⑫

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

R&D STATUS REPORT
SDC -- A BURROUGHS COMPANY

AD-A175 283

ARPA ORDER NUMBER: 5262
PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)
CONTRACTOR: System Development Corporation   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: 5/1/86 - 7/31/86

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

CONTRACT DEPARTMENT

AUG 07 1986

SDC · A BURROUGHS CO.
CUSTOM PRODUCTS GROUP.

DTIC
ELECTE
DEC 22 1986
S   D
E

86 12 04 070

1. Description of Progress

1.1. Grammar

Conjunction

The conjunction mechanism has been extended to allow semantic as well as syntactic processing of conjoined structures. The meta-conjunction component, which automatically generates conjunction rules for a grammar, has been enhanced so that it also generates rules describing the appropriate input to semantics, the intermediate syntactic representation or ISR, for each conjoined structure that is parsed. The modules which interpret and simplify the ISR have been extended so that the ISR for elements which are syntactically null under conjunction, but which have non-null semantics, can be computed. For example, in *normal sac lube oil pressure and temperature*, the left modifiers of *temperature* are syntactically null, but the intended interpretation of this noun phrase is "normal sac lube oil pressure and normal sac lube oil temperature." The ISR modules fill in this missing information so that the input to semantics is fully specified.

The control structure for the semantics component has been changed to allow the components to recurse on conjoined structures to interpret each conjunct. For conjoined noun phrases, a discourse entity is created for the referent of each conjunct as well as for the entire noun phrase. This is necessary because it is possible to follow a sentence like, *Measured sac lube oil pressure and temperature.* with a reference to the pressure and temperature together, as in *They were within normal range.* or with a reference to just one of the entities, as in *Pressure was 60 psig.*

1.2. Semantics

Interpreter Modifications

The semantics interpreter has been given a recursive structure so it can call itself to analyze the separate conjuncts of a conjoined structure, as in *Unit has excessive wear on inlet impellor assembly and shows high usage of oil.* or to analyze the clausal argument of a verb, as in *Believe the coupling from diesel to sac lube oil pump to be sheared.*

Coverage

The classes of words with lexical decompositions has been extended to include several classes of verbs which take propositional arguments: verbs which function as temporal operators (e.g., *occur*); verbs whose meaning includes temporal information (e.g., *result*); both verbs and adjectives where an animate experiencer is in some cognitive or perceptual state with respect to a propositional argument (e.g., *believe, evident*); and finally, verbs where an animate actor or experiencer has an active relation to a propositional argument (e.g., *conduct, experience*). Also, semantics has an existential predicate for interpreting noun phrase fragments like *failure of sac.*

1.3. Pragmatics

Time Module

The time module implemented last quarter has been fine-tuned and upgraded. Both the semantic model of tense and aspect and the procedure for consulting this model were revised. The semantic model takes as input some combination of tense and aspect values to produce its output. The model was streamlined so at each point of the procedure where the model is consulted, only the critical information is modelled. Refinement of the temporal analyzer was accomplished by modular separation of three temporal relations: 1) the relation of the state-of-affairs to the time of the report; 2) relations established by the use of the perfect; and finally, 3) relations established by temporal connectives and their complements. Computation of the temporal relations of a state-of-affairs requires close cooperation among these three relational modules. For example, a sentence with both a perfect and a temporal adverbial requires that a reference point established by the perfect be equated with a reference point established by the adverb. In the following sentence, the failure event is before some moment M, where M is equated with the moment when the pump seized.

Sac had failed when pump seized.

Prolog unification makes it possible to find the relevant time arguments that need to be passed from one relational module to the next.

Several updates to the time module have been completed which extend its coverage. By examining both the surface verb and the output of the semantic analysis of a clause, the time component can now distinguish between verbs which directly denote a particular state-of-affairs (e.g., *fail*) -- first-order verbs -- and those which supply temporal information about a state-of-affairs mentioned as an argument to the verb (e.g., *occur*) -- second order verbs. This has made it possible to provide a general treatment of second-order temporal operator verbs.

A second update has been to add temporal inferences to some verbs using the same type of input which helps distinguish between first-order and second-order verbs. For example, the semantic decomposition of the verb *result* specifies a causal relationship between two states-of-affairs. The time module adds the temporal inference that the causing event must precede the resultant state or event.

Finally, the procedure for finding the tense of the main clause has been changed to make it possible to handle tenseless embedded propositions in a uniform way, whether they are syntactically in the form of nominalizations or non-finite clauses.

1.4. Environment

Some switches (run/environment options) have been added to the PUNDIT system, the most notable of which is a switch that will allow us to send off an entire text to be processed at once. Also, the lexical reader has been modified to allow us to recognize abbreviations. We previously had counted sentences with abbreviations as sentences we could not parse.

1.5. Facilities

We have successfully ported the NYU domain model to our Symbolics machines. Further, we have found out how to extract part/whole information from the model that can be used as selectional information during semantic analysis. This is preparatory to integrating Pundit's semantics module with NYU's domain model. Another piece of selectional information that looks like a a valuable point of interaction between the semantics and the domain model is information about actual connections between different components of a SAC.

Also a facility has been added to the Symbolics for displaying the parse trees generated by the grammar. It has a flexible design structure permitting any tree or portion of a tree to be displayed. When integrated into PUNDIT, the tree-display facility will enhance the grammar development environment by making it possible to compare trees and subtrees.

Efforts to port the newest version of PUNDIT, which includes the ISR, (intermediate syntactic representation), continue. We have had problems with operator declarations, the "call" procedure, and most recently, the distinction (or lack thereof) between prolog predicates and prolog lists.

2. Change in Key Personnel

In May, Francois Lang joined the Natural Language Processing group, to work primarily on syntax and morphology. Mr. Lang comes to SDC from the doctoral program in Computer Science at the University of Pennsylvania where he has been specializing natural language processing and artificial intelligence.

3. Summary of Substantive Information from Meetings and Conferences

3.1. Professional Meetings Attended

3.1.1. ISI

Lynette Hirschman and Martha Palmer attended a two-day meeting of DARPA Strategic Computing Natural Language contractors, held at ISI. We presented an hour overview of SDC's PUNDIT Text Processing System, which was very well received.

3.1.2. ACL

All members of the Natural Language Processing group attended the 1986 ACL Meeting at Columbia University. During the meeting, Martha Palmer presented SDC's paper, "Recovering Implicit Information", which was well received. We also held two demonstrations of the PUNDIT system, running on a Xerox 1109. We had approximately 30 people at two demonstrations, with many interested questions and requests for literature describing the

system.

3.1.3. SNUG: Symbolics National User's Group

John Dowding and Leslie Riley attended the Second Annual Symbolics LISP User's Group National Symposium from June 2 - June 6, 1986. The focus of the symposium was the new release 7 system.

3.1.4. Lexicon Workshop

Martha Palmer attended several meetings of the Lexicon Workshop held at Columbia University concurrently with the summer LSA. The workshop included computer scientists, linguists, and psychologists from Bellcore, Bell Labs, MIT, Princeton and IN IBM. The first session was devoted to reviewing the previous workshop held in Grossetto, Italy. Larry Urdang attended one meeting in which he summarized 18 features to be included in a lexical entry that had been generally accepted by the Grossetto group. Martha Palmer presented several sample lexical entries in the style used in the PUNDIT system which favorably impressed the other workshop participants.

3.1.5. Minnowbrook DB Workshop

Lynette Hirschman presented an overview of SDC's work in Artificial Intelligence and Database, as well as an overview on SDC's work on Natural Language interfaces to databases at the Minnowbrook Workshop on Databases, July 22-25. This included mention of the DARPA contract and PUNDIT's use to *create* databases and knowledge bases from text input in the form of messages.

3.2. SDC/NYU Meeting #9 (June 26, SDC, Paoli, PA)

Ralph Grishman, Tomasz Ksiezyk, Ngo Thanh Nhan, Ping Peng, and Michael Moore came to Paoli to meet with Lynette Hirschman, Martha Palmer, Chris Andrews, Deborah Dahl, John Dowding, Leslie Riley, and Rebecca Schiffman. Ping presented preliminary results on designing a method to recover from a failed parse by looking for possible semantic relationships between the largest potential constituents. General discussion focussed on strategies for syntax/semantics interaction. Tomasz and Chris discussed porting the NYU domain model to SDC's Symbolics machine. Martha and Ralph discussed the PUNDIT verb representations and how NYU might make use of them.

4. Problems Encountered and/or Anticipated

We continue to have many difficulties with the Symbolics Prolog environment.

5. Action Required by the Government

We are awaiting some more firm guidelines for anticipated levels of funding for the two-year follow-on contract. SDC and NYU have submitted some preliminary estimates of work that could be accomplished under various funding scenarios.

6. Fiscal Status

(1) Amount currently provided on contract:
$ 672,833 (funded) $683,105 (contract value)

(2) Expenditures and commitments to date:
$ 389,573

(3) Funds required to complete work:
$ 283,260