

Acknowledgments
This project was funded by the Air Force Research Labs.

References

Allen, J. and M. Core. 1997. Draft of DAMSL: Dialog Act Markup in Several Layers. www.cs. rochester.edu/research/cisd/resources/damsl/

Austin, J. L. 1962. How to do Things with Words. Clarendon Press, Oxford.

Banks, J., & Bowman, N. D. 2014. Avatars are (sometimes) people too: Linguistic indicators of parasocial and social ties in player-avatar relationships. New Media & Society, 16. Doi: 10.1177/1461444814554898

Bass, B. M., & Avolio, B. J. 1995. The Multifactor Leadership Questionnaire (MLQ). Redwood City, CA: Mind Garden.

Beebe, S. A., & Masterson, J. T. 2006. Communicating in small groups: Principles and practices. Boston, MA: Pearson/Allyn and Bacon

Blaylock, N. 2002. Managing Communicative Intentions in Dialogue Using a Collaborative Problem-Solving Model. Technical Report 774, University of Rochester, CS Dept.

Bracewell, D. B., M. Tomlinson, Y. Shi, J. Bensley, and M. Draper. 2011. Who's playing well with others: Determining collegiality in text. In *Proceedings of the 5th IEEE International Conference on Semantic Computing (ICSC 2011).*

Bradford, L. P. 1978. Group development. La Jolla, Calif., University Associates.

Broadwell, George Aaron,, Jennifer Stromer-Galley, Tomek Strzalkowski, Samira Shaikh, Sarah Taylor, Umit Boz, Alana Elia, Laura Jiao, Ting Liu and Nick Webb. 2012. Modeling Socio-Cultural Phenomena in Discourse. Journal of Natural Language Engineering, Cambridge Press.

Bunt, H. 1994. Context and Dialogue Control. Think Quarterly 3(1), 19-31.

Carberry, S. and L. Lambert. 1999. A Process Model for Recognizing Communicative Acts and Modeling Negotiation Dialogue. Computational Linguistics, 25(1), pp. 1-53

Carlson, L. 1983. Dialogue Games: An Approach to Discourse Analysis. D. Reidel.

Ellis, D. G., & Fisher, B. A. 1994. Small group decision making: Communication and the group process. New York: McGraw-Hill

Freedman, Marjorie., Alex Baron, Vasin Punyakanok and Ralph Weischedel. 2011. Language Use: What it can tell us? In *Proceedings of the Association of Computational Linguistics. Portland, Oregon.*

Givon, T. 1983. Topic continuity in discourse: A quantitative cross-language study. Amsterdam: John Benjamins.

Herring, S. C. 1994. Politeness in Computer Culture: Why Women Thank and Men Flame. In: *Cultural Performances: Proceedings of the Third Berkeley Women and Language Conference (pp. 278-294). Berkeley Women and Language Group.*

Huffaker, D. 2010. Dimensions of leadership and social influence in online communities. Human Communication Research, 36, 596-617.

Jang, Y., & Ryu, S. 2011. Exploring game experiences and game leadership in massively multiplayer online role-playing games. British Journal of Educational Technology, 32, 616-623.

Ji, G. and J. Bilmes. 2006. Backoff Model Training using Partially Observed Data: Application to Dialog Act Tagging. *In Proceedings of the Human Language Technology/ American chapter of the Association for Computational Linguistics (HLT/NAACL'06)*

Jurafsky, D., E. Shriberg and D. Biasca. 1997. Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual. http://stripe.colorado.edu/ ~jurafsky/ manual.august1.html

Krippendorff, K. 2003. Content analysis. Thousand Oaks, CA.

Levin, L., A. Thyḿe-Gobbel, A. Lavie, K. Ries, and K. Zechner. 1998. A Discourse Coding Scheme for Conversational Spanish. In *Proceedings of the International Conference on Speech and Language Processing.*

Linell, P. 1990. The power of dialogue dynamics. In Ivana Markov´a and Klaus Foppa, editors, The Dynamics of Dialogue. Harvester, 147–177.

Messinger, P.R., Ge, X., Stroulia, E., Lyons, K., Smirnov, K., & Bone, M. 2008. On the Relationship Between My Avatar and Myself. Journal of Virtual Worlds Research, 1, 2.

Poesio, M. and A. Mikheev. 1998. The predictive power of game structure in dialogue act recognition. International Conference on Speech and Language Processing (ICSLP-98).

Posner, B. Z., & Kouzes, J. M. 1988. Development and validation of the leadership practices inventory. *Educational and Psychological Measurement, 48,* 483-496.

Samuel, K., S. Carberry, and K. Vijay-Shanker. 1998. Dialogue Act Tagging with Transformation-Based Learning. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Montreal.*

Sanders, R. E., Pomerantz, A., Stromer-Galley, J. 2010. Some ways of taking issue with another participant during a group deliberation. *Paper presented at the annual meeting of the National Communication Association, San Francisco, CA*

Searle, J. R. 1969. Speech acts: An essay in the philosophy of language. Cambridge: The University Printing House.

Shaikh, S., T. Strzalkowski, S. Taylor and N. Webb. 2010. MPC: A Multi-Party Chat Corpus for Modeling Social Phenomena in Discourse, in proceedings of the 7th International Conference on Language Resources and Evaluation (LREC2010), Valletta, Malta. 2010.

Stolcke, A., K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C. Van Ess-Dykema, and M. Meteer. 2000. Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. Computational Linguistics 26(3), 339–373.

Stromer-Galley, J., & Martey, R. M. 2009. Visual spaces, norm governed places: The influence of spatial context online. *New Media & Society*, 11, 1041-1060.

Strzalkowski, T., Broadwell, G. A., Stromer-Galley, J., Shaikh, S., Taylor, S., & Webb, N. 2010. Modeling socio-cultural phenomena in discourse. In *Proceedings of The 23rd International Conference on Computational Linguistics. Beijing, China.*

Xanthopoulou, D., & Papagiannidis, S. 2012. Play online, work better? Examining the spillover of active learning and transformational leadership. Technological Forecasting & Social Change, 79, 1328-1339. Doi: doi:10.1016/j.techfore.2012.03.006

Yee, N., Bailenson, J.N., Urbanek, M., Chang, F., and Merget, D. 2007. The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. Cyberpsychology & Behavior, 10, 1. pp. 115-21.

Yee, N., Ducheneaut, N., Nelson, L., Likarish, P. 2011. Introverted elves and conscientious gnomes: The expression of personality in World of Warcraft. In *Proceedings of CHI 2012, Vancouver, BC, Canada.* May 7-12.