

New York University

Principal investigator: Ralph Grishman

Natural Language Understanding

Our task, broadly stated, is the development of systems for the understanding of narrative messages in limited domains. Improving the current state-of-the-art for such systems will require a better understanding of how to capture and utilize domain information, and how to effectively combine the various sources of information (syntactic, semantic, and discourse) to create a robust language analyzer.

For our study of the structure and use of domain information, we have focussed on CASREP messages describing the failure, diagnosis, and attempted repair of shipboard equipment. Because much of the information -- particularly the relation between the individual events in the narrative -- is implicit, a thorough understanding of these messages requires substantial knowledge of the equipment involved. We have captured this knowledge for one piece of equipment -- the starting air system for gas turbine propulsion systems -- through a model which incorporates structural and functional information about the equipment. We have then developed a natural language analyzer which uses the simulation capabilities of this model to determine the implicit causal relations between the events in a message. For example, if a message states "Compressor won't start. Shaft was sheared." the language analyzer could use the model to verify that the shearing was a plausible cause of the failure of the compressor.

For our study of the effective integration of syntactic, semantic, and discourse information, we are developing an analyzer for ship sighting messages (RAINFORMs). These messages use very compressed natural language, where many of the participants in an event are implicit; the challenge for the analyzer is to recover this information. To parse these messages, we have extended a general English grammar to allow the omission of case and complement markers ("of", "at", "to", "as"), copulas ("be"), and a few other constituents. Our parser does a best-first search to obtain the analysis which is consistent with the semantic constraints of the domain and has the fewest omissions. To recover missing operands, we are experimenting with a measure of message coherence based on cause/effect and enabling condition/action relations between events in the message. When semantic case frame constraints alone would allow several possible fillers for a missing operand, our system prefers the analysis yielding the highest message coherence.

Considerable additional experimentation is required to evaluate the role of this coherence measure, and to balance the syntactic, semantic, and discourse coherence constraints. We expect that the data for MUCK-II will provide some initial evaluation of these techniques.