

11

INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT Unisys/Defense Systems

ARPA ORDER NUMBER: 5282

PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)

CONTRACTOR: Unisys/Defense Systems

CONTRACT AMOUNT: \$683,105

CONTRACT NO: N00014-85-C-0012

EFFECTIVE DATE OF CONTRACT: 4/29/85

EXPIRATION DATE OF CONTRACT: 4/29/87

PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman

PHONE NO. (215) 648-7554

SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: 11/1/86 - 1/31/87

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

DTIC FILE COPY

AD-A178 411

DTIC
ELECTE
MAR 27 1987
A

This document has been approved for public release and sale; its distribution is unlimited.

87 3 11 114



**System Development
Corporation**
A Burroughs Company

System Development Corporation
PO Box 517
Paoli PA 19301
Telephone 215 648 7200

HDT-87-040
27 February 1987

Office of Naval Research
Department of the Navy
800 N. Quincy Street
Arlington, VA 22217-5000

Attention: D. Alan Meyrowitz, Code 1133

Reference: DARPA Contract No. N00014-85-C-0012

Subject: Status Report of the "Darpa Natural Language
Understanding Program"
Reporting Period 11/1/86 - 1/31/87

Gentlemen:

In accordance with the referenced contract requirements, we are pleased to submit an R & D Status Report for the DARPA Natural Language Understanding Program.

For any questions, please feel free to contact either Dr. Lynette Hirschman, Principal Investigator (215/648-7554) or the undersigned (215/648-2263).

Very truly yours,

Unisys, Defense Systems
(formerly System Development Corporation)

H. D. Tuck
Contract Manager

HDT/d

cc: See Attached Distribution

Distribution

2 copies

LCOL Robert Simpson
IPTO
Defense Advanced Research Projects
Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

1 copy
(cover letter only)

Defense Contract Administration Services
Management Area - Philadelphia
P.O. Box 7699
Phila., PA 19101-7699
Attn: Mr. A. Stein, DCASR-PHI-GAAD-E1

6 copies

Director, Naval Research Laboratory
Attn: Code 2627
Washington, D.C. 20375

12 copies

Defense Technical Information Center
Bldg. 5, Cameron Station
Alexandria, Virginia 22314

1. Description of Progress

1.1. Grammar

1.1.1. Intermediate Syntactic Representation

Rules have been added in the Intermediate Syntactic Representation (ISR) component to produce a regularised representation for the modal verbs *will/would*, *shall/should*, *can/could*, and *may/might*, as in *Pump will not turn*. Modal verbs are represented in the ISR as operators in the same way that tense and aspect are represented. The semantically relevant verb, for example, *turn* in the sentence above, appears in the ISR position for the main verb of a clause. This change is illustrated by comparing the ISR for *Pump turns*,

[present,turn,[tpos(\square),[nvar([pump,singular,X])]]],

with that of *Pump will turn*,

[present,will,turn,[tpos(\square),[nvar([pump,singular,X])]]].

This provides a basis for a semantic treatment of the modal concepts of possibility and necessity, as well as for a treatment of the future tense, as in *New sac will arrive*.

1.2. Syntax/Semantics Interaction

An agenda for the work on syntax/semantics interaction has been agreed on, with three main goals: I) a first pass at an implementation where semantics interacts with syntax with respect to transitivity alternation verbs; II) a first pass at finding a correlation between the grammatical terms and the syntactic terms used by the verb mapping rules; III) a survey of relevant linguistic literature. A procedure for cooperative syntactic and semantic analysis of transitivity alternation verbs (e.g., *operate* as in *the ship's force operated the engine* [tr.] versus *the engine operated* [intr.]) has been designed. A restriction in the grammar will call the semantic clause-analyser during the parse of such verbs and the control structure of the semantic analyser will be changed in order to return the appropriate information to the parser, thus refocusing the search for the preferred type of object.

Because there appear to be many constraints on the instantiation of thematic roles, this kind of interaction between syntax and semantics may be possible with many verbs. Addressing this problem will serve to extend the current coverage and also restrict experimentation with the new control strategy to a carefully delimited arena.

The goal is to complete implementation of these changes during the coming quarter.

1.3. Semantics

1.3.1. Temporal Analysis

The temporal analysis component has been modified to consult a domain model before embarking on a full-scale analysis of a possible temporal adverbial. Previously, the time component only inspected the introductory preposition or conjunction of such an adverbial, but not its complement. However, some prepositions, such as *for*, *may* or *may not* introduce a temporal adverbial. For these cases, a domain model has been designed which the time component can use to determine whether the complement does contain temporal information. If it does, then temporal analysis continues. Otherwise, no temporal analysis is done.

The procedure which displays the temporal relations has been updated. A paper describing the temporal analysis component has been written and submitted for presentation at the annual meeting of the Association for Computational Linguistics.

1.3.2. Editor for Semantics Rules

The semantic rule editor described in the previous quarterly progress report has been fully implemented and tested. This editor allows a user to specify verb decompositions, syntax rules to map the abstract semantic arguments of decompositions to syntactic constituents, and semantics rules to check the consistency of syntactic fillers for semantic arguments with the domain model. The editor also enforces consistency among the rules which pertain to a particular verb. The editor greatly aids the task of porting to new domains as well as the task of continuing the development of the current domain. Testing led to further refinements in the design which permit the user to specify matching variables, even though Prolog assigns arbitrary names to variables. In order to prevent the editor from slowing the performance times, it does not reside permanently in the system, but is rather compiled in when needed.

1.3.3. Nominalisations

A paper describing the treatment of nominalisations has been written and submitted for presentation at the annual meeting of the Association for Computational Linguistics.

1.3.4. PUNDIT Applications

We have been in touch with the FRESH people at TI about the feasibility of using PUNDIT to automatically process CASREPS for input to FRESH. We are trying to get copies of sanitised FRESH messages from CINCPACFLT in Honolulu. We are also investigating an application involving analysis of some Trident messages which are processed under a Sperry contract.

1.4. Environment

1.4.1. Testing Procedure

Because of the size of PUNDIT and the number of people working on it, we have found it necessary to develop a procedure that will extensively test updates to system. The implementation is nearly complete. Each member of the group has designed input that will test a specific part of PUNDIT. The testing procedure is designed to handle switches, so that the test may be tailored to each input. The output of the testing procedure is a file which contains comments as to what, if anything, has changed since the last time the test was run.

1.4.2. Port to the TI Explorer

Work continues on fully porting PUNDIT to the TI Explorer. The interface to menus and user I/O has been modularised, and code developed on the Symbolics for menu interfacing has been modified to run on the Explorer. We have also begun to port the latest version of the NYU domain model, which was developed on the Symbolics, to the Explorer.

1.5. Facilities

We have installed Release 7.0 of the Symbolics operating system.

Four Sun workstations are now being used by members of the group. Development is now carried out on the Suns as well as on the Vax.

2. Change in Key Personnel

None.

3. Summary of Substantive Information from Meetings and Conferences

3.1. Professional Meetings Attended

3.1.1. Linguistic Society of America

Deborah Dahl and Marcia Linebarger attended the Winter Meeting of the Linguistic Society of America in New York in December.

3.1.2. Theoretical Issues in Natural Language Processing

Deborah Dahl and Bonnie Webber attended and participated in a panel discussion at a workshop on Theoretical Issues in Natural Language Processing (TINLAP-3) in Las Cruces, New Mexico, in January.

3.2. SDC/NYU Meetings

Lynette Hirschman, Martha Palmer, Deborah Dahl, Francois Lang, Leslie Riley and Rebecca Schiffman attended a meeting at NYU on January 30 with Ralph Grishman, Ngo Thanh Nhan, Tomasz Ksiesyk, Leo Joskowicz, Michael Moore and John Sterling. Palmer gave a presentation on the SDC treatment of nominalizations. Lang discussed the mechanism for the collection of selectional patterns. Sterling described the NYU treatment of the semantics of noun phrases. Nhan presented the NYU implementation of clause semantics. Grishman described the NYU plans for the period until the end of contract on May 1 and additional plans for the following few months.

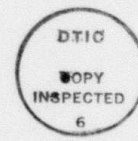
4. Problems Expected or Anticipated

The status of the follow-on contract needs to be confirmed.

5. Action Required by the Government

6. Fiscal Status

- (1) Amount currently provided on contract:
 - \$ 672,833 (funded) \$683,105 (contract value)
- (2) Expenditures and commitments to date:
 - \$ 521,766
- (3) Funds required to complete work:
 - \$ 151,067



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
AI	