Improving Translation through Contextual Information

Maite Taboada[*]
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
taboada+@cmu.edu

Abstract

This paper proposes a two-layered model of dialogue structure for task-oriented dialogues that processes contextual information and disambiguates speech acts. The final goal is to improve translation quality in a speech-to-speech translation system.

1 Ambiguity in Speech Translation

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. "Twelve fifteen" could be the time "a quarter after twelve", the price "one thousand two hundred and fifteen", the room number "one two one five", and so on. Although English can conflate all those possible meanings into one expression, the translation into other languages usually requires more specificity.

If this is a problem for any human listener, the problem grows considerably when it is a parser doing the disambiguation. In this paper, I explain how we can use discourse knowledge in order to help a parser disambiguate among different possible parses for an input sentence, with the final goal of improving the translation in an end-to-end speech translation system.

The work described was conducted within the JANUS multi-lingual speech-to-speech translation system designed to translate spontaneous dialogue in a limited domain (Lavie et al., 1996). The machine translation component of JANUS handles these problems using two different approaches: the Generalized Left-to-Right parser GLR* (Lavie and Tomita, 1993) and Phoenix, the latter being the focus of this paper.

2 Disambiguation through Contextual Information

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 set of possible parses provided by the Phoenix parser for an input string. The discourse module selects 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 a dialogue into smaller, self-contained units that provide information on what phases are over or yet to be covered: Are we past the greeting phase? If a flight was reserved, should we expect a payment phase at some point in the rest of the conversation?
3. The structure of adjacency pairs (Schegloff and Sacks, 1973), together with the responses to speech functions (Halliday, 1994; Martin, 1992). If one speaker has uttered a request for information, we expect some sort of response to that — an answer, a disclaimer or a clarification.

The domain of the dialogues, named *travel planning domain*, consists of dialogues where a customer makes travel arrangements with a travel agent or a hotel clerk to book hotel rooms, flights or other forms of transportation. They are *task-oriented dialogues*, in which the speakers have specific goals of carrying out a task that involves the exchange of both information and services.

Discourse processing is structured in two different levels: the context module 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

[*] The author gratefully acknowledges support from "la Caixa" Fellowship Program, ATR Interpreting Laboratories, and Project Enthusiast.

any adjacency pair. such as a question-answer sequence. The model adopted here is that of a two-layered finite state machine (henceforth FSM). and the approach is that of *late-stage disambiguation*. where as much information as possible is collected before proceeding on to disambiguation. rather than restricting the parser's search earlier on.

3 Representation of Speech Acts in Phoenix

Writing the appropriate grammars and deciding on the set of speech acts for this domain is also an important part of this project. The selected speech acts are encoded in the grammar — in the Phoenix case. a semantic grammar — the tokens of which are concepts that the segment in question represents. Any utterance is divided into SDUs — Semantic Dialogue Units — which are fed to the parser one at a time. SDUs represent a full concept. expression. or thought. but not necessarily a complete grammatical sentence. Let us take an example input, and a possible parse for it:

```
(1) Could you tell me the prices at the Holiday Inn?
([request] (COULD YOU
([request-info] (TELL ME
([price-info] (THE PRICES
([establishment] (AT THE
([establishment-name] (HOLIDAY INN))))))))))
```

The top-level concepts of the grammar are speech acts themselves. the ones immediately after are further refinements of the speech act. and the lower level concepts capture the specifics of the utterance. such as the name of the hotel in the above example.

4 The Discourse Processor

The discourse module processes the global and local structure of the dialogue in two different layers. The first one is a general organization of the dialogue's subparts: the layer under that processes the possible sequence of speech acts in a subpart. The assumption is that negotiation dialogues develop in a predictable way — this assumption was also made for scheduling dialogues in the Verbmobil project (Maier. 1996) —. with three clear phases: *initialization. negotiation.* and *closing.* We will call the middle phase in our dialogues the *task performance phase.* since it is not always a negotiation *per se.* Within the task performance phase very many subdialogues can take place. such as information-seeking, decision-making, payment. clarification. etc.

Discourse processing has frequently made use of sequences of speech acts as they occur in the dialogue. through bigram probabilities of occurrences. or through modelling in a finite state machine. (Maier. 1996; Reithinger et al.. 1996; Iida and Yamaoka. 1990; Qu et al.. 1996). However. taking into account only the speech act of the previous segment might leave us with insufficient information to decide — as is the case in some elliptical utterances which do not follow a strict adjacency pair sequence:

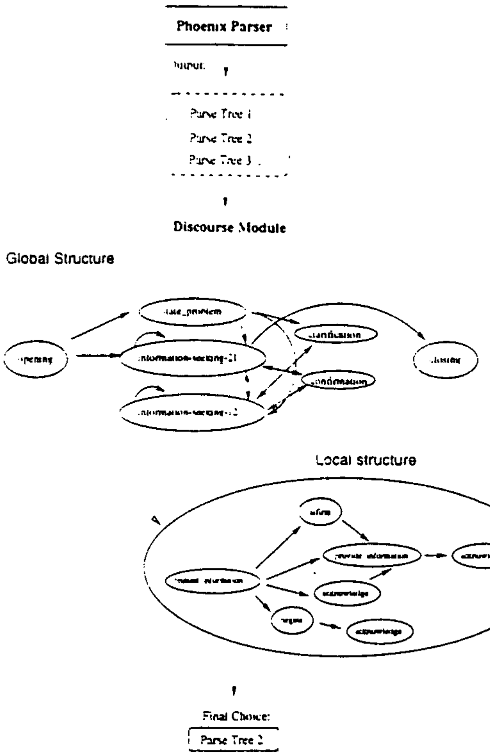

Figure 1: The Discourse Module

(2) (*talking about flight times...*)
S1 I can give you the arrival time. Do you have that information already?
S2 No. I don't.
S1 It's twelve fifteen.

If we are in parsing the segment "It's twelve fifteen". and our only source of information is the previous segment. "No. I don't". we cannot possibly find the referent for "twelve fifteen". unless we know we are in a subdialogue discussing flight times. and arrival times have been previously mentioned.

Our approach aims at obtaining information both from the subdialogue structure and the speech act sequence by modelling the global structure of the dialogue with a FSM. with opening and closing as initial and final states. and other possible subdialogues in the intervening states. Each one of those states contains a FSM itself. which determines the allowed speech acts in a given subdialogue and their sequence. For a picture of the discourse component here proposed. see Figure 1.

Let us look at another example where the use of information on the previous context and on the speaker alternance will help choose the most appropriate parse and thus achieve a better translation.

The expression "okay" can be a prompt for an answer (3), an acceptance of a previous offer (4) or a backchanneling element, i.e., an acknowledgement that the previous speaker's utterance has been understood (5).

(3) S1 So we'll switch you to a double room, okay?

(4) S1 So we'll switch you to a double room.
S2 Okay.

(5) S1 The double room is $90 a night.
S2 Okay, and how much is a single room?

In example (3), we will know that "okay" is a prompt, because it is uttered by the speaker after he or she has made a suggestion. In example (4), it will be an acceptance because it is uttered after the previous speaker's suggestion. And in (5) it is an acknowledgment of the information provided. The correct assignment of speech acts will provide a more accurate translation into other languages.

To summarize, the two-layered FSM models a conversation through transitions of speech acts that are included in subdialogues. When the parser returns an ambiguity in the form of two or more possible speech acts, the FSM will help decide which one is the most appropriate given the context.

There are situations where the path followed in the two layers of the structure does not match the parse possibility we are trying to accept or reject. One such situation is the presence of clarification and correction subdialogues at any point in the conversation. In that case, the processor will try to jump to the upper layer, in order to switch the subdialogue under consideration. We also take into account the situation where there is no possible choice, either because the FSM does not restrict the choice — i.e., the FSM allows all the parses returned by the parser — or because the model does not allow any of them. In either of those cases, the transition is determined by unigram probabilities of the speech act in isolation, and bigrams of the combination of the speech act we are trying to disambiguate plus its predecessor.

5 Evaluation

The discourse module is being developed on a set of 29 dialogues, totalling 1,393 utterances. An evaluation will be performed on 10 dialogues, previously unseen by the discourse module. Since the module can be either incorporated into the system, or turned off, the evaluation will be on the system's performance with and without the discourse module. Independent graders assign a grade to the quality of the translation[1]. A secondary evaluation will be based on the quality of the speech act disambiguation itself, regardless of its contribution to translation quality.

6 Conclusion and Future Work

In this paper I have presented a model of dialogue structure in two layers, which processes the sequence of subdialogues and speech acts in task-oriented dialogues in order to select the most appropriate from the ambiguous parses returned by the Phoenix parser. The model structures dialogue in two levels of finite state machines, with the final goal of improving translation quality.

A possible extension to the work here described would be to generalize the two-layer model to other, less homogeneous domains. The use of statistical information in different parts of the processing, such as the arcs of the FSM, could enhance performance.

References

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

Hitoshi Iida and Takyuki Yamaoka. 1990. Dialogue Structure Analysis Method and Its Application to Predicting the Next Utterance. Dialogue Structure Analysis. German-Japanese Workshop, Kyoto, Japan.

Alon Lavie, Donna Gates, Marsal Gavaldà, Laura Mayfield, Alex Waibel, Lori Levin. 1996. Multi-lingual Translation of Spontaneously Spoken Language in a Limited Domain. In *Proceedings of COLING 96.* Copenhagen.

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

Elisabeth Maier. 1996. Context Construction as Subtask of Dialogue Processing: The Verbmobil Case. In *Proceedings of the Eleventh Twente Workshop on Language Technology. TWLT 11.*

James Martin. 1992. *English Text: System and Structure.* John Benjamins. Philadelphia/Amsterdam.

Yan Qu, Barbara Di Eugenio, Alon Lavie, Lori Levin. 1996. Minimizing Cumulative Error in Discourse Context. In *Proceedings of ECAI 96*, Budapest, Hungary.

Norbert Reithinger, Ralf Engel, Michael Kipp, Martin Klesen. 1996. Predicting Dialogue Acts for a Speech-to-Speech Translation System. In *Proceedings of ICSLP 96*, Philadelphia, USA.

Emmanuel Schegloff and Harvey Sacks. 1973. Opening up Closings. *Semiotica 7*, pages 289-327.

Wayne Ward. 1991. Understanding Spontaneous Speech: the Phoenix System. In *Proceedings of ICASSP 91.*

[1] The final results of this evaluation will be available at the time of the ACL conference.