

# NATURAL LANGUAGE RESEARCH

*Aravind K. Joshi (PI), Mitch Marcus (Co-PI),  
Mark Steedman (co-PI), and Bonnie Webber (co-PI)*

Department of Computer and Information Science  
University of Pennsylvania  
Philadelphia, PA 19104

## PROJECT GOALS

Our 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; and (4) investigating how structural and statistical information can be combined, both for processing, as well as for acquisition.

## RECENT RESULTS

- Recently we have developed and implemented a new predictive left-to-right parser for tree adjoining grammars (TAG). This parser does not preserve the so-called valid prefix property and thereby achieves efficiency. A key discovery was that the valid prefix property does not necessarily hold for a parser for non-context-free grammars, although it holds trivially for context-free grammars (CFG).
- A new shift reduce parser for CFG, based on the shared forest approach, was designed and implemented. This is the first such parser whose performance can be precisely defined. Previous approaches to parsing by the use of the shared forest approach by other investigators elsewhere have not been successful in giving a precise characterization.
- TAGs are tree-based systems and have a domain of locality larger than CFGs. It was always understood that a compositional semantics can be defined for a TAG based on the object language trees, but there is no point in doing this because this approach ignores the TAG derivation tree completely. Hence, a formalism, called synchronous TAG, for integrating syntax and semantics using the derivation trees of TAG, was developed. This work was carried out in collaboration with Stuart Shieber (Harvard University).
- Categorical theory has been successfully applied to the problem of computer synthesis of contextually appro-

priate intonation in spoken language.

- Implementation of the theory of tense and aspect using an event calculus was initiated.
- Extension of a formal theory unifying intonational structures, discourse information structures and surface syntactic structures within a categorial framework to a wide range of coordinate and relativized constructions.
- The distribution of verb phrase ellipses with respect to antecedent location as preface to a focus analysis of the phenomenon was carried out using the Brown corpus.
- A study of free-adjunct-and purpose clauses was carried out in a corpus of natural language instructions as a first step in developing an adequate representational formalism for action description in natural language.

## PLANS FOR THE COMING YEAR

- Complete implementation of the shift reduce parser for TAG, based on the shared forest approach.
- Complete the work on coordination in TAG and begin implementation of this approach in the TAG parser.
- Implement the synchronous TAG formalism for integrating syntax and semantics of TAG and develop a small machine translation system.
- Begin exploration of statistical techniques for both parsing and acquisition of TAGs.
- Further development of the event calculus implementation of tense and aspect.
- Continue development of the categorial grammar-driven theory of computer synthesis of contextually appropriate intonation.
- Develop an algorithm for resolving instances of verb phrase ellipses. This algorithm will be focus-based and will build up our earlier work on *centering* which was developed for the interpretation of definite noun-phrases in discourse.