

Project Summary:

Natural Language Understanding: Integrating Syntax, Semantics, and Discourse.

Institution: UNISYS

Principal Investigators: Dr. Lynette Hirschman, Dr. Martha Palmer

Technical Summary: The focus of the UNISYS research is on integrating multiple knowledge sources, including syntax, semantics, pragmatics, a domain model, and both domain-specific and domain-independent knowledge sources, to produce a system capable of understanding messages in a restricted domain.

Objectives:

- (1) Creation of an integrated, portable system for the detailed understanding of multi-paragraph text and messages.
- (2) Develop a principled computational treatment of semantic and discourse phenomena, including treatment of reference and temporal information.
- (3) Development of an integrated logic-based analysis algorithm which uses syntactic, semantic and discourse information to maximize robustness and search focus.
- (4) Build a spoken language system through integration of a large natural language system with a speech recognition system.

Recent Accomplishments:

- **System:** Development of modular, interleaved system architecture, enhancing portability by enforcing segregation of core PUNDIT from domain-specific modules. Demonstrated by port to domain of Navy Trouble Failure Reports.
- **Applications:** Processing of CASREP messages, producing tabular summary; Navy RAINFORM (sighting messages), and Trouble and Failure reports, producing database updates.
- **Technology Transfer Activity:** PUNDIT User's Guide released, system installed at National Library of Medicine, Unisys applications group; system requested by RADC, CECOM, Harvard, U. Penn., Cambridge U., Swedish Inst. of Computer Science.
- Treatment of fragmentary input as found in message traffic, by means of minimal extensions to the syntactic, semantic and pragmatic components.
- Semantic coverage of nominalized verbs, adjectival participles and noun predicates, based on the treatment of the underlying predicate, using reference resolution and temporal processing to complete the interpretation.
- Interactive selection component, interleaved with syntax, uses semantic (selectional) information to filter parses, producing dramatic, 8-fold, decrease in number of parses.
- Modules for processing intra-sentential temporal information and referring expressions (including definite and indefinite noun phrases, phrases with omitted determiners, and reference to events).
- Integration of contextual information from fixed message fields with that from free-text fields: system now handles complex multi-paragraph message formats.

Plans for next year:

- Demonstrate extensibility of PUNDIT to new messages, showing correct information capture by filling in 70% of target DB fields with a false positive rate of less than 5%.
- Integrate PUNDIT with knowledge representation and reasoning systems to provide inferencing capabilities for correct filling of DB fields.
- Treatment of intersentential temporal relations and implementation of a domain-independent discourse management component, for handling a variety of message formats.
- Demonstrate PUNDIT on Resource Management domain, in preparation for Spoken Language Understanding.