

Natural Language Research
Aravind Joshi, Mitch Marcus,
Mark Steedman, and Bonnie Webber
University of Pennsylvania

OBJECTIVE

The main objective is basic research and system development leading to (1) characterization of information carried by (a) syntax, semantics, and discourse structure, (b) their relation to information carried by intonation, and (c) development of methods for using this information for generation and understanding; (2) development of architectures for integration of utterance planning with lexical, syntactic and intonational choice; (3) development of incremental strategies for using syntactic, semantic, and pragmatic knowledge in understanding and generating language.

RECENT ACCOMPLISHMENTS

- Completed an implementation of a parser for the Tree Adjoining Grammar (TAG).
- Experimental evaluation of the TAG parser showed that the two-pass strategy in the current parser led to 90% reduction in the number of states and 85% reduction in the number of elementary trees, as compared to the one-pass strategy.
- Developed a formal theory unifying intonational structure, discourse information structures and surface syntactic structures within the categorial framework.
- Extended earlier characterization of natural language tense and aspect in terms of event ontologies and discourse structure.
- Learning — See adjoining report on “Large Corpus Development.”
- Completed a pilot implementation of the multi-functional cooperative response system.
- Developed a novel approach for integrating syntax and semantics in TAGs, and explored its application to generation and translation.

PLANS

- Develop applications of categorial theory to the problem of computer synthesis of contextually appropriate intonation in spoken language.
- Begin implementation of the theory of tense and aspect using an event calculus.
- Complete the work on the lexicon for the TAG parser and test the parser.
- Complete the development of a deterministic parser for TAGs.
- Develop the formalism of synchronous TAGs (for integrating syntax and semantics), develop a parser for synchronous TAGs and explore further its application to machine translation.
- Learning — See adjoining report on “Large Corpus Development.”