TACITUS: A Message Understanding System

Jerry R. Hobbs, Douglas Appelt, John Bear,
Mark Stickel, and Mabry Tyson
SRI International
Menlo Park, California
(415) 859-2220

TACITUS is a general and domain-independent natural language processing system, used so far primarily for message processing. It performs a syntactic analysis of the sentences in the text, producing a logical form. Next, inferential pragmatics processing is applied to the logical form to solve problems of schema recognition, reference resolution, metonymy resolution, and the interpretation of vague predicates. An analysis component then produces the desired output for the application. TACITUS has been applied to several quite different domains, including naval equipment failure reports, naval operations reports, and terrorist reports.

The syntactic component is the DIALOGIC system, developed originally for the TEAM transportable, natural language interface to databases. The parser is bottom-up and produces all the parses at once, together with their logical forms. Its grammar is among the largest computer grammars of English in existence, giving nearly complete coverage of such phenomena as sentential complements, relative clauses, adverbials, sentence fragments, and the most common varieties of conjunction. Selectional constraints are applied, and there are a large number of heuristics for selecting the preferred parses of ambiguous sentences. The logical form produced is an "ontologically promiscuous" version of first-order predicate calculus, in which relations of grammatical subordination are represented. Optionally and where possible, the logical forms for different parses are merged into a neutral representation.

Pragmatics processing is based on abductive inference, implemented in the Prolog Technology Theorem Prover (PTTP), using a knowledge base encoding commonsense and domain-specific knowledge in the form of predicate-calculus axioms. The fundamental idea is that the interpretation of a sentence is the minimal proof from the knowledge base of the logical form of the sentence together with the constraints predicates impose on their arguments, allowing for coercions, where one merges redundancies where possible and makes assumptions where necessary. This formulation leads to an elegant, unified solution to the problems of schema recognition, reference resolution, metonymy resolution, and the interpretation of vague predicates. The output of this component is an elaborated logical form with the relevant inferences drawn and the identities of entities explicitly encoded.

Finally, an analysis component takes the interpretation produced by the pragmatics component and generates the required output. For the equipment failure reports, this is a diagnosis of the problem described. For the naval operations reports and the terrorist reports, this is entries for a database. With very little effort, analysis components could be constructed for a number of other applications, such as message routing and message prioritizing.

A number of convenient knowledge-acquisition facilities have been implemented for TACITUS. These include a menu-based lexical acquisition component, a sort hierarchy editor, and a component allowing entry of axioms in a subset of English.