# Towards Automated Analysis of Spoken Dialogue Using Discourse Topology

**Susann LuperFoy**
The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102 USA
luperfoy@mitre.org
and
Georgetown University
Department of Linguistics
Washington, DC 22057 USA

**David Duff**
The MITRE Corporation
1820 Dolley Madison Blvd.
McLean, VA 22102 USA
duff@mitre.org

## Abstract

This paper describes our current efforts in empirical analysis of human-human dialogue interaction data. The methods we propose abstracts away from the linguistic content of a dialogue to analyze acoustic and interaction phenomena directly. The focus is on properties of the speech signal and on language-independent interaction behavior as opposed to information content of the utterances exchanged between dialogue participants. We are exploring machine learning techniques for ways to convert our algorithms to trainable or adaptable system components.

## Motivation: Robust Dialogue Analysis

Manual and automated analysis of empirical dialogue data are used to develop models for human-machine interaction. The study of human-human spoken dialogue helps user interface designers develop improved models for communicative behavior in computer-human interaction, providing insights into the needs and propensities of the human dialogue partner (e.g., Di Eugenio, et al., 1997; Walker, 1994). We are engaged in an ongoing project, the ultimate goal of which is a suite of automatic methods for analysis of spoken monologue and multiparty discourse processing (Duff et al. 1996, LuperFoy et al. 1997).

An alternative motivation, and the one pursued here, is the analysis of large amounts of human-human dialogue behavior as an end in itself. An instance of this sort of effort is the Discourse Resource Initiative (Allen, et al., 1997; DRI, 1996) the aim of which is to facilitate cooperation among discourse researchers through establishment of common data annotation schemata and through sharing of discourse corpora, analysis tools, and statistical results. The empirical analysis task we have undertaken is described in terms of (1) characteristics of the raw data under study, (2) input to the discourse-level analyzer from remaining software components, (3) constraints stemming from system-level requirements, i.e., the purpose of the overall analysis task, and (4) an emphasis on generalizability, scalability and portability of software development results.

The data we study contain large amounts of spontaneous spoken dialogue with an unrestricted vocabulary. These data contain false starts, filled pauses, self-corrections, backchannels, overlapping speech between participants, and occasional non-verbal vocalizations such as laughter, singing, and sighs in addition to dialectal variation, and ungrammaticality and fragmentary sentences common in conversational discourse.

Ideally, we want our system to be universal, working with languages for which no large vocabulary continuous speech recognition software exists. Even for English and other languages for which recognizers exist, performance on spontaneous unrestricted dialogue is not yet accurate enough to support reliable natural language processing, and the quantities of data involved make substitution of speech recognition with manual transcription too difficult or too expensive for the large scale tasks we are pursuing.

So a set of software engineering constraints stem from the limitations of the containing system that invokes our discourse-level analyzer. That system is assumed to lack high accuracy speech recognition and "upstream" modules of morphology, syntax, and sentential semantics on which traditional discourse semantic processing tends to rely. To cope with these constraints, the current effort began with the extraction from spoken discourse corpora of features we call the "discourse topology". These are the discourse-level properties of the data that can be extracted, measured, or inferred in the absence of output from remaining components of natural language processing systems (morphology, syntax, sentential semantics).

We are testing the hypothesis that useful information can be inferred about the structure and content of a discourse by looking at these topological features, i.e., that these measures can inform discourse interpretation tasks such as topic segmentation, characterization of discourse genre, characterization of speaker, relative social roles of participants in a multiparty discourse, assignment of speech acts, assignment of conventional structure to various types of stylized discourses, and more.

The application task we addressing is the search, in a single dialogue or in a corpus, for certain dialogue-level events which can be recognized independently of the lexical content or information content of the conversation. For example, for some information retrieval applications it is useful to detect types of discourse segment, e.g., speakers negotiating, giving instructions or training, arguing, interviewing, etc. We claim that such dialogue

patterns can be, in some cases, identified without recognizing the words of the dialogue and without the depth of traditional natural language understanding methods.

In other applications, we want to extract high-level discourse structure in order to index a corpus for subsequent tasks of search and retrieval keyed on occurrences of specific structural patterns. Yet another type of task calling for dialogue topological analysis is the classification of an entire dialogue or clustering of a corpus of dialogues according to global parameters such as purpose or interaction style. For many such applications, the dialogue topological properties of the dialogue are at least as important as the lexical information in a dialogue for characterizing the nature of the interaction; we often gain more from knowing the amount of overlapping speech, range of voice frequency and amplitude for the two speakers, and occurrence of questions, the detection of which may benefit more from topology than grammatical analysis.

A final set of constraints stem from our goal to build methods and software systems that are trainable, either automatically or manually, on new data and for novel information retrieval tasks. That is, we want to construct generic analysis tools without knowing to which languages they will ultimately be applied nor what questions will be asked concerning the eventual discourse data sources.

## Approach: Dialogue Topology

The *dialogue topology* approach to dialogue analysis is based exclusively on high-level features of the speech signal, such as timing and prosodic information. Our goal is to discover the mechanism that allows a human, overhearing a conversation between two speakers in a language that one does not understand. One is able to tell a lot about what was going on between the two speakers (through prosody and timing information), even without understanding any of the actual words that they utter. We have implemented a tool for extracting dialogue topology from the speech signal in the absence of semantic or information content from linguistic processing: lexical, morphological, syntactic, and sentential semantic.

Thus far, our experimental data has been spoken dialogue corpora obtained from the Linguistic Data Consortium, specifically the Switchboard (LDCa) and Callhome (LDCb) corpora. The LDC data we have used were annotated manually, or via a semi-automated process, although we believe that with modest effort, it will be possible to automate parts of the data collection process. We were granted access to a database of prosodic information developed for Switchboard data (Shriberg, et al., 1997) as part of an effort to improve the word error rate of large vocabulary conversational speech recognition (LVCSR) via language models that incorporate discourse-level information.

Thus far we have primarily explored timing information, specifically the start and stop times of utterances of individual speakers as marked in the LDC data sources. Derived from this low-level information is an intermediate level of analysis that identifies key discourse events and features such as pauses and overlaps, and assign them properties according to their duration and their location in the dialogue. At the next higher level of abstraction, we have begun to identify patterns of dialogue interaction. For example, from timing information we locate pauses and overlapping speech from which we can then construct patterns of speaker interruption and competition for the floor. We characterize these patterns based on the duration of overlap and how the competition is resolved—i.e., which speaker takes over the floor. It is from these types of patterns that the name "topology" emerged, since they are indicators of the general "shape" of the exchange as opposed to the actual semantic content.

Topological processing can accumulate features at any of several granularity points:

- *Utterance or segment.* Features can be extracted to classify specific utterances or groups of utterances in a dialogue. Dialogue segment boundaries may be predetermined in the input data or, alternatively, topological analysis can assign segmentation boundaries. Segments themselves can then be assigned features based on topology or based on semantic content.
- *Entire dialogue.* Features can be collected to serve as an overall classification of a dialogue. For example, the relative roles of speakers, their relationship to one another (formal, familiar, friendly, antagonistic), the total length of the dialogue, percentage of time the individual participants were talking, etc.
- *Dialogue participants across sessions.* Information can be accumulated on a particular participant in a corpus of dialogues, e.g., to determine a speaker's characteristic speech patterns, their preferences, their personality or dialect, or their function in a management hierarchy of an organization.

Our work thus far has focused primarily on utterance and dialogue-level feature accumulation as we sought indicators of features such as question/answer pairs in dialogue, backchannel utterances, familiar versus formal dialogues, etc. in timing and prosodic data

To support our own exploration of candidate dialogue features to measure, we have built a preliminary version of a dialogue visualization tool that allows us to explore the data generated through topological extraction. It presents in parallel the "thumbnail" diagrams of several dialogues, allowing for quick viewing of topological features by the developer. Contributions by Speaker A, Speaker B, pauses, and overlapping speech are indicated with different colors. The developer can manipulate their view of the data by zooming in, panning, setting the number of dialogues to visualize at a time, and defining thresholds for distinguishing primary utterances from backchannels. This improves our ability to search for relationships between patterns of features and salient discourse properties. For example, a given pattern of backchanneling may be thought to indicate greater support from the hearer for the speaker, and certain patterns of pauses and overlapping speech can indicate a level of formality of the discourse setting, or familiarity of the speakers, and thus allow one to compare dialogue segments. Our visualization tools

helps us in rapid hypothesis formulation, to quickly characterize and contrast a set of dialogues prior to investing in an in-depth analysis.

Our next step is to develop adaptable versions of these techniques by using manually annotated discourse selections to assess the strength of correlations discovered in this way, so that they can be developed into automatically acquired models. We are interested in using machine learning techniques in order to accomplish the fourth of our objectives outlined in the introductory section of this paper, that is, to make our algorithms flexible and easy to apply to new dialogue patterns of interest.

Several key questions affecting the success of this approach remain to be answered. One is whether it will be possible to build and train language models which can detect the features of interest with any generality and reliability. This, in turn, would seem to depend on our success in finding learning algorithms that are amenable to the particular characteristics of our data and of the features we are trying to extract (thus, our interest in this workshop). Another is whether there will be any language universality to the features, that is, whether a pattern for detecting a certain feature in English dialogues (an argument, a backchannel utterance, etc.) would necessarily work for dialogues in other languages.

## Acknowledgments

We acknowledge the significant contribution of participants in the Dialogue Modeling team at the 1997 Summer Workshop lead by Dan Jurafsky. We received data and useful ideas from Liz Shriberg, Andreas Stolke, and Klaus Ries, and from MITRE Discourse team members Lisa Harper, Daniel Loehr, Keith Miller and Florence Reeder.

## References

Allen, J., and Core, M. Draft of DAMSL: Dialogue Act Markup in Several Layers. 1997 Available from http://www.cs.rochester.edu:80/research/trains/annotation.

Di Eugenio, B., Jordan, P. W., Thomason, R. H., and Moore, J. D. Reconstructed Intentions in Collaborative Problem Solving Dialogues. *AAAI Fall Symposium on Communicative Action in Humans and Machines, 1997.*

DRI (Discourse Resource Initiative). See webpages at <http://www.georgetown.edu/luperfoy/Discourse-Treebank/dri-home.html>.

Duff, D. , B. Gates, and S. LuperFoy "An Architecture for Spoken Dialogue Management." In Proceedings of the 1996 International Conference on Speech and Language Processing (ICSLP). 1996.

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P., and Van Ess-Dykema, C. "Johns Hopkins LVCSR Workshop-97 SWBD Discourse Language Modeling Project Final Project Report. 1997

LDCa (Linguistic Data Consortium). The Callhome Speech Corpus. <http://www.ldc.upenn.edu/ldc/catalog-/html/package/ch.html>

LDCb (Linguistic Data Consortium). The Switchboard-1 Telephone Speech Corpus, <http://www.ldc.upenn.edu-/ldc/catalog/html/speech_html/SWBREV.html>

LuperFoy, S. and D. Duff "A Centralized Troubleshooting Mechanism for a Spoken Dialogue Interface to a Simulation Application." In proceedings of the 1996 International Symposium on Spoken Dialogue (ISSD). 1996.

LuperFoy, S. Summary of Multiparty Dialogue Sessions at first Discourse Resource Initiative Workshop at University of Pennsylvania Institute for Research in Computer Science. 1996. Available from: <http://epsilon3.-georgetown.edu/~discours/sigdial/sigdial

Shriberg, E., Bates, R., Taylor, P., Stolcke, A., Jurafsky, D., Ries, K., Coccaro, N., Martin, R., Meteer, M., and Van Ess-Dykema, C. "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?" 1997.

Walker, M. A., "Experimentally Evaluating Communicative Strategies: The Effect of the Task>" In proceedings AAAI annual meeting. 1994.