Towards Reducing and Managing Uncertainty in Spoken Dialogue Systems

Raquel Fernández, Andrea Corradini, David Schlangen, Manfred Stede

Department of Linguistics
University of Potsdam, Germany
{raquel,andrea,das,stede}@ling.uni-potsdam.de

1 Motivation

Spoken dialogue systems (SDSs)—computational systems that can engage in a dialogue with a human user about a restricted topic by understanding and reacting to spoken natural language—have a valuable potential not only as commercial systems that can handle useful tasks in real situations, but also as a test bed for semantic and pragmatic theories of dialogue interaction.

Besides the difficulties inherent to any natural language understanding system (like e.g. ambiguity and context-dependence), SDSs are faced with additional challenges that often derive from the lack of technical accuracy of their components, most prominently the automatic speech recogniser (ASR). For this reason, SDSs are confronted with a great degree of uncertainty when processing user utterances. In those cases where understanding does not fail completely, a system will typically be able to form some hypotheses about the input received from the user. However, judging the quality of these hypotheses is itself a highly uncertain task, and finding answers to questions such as *'Did the user really say X?'*, *'Did the user mean Y?'* or *'Is Z what the user intended me to do?'* can be a very hard enterprise for a dialogue system.

One of the crucial aspects that contributes to reducing uncertainty is the use of meaningful clarification and grounding strategies that are able to tackle the problems that the system encounters and, when appropriate, give feedback to the user about its internal representations.

In this paper we describe ongoing work carried out within the project "DEAWU: DEAling With Uncertainty in Spoken Dialogue Systems".[1] The

[1]DEAWU is a Marie Curie Transfer of Knowledge project funded by the Eu-

main aims of the project are (1) to improve the *self-awareness* of SDSs with respect to the quality of their hypotheses, (2) to investigate the strategies employed by humans to deal with understanding problems, and (3) to explore how these strategies can be adapted to and implemented in SDSs.

2 General Approach

To investigate the issues outlined in the previous section, we have chosen a restricted domain based on a game commonly known as *Pentomino*, where 12 different pieces made up of 5 squares can be placed together on a board to form a predefined figure or puzzle. The game is adapted to a communicative situation by having two participants with two distinct roles, *player* and *executor* (that in an SDS correspond to user and system, respectively), who collaborate on building the puzzle. This setting, which is comparable to the one used in the classic Tangram experiments of [2], gives rise to interactions common in instructional command-and-control-style dialogue, where the player holds the initiative and gives instructions to the executor in order to select appropriate pieces and place them on a gridded board. The nature of the task thus brings in rich and complex reference resolution phenomena.

A prototype of a system that is able to handle this task is currently being implemented using an agent-based architecture with Sphinx-4 [6] as ASR and DIPPER [1] as dialogue management framework. The prototype is currently able to understand and execute simple user commands, and will in the future be capable of interpreting fine-grained referring expressions and show relevant clarification and grounding behaviour.

In parallel to the system development work, we follow a data-driven approach whereby we collect and analyse human-human dialogues to investigate how humans deal with understanding problems and how this knowledge can be transferred to SDSs. In order to bypass the differences observed between human-human and human-computer interaction [4], the approach we favour (similar in spirit to the experiments of [5]) is to collect human-human dialogues under strictly controlled conditions. In particular, we distort a conversational setting in a controlled and systematic way to study the strategies humans use to deal with restrictions that are typically present in SDSs. In this respect, in recent work [3] we reported on an experiment where we compared free turn-taking with a push-to-talk condition that simulates a restricted turn-taking policy commonly used in SDSs. Interestingly,

ropean Union, grant EU #FP6-2002-Mobility 3014491. For further details see http://www.ling.uni-potsdam.de/DEAWU

we found that push-to-talk dialogues were in fact more efficient than free turn-taking ones (the same goal was reached using fewer words). This was achieved by a different macro-pattern of dialogue acts, with push-to-talk dialogues being more focussed on the task and containing less interaction management actions such as acknowledgements and backchannels.

3 Current and Future Work

We are currently running a *noisy channel* experiment in which the input received by the executor is partially blocked by random noise. This distortion increases the need for clarification and motivates the development of global strategies to deal with defective acoustic channels. We expect to be able to present the results of this study in the near future.

We have observed that in the chosen *Pentomino* domain participants can solve the task without attempting to correctly interpret their partner's utterances in every situation. In order to evaluate to what extent the strategies developed by the subjects are dependent on the task at hand, we plan to experiment with other tasks as well. We are currently running pilot studies that explore the use of a *dictation task* where correct understanding is mandatory.

In addition to this, we also plan to conduct a Wizard-of-Oz experiment to collect a corpus of referring expressions that users are likely to use with our prototype system.

References

[1] J. Bos, E. Klein, O. Lemon, and T. Oka. DIPPER: Description and Formalisation of an Information-State Update Dialogue System Architecture. In *Proc. of the 4th SIGdial Workshop on Discourse and Dialogue*, Sapporo, Japan, 2003.

[2] H. Clark and Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[3] R. Fernández, T. Lucht, K. Rodríguez, and D. Schlangen. Interaction in task-oriented human-human dialogue: The effects of different turn-taking policies. In *Proc. of the first International IEEE/ACL Workshop on Spoken Language Technology*, Palm Beach, Aruba, 2006.

[4] N.M. Fraser and G.N. Gilbert. Simulating speech systems. *Computer Speech and Language*, 5:81–99, 1991.

[5] G. Skantze. Exploring human error recovery strategies: Implications for spoken dialogue systems. *Speech Communication*, 45(3):325–341, 2005.

[6] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel. Sphinx-4: A flexible open source framework for speech recognition. Technical Report SMLI TR2004-0811, Sun Microsystems Inc., 2004.