Contextual Information for Disambiguation in a Speech-to-Speech Translation System

Maite Taboada
Carnegie Mellon University
and
Universidad Complutense de Madrid
taboada+@cmu.edu

15 November 1996

For any given utterance out of what we can loosely call *context*, there is usually more than one possible interpretation. A speaker's utterance of an elliptical expression, like the figure "twelve fifteen," might have a different meaning depending on the context of situation, the way the conversation has evolved until that point, and the previous speaker's utterance.

In this paper I will explore those three strata of discourse in order to gain on understanding of how speakers *mean* in a given situation. This work is part of the JANUS multi-lingual speech-to-speech translation system designed to translate spontaneous dialogue in a limited domain [Lavie 96]. JANUS is designed to deal with the kind of problems that naturally occur in spontaneous speech — such as mispronunciations, restarts, noises, slightly ungrammatical input, and the lack of clear sentence boundaries — with additional errors introduced by the speech recognizer. The machine translation component of JANUS handles these problems using two different approaches: GLR* and Phoenix. The GLR* parser [Lavie and Tomita 93] is designed to be more accurate, whereas the Phoenix parser [Ward 91] is more robust. Both are language-independent and follow an interlingua-based approach. The current system translates spontaneous dialogues in the scheduling domain, with English, Spanish, and German as both source and target languages.

This project addresses the problem of choosing the most appropriate semantic parse for any given input. The approach is to combine discourse information with the output of the Phoenix parser, a set of possible parses for an input string. There might be more than one acceptable semantic parse for an input. The discourse module interacts with the parser, selecting one of these possibilities. The decision is to be based on

1. The domain of the dialogue. JANUS deals with dialogues restricted to

a domain, such as scheduling an appointment, or making travel arrangements. The general topic provides some information about what types of exchanges, and therefore speech acts, can be expected.

2. The macro-structure of the dialogue up to that point. We can divide any dialogue in smaller, self-contained units that provide information on what phases are over or yet to be covered: Are we past the greeting phase? Has there been any agreement on acceptance of services? If one of the speakers has reserved a flight, should we expect a payment phase at some point in the rest of the conversation?

3. The structure of adjacency pairs [Schegloff and Sacks 73], together with the responses to speech functions [Halliday 94, Martin 92]. If one speaker has uttered a request for information, we should expect some sort of response to that — an answer, a disclaimer, even a clarification subdialog.

The context module in the system keeps a global history of the conversation, from which it will be able to estimate, for instance, the likelihood of a greeting once the opening phase of the conversation is over. A more local history predicts the expected response in any adjacency pair, such as a question-answer sequence.

References

[Halliday 94] M.A.K. Halliday. *An Introduction to Functional Grammar*, Edward Arnold, London, 1994 (2nd edition).

[Lavie 96] A. Lavie, D. Gates, M. Gavaldà, L. Mayfield, A. Waibel, L. Levin. Multi-lingual Translation of Spontaneously Spoken Language in a Limited Domain. In *Proceedings of COLING 96*. Copenhagen. 1996.

[Lavie and Tomita 93] A. Lavie and M. Tomita. GLR*: An Efficient Noise Skipping Parsing Algorithm for Context Free Grammars. *Proceedings of the Third International Workshop on Parsing Technologies, IWPT 93*, Tilburg, The Netherlands, 1993.

[Martin 92] J. Martin *English Text: System and Structure.* John Benjamins. Philadelphia/Amsterdam. 1992.

[Schegloff and Sacks 73] E. Schegloff and H. Sacks. Opening up Closings. *Semiotica 7*, 289-327, 1973.

[Ward 91] W. Ward. Understanding Spontaneous Speech: the Phoenix System. In *Proceedings of ICASSP 91*, 1991.