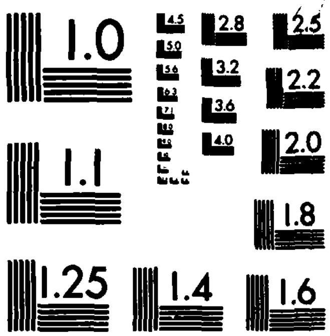
AD-A131 722 PREREQUISITES FOR DERIVING FORMAL SPECIFICATIONS FROM NATURAL LANGUAGE RE..(U) DELAWARE UNIV NEWARK DEPT OF COMPUTER AND INFORMATION SCIENCE.. R M WEISCHEDEL 1/1

UNCLASSIFIED APR 83 AFOSR-TR-83-0687 F49620-79-C-0131 F/G 9/2 NL

END

FILMED

DTIC


1.0
4.5
5.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

(12)

Prerequisites for Deriving Formal Specificaitons from
Natural Language Requirements:
Final Report*

by

Ralph M. Weischedel
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19711

April, 1983

* Research sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, Under grant no. AFOSR-80-0190. The United States Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation herein.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0687	2. GOVT ACCESSION NO. AD-A131 722	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Prerequisites to Deriving Formal Specifications from Natural Language Requirements: [redacted]		5. TYPE OF REPORT & PERIOD COVERED FINAL 9/1/81-2/28/83
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Ralph M. Weischedel		8. CONTRACT OR GRANT NUMBER(s) Grant No. AFOSR-80-0190
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Computer & Information Sciences University of Delaware Newark, DE 19711		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A2
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB Washington, DC 20332		12. REPORT DATE April, 1983
		13. NUMBER OF PAGES 12
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

COPY INSPECTED 7

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

formal specifications, English specifications, modules, software design, natural language processing, software engineering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Since English specifications and formal specifications of modules are complementary and since formal specifications require so much effort to write, our workhas investigated application of artificial intelligence techniques to aid in the software specification process.

The effort for this year concentrated on constructing a small prototype of a system that transforms English descriptions to formal specifications under significant user assistance.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

0. Organization of the Report

This final report of grant AFOSR-80-0190 covers the second year of the grant and is divided into five sections. Section one gives a brief summary of the motivation of the effort. Section two reports our progress this year. Section three summarizes the report. The references are collected in section four; a list of publications and presentations during the first year's effort appears in section five; one for this year's effort appears in section six. Related work is cited throughout section two; however, a complete discussion of related work appears in Weischedel (1980) and in Weischedel and Chester (1983).

1. Motivation of the Work

The technique of formal module specifications seems to offer much toward alleviating many problems of large software systems (those systems requiring at least 25 programmers for development and at least 30,000 lines of source code). The high cost of software maintenance, the predominance of design errors, the difficulty in modifying software, and the difficulty and cost of diagnosing and correcting design errors are some of the problems addressed by formal specifications based on the information-hiding principle. Yet, the creating of formal specifications is very difficult, requiring much upfront effort. For instance, Parnas (1976, p. 7) states, "Experience has shown that the effort involved in writing the set of specifications can be greater than the effort it would take to write one complete program." It is also generally agreed that formal specifications are difficult to understand.

Our work, which has been funded by AFOSR under contract number F49620-79-C-0131 and grant number AFOSR-80-0190, has had both long-term and short-term goals. The long-term objectives of the final year of this effort have been a preliminary, fundamental study of the problems involved in understanding software system requirements written in English and transforming them into formal specifications of a software module. The research has culminated in a prototype system. A short-term objective of this final year of the effort has been to seek a more understandable alternative to first-order quantification in formal specification languages.

The motivation of the work is covered at greater length in the proposal for the grant and the contract mentioned above.

2. Results

The proposal for this present year's effort included three areas; one in short-term-goals and two in long-term goals.

2.1 Short-term Goals

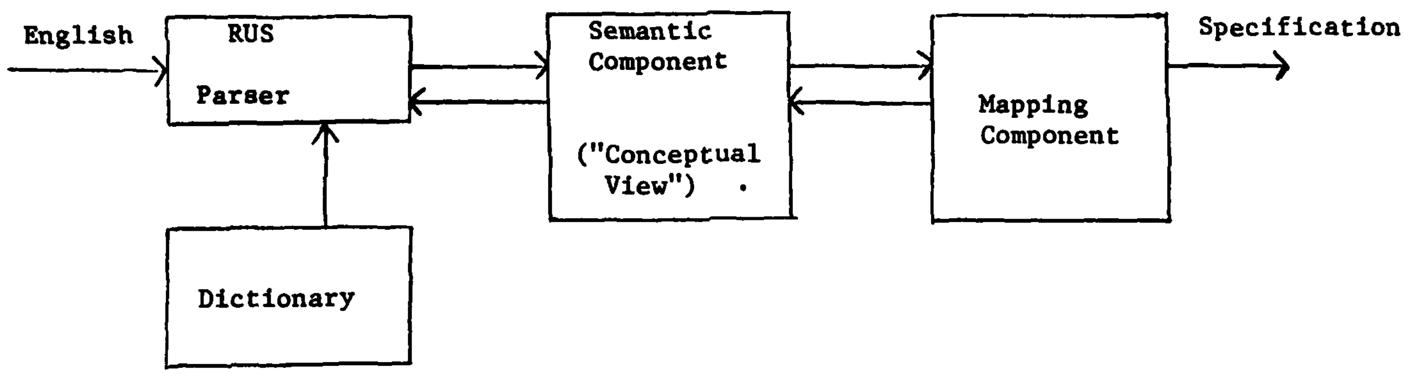
The formal specification of a module must be understandable if it is to achieve its purpose, for it acts as a contract between designers and programming team, stating exactly what the programming team's product must do (Parnas, 1977). Unless they are understandable, 1) programmers will not know what the module they are to implement is to do nor how to use other modules, and 2) designers will not be able to detect design errors nor easily confirm that their design satisfies user requirements. Also, if one is to use a reference library of formal specifications, they must be understandable, for if the designer cannot understand the alternative specifications, how can an intelligent choice be made among the alternatives? If formal specifications of module interfaces are to become widely used, they must be understandable.

The proposal for this year's work identified quantification in formal logic both as a source of difficulty in understanding formal specifications and as a mechanism for which an alternative could be found. Though we originally anticipated that a new specification language would be needed, we found an interesting alternative, namely, Horn clauses, which had never before been applied for software specification, though they had been used in logic programming. The definition of Horn clauses and a discussion of their applicability is presented in Weischedel and Chester (1983); a review of the use of Horn clauses is given in Chester (1982). We do not believe them to be the ultimate specification language; nevertheless they have many advantages, including a highly understandable treatment of quantification.

2.2 Long-term Goals

The two sections that follow deal with the long-term goals. In particular, Weischedel (1980) identifies the form a tool might have, which, under significant user direction, would derive formal specifications from English requirements. A second long-term goal is anaphora usage, a well-known unsolved problem in natural language processing. We have developed a small prototype of such a system.

The structure we have chosen for the prototype tool is given in Figure 1. All arrows represent data flow.



SYSTEM STRUCTURE

Figure 1

Given an English input, the RUS parser (Bobrow, 1978) recognizes the syntactic structure of it. Using a lexicon (a highly formalized dictionary), the parser generates a semantic representation for the input. Since the syntax of English is highly ambiguous, the parser interacts with the knowledge base in the "conceptual view" as a means of using factual knowledge to reduce the number of semantically senseless parses tried by the parser. The parser and dictionary constitute the syntactic component of the system.

The semantic component applies simple semantic constraints to filter out anomalous interpretations during parsing. For example, in this environment, a process or a person may delete a data entity from a data strucutre; any parse of a clause involving deletion must conform to that sense. The operation of the parser and semantic component, as well as our experience with them are reported in Weischedel and Chester (1983).

Much of our effort in semantics has been designing the mapping component and the input and output representations for it. The input to the mapping component is a set of Horn clauses whose predicates, functions, and constants are those of the user. The motivation for having two levels of representation is detailed in Weischedel and Chester (1983). The design of the mapping component is one of the most important contributions of our research, since few have considered these issues before. Our work has provided a basis for solutions to problems in interpreting spatial metaphor, determining the meaning of known words in new environments, and recognizing semantic paraphrases. The results are reported in Weischedel and Chester (1983).

Though we had originally planned a small effort in generating clarifying questions, we decided the effort would be better spent on the more unique aspects of our research. Several larger research efforts (Mann, et al. 1981) are underway in natural language generation, including one (Swartout, 1982) in generating English paraphrases of formal specification.

2.2 Anaphora

In addition to the semantic analysis discussed above, a separate problem in semantics was studied. Determining what a pronoun or noun phrase refers to has been a topic of much interest recently in artificial intelligence; it is called definite anaphora resultion. Sidner (1979) presents a heuristic for detecting clues in dialogue for the focus of attention, and explains how that can be used for determining reference. The consultant on this grant has been modifying this heuristic for texts of English specifications rather than dialogue; see Joshi and Weinstein (1982) and Grosz, et al. (1983). In doing this, he has found that Sidner's notion of focus is inadequate. Rather, two objects seem to serve as centers of attention. A backward center serves as a focus in a single sentence and corresponds to previous objects in the text. A forward center serves as a focus for later references and introduces new objects. In the sentence "Processes must execute in a single fork", "processes" serves as the backward center, and "a single fork" is the forward center.

3. Conclusions

The effort for this year involved three areas of work. In the first area we investigated Horn clauses as a more understandable alternative to logical quantifiers. The second area was developing an experimental software tool to test our solutions to problems in transforming English descriptions to formal specifications under significant user assistance. A third problem is anaphora resolution; the heuristics of Sidner (1979) have been extended.

This application of natural language processing has the following open problems:

1. formal specification languages with richer semantics so that the level of the target language is closer to that of natural languages,

2. development of more flexible, forgiving natural language interfaces (Weischedel 1983) that have partial understanding even of poorly formed input,

3. extension of the technology to broad areas of specification,

4. continued development of high quality English generation components both for creating a paraphrase of the formal specification and for generating questions to clarify ambiguous or vague aspects of the English definitions, and

5. further development of the mapping phase.

4. References

Bobrow, R., "The RUS System", in Research in Natural Language Understanding, by B. Webber and R. Bobrow, BBN Report No. 3878, 1978.

Chester, Daniel L. "Elements of Knowledge-Based Expert Systems". Proceedings of Micro-Delcon '82, IEEE Computer Society, 1982.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstein. "Providing a Unified Account of Definite Noun Phrases in Discourse". To appear in Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, 1983.

Joshi, Aravind K. and Scott Weinstein. "Control of Inference: Role of Some Aspects of Discourse Structure - Centering". Proceedings of the Seventh International Joint Conference on Artificial Intelligence, American Association for Artificial Intelligence, Mento Park, CA, 1981, 385-387.

Mann, William C., Madeleine Bates, Barbara J. Grosz, David D. McDonald, Kathleen R. McKeown, and William R. Swartout. "Text Generation: The State of the Art and the Literature", Technical Report ISI/RR-81-9, USC/Information Sciences Institute, 1981.

Parnas, D. L., "On the Design and Development of Program Families," IEEE Transactions on Software Engineering, SE-2, No. 1, March, 1976, 1-8.

Parnas, David L., "The Use of Precise Specifications in the Development of Software," Information Processing 77, B. Gilchrist, (ed.), North-Holland Publishing Company, New York, 1977.

Sidner, Candace Lee, "Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse," AI-TR 537, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA: Harvard University Press, 1972.

Swartout, William B., "GIST English Generator", Proceedings of the National Conference for Artificial Intelligence, August, 1982, 404-409.

Weischedel, Ralph M. "Prerequisites to Deriving Formal Specifications from Natural Language Requirements: Final Report", Dept. of Computer & Information Sciences, University of Delaware, Newark, DE, 1980.

Weischedel, Ralph M. "Introduction to Handling Ill-Formed Input", Proceedings of the Conference on Applied Computational Linguistics, February, 1983, 3-16.

Weischedel, Ralph M. "Design of a System that Understands Informal Specifications", Technical Report, Dept of Computer &

Information Sciences, University of Delaware, 1983.

5. List of Presentations and Reports Generated the First Year

Joshi, Aravind K. and Scott Weinstein. "Control of Inference: Role of Some Aspects of Discourse Structure - Centering". Proceedings of the Seventh International Joint Conference on Artificial Intelligence, American Association for Artificial Intelligence, Mento Park, CA, 1981, 385-387.

Weischedel, Ralph M. "The Problem with Software Specifications & Two Emerging Solutions", Presented to the Computer Science Research Group, Ford Aerospace & Communications Corporation, Palo Alto, CA, August, 1980.

Weischedel, Ralph M. "Two Possible Uses", First KL-one Workshop, Jackson, NH, October 15-17, 1981. (Oral presentation only; no proceedings was published.)

Weischedel, Ralph M. "Reducing the Effort in Creating Formal Specifcations of Software Modules," Proceedings of the Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, March 26-28, 1981, 86-90.

Weischedel, Ralph M. "Practical Issues in having a usable Library of Software Specifications", Department of Computer & Information Sciences, University of Delaware, Newark, DE 1981b.

Additionally, Ralph M. Weischedel was a participant in the Sublanguages Panel of the Applied Computational Linguistics in Perspective Workshop, June 26-27, 1981. The workshop was sponsored by the Office of Naval Research and the National Science Foundation. His experience in the style of English in software specifications was presented there. An overview of the workshop as a whole appears as a paper by Carroll Johnson and Joan Bachenko, "Applied Computational Linguistics in Perspective: Proceedings of the Workshop", American Journal of Computational Linguistics Volume 8, Number 2, 1982.

6. Presentations and Reports Generated this Year

Weischedel, Ralph M., "Prerequisites to Deriving Formal Specifications from Natural Language Requirements", Presented at General Electric Research Labs, January, 1982.

Chester, Daniel L., "Elements of Knowledge-Based Expert Systems", Proceedings of Micro-Delcon '82, IEEE Computer Society, March, 1982.

Weischedel, Ralph M., "Design of a System for Deriving Formal Specifications from Natural Language Requirements", Presented at the Ohio State University, October, 1982.

Weischedel, Ralph M., "Design of a System for Deriving Formal Specifications from Natural Langauge Requirements", Presented at the Naval Research Laboratory Artificial Intelligence Center,

November, 1982.

Weischedel, Ralph M. and Daniel L. Chester. "Design of a System that Understands Informal Specifications", Technical Report, Dept of Computer & Information Sciences, University of Delaware, 1983.

Weischedel, Ralph M. "Mapping Between Semantic Representations using Horn Clauses", Submitted to the Third National Conference on Artificial Intelligence, 1983.

Grosz, Barbara J., Aravind K. Joshi, and Scott Weinstiein, "Providing a Unified Account of Definite Noun Phrases in Discourse", To apperar in Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics, June, 1983.