

Research and Development in Natural Language Understanding

Ralph Weischedel
BBN Systems and Technologies Corporation

Duration: December 21, 1984 - September 30, 1989

Brief Summary of Objectives: There are three objectives of the contract: to perform research and development in parallel parsing, semantic representation, ill-formed input, discourse, and tools for linguistic knowledge acquisition, and to integrate software components from BBN and elsewhere to produce Janus, DARPA's New Generation Natural Language Interface, and to demonstrate state-of-the-art natural language technology in DARPA applications. The following software has been distributed: natural language system; IRACQ, knowledge acquisition system; System components and knowledge bases of Janus; KL-TWO knowledge representation and inference system integrated with Janus; various components for DARPA's Spoken Language Systems Project at BBN.

Summary of accomplishments:

1. The first 20 months of the effort were devoted to technology transfer of IRUS, the understanding component of Janus to the Government. To this end, BBN delivered IRUS, its knowledge acquisition tools (IRACQ and KREME), and knowledge bases for lexical semantics, lexical syntax, a domain model, and transformation rules to data base structure, to Texas Instruments for integration in DARPA's Fleet Command Center Battle Management Program (FCCBMP). Working with the Naval Ocean Systems Center, BBN officially demonstrated IRUS-86 in summer 1986 as part of the FCCBMP to representatives of CINCPACFLT, DARPA, SPAWAR, and NOSC. As of August, 1986, all of the components were transferred to NOSC, and BBN began focus on the component research goals and on the system integration goals of Janus.
2. In conjunction with the USC/Information Sciences Institute, USC/ISI's Penman language generation component was integrated with IRUS, to provide paraphrases and answers in English. This was demonstrated in May, 1987.
3. In the second version of Janus, the Penman generation component was replaced by the Mumble grammar for generation and BBN's Spokesman text planner. This was delivered to Lockheed for integration with DARPA's AirLand Battle Management Program in late 1987. We estimate that this second version of Janus had greater functionality (e.g., not just paraphrase and answer generation, but also multi-paragraph, multi-page output generation), and that the generator was 2-10 times faster than the first version.
4. Our component research made several direct contributions of technology to BBN's Spoken Language Systems effort, including the initial grammar (a unification-based grammar), the semantic representation language, a parallel parsing algorithm, and components for mapping from the semantic representation (an intensional logic) to code for one or more application systems. In addition, we are in process of publishing our research results on dealing with errorful, novel, or unclear language; clarification dialogue; centering algorithms for reference resolution; knowledge acquisition; and a hybrid semantic representation based on intensional logic and a terminological knowledge representation.
5. Our design for seamless interfaces has been adopted in the human-machine interface for DARPA's CASES expert system.
6. Software resulting from our research and development effort has now been distributed (for R&D purposes) to the FCCBMP at the Pacific Fleet Command Center, NOSC, the University of Pennsylvania, the University of Massachusetts, Harvard University, USC/ISI, Texas Instruments, Lockheed Austin Division, RADC, and NSA.