Using Speech Act Theory to Model Conversations for Automated Classification and Retrieval

Douglas P. Twitchell, Mark Adkins, Jay F. Nunamaker Jr.,
Judee K. Burgoon
University of Arizona, USA
{dtwitchell, madkins, jnunamaker, jburgoon}@cmi.arizona.edu

Abstract

Instant messaging, chat rooms and other forms of synchronous computer-mediated communication (CMC) are increasing in use in the business, military, and consumer world. The language action perspective provides methods for analyzing and modelling repeated business conversations including synchronous CMC. This paper describes a method for creating a profiles for large amounts of synchronous CMC conversation after it has occurred. Called a speech act profile, it is based on speech act theory. The profiles can be used either as patterns for classifying conversations or for creating visual maps of the conversations themselves. Application of the profiles in information retrieval and deception detection are discussed.

Portions of this research were supported by funding from the U. S. Air Force Office of Scientific Research under the U. S. Department of Defense University Research Initiative (Grant \#F49620-01-1-0394). The views, opinions, and/or findings in this report are those of the authors and should not be construed as an official Department of Defense position, policy, or decision.

Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004)

Rutgers University, The State University of New Jersey, New Brunswick, NJ, USA, June 2-3, 2004
(M. Aakhus, M. Lind, eds.)
www.scils.rutgers.edu/lap04/lap04.htm

1. Introduction

Online conversations are becoming more and more commonplace. In 2003, worldwide use of instant messaging among consumers was 188 million users (King, 2004). More importantly, however, the business users of instant messaging numbered 22 million and are expected to increase to over 100 million in 2006. Some of those users, such as the National Association of Securities Dealers are required by government regulation to save their messages for three years (NASD, 2003). A case study by Adkins and Kruse (2004) on the U.S. Navy's Commander Task Force 50 during operation Enduring Freedom found extensive use of text-based communication for distributed interaction between ships, services, and assets ashore. This ever-growing archive of conversations represents an opportunity for both language action researchers and business managers. Researchers can use this data to study human communication in chat/IM conditions. Managers can use the data to detect fraud, retrieve conversations of interest, and direct training.

The Language Action Perspective (LAP) of business conversations is founded on two theoretical basis. First, from speech act theory comes the assertion that with each utterance in a conversation an action is performed by the speaker. Second, these actions (or speech acts) are organized into conversations according to predefined patterns (GoldKuhl, 2003). Winograd and Flores' (1986) conversation for action is an example of a pattern of speech acts organized together to create a specific type of conversation. A number of systems have been created by the Language Action Perspective community to facilitate, design, and model business conversations and processes as conversations. Some of these systems, such as SAMPO (Auramäki, Lehtinen, & Lyytinen, 1988), DEMO (Dietz, 1994, 2003) and BAT (GoldKuhl, 1996) provide a means for diagramming and modeling business processes according to the conversations that occur during those business processes. Other systems such as The Coordinator (LAP's original system) (Winograd, 1987), MILANO (De Michelis & Grasso, 1994; Agostini, De Michelis, & Grasso, 1997) and Negoisst (Schooop, Jertila, & List, 2003) are actual pieces of software that facilitate business conversations and negotiations. The methods presented in this paper are partially actual software and partially future software. Unlike the LAP systems mentioned above, however, the methods presented here are analysis methods for use mainly *after* the business conversations take place. Furthermore, the methods are meant for free-flowing natural language conversations such as those found in instant messaging and email.

Instead of predefining conversation types and attempting to design or facilitate business conversations according to the categories as done by traditional LAP systems, speech act profiling looks at business conversations as existing empirical data that are categorized during or after they actually take place. This approach might partially answer Goldkuhl's (2003) question: "Can a more inductive way of investigating business interaction be performed with less use of pre-defined communicative patterns?" by using existing business conversations as data, profiling them, and, finally, automatically classifying them according to patterns of interest—patterns that are either predefined or created *post hoc*.

There are several reasons for attempting to automate the classification of conversations. Information retrieval could benefit from the automated classification of speech acts. Archived conversations could have intent profiles attached to them that indicate the overall intention of the conversation. For example, a web searcher might be interested in conversations that contain critiques of the Vietnam war. Using current search engines, the searcher could search for the words *Vietnam*, *war*, and *critique*. However, many critiques of the war might not contain the word *critique*, and would thus be lost (or receive a low ranking) in such a search. If the searcher was able to issue a query such as *Vietnam war (critique)* where *critique* is the purpose of at least one participant in the conversation, she would likely get better results. The search for the semantic meaning of the words *Vietnam war* using conventional searching techniques would then be combined with the search for the pragmatic *force* of the word *critique*, yielding a search result with higher precision than searching on semantic meaning alone. Conversations could also be classified and retrieved according to deceptive intent. Deception, a perlocutionary act intended to create a false impression in a hearer, can be made up of a number of speech acts each with their own intention. For example, because deceivers are loath to be caught in their deception, they will often put on a submissive front. The expression of fewer assertions and more expressives (which include agreements) could indicate submissiveness. Furthermore, because deceivers usually express some uncertainty as a way to hedge their deception, those looking for deception could issue a query for conversations where a participant was being submissive or showing uncertainty.

1.1 Speech Act Profiling

The classification of conversations can be based on speech act profiling. Speech act profiling (Twitchell & Nunamaker Jr., 2004) is a method for analyzing and visualizing conversations and their participants. It extends Stolcke et. al.'s (2000) method of dialog act modeling and combines it with Alston's (2000) idea of illocutionary act potential, which is realized with Subasic and Huettner's (2001) fuzzy typing. These methods are used to create a set of summed probabilities for each speech act type during a conversation, which are then subtracted from the training corpus average to obtain divergences from normal speech. The speech acts used are the 42 dialogue acts in the modified SWBD-DAMSL tag set (Jurafsky, Shriberg, & Biasca, 1997). With the speech act potential probabilities and the categories of speech acts defined, a visual representation of the speech act profile can be created. Like Subasic and Huettner's (2001) visualization scheme, the fuzzy speech act profile visualization consists of a radar graph with all of the potential dialogue categories, acts, and their relative values.

The speech act profile in Figure 1 (taken from (Twitchell & Nunamaker Jr., 2004)) shows conversation taken from the SwitchBoard corpus. Here Speaker B is questioning Speaker A as indicated by the greater than normal number of WH-QUESTIONS (qw) and YES-NO-QUESTIONS (qy) by Speaker B and the greater than normal number of STATEMENTS (sd) by Speaker A. In this conversation, Speaker B is asking a multitude of questions, and Speaker A is merely replying making the conversation look like an interview. The excerpt from the conversation found in Table 1 shows that, indeed, Speaker B is asking numerous questions and Speaker A is responding to the questions with statements. The remainder of the conversation follows a similar pattern.

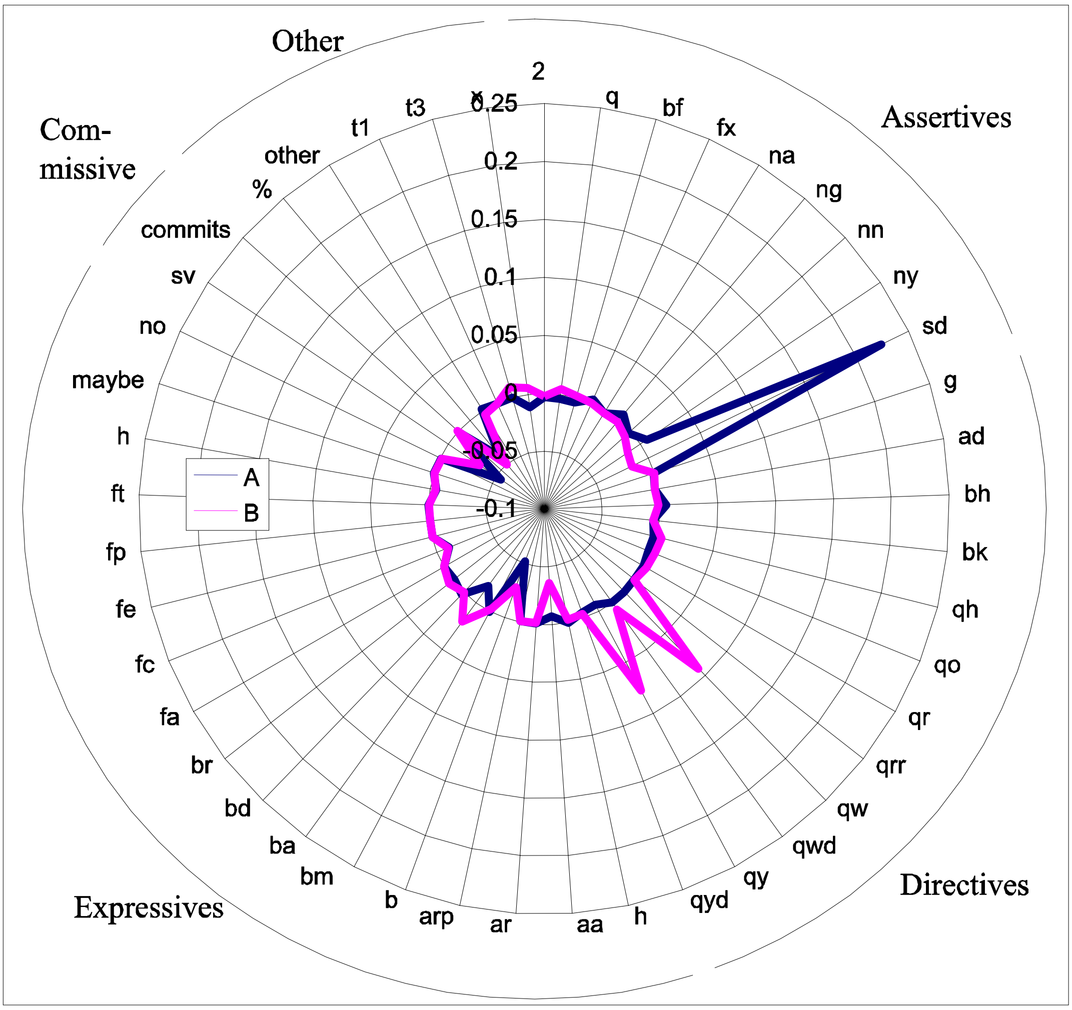

Figure 1: Sample fuzzy speech act profile showing interview-like behavior

1.2 Classifying Conversations

The set of summed probabilities that makes up the speech act profile can be used as features for classification. The conversation profiled in Figure 1 represents an interview-like conversation, and similar conversations could be automatically classified as such if their statements and questions probabilities exceeded certain thresholds. This simple type of classification only classifies conversations into two classes: interview-like and not interview-like. Recent research using speech act profiling was able to identify uncertainty in chat participants who were instructed to be deceptive (Twitchell, Nunamaker, & Burgoon, 2004). Speech act profiling's ability to classify participants as uncertain may aid those searching for deceptive behavior.

Table 1: Excerpt of conversation represented by the speech act profile in Figure 1

Speaker	Utterance
B	Where are you going to school?
A	N C State.
B	What's that?
A	Uh, North Carolina State.
B	So you're on Spring Break?
A	Not yet.
A	Ours don't start until, uh, next week.
B	So where are you?
A	Where am I?
B	Yeah.
A	What do you mean, where?
A	Oh,
A	in Raleigh.

Another interesting and useful problem is classifying the conversations into a taxonomy. This could be accomplished given a set of conversations manually classified using that taxonomy that can be used for training. One such taxonomy was introduced by Winograd (1987) in his work on the language action perspective of software design. He categorizes conversations into four categories: conversations for action, conversations for clarification, conversations for possibilities, and conversations for orientation. This taxonomy is based on the purpose or intention of the conversation and the reason the instigator has for beginning the conversation. Conversations for action usually begin with a request or an offer and the purpose of the conversation is some action to be taken either by the requestor, in the case of an offer, or by the requestee, in the case of a request. Conversations for clarification center of obtaining more information about something already said or a previous conversation. Conversations for possibilities have the purpose of creating ideas and settling on several promising ones for further analysis. Finally, conversations for orientation begin as a way for participants to exchange information about each other or a situation or for one participant to gain information from the other.

There are a number of classification methods that can be used to classify conversations given the features produced by the speech act profile, a taxonomy of conversations, and a manually classified training set. Each classifier has its advantages and disadvantages. For example, C4.5 decision trees (Quinlan, 1993) attempt to use the training set to build a tree that minimizes information entropy. Decision trees have the advantage of being readily interpretable. To find out why a conversation was classified a certain way, one only has to follow the path from the root of the decision tree to the point of classification. Neural networks, on the other hand, are not as interpretable, but have been shown to be in many instances more accurate (Zhou, Twitchell, Qin, Burgoon, & Nunamaker Jr., 2004).

Another classification method uses a set of hidden Markov models, each of which represents one conversation type. Conversations with each speech act manually annotated would be manually categorized, and the conversations from each category used to train the corresponding hidden Markov model in a process similar to the one used in speech act profiling described in (Twitchell & Nunamaker Jr., 2004) and (Stolcke et al., 2000). New conversations to be classified are run through each hidden Markov model using the forward-backward algorithm for hidden Markov Models (Rabiner, 1989). The classification represented by the model yielding the greatest probability is assigned to the conversation. Figure 2 is a state diagram representing Winograd and Flores' (Winograd & Flores, 1986) conversation for action. The state diagram can be turned into a hidden Markov model using the training process to assign probabilities to each of the arcs and probabilities to possible utterances.

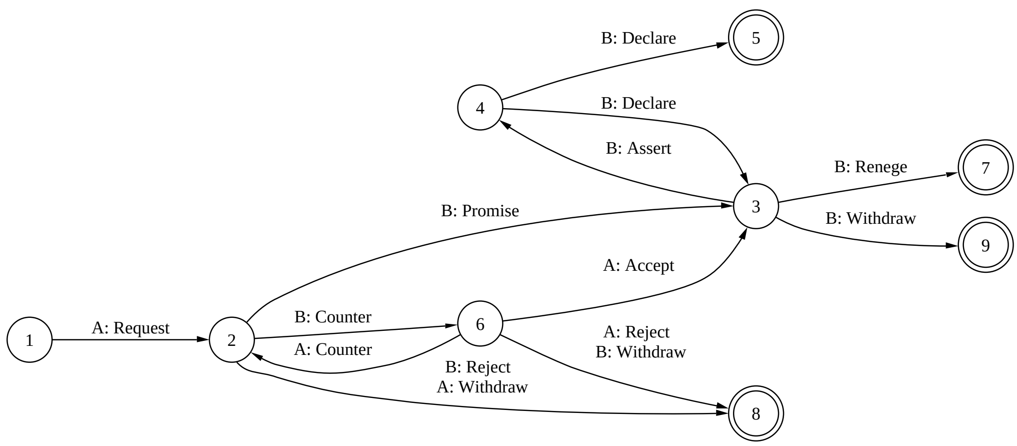

Figure 2: State diagram of a conversation for action (Winograd & Flores, 1986)

2. Conclusions and FutureWork

The study of the Language Action Perspective on communication modeling has shed light on how speech act theory can be applied to communications in organizations. It also has provided a number of systems for modeling and supporting such communication. This paper attempts to augment these systems and models with a method for automated classification of business conversations—classifications that can be used for retrieval.

Speech act profiling is one method that has been shown to be useful in discriminating between types of conversations including conversations with deception- induced uncertainty. It, along with other similar Markov-model-based techniques, may be useful in classifying conversations into more general framework, giving researchers and managers the tools to manage the millions of conversations produced each day. For researchers, not having to code conversations for intent by hand is an important labor-saving feature of conversation classification.

References

Adkins, M., & Kruse, J. (2004). *Case study: Network centric warfare in the u.s. navy's fifth fleet: Web-supported operational level command and control in operation enduring freedom* (Tech. Rep.). Center for the Management of Information.

Agostini, A., De Michelis, G., & Grasso, M. A. (1997). Rethinking CSCW systems: The architecture of MILANO. In *Proceedings of the fifth european conference on computer supported cooperative work* (p. 33-48). Lancaster, UK: Kluwer Academic Publishers.

Alston, W. P. (2000). *Illocutionary acts and sentence meaning*. Ithaca, N.Y.: Cornell University Press.

Auramäki, E., Lehtinen, E., & Lyytinen, K. (1988). Speech-act-based office modeling approach. *ACM Transactions on Office Information Systems*, 6(2), 126- 152.

De Michelis, G., & Grasso, M. A. (1994). Situating conversations within the language/action perspective: The milan conversation model. In R. Furuta & C. Neuwirth (Eds.), *Proceedings of the conferenceon computer supported cooperative work* (p. 89-100). Chapel Hill, North Carolina: ACM Press.

Dietz, J. L. G. (1994). Business modelling for business redesign. In *Proceedings of the twenty-seventh annual Hawaii international conference on system sciences*. Maui, Hawaii: IEEE Computer Society.

Dietz, J. L. G. (2003). The atoms, molecules and fibers of organizations. *Data and Knowledge Engineering*, 47(3), 301-325.

GoldKuhl, G. (1996). Generic business framworks and action modelling. In *Proceedings of the first international workshop on the language/action perspective* (LAP 96). Tilburg, The Netherlands: Springer Verlag.

GoldKuhl, G. (2003). Conversational analysis as a theoretical foundation for language action aproaches? In H. Weigand, G. GoldKuhl, & A. de Moor (Eds.), *Proceedings of the 8th international working conference on the languageaction perspective on communication modelling*. Tilburg, The Netherlands.

Jurafsky, D., Shriberg, E., & Biasca, D. (1997). Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, Draft 13.

King, J. (2004). Chat gets serious. http://www.computerworld.com/ softwaretopics/ software/groupware/story/0,10801,90311,00.html. *ComputerWorld*.

NASD. (2003). Instant messaging: Clarification for members regarding supervisory obligations and recordkeeping requirements for instant messaging. http://www.nasdr.com/pdf-text/0333ntm.pdf.

Quinlan, J. R. (1993). *Programs for machine learning*. Morgan Kauffman.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257 - 286.

Schooop, M., Jertila, A., & List, T. (2003). Negoisst: a negotiation support system for electronic business-to-business negotiations in e-commerce. *Data and Knowledge Engineering*, 47(3), 371-401.

Stolcke, A., Reis, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Van Ess-Dykema, C., Martin, R., & Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 339-373.

Subasic, P., & Huettner, A. (2001). Affect analysis of text using fuzzy semantic typing. *IEEE Transactions on Fuzzy Systems*, 9(4), 483-496.

Twitchell, D. P., Nunamaker, J. F. J., & Burgoon, J. K. (2004). Using speech act profiling for deception detection. In *Lecture notes in computer science: Intelligence and security informatics: Proceedings of the second NSF/NIJ symposium on intelligence and security informatics*. Tucson, Arizona (In Press).

Twitchell, D. P., & Nunamaker Jr., J. F. (2004). Speech act profiling: A probabilistic method for analyzing persistent conversations and their participants. In *Thirty-seventh annual Hawaii international conference on system sciences (CD/ROM)*. Big Island, Hawaii: IEEE Computer Society Press.

Winograd, T. (1987). A language/action perspective on the design of cooperative work. *Human Computer Interaction*, 3(1), 3-30.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood, New Jersey: Ablex Publishing Corporation.

Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker Jr., J. F. (2004). Toward the automatic prediction of deception - an empirical comparison of classification methods. *Journal of Management Information Systems*, 20(4), 139-166.