

Reacting to Agreement and Error in Spoken Dialogue Systems Using Degrees of Groundedness

Antonio Roque

University of Southern California - Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
roque@ict.usc.edu

Abstract

Computational models of grounding are extended to include representations of degrees of groundedness. These representations are then used for decision-making in dialogue management for spoken dialogue systems. Several domains will be explored with this model, and an implementation will be tested and evaluated.

Background

The general public is becoming increasingly aware of spoken dialogue systems: as telephone-based Interactive Voice Response systems, for in-car navigation, or for personal computer control, for example. Researchers study other domains such as tutoring, collaborative planning, and interacting with virtual humans. However, both commercial and research systems must become more robust before they will be widely accepted. This research proposes to extend ideas from artificial intelligence and psychology of language to build dialogue systems that model *degrees of groundedness* to track and react to the extent to which the system and human user are in agreement.

Most models of discourse, be they from philosophy, linguistics, artificial intelligence, or psychology, use some version of a notion of *common ground*, the beliefs that the participants of that discourse have in common (Clark and Marshall, 1981). (Clark and Schaefer, 1989) demonstrates the importance of studying how material is added to that common ground, and defines grounding as a collaborative activity between participants who work towards establishing a set of mutually-held beliefs. Behavior such as confirmations ("OK, you said seven o'clock"), backchannelling ("uh-huh"), and corrections ("I said ten, not seven") can be explained by models of grounding. Participants work towards the *grounding criterion*, at which point both parties are satisfied that they agree on the grounded material. These concepts are further developed in (Traum, 1994), which presents a computational model of grounding implemented in a planning assistant for railroad

resource allocation. The computational model of (Traum, 1994) uses a binary model of groundedness - something is either grounded or it is not - although the currently undeveloped notion of *degrees of groundedness* - the extent to which material is grounded - also appears to be useful, as described in (Traum and Dillenbourg, 1998).

This research proposes to formalize the concept of degrees of groundedness, using it to track the extent to which the user of a spoken dialogue system is being understood, and thereby enabling the system to respond appropriately. Current spoken dialogue systems use Natural Language Processing techniques to understand what a user is saying, but they usually fail to effectively monitor the evidence they have for how well the system and the human user are understanding each other. Research such as (Paek and Horvitz, 1999, 2000) and (Bohus and Rudnicky, 2005) addresses mutual understanding and error correction from various directions. This approach to modeling degrees of groundedness comes from a rich heritage of dialogue research not only in error handling in spoken dialogue systems, but also in representations of common ground as developed in the fields of philosophy, linguistics, and psychology. It promises to contribute to our understanding of how humans and machines communicate, as well as to the long-term goal of making generally usable spoken dialogue systems.

Approach

This research was motivated by problems addressed during the development of a spoken dialogue system, Radiobot-CFF, for radio-based artillery fire request training (Roque et al., 2006a). The initial version of Radiobot-CFF immediately used information received from the human operator without tracking how much evidence it had that it had understood the human correctly. If the human operator corrected it, Radiobot-CFF would update its information representation, but by that time, it might have committed actions in its simulated world that could not be undone.

Not coincidentally, the dialogues in the radio-based artillery fire request are highly structured, with confirmations at every turn and standard procedures for self- and other-correction. This presented an opportunity to develop a model of grounding, with an eye to later extending it to other domains.

The corpus of dialogues produced by the evaluation of Radiobot-CFF guided the initial model of degrees of groundedness. Building this model involved identifying the different kinds of *evidence of understanding* - the phenomena that dialogue participants use to determine how grounded information is. An analysis of the patterns of evidence of understanding given over series of turns produced a set of degrees of groundedness: that is, how grounded the material was after a sequence of dialogue moves containing evidence of understanding. This analysis also identified the grounding criteria for this domain: the degree of groundedness that each bit of information had to reach before it could be said to be sufficiently grounded. The final goal is a mechanism for deciding, at any point of the conversation, what kind of evidence of understanding to use in a reply based on the material being dealt with, its current degree of groundedness, and its grounding criteria.

Current Status

I am currently conducting an off-line evaluation to quantify this model's success in covering the domain; similar off-line evaluations for this domain have previously been conducted in (Roque et al., 2006b). After reviewing my results, I plan to adapt the initial Radiobot-CFF implementation to include my model, and evaluate its effectiveness by running a series of training sessions and analyzing the resulting data.

I am currently also considering how to make the model domain-independent. My plan is to quantify the domain knowledge needed for the grounding criteria, represent the beliefs of the participants, and using weighed abduction to link the evidence of understanding to the grounding criteria. I plan to use air traffic control ground-air communications as a second domain, and am considering several other corpora for other domains, including at least one general conversation domain. An important component of this effort is identifying an evaluation methodology that helps me quantify the benefits of this approach. My long-term plan is to develop the principles of groundedness in a way that can be used for analyzing dialogues and for building dialogue managers that can effectively handle conversations that are problematic for current spoken dialogue systems.

Acknowledgments

This work has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

- Bohus, D., and Rudnicky, A. 2005. Error Handling in the RavenClaw Dialog Management Architecture. In Proceedings of HLT-EMNLP-2005, Vancouver, Canada
- Clark, H., and Marshall, C. 1981. Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 10-63. Cambridge.
- Clark, H., and Schaefer, E. 1989. Contributing to discourse. *Cognitive Science*, 13, 259-294.
- Paek, T., and Horvitz, E. 2000. Grounding criterion: Toward a formal theory of grounding. Microsoft Technical Report, MSR-TR-2000-40.
- Paek, T. and Horvitz, E. 1999. Uncertainty, utility, and misunderstanding: A decision-theoretic perspective on grounding in conversational systems. In Proceedings of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems.
- Roque, A; Leuski, A; Rangarajan, V; Robinson, S; Vaswani, A; Narayanan, S.; and Traum, D. 2006. Radiobot-CFF: A Spoken Dialogue System for Military Training. In Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP).
- Roque, A.; Ai, H.; and Traum, D. 2006. Evaluation of an Information State-Based Dialogue Manager, In Proceedings of Brandial 2006: The 10th Workshop on the Semantics and Pragmatics of Dialogue, University of Potsdam, Germany.
- Traum, D., 1994. A Computational Theory of Grounding in Natural Language Conversation, TR 545 and Ph.D. Thesis, Computer Science Dept., U. Rochester.
- Traum, D., and Dillenbourg, P., 1998. Towards a normative model of grounding in collaboration, In Proceedings of the ESSLI98 Workshop on Mutual Knowledge, Common Ground, and Public Information.