

## ABSTRACTS OF CURRENT LITERATURE

For copies of the following papers on Project SEMSYN, write to

Frau Hörmann  
c/o Project SEMSYN  
Institut für Informatik  
Universität Stuttgart  
Azenbergstrasse 12  
D-7000 Stuttgart 1, West Germany

### **When Mariko Talks to Siegfried – Experiences from a Japanese/German Machine Translation Project**

*Dietmar Rösner*

In this paper we report on experiences from a 2 1/2 year project that designed and implemented a prototypical Japanese to German machine translation system for titles of Japanese scientific papers. The analysis of the Japanese input and its transformation into a semantic representation is done by FUJITSU's system ATLAS/II. SEMSYN's part is to produce a correct and understandable German text from these interface structures.

### **Linguistic Tools and Software Tools of the SEMSYN Project**

*Dietmar Rösner*

This paper gives an overview of the tools that have been developed during the implementation of the SEMSYN generator for German on a Symbolics Lisp machine. These tools include:

- **interface tools** that provide easy and comfortable communication with the system;
- **experimentation tools**: e.g., SEMNET-EDIT, a tool for interactively editing or creating semantic nets and generating German from them;
- **lexicon tools**: e.g., menu-based interfaces for lexicon maintenance;
- **linguistic tools**: e.g., a formalism for specifying the intended utterances as functional structures and a language for manipulating those structures.

### **From Titles to Text – The Development of the Text Generator SEMTEX (in German)**

*Dietmar Rösner*

SEMTEX is an implemented system that generates "realistic" German newspaper stories like the following:

#### **Geringfügige Reduzierung der Arbeitslosenzahl.** NUERNBERG/BONN (cpa)

Die Zahl der Arbeitslosen in der Bundesrepublik Deutschland hat sich während des Oktober nur sehr wenig verringert. Sie ist von 2151600 auf 2148800 zurückgegangen. Die Arbeitslosenquote hatte Ende Oktober einen Wert von 8.6 Prozent. Sie hatte am Ende des Vergleichszeitraumes des Vorjahrs ebenfalls bei 8.6 Prozent gelegen. Regierungssprecher Ost bewertet die Verringerung der Arbeitslosenzahl positiv. Der stellvertretende DGB-Vorsitzende Muhr erklärt daß der Rückgang der Zahl der Arbeitslosen nicht darüber hinwegtäuschen dürfe, daß sie jetzt unverändert unerträglich hoch sei.

Starting point for this application is just the data from the monthly job market report (numbers of unemployed, open jobs, ...). A rudimentary "text planner" takes these data and those of relevant previous months, checks for changes and significant developments, simulates possible comments of various political speakers with regard to these developments, and finally creates an ordered list of frames as representation for the content of the intended text. SEMTEX then converts this list into a newspaper story in German, using an extended version of the generator of the SEMSYN project.

Following are the latest reports issued by the HAM-ANS project. For single copies, please write to

Universität Hamburg  
Fachbereich Informatik  
Attn.: Ms. Catharina Helbig

Postfach 30 27 62  
D-2000 Hamburg 36, West Germany

**User Modeling, Dialog Structure, and Dialog Strategy in HAM-ANS**

**Katharina Mork**

*Report ANS-31, March 1985 (19 pages)*

To appear in *Proceedings of the 2nd EACL Conference*, Geneva, 1985

AI dialog systems are now evolving from question-answering systems toward advising systems. This includes:

- structuring dialog,
- understanding and generating a wider range of speech acts than simply information request and answer,
- user modeling.

User modeling in HAM-ANS is closely connected to dialog structure and dialog strategy. In advising the user, the system generates and verbalizes speech acts. The choice of the speech act is guided by the system's user profile and dialog strategy.

**Representing and Processing Copula and Full-Verb Sentences in HAM-ANS**

**Stephan Busemann, Wolfgang Hoepfner, Heinz Marburger, Katharina Morik**

*Report ANS-32, September 1985.*

To appear in Stoyan, H., Ed., *GWAI-85*, 9th German Workshop on Artificial Intelligence

We first introduce the domains HAM-ANS operates in, illustrating which types of verbs (and actions) may occur. We then argue that two different kinds of representation are necessary in order to process these verbs in an adequate manner. How this is done by the major components without incurring the expense of duplication is described in the following sections, revealing relations between linguistic and domain specific verb properties and dependencies between the model of the respective domain and the depth of the verb's semantic representation.

**Cooperativeness in Natural Language Access Systems**

**Heinz Marburger**

In Brauer, W., Radig, B., Eds., *Wissensbasierte Systeme, GI-Kongress 1985. Informatik Fachberichte 112*. Berlin: Springer 1985: 135-144 (in German)

Natural language access systems must be able to react cooperatively on various levels in order to gain acceptance from users. For this reason this requirement has attracted increasing amounts of attention from researchers in recent years. The representation offers an overview of cooperative behavior in a number of existing systems and others presently under development, concentrating on the necessary additional knowledge sources and processes.

The dollar figure given for each of the following technical reports covers the cost of copying and postage. Please make check payable to Boston University; mail to

Computer Science Department  
Boston University  
771 Commonwealth Avenue  
Boston, MA 02215

**The Weak Generative Capacity of Parenthesis-Free Categorical Grammars**

**Joyce Friedman, Dawai Dai,**

**Weiguo Wang**

*BUCS Tech Report 86-001, January 1986 (20 pages) \$2.50*

We study the weak generative capacity of a class of parenthesis-free categorical grammars derived from those of Ades and Steedman by varying the set of reduction rules. With forward cancellation as the only rule, the grammars are weakly equivalent to context-free grammars. When a backward combination rule is added, it is no longer possible to obtain all the context-free languages. With suitable restriction of the forward partial rule, the languages are still context-free and a push-down automaton can be used for recognition. Using the unrestricted rule of forward partial combination, a context-sensitive language is obtained.

**Phonological Analysis for French Dictation: Preliminaries to an Intelligent Tutoring System**

**Joyce Friedman, Carol Neidle**

*BUCS Tech Report 86-004, April 1986 (17 pages) \$2.50*

A set of programs for the phonological analysis of French dictation exercises has been written as a preliminary step in the development of an Intelligent Tutoring System for French. In this paper, we describe and illustrate the programs to date and give an overview of the total system as envisaged.

**Categorical and Non-Categorical Languages**

**Joyce Friedman, Ramarathnam**

**Venkatesan**

We study the formal and linguistic properties of a class of parenthesis-free categorical grammars derived from those of Ades and Steedman by varying the set of reduction rules. We characterize the reduction rules capable of

*BUCS Tech Report 86-005, April 1986*  
(5 pages) \$1.50

generating context-sensitive languages as those having a partial combination rule and a combination rule in the reverse direction. We show that any categorial language is a permutation of some context-free language, thus inheriting properties dependent on symbol counting only. We compare some of their properties with other contemporary formalisms.

The following abstracts are from the Proceedings of the Conference, 24th Annual Meeting of the Association for Computational Linguistics, 10-13 June 1986 (Columbia University, New York). Proceedings are \$20 to ACL members, \$25 to non-members; for first-class mailing in the U.S., Canada and Mexico, add \$8; for air-printed matter delivery elsewhere, add \$16). To obtain a copy, use the order form at the back of this issue or send a request, with check payable to ACL, to

Dr. Donald E. Walker (ACL)  
Bell Communications Research  
445 South Street, MRE 2A379  
Morristown, NJ 07970 USA

**Bringing Natural language Processing to the Microcomputer Market:**

**The Story of Q&A**

*Gary G. Hendrix*

Symantec Corporation  
10201 Torre Avenue  
Cupertino, CA 95014

*p. 2*

This is the story of how one of the new natural language processing products reached the marketplace. On the surface, it is the story of one NL researcher-turned-entrepreneur (yours truly) and of one product, Q&A. But this is not just my story: It is in microcosm the story of NL emerging from the confines of the academic world, which in turn is an instance of the old theme "science goes commercial."

**Time and Tense in English**

*Mary P. Harper, Eugene Charniak*

Department of Computer Science  
Brown University  
Box 1910

Providence, RI 02912

*pp. 3-9*

Tense, temporal adverbs, and temporal connectives provide information about when events described in English sentences occur. To extract this temporal information from a sentence, it must be parsed into a semantic representation which captures the meaning of tense, temporal adverbs, and temporal connectives. Representations were developed for the basic tenses, some temporal adverbs, as well as some of the temporal connectives. Five criteria were suggested for judging these representations, and based on these criteria the representations were judged.

**Recovering Implicit Information**

*Martha S. Palmer, Deborah A. Dahl,  
Rebecca J. Schiffman, Lynette Hirschman,  
Marcia Linebarger, John Dowding*

Research and Development Division  
SDC - A Burroughs Company  
P. O. Box 517

Paoli, PA 19301

*pp. 10-19*

This paper describes the SDC PUNDIT (Prolog UNDERstands Integrated Text) system for processing natural language messages. PUNDIT, written in PROLOG, is a highly modular system consisting of distinct syntactic, semantic, and pragmatic components. Each component draws on one or more sets of data, including a lexicon, a broad-coverage grammar of English semantic verb decompositions, rules mapping between syntactic and semantic constituents, and a domain model.

This paper discusses the communication between the syntactic, semantic, and pragmatic modules that is necessary for making implicit linguistic information explicit. The key is letting syntax and semantics recognize missing linguistic entities as implicit entities, so that they can be labelled as such, and reference resolution can be directed to find specific referents for the entities. In this way the task of making implicit linguistic information explicit becomes a subset of the tasks performed by reference resolution. The success of this approach is dependent on marking missing syntactic constituents as elided and missing semantic rules as ESSENTIAL, so that reference resolution can know when to look for referents.

**Semantic Acquisition in TELI: A Transportable, User-Customized Natural Language Processor**

*Bruce W. Ballard, Douglas E. Stumberger*

AT&T Bell Laboratories

We discuss ways of allowing the users of a natural language processor to define, examine, and modify the definitions of any domain-specific words or phrases known to the system. An implementation of this work forms a critical portion of the knowledge acquisition component of our Transportable English-Language Interface (TELI), which answers English questions

600 Mountain Avenue  
Murray Hill, NJ 07974  
pp. 20-29

### Computational Complexity of Current GPSG Theory

*Eric Sven Ristad*

MIT Artificial Intelligence Lab  
545 Technology Square  
Cambridge, MA 02139 and  
Thinking Machines Corporation  
245 First Street  
Cambridge, MA 02142  
pp. 30-39

### Defining Natural Language Grammars in GPSG

*Eric Sven Ristad*

MIT Artificial Intelligence Lab  
545 Technology Square  
Cambridge, MA 02139 and  
Thinking Machines Corporation  
245 First Street  
Cambridge, MA 02142  
pp. 40-44

### Constraint Propagation in Kimmo Systems

*G. Edward Barton, Jr.*

MIT Artificial Intelligence Lab  
545 Technology Square  
Cambridge, MA 02139  
pp. 45-52

about tabular (first normal-form) data files and runs on a Symbolics Lisp Machine. However, our techniques enable the design of customization modules that are largely independent of the syntactic and retrieval components of the specific system they supply information to. In addition to its obvious practical value, this area of research is important because it requires careful attention to the *formalisms* used by a natural language system and to the *interactions* among the modules based on those formalisms.

An important goal of computational linguistics has been to use linguistic theory to guide the construction of computationally efficient real-world natural language processing systems. At first glance, generalized phrase structure grammar (GPSG) appears to be a blessing on two counts. First, the precise formalisms of GPSG might be a direct and transparent guide for parser design and implementation. Second, since GPSG has weak context-free generative power and context-free languages can be parsed in  $O(n^3)$  by a wide range of algorithms, GPSG parsers would appear to run in polynomial time. This widely-assumed GPSG “efficient parsability” result is misleading: here we prove that the universal recognition problem for current GPSG theory is exponential-polynomial time hard, and assuredly intractable. The paper pinpoints sources of complexity (e.g., metarules and the theory of syntactic features) in the current GPSG theory and concludes with some linguistically and computationally motivated restrictions on GPSG.

Three central goals of work in the generalized phrase structure grammar (GPSG) linguistic framework, as stated in the leading book *Generalized Phrase Structure Grammar* (Gazdar et al. 1985), are (1) to characterize all and only the natural language grammars, (2) to algorithmically determine membership and generative power consequences of GPSGs, and (3) to embody the universalism of natural language entirely in the formal system, rather than by statements made in it.

These pages formally consider whether GPSG’s weak context-free generative power (wcfgp) will allow it to achieve the three goals. The centerpiece of this paper is a proof that it is undecidable whether an arbitrary GPSG generates the nonnatural language  $\Sigma^*$ . On the basis of this result, I argue that GPSG fails to define the natural language grammars, and that the generative power consequences of the GPSG framework cannot be algorithmically determined, contrary to goals one and two. In the process, I examine the linguistic universalism of the GPSG formal system and argue that GPSGs can describe an infinite class of nonnatural context-free languages. The paper concludes with a brief diagnosis of the result and suggests that the problem might be met by abandoning the weak context-free generative power framework and assuming substantive constraints.

Taken abstractly, the two-level (Kimmo) morphological framework allows computationally difficult problems to arise. For example,  $N+1$  small automata are sufficient to encode the Boolean satisfiability problem (SAT) for formulas in  $N$  variables. However, the suspicion arises that natural-language problems may have a special structure – not shared with SAT – that is not directly captured in the two-level model. In particular, the natural problems may generally have a modular and local nature that distinguishes them from more “global” SAT problems. By exploiting this structure, it may be possible to solve the natural problems by methods that do not involve combinatorial search.

We have explored this possibility in a preliminary way by applying *constraint propagation* methods to Kimmo generation and recognition.

**Computational Complexity in Two-Level Morphology**

*G. Edward Barton, Jr.*

MIT Artificial Intelligence Lab  
545 Technology Square  
Cambridge, MA 02139  
*pp. 53-59*

Constraint propagation can succeed when the solution falls into place step-by-step through a chain of limited and local inferences, but it is insufficiently powerful to solve unnaturally hard SAT problems. Limited tests indicate that the constraint-propagation algorithm for Kimmo generation works for English, Turkish, and Warlpiri. When applied to a Kimmo system that encodes SAT problems, the algorithm succeeds on "easy" SAT problems but fails (as desired) on "hard" problems.

**Parsing a Free-Word Order Language: Warlpiri**

*Michael B. Kashket*

MIT Artificial Intelligence Laboratory  
545 Technology Square, room 823  
Cambridge, MA 02139  
*pp. 60-66*

Morphological analysis must take into account the spelling-change processes of a language as well as its possible configurations of stems, affixes, and inflectional markings. The computational difficulty of the task can be clarified by investigating specific models of morphological processing. The use of finite-state machinery in the "two-level" model by Kimmo Koskeniemi gives it the appearance of computational efficiency, but closer examination shows the model does not guarantee efficient processing. Reductions of the satisfiability problem show that finding the proper lexical/surface correspondence in a two-level generation or recognition problem can be computationally difficult. The difficulty increases if unrestricted deletions (null characters) are allowed.

Free-word order languages have long posed significant problems for standard parsing algorithms. This paper reports on an implemented parser, based on Government-Binding theory (GB; Chomsky 1981, 1982), for a particular free-word order language, Warlpiri, an aboriginal language of central Australia. The parser is explicitly designed to transparently mirror the principles of GB.

The operation of this parsing system is quite different in character from that of a rule-based parsing system, e.g., a context-free parsing method. In this system, phrases are constructed via principles of selection, case-marking, case-assignment, and argument-linking, rather than by phrasal rules.

The output of the parser for a sample Warlpiri sentence of four words in length is given. The parser was executed on each of the 23 other permutations of the sentence, and it output equivalent parses, thereby demonstrating its ability to correctly handle the highly scrambled sentences found in Warlpiri.

**The Relationship Between Tree Adjoining Grammars and Head Grammars**

*D. J. Weir, K. Vijay-Shanker, A. K. Joshi*

Department of Computer and Information Science  
University of Pennsylvania  
Philadelphia, PA 19104  
*pp. 66-74*

We examine the relationship between the two grammatical formalisms: Tree Adjoining Grammars and Head Grammars. We briefly investigate the weak equivalence of the two formalisms. We then turn to a discussion comparing the linguistic expressiveness of the two formalisms.

**Categorial and Non-categorial Languages**

*Joyce Friedman, Tamarathnam Venkatesan*

Computer Science Department  
Boston University  
111 Cummington Street  
Boston, MA 02215  
*pp. 75-77*

We study the formal and linguistic properties of a class of parenthesis-free categorial grammars derived from those of Ades and Steedman by varying the set of reduction rules. We characterize the reduction rules capable of generating context-sensitive languages as those having a partial combination rule and a combination rule in the reverse direction. We show that any categorial language is a permutation of some context-free language, thus inheriting properties dependent on symbol counting only. We compare some of their properties with other contemporary formalisms.

**Parsing Conjunctions Deterministically**

*Donald W. Kosy*

The Robotics Institute  
Carnegie-Mellon University

Conjunctions have always been a source of problems for natural language parsers. This paper shows how these problems may be circumvented using a rule-based, wait-and-see parsing strategy. A parser is presented which analyzes conjunction structures deterministically, and the specific rules it

Pittsburgh, PA 15213

pp. 78-84

### **Copying in Natural Languages, Context-Freeness, and Queue Grammars**

*Alexis Manaster-Ramer*

University of Michigan

2236 Fuller Road #108

Ann Arbor, MI 48105

pp. 85-89

uses are described and illustrated. This parser appears to be faster for conjunctions than other parsers in the literature and some comparative timings are given.

The documentation of (unbounded-length) copying and cross-serial constructions in a few languages in the recent literature is usually taken to mean that natural languages are slightly context-sensitive. However, this ignores those copying constructions which, while productive, cannot be easily shown to apply to infinite sublanguages. To allow such finite copying constructions to be taken into account in formal modeling, it is necessary to recognize that natural languages cannot be realistically represented by formal languages of the usual sort. Rather, they must be modeled as families of formal languages or as formal languages with indefinite vocabularies. Once this is done, we see copying as a truly pervasive and fundamental process in human language. Furthermore, the absence of mirror-image constructions in human languages means that it is not enough to extend Context-free Grammars in the direction of context-sensitivity. Instead, a class of grammars must be found which handles (context-sensitive) copying but not (context-free) mirror images. This suggests that human linguistic processes use queues rather than stacks, making imperative the development of a hierarchy of Queue Grammars as a counterweight to the Chomsky Grammars. A simple class of Context-free Queue Grammars is introduced and discussed.

### **A Model of Revision in Natural Language Generation**

*Marie M. Vaughan, David D. McDonald*

Department of Computer and

Information Science

University of Massachusetts

Amherst, MA 01003

pp. 90-96

We outline a model of generation with revision, focusing on improving textual coherence. We argue that high quality text is more easily produced by iteratively revising and regenerating, as people do, rather than by using an architecturally more complex single pass generator. As a general area of study, the revision process presents interesting problems. Recognition of flaws in text requires a descriptive theory of what constitutes well written prose and a parser which can build a representation in those terms. Improving text requires associating flaws with strategies for improvement. The strategies, in turn, need to know what adjustments to the decisions made during the initial generation will produce appropriate modifications to the text. We compare our treatment of revision with those of Mann and Moore (1981), Gabriel (1984), and Mann (1983).

### **The ROMPER System: Responding to Object-Related Misconceptions Using Perspective**

*Kathleen F. McCoy*

Department of Computer and

Information Science

University of Delaware

Newark, DE 19716

pp. 97-105

As a user interacts with a database or expert system, s/he may reveal a misconception about the objects modeled by the system. This paper discusses the ROMPER system for responding to such misconceptions in a domain independent and context sensitive fashion. ROMPER reasons about possible sources of this misconception. It operates on a model of the user and generates a cooperative response based on this reasoning. The process is made context sensitive by augmenting the user model with a new notion of object perspectives which highlights certain aspects of the user model due to previous discourse.

### **Encoding and Acquiring Meanings for Figurative Phrases**

*Michael G. Dyer, Uri Zernik*

Artificial Intelligence Laboratory

Computer Science Department

3531 Boelter Hall

University of California

Los Angeles, CA 90024

pp. 106-111

Here we address the problem of mapping phrase meanings into their conceptual representations. Figurative phrases are pervasive in human communication, yet they are difficult to explain theoretically. In fact, the ability to handle idiosyncratic behavior of phrases should be a criterion for any theory of lexical representation. Due to the huge number of such phrases in the English language, phrase representation must be amenable to parsing, generation, and also to learning. In this paper we demonstrate a semantic representation which facilitates, for a wide variety of phrases, both learning and parsing.

### **Semantically Significant Patterns in**

Natural language processing systems need large lexicons containing explicit

**Dictionary Definitions****Judith Markowitz**Computer Science Department  
De Paul University  
Chicago, IL 60604**Thomas Ahlswede, Martha Evens**Computer Science Department  
Illinois Institute of Technology  
Chicago, IL 60616*pp. 112-119*

information about lexical-semantic relationships, selection restrictions, and verb categories. Because the labor involved in constructing such lexicons by hand is overwhelming, we have been trying to construct lexical entries automatically from information available in the machine-readable version of *Webster's Seventh Collegiate Dictionary*. This work is rich in implicit information; the problem is to make it explicit. This paper describes methods for finding taxonomy and set-membership relationships, recognizing nouns that ordinarily represent human beings, and identifying active and stative verbs and adjectives.

**Computer Methods for Morphological Analysis****Roy J. Byrd, Judith L. Klawans**IBM Thomas J. Watson Research Center  
Yorktown Heights, NY 10598**Mark Aronoff, Frank Anshen**SUNY  
Stony Brook, NY 11794*pp. 120-127*

This paper describes our current research on the properties of derivational affixation in English. Our research arises from a more general research project, the Lexical Systems project at the IBM Thomas J. Watson research laboratories, the goal for which is to build a variety of computerized dictionary systems for use both by people and by computer programs. An important sub-goal is to build reliable and robust word recognition mechanisms for these dictionaries. One of the more important issues in word recognition for all morphologically complex languages involves mechanisms for dealing with affixes.

**Bulk Processing of Text on a Massively Parallel Computer****Gary W. Sabot**Thinking Machines Corporation  
245 First Street  
Cambridge, MA 02142*pp. 128-135*

Dictionary lookup is a computational activity that can be greatly accelerated when performed on large amounts of text by a parallel computer such as the Connection Machine™ Computer (CM). Several algorithms for parallel dictionary lookup are discussed, including one that allows the CM to look up words at a rate 450 times that of look-up on a Symbolics 3600 Lisp Machine.

**The Intonational Structuring of Discourse****Julia Hirschberg, Janet Pierrehumbert**AT&T Bell Laboratories  
600 Mountain Avenue  
Murray Hill, NJ 07974*pp. 136-144*

We propose a mapping between prosodic phenomena and semantico-pragmatic effects based upon the hypothesis that intonation conveys information about the intentional as well as the attentional structures of discourse. In particular, we discuss how variations in pitch range and choice of accent and tune can help to convey such information as: discourse segmentation and topic structure, appropriate choice of referent, the distinction between "given" and "new" information, conceptual contrast or parallelism between mentioned items, and subordination relationships between propositions salient in the discourse. Our goals for this research are practical as well as theoretical. In particular, we are investigating the problem of intonational assignment in synthetic speech.

**The Contribution of Parsing to Prosodic Phrasing in an Experimental Text-to-Speech System****Joan Bachenko, Eileen Fitzpatrick, C.E. Wright**AT&T Bell Laboratories  
600 Mountain Avenue  
Murray Hill, NJ 07974*pp. 145-155*

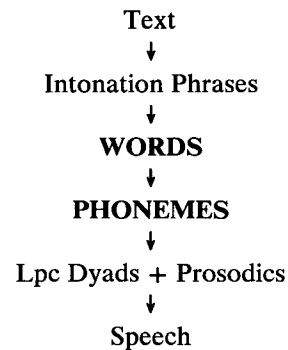
While various aspects of syntactic structure have been shown to bear on the determination of phrase-level prosody, the text-to-speech field has lacked a robust working system to test the possible relations between syntax and prosody. We describe an implemented system which uses the deterministic parser Fidditch to create the input for a set of prosody rules. The prosody rules generate a prosody tree that specifies the location and relative strength of prosodic phrase boundaries. These specifications are converted to annotations for the Bell Labs text-to-speech system that dictate modulations in pitch and duration for the input sentence.

We discuss the results of an experiment to determine the performance of our system. We are encouraged by an initial 5 percent error rate and we see the design of the parser and the modularity of the system allowing changes that will upgrade this rate.

**Morphological Decomposition and Stress Assignment for Speech Synthesis****Kenneth Church**

A speech synthesizer is a machine that inputs a stream of text and outputs a speech signal. This paper will discuss a small piece of how words are converted to phonemes.

AT&T Bell Laboratories  
600 Mountain Avenue  
Murray Hill, NJ 07974  
*pp. 156-164*



Typically, words are converted to phonemes in one of two ways: either by looking the words up in a dictionary (with possibly some limited morphological analysis), or by sounding the words out from their spelling using basic principles.

### **A Sentence Analysis Method for a Japanese Book Reading Machine for the Blind**

*Yutaka Ohyama, Toshikazu Fukushima,  
Tomoki Shutoh, Masamichi Shutoh*

C&C Systems Research Laboratories  
NEC Corporation

1-1, Miyazaki 4-chome, Miyamae-ku  
Kawasaki-city, Kanagawa 213, Japan  
*pp. 165-172*

The following proposal is for a Japanese sentence analysis method to be used in a Japanese book reading machine. This method is designed to allow for several candidates in case of ambiguous characters. Each sentence is analyzed to compose a data structure by defining the relationship between words and phrases. This structure (called *network structure*) involves all possible combinations of syntactically correct phrases. After *network structure* has been completed, heuristic rules are applied in order to determine the most probable way to arrange the phrases and thus organize the best sentence. All information about each sentence – the pronunciation of each word with its accent and the structure of phrases – will be used during speech synthesis. Experiment results reveal that 99.1% of all characters were given their correct pronunciation. Using several recognized character candidates is more efficient than only using first ranked characters as the input for sentence analysis. Also this facility increases the efficiency of the book reading machine in that it enables the user to select other ways to organize sentences.

### **Japanese Prosodic Phrasing and Intonation Synthesis**

*Mary E. Beckman, Janet B. Pierrehumbert*

Linguistics and Artificial Intelligence  
Research

AT&T Bell Laboratories  
600 Mountain Avenue  
Murray Hill, NJ 07974  
*pp. 173-180*

A computer program for synthesizing Japanese fundamental frequency contours implements our theory of Japanese intonation. This theory provides a complete qualitative description of the known characteristics of Japanese intonation, as well as a quantitative model of tone-scaling and timing precise enough to translate straightforwardly into a computational algorithm. An important aspect of the description is that various features of the intonation pattern are designated to be phonological properties of different types of phrasal units in a hierarchical organization. This phrasal organization is known to play an important role in parsing speech. Our research shows it also to be one reflex of intonational prominence, and hence of focus and other discourse structures. The qualitative features of each phrasal level and their implementation in the synthesis program are described.

### **FORUM ON CONNECTIONISM**

#### **Questions about Connectionist Models of Natural Language**

*Mark Liberman*

AT&T Bell Laboratories  
600 Mountain Avenue  
Murray Hill, NJ 07974  
*pp. 181-183*

**MODERATOR STATEMENT** (abridged) My role as interlocutor for this ACL Forum on Connectionism is to promote discussions by asking questions and making provocative comments. I will begin by asking some questions that I will attempt to answer myself, in order to define some terms. I will then pose some questions for the panel and the audience to discuss, if they are interested, and I will make a few critical comments on the abstracts submitted by Waltz and Sejnowski, intended to provoke responses from them. ...



### Language Learning in Massively-Parallel Networks

*Terrence J. Sejnowski*

Biophysics Department  
Johns Hopkins University  
Baltimore, MD 21218

*p. 184*

### Connectionist Models for Natural Language Processing

*David L. Waltz*

Thinking Machines Corporation  
245 First Street  
Cambridge, MA 02142 *and*  
Program in Linguistics & Cognitive Science  
Brandeis University  
Brown 125  
Waltham, MA 02254

*p. 185*

*...end of forum...*

### Donnellan's Distinction and a Computational Model of Reference

*Amichai Kronfeld*

Artificial Intelligence Center  
SRI International  
333 Ravenswood Avenue  
Menlo Park, CA 94025  
*and* Center for the Study of Language  
and Information  
Stanford University  
Stanford, CA 94305

*pp. 186-191*

### The Detection and Representation of Ambiguities of Intension and Description

*Brenda Fawcett, Graeme Hirst*

Department of Computer Science  
University of Toronto  
Toronto, Ontario, Canada M5S 1A4

*pp. 192-199*

### A Property-Sharing Constraint in Centering

*Megumi Kameyama*

Department of Computer and Information  
Science  
The Moore School of  
Electrical Engineering/D2  
University of Pennsylvania  
Philadelphia, PA 19104

**PANELIST STATEMENT** (abridged) Massively-parallel connectionist networks have traditionally been applied to constraint-satisfaction in early visual processing (Ballard, Hinton, and Sejnowski 1983), but are now being applied to problems ranging from the Traveling Salesman Problem to language acquisition (Rumbelhart and McClelland 1986). In these networks, knowledge is represented by the distributed pattern of activity in a large number of relatively simple neuron-like processing units, and computation is performed in parallel by the use of connections between the units. ...

**PANELIST STATEMENT** (abridged) After an almost twenty-year lull, there has been a dramatic upsurge of interest in massively-parallel models for computation, descendants of perceptron and pandemonium models, now dubbed "connectionist models". Much of the connectionist research has focussed on models for natural language processing. There have been three main reasons for this increase in interest:

1. Scientific adequacy of the models.
2. The availability of fine-grained parallel hardware to run the models.
3. The demonstration of powerful connectionist learning models.

...

In this paper I describe how Donnellan's distinction between referential and attributive uses of definite descriptions should be represented in a computational model of reference. After briefly discussing the significance of Donnellan's distinction, I reinterpret it as being three-tiered, relating to object representation, referring intentions, and choice of referring expression. I then present a cognitive model of referring, the components of which correspond to this analysis, and discuss the interaction that takes place among those components. Finally, the implementation of this model, now in progress, is described.

Ambiguities related to intension and their consequent inference failures are a diverse group, both syntactically and semantically. One particular kind of ambiguity that has received little attention so far is whether it is the speaker or the third party to whom a description in an opaque third-party attitude report should be attributed. The different readings lead to different inferences in a system modeling the beliefs of external agents.

We propose that a unified approach to the representation of the alternative readings of intension-related ambiguities can be based on the notion of a **descriptor** that is evaluated with respect to intensionality, the beliefs of agents, and a time of application. We describe such a representation, built on a standard modal logic, and show how it may be used in conjunction with a knowledge base of background assumptions to license restricted substitution of equals in opaque contexts.

A constraint is proposed in the Centering approach to pronoun resolution in discourse. This "property-sharing" constraint requires that two pronominal expressions that retain the same Cb across adjacent utterances share a certain common grammatical property. This property is expressed along the dimension of the grammatical function SUBJECT for both Japanese and English discourses, where different pronominal forms are primarily used to realize the Cb. It is the zero pronominal in Japanese, and the (unstressed) overt pronoun in English. The resulting constraint complements the original Centering, accounting for its apparent violations and

*pp. 200-206*

**A Model of Plan Inference that Distinguishes between the Beliefs of Actors and Observers**

*Martha E. Pollack*

Artificial Intelligence Center *and* Center  
for the Study of Language and Information  
SRI International  
333 Ravenswood Avenue  
Menlo Park, CA 94025  
*pp. 207-214*

providing a solution to the interpretation of multi-pronominal utterances. It also provides an alternative account of anaphora interpretation that appears to be due to structural parallelism. This reconciliation of centering/focusing and parallelism is a major advantage. I will then add another dimension called the "speaker identification" to the constraint to handle a group of special cases in Japanese discourse. It indicates a close association between centering and the speaker's viewpoint, and sheds light on what underlies the effect of perception reports on pronoun resolution in general. These results, by drawing on facts in two very different languages, demonstrate the cross-linguistic applicability of the centering framework.

Existing models of plan inference (PI) in conversation have assumed that the agent whose plan is being inferred (the actor) and the agent drawing the inference (the observer) have identical beliefs about actions in the domain. I argue that this assumption often results in failure of both the PI process and the communicative process that PI is meant to support. In particular, it precludes the principled generation of appropriate responses to queries that arise from invalid plans. I describe a model of PI that abandons this assumption. It rests on an analysis of plans as mental phenomena. Judgments that a plan is invalid are associated with particular discrepancies between the beliefs that the observer ascribes to the actor when the former believes that the latter has some plan, and the beliefs that the observer herself holds. I show that the content of an appropriate response to a query is affected by the types of any such discrepancies of belief judged to be present in the plan inferred to underlie that query. The PI model described here has been implemented in SPIRIT, a small demonstration system that answers questions about the domain of computer mail.

**Linguistic Coherence: A Plan-Based Alternative**

*Diana J. Litman*

AT&T Bell Laboratories  
600 Mountain Avenue, 3C-408A  
Murray Hill, NJ 07974  
*pp. 215-223*

To fully understand a sequence of utterances, one must be able to infer implicit relationships between the utterances. Although the identification of sets of utterance relationships forms the basis for many theories of discourse, the formalization and recognition of such relationships has proven to be an extremely difficult computational task.

This paper presents a plan-based approach to the representation and recognition of implicit relationships between utterances. Relationships are formulated as discourse plans, which allows their representation in terms of planning operators and their computation via a plan recognition process. By incorporating complex inferential processes relating utterances into a plan-based framework, a formalization and computability not available in the earlier works is provided.

**The Structure of User-Adviser Dialogues: Is there Method in their Madness?**

*Raymonde Guindon*

Microelectronics and Computer Technology  
Corporation – MCC

*Paul Sladky*

University of Texas, Austin *and* MCC

*Hans Brunner*

Honeywell – Computer Sciences Center

*Joyce Conner*

MCC

*pp. 224-230*

Novice users engaged in task-oriented dialogues with an adviser to learn how to use an unfamiliar statistical package. The users' task was analyzed and a task structure was derived. The task structure was used to segment the dialogue into subdialogues associated with the subtasks of the overall task. The representation of the dialogue structure into a hierarchy of subdialogues, partly corresponding to the task structure, was validated by three converging analyses. First, the distribution of non-pronominal noun phrases and the distribution of pronominal noun phrases exhibited a pattern consistent with the derived dialogue structure. Non-pronominal noun phrases occurred more frequently at the beginning of subdialogues than later, as can be expected since one of their functions is to indicate topic shifts. On the other hand, pronominal noun phrases occurred less frequently in the first sentence of the subdialogues than in the following sentences of the subdialogues, as can be expected since they are used to indicate topic continuity. Second, the distributions of the antecedents of pronominal noun phrases and of non-pronominal noun phrases showed a

### Commonsense Metaphysics and Lexical Semantics

*Jerry R. Hobbs, William Croft, Todd Davies, Douglas Edwards, Kenneth Laws*  
Artificial Intelligence Center  
SRI International  
pp. 231-240

pattern consistent with the derived dialogue structure. Finally, distinctive clue words and phrases were found reliably at the boundaries of sub-dialogues with different functions.

In the TACITUS project for using commonsense knowledge in the understanding of texts about mechanical devices and their failures, we have been developing various commonsense theories that are needed to mediate between the way we talk about the behavior of such devices and causal models of their operation. Of central importance in this effort is the axiomatization of what might be called "commonsense metaphysics". This includes a number of areas that figure in virtually every domain of discourse, such as scalar notions, granularity, time, space, material, physical objects, causality, functionality, force, and shape. Our approach to lexical semantics is then to construct core theories of each of these areas, and then to define, or at least characterize, a large number of lexical items in terms provided by the core theories. In the TACITUS system, processes for solving pragmatics problems posed by a text will use the knowledge base consisting of these theories in conjunction with the logical forms of the sentences in the text to produce an interpretation. In this paper we do not stress these interpretation processes; this is another, important aspect of the TACITUS project, and it will be described in subsequent papers.

### A Terminological Simplification Transformation for Natural Language Question-Answering Systems

*David G. Stallard*  
BBN Laboratories Inc.  
10 Moulton Street  
Cambridge, MA 02238  
pp. 241-246

A new method is presented for simplifying the logical expressions used to represent utterance meaning in a natural language system. This simplification method utilizes the encoded knowledge and the limited inference-making capability of a taxonomic knowledge representation system to reduce the constituent structure of logical expressions. The specific application is to the problem of mapping expressions of the meaning representation language to a database language capable of retrieving actual responses. Particular account is taken of the model-theoretic aspects of this problem.

### Some Uses of Higher-Order Logic in Computational Linguistics

*Dale A. Miller, Gopalan Nadathur*  
Computer and Information Sciences  
University of Pennsylvania  
Philadelphia, PA 19104-3897  
pp. 247-256

Consideration of the question of meaning in the framework of linguistics often requires an allusion to sets and other higher-order notions. The traditional approach to representing and reasoning about meaning in a computational setting has been to use knowledge representation systems that are based on first-order logic or that use mechanisms whose formal justifications are to be provided after the fact. In this paper we shall consider the use of a higher-order logic for this task. We first present a version of definite clauses (positive Horn clauses) that is based on this logic. Predicate and function variables may occur in such clauses and the terms in the language are the typed  $\lambda$ -terms. Such term structures have a richness that may be exploited in representing meanings. We also describe a higher-order logic programming language, called  $\lambda$ Prolog, which represents programs as higher-order definite clauses and interprets them using a depth-first interpreter. A virtue of this language is that it is possible to write programs in it that integrate syntactic and semantic analyses into one computational paradigm. This is to be contrasted with the more common practice of using two entirely different computation paradigms, such as DCGs or ATNs for parsing and frames or semantic nets for semantic processing. We illustrate such an integration in this language by considering a simple example, and we claim that its use makes the task of providing formal justifications for the computations specified much more direct.

### A Logical Semantics for Feature Structures

*Robert T. Kasper, William C. Rounds*  
Electrical Engineering and Computer

Unification-based grammar formalisms use structures containing sets of features to describe linguistic objects. Although computational algorithms for unification of feature structures have been worked out in experimental research, these algorithms become quite complicated, and a more precise

Science Department  
University of Michigan  
Ann Arbor, Michigan 48109  
*pp. 257-266*

description of feature structures is desirable. We have developed a model in which descriptions of feature structures can be regarded as logical formulas, and interpreted by sets of directed graphs which satisfy them. These graphs are, in fact, transition graphs for a special type of deterministic finite automaton.

This semantics for feature structures extends the ideas of Pereira and Shieber, by providing an interpretation for values which are specified by disjunctions and path values embedded with disjunctions. Our interpretation differs from that of Pereira and Shieber by using a logical model in place of a denotational semantics. This logical model yields a calculus of equivalences, which can be used to simplify formulas.

Unification is attractive, because of its generality, but it is often computationally inefficient. Our model allows a careful examination of the computational complexity of unification. We have shown that the consistency problem for formulas with disjunctive values is NP-complete. To deal with this complexity, we describe how disjunctive values can be specified in a way which delays expansion to disjunctive normal form.

#### FORUM ON MACHINE TRANSLATION

##### What Should Machine Translation Be?

*John S. White*

Siemens Information Systems  
Linguistics Research Center  
P.O. Box 7247, University Station  
Austin, TX 78712

*p. 267*

**MODERATOR STATEMENT** (abridged) After a considerable hiatus of interest and funding, machine translation has come in recent years to occupy a significant place in the discipline of natural language processing. It has also become one of the most visible representations of natural language processing to the outside world. Machine translation systems are relatively unique with respect to the extent of the coverage they attempt, and correspondingly, the size of the grammatical and lexical corpora involved. Adding to this the complexity introduced by multiple language directions into the same system design (and the enormous procedural problems imposed by simultaneous development in several sites) gives some clue as to the optimism which presently exists for machine translation. ...

##### Machine Translation Will Not Work

*Martin Kay*

Xerox Palo Alto Research Center  
3333 Coyote Hill Road  
Palo Alto, CA 94304

*p. 268*

**PANELIST STATEMENT** (abridged) Large expenditures on fundamental scientific research are usually limited to the hard sciences. It is therefore entirely reasonable to suppose that, if large sums of money are spent on machine translations, it will be with the clear expectation that what is being purchased is principally development and engineering, and that the results will contribute substantially to the solution of some pressing problem.

Anyone who accepts large (or small) sums on this understanding is either technically naive or dangerously cynical. ...

##### Machine Translation Already Does Work

*Margaret King*

ISSCO  
54, rte des Acacias  
CH-1227 Geneva, Switzerland

*pp. 269-270*

**PANELIST STATEMENT** (abridged) The first difficulty in answering a question like "Does machine translation work" is that the question itself is ill-posed. It takes for granted that there is one single thing called machine translation and that everyone is agreed about what it is. But in fact, even a cursory glance at the systems already around, either in regular operational use or under development, will reveal a wide range of different types of systems. ...

*...end of forum...*

#### Selected Dissertation Abstracts

Compiled by:

Susanne M. Humphrey, National Library of Medicine, Bethesda, MD 20894

Bob Krovetz, University of Massachusetts, Amherst, MA 01002

The following are citations selected by title and abstract as being related to computational linguistics or knowledge representation, resulting from a computer search, using the BRS Information Technologies retrieval service, of the Dissertation Abstracts International (DAI) data base produced by University Microfilms International.

Included are the title; author; university, degree, and, if available, number of pages; DAI subject category chosen by the author of the dissertation; UM order number and year-month of entry into the data base; and abstract. References are sorted first by DAI subject category and second by author. Citations denoted by an MAI reference do not yet have abstracts in the data base and refer to abstracts in the published Masters Abstracts International.

Unless otherwise specified, paper or microform copies of dissertations may be ordered from

University Microfilms International

Dissertation Copies

Post Office Box 1764

Ann Arbor, MI 48106

telephone for U.S. (except Michigan, Hawaii, Alaska): 1-800-521-3042

for Canada: 1-800-268-6090.

Price lists and other ordering and shipping information are in the introduction to the published DAI. An alternate source for copies is sometimes provided at the end of the abstract.

The dissertation titles and abstracts contained here are published with permission of University Microfilms International, publishers of Dissertation Abstracts International (copyright 1985, 1986 by University Microfilms International), and may not be reproduced without their prior permission.

### **An Instrument for Evaluating the Use of Speech in Computer Communication**

*Janan Arif Al-Awar*

The Johns Hopkins University Ph.D. 1985,

227 pages

Computer Science

*University Microfilms International*

*ADG85-18472. 8512*

The thesis of this research is that the usefulness of voice Input/Output (I/O) for computer systems can be determined through a set of defining characteristics. These characteristics can be divided into two main groups: (a) those conditions necessary for the viability of any voice communication system, and (b) additional features that may not be vital for the system, but, if implemented, would greatly enhance its usability.

The purpose of the research was to develop a paper and pencil instrument for evaluating the usefulness of speech in user-computer communication. The items in the instrument were obtained by identifying features that are considered imperative and/or desirable in either a voice input or voice output system. These features were assembled in three groups: items related to the users, items related to the tasks, and items related to the environment in which the system would be present. They were validated by an analysis of their contents, an empirical content validation study, and field observation studies of two operational voice systems. Finally, the effects of voice feedback on performance were investigated in a laboratory simulation experiment.

### **Extraction and Generalization of Expert Advice**

*David Paul Benjamin*

New York University Ph.D. 1985, 172 pages

Computer Science

*University Microfilms International.*

*ADG85-22013. 8602*

This work describes a method for representing knowledge in production systems which makes use of the conflict set. This permits a rich description of task situations, and allows the use of control productions to effect conflict resolution. A set of extensions to the OPS5 production system is described which facilitates the implementation of this approach within OPS5. This extended system is then used to implement a multi-level, goal-directed production system for the construction of expert systems, CAMERA, in which control information is automatically built from the actions of an expert trainer. This control information consists of sequencing and goal information which is interactively extracted from the trainer by CAMERA, and generalized by DISC, which models generalization as the process of finding "discriminating" features, which are those features of a situation that cause a particular method to be chosen, and then constructing a description of those features. When solving a task, CAMERA examines only the discriminating features specified in the generalized control rules. Thus, instead of matching all the productions against the working memory, CAMERA considers only the relevant rules. Experiments with the system are described.

### **Entity-Based Data Base Management Systems**

*Robert Pershing Brazile*

The University of Texas at Dallas Ph.D. 1985,

A data base management system (DBMS) which supports one of the three popular data models (hierarchical, network and relational) uses a record-based technique for implementing a data base design. This is advantageous when the data being stored is homogeneous, but the technique is

103 pages  
 Computer Science  
*University Microfilms International.*  
 ADG85-16506. 8512

not as well suited towards implementing data bases with particularly diverse properties and data types. An alternative technique, the entity-based technique, is presented which allows the design of the data base to be described and implemented as entities and relationships.

The entities and relationships may be presented to a DBMS which will have the capability of deciding whether to store them as individual entities or group them into records. The advantages of defining and maintaining the conceptual view of the data base with entities and relationships are discussed and examples are given to illustrate these advantages.

A system to support the entity-based design technique can be implemented in two ways: (1) Write an interface to an existing data base management system; (2) Implement a data base management system explicitly for the entity-based technique. A set of primitive binary relations is presented which will support the concept of entity-based design. By implementing procedures to create these primitive relations, an entity-based design may be stored and populated as a data base. Algorithms for implementing such procedures for both the above-mentioned cases are included.

Also presented are several examples of traditional data base applications and how they might be implemented using the entity-based technique. In addition to the traditional applications, the entity-based technique can support new applications. There is currently much activity in research concerning knowledge representation in Artificial Intelligence (AI) applications. An example is given showing that the entity-based technique is flexible enough to support many different knowledge representations. By using this data base design technique, the AI researcher has available the features and capabilities of a data base management system. An interface to define the knowledge representation to the DBMS is presented and an example using the interface is given.

### **Queries in Deductive Database Systems**

*Ey-Chih Chow*

University of California, Berkeley, Ph.D.  
 1984, 113 pages  
 Computer Science  
*University Microfilms International.*  
 ADG85-12783. 8601

The problem of efficiently evaluating queries in deductive, relational database systems is examined. The major areas investigated are: (1) appropriate arrangements of data, intensional and extensional, in deductive database systems, (2) efficient sequential evaluation strategies that take advantage of the underlying data characteristics and query semantics, and (3) efficient parallel evaluation strategies that are tailored to distinct types of queries.

First, the database formalism of PROLOG logic is argued to be appropriate for such systems. Query evaluation strategies based on the traditional PROLOG deductive mechanism are then examined to determine their efficiency under different environments. This examination is done through two distinct categories of systems: a simplified model of a conventional database system and a purely deductive database system. These two systems are distinct in their underlying data characteristics and query semantics. Their different data characteristics are due to different data-saving schemes, which achieve storage saving and high expressive power, respectively, under varying applications. Differences in data characteristics include relative amount of extensional vs. intensional data, the order of each relation, the number of clauses per relation, and the overall syntactic structure of data per relation. Semantics of queries, on the other hand, deal with the number of qualifying instantiations required to answer a query.

The analysis shows that, for sequential evaluation in the simplified model of a conventional database system, the traditional PROLOG deductive mechanism is able to provide only primitive techniques for trimming down the cross products of multi-relation queries. This turns out to be the most expensive part of query evaluation in such systems. For sequential evaluation in the purely deductive database system, by contrast, the

PROLOG deductive mechanism is adequate for efficient evaluation of queries provided the system has a sophisticated technique for ordering predicates in the qualifications of queries being evaluated. For parallel evaluation, on the other hand, the maximal efficient strategies in the simplified model of conventional database systems are characterized by dividing-by-relation. Those in the purely deductive database systems are characterized by dividing-by-clause. (Abstract shortened with permission of author.)

**A Connectionist Approach to Word Sense Disambiguation**

*Garrison Weeks Cottrell*

The University of Rochester Ph.D. 1985,  
145 pages

Computer Science

*University Microfilms International*  
ADG85-16464. 8512

A new architecture for representing parsing of natural language is described which conforms to psycholinguistic, neurolinguistic and computational constraints. The parsing model uses a particular spreading activation or neural network scheme called connectionism which entails a massive number of appropriately connected computing units that communicate through weighted levels of excitation and inhibition. Such an architecture adds considerable constraints of its own which serve to explain some constraints at the functional level. The model accounts for psycholinguistic data on the access of word meanings, recent neurolinguistic data on agrammatism, and some of the apparent parsing strategies of normals.

**Syntactic Clues to Discourse Structure:  
A Case from Journalism**

*Nan Decker*

Brown University Ph.D. 1985, 297 pages  
Computer Science. Language, Linguistics

*University Microfilms International*  
ADG85-19827. 8601

This dissertation describes a syntax-based technique for processing newspaper reports. The DUMP (Discourse Understanding Model Program) program creates summaries of news reports and labels the kinds of information delivered by clauses in the text based solely on their syntactic form. DUMP's approach is a departure from the knowledge-based, relatively syntax-free methods of summary creation described in DeJong (1979) and also from information formatting techniques in sublanguage research which rely on word co-occurrence classes in texts from restricted semantic domains (Hirschman and Sager 1982).

The syntactic rules used by DUMP reflect grounding principles found universally in discourse. Certain assertional syntactic structures in the text deliver foreground information which tells the events in the story. These events comprise a summary of the report. Less assertional structures are used to express background, supportive information which fleshes out the skeleton provided in the foreground. The strong correlation between syntactic form and information type of this background information allows DUMP to subcategorize it into the following classes: plans, background events and processes, current state, secondary information, identifications, import, effects of actions, related episodes, comments and collateral.

Specific syntactic structures also mark the beginning of new episodes, characterized by a change in participants, setting and/or time.

The principles embodied by DUMP are useful to both sublanguage and AI research. In the former case, they allow the automatic processing of texts regardless of the breadth of their semantic field, since all texts achieve texture by the division of information into foreground and background. (The degree of correlation between syntactic form and information type of the supportive material in other genres is suggested as a topic for further investigation.) In the latter case, DUMP could serve as a support to knowledge-based programs which develop conceptual representations of text.

**A Mechanism for Natural Language Database**

*Richard Allen Feinauer*

University of Cincinnati Ph.D. 1985,  
367 pages

Computer Science

*University Microfilms International*

The purpose of this dissertation is to investigate the capabilities of a transportable natural language database query methodology that has only a surface level understanding of the user's query and uses a relational logical schema as the basis of its world model. A secondary goal of this dissertation is to explore the usefulness of explicit optimization techniques in a natural language database query methodology.

The basic features of the methodology described in this dissertation and

ADG85-18098. 8512

implemented in a test system called DRIVER are an Analyzer that converts the user's query into a relational algebra statement and an Evaluator which converts the relational algebra statement into the data manipulation language of the target database management system and presents the answer to the user. The Analyzer contains five components. They are: a Word Role Identifier, a Phrase Segmenter, a Phrase Analyzer, a Query Generator, and a User Dialog. Each component transforms the query into a form which is closer to the relational algebra statement than its input. The Analyzer has four external sources of information. They are: a query grammar, a world model, a query complexity measure, and the user. The world model is based on a relational logical schema of the target database domain. The physical database may have any organization provided that a relational schema can be mapped onto it. Both the query grammar and the complexity measure make extensive use of the logical schema.

The investigative methodology was evaluated using 640 test queries. Four hundred and four (63.1%) of those queries were interpreted correctly. One hundred and fourteen (17.9%) of the queries were interpreted substantially correctly (the interpretation was correct but unfriendly or it provided a super set of the desired information). One hundred and twenty-two (19.0%) of the queries were not interpreted correctly. Three hundred and thirty-three of the 404 correctly interpreted queries and 85 of the 114 substantially correctly interpreted queries had only a single interpretation. For the remaining queries two or more interpretations were produced and the user had to select the correct interpretation.

This dissertation makes contributions to the following aspects of natural language database query research: improved understanding of the capabilities and limitations of methodologies that have only a surface level understanding of the query, improved understanding of the limitations and capabilities of the logical schema as the basis of a world model, and a demonstration of the usefulness of explicit optimization techniques in natural language research. This dissertation also develops a powerful disambiguation tool called the complexity measure.

**A Study of Parallelism in the Classifier System and its Application to Classification in KL-ONE Semantic Networks**

*Stephanie Forrest*

The University of Michigan Ph.D. 1985,  
244 pages

Computer Science

*University Microfilms International*

ADG85-20897. 8601

Current techniques for knowledge representation in artificial intelligence limit their applicability in many domains. One reason for this limitation is the large amount of computation involved in processing reasonably-sized knowledge-bases. Current research in parallelism suggests that one promising direction is the development of parallel architectures that are designed for applications in artificial intelligence.

In the dissertation, I show how one model of fine-grained parallelism, the Classifier System, can be used to implement a set of useful operations for the classification of knowledge in semantic networks. The Classifier System appears amenable to hardware implementation, but for the dissertation, a software simulation was written. The "classification" problem was selected as the focus of the investigation because it is a central problem for many knowledge-based systems. Of the various knowledge representation paradigms in use today, the KL-ONE family has addressed the problem of classification most directly. I therefore, have used a subset of this language for my investigations.

I have implemented a compiler that translates KL-ONE definitions into a Classifier System representation. In addition, I have developed a group of parallel algorithms that uses the Classifier System representation to decide where an incoming concept should be classified in an existing KL-ONE network. A significant part of the work on this project has consisted of developing a collection of more general algorithms for the Classifier System (set operations, numerical processing, and the construction of default hierarchies) that have formed the basis for the classification algorithms.



**The Inclusion of Expertise in a Decision Support System for Strategic Decision Making**

*Kenneth Michael Goul*

Oregon State University Ph.D. 1985,  
236 pages

Computer Science

*University Microfilms International*

*ADG85-18598. 8512*

**Aspects of the Implementation of Type Theory**

*Robert William Harper, Jr.*

Cornell University Ph.D. 1985, 234 pages  
Computer Science

*University Microfilms International*

*ADG85-16933. 8601*

**A Distributed Knowledge-Based Learning System for Information Retrieval**

*Uttam Mukhopadhyay*

The study was divided into three major phases: designing and implementing the general operations for controlling the Classifier System, reformulating the KL-ONE formalism in terms of these operations, and analyzing the efficiency of the parallel algorithms with respect to the inherent computational trade-offs among the number of processors, length of computation, and degree of inter-processor communication. The study concludes that architectures of this type are capable of significantly reducing the time complexity of common semantic network operations.

A decision support system (DSS) incorporating domain expertise guides, tutors, and consults a decision maker in opportunity, problem, and crisis identification activities. The objective for the system is to promote improved decision making. Using an "Independent Groups" design, an experimental study was conducted to investigate the effects of DSS use on performance in the assessment phase of the strategic planning process.

The findings of the study indicate that decision support systems incorporating expertise can improve the effectiveness of problem recognition in unstructured environments. Experimental treatments consisted of (1) use of a DSS with a complete rule base, (2) use of a DSS with a 10% subset of the complete rule base, and (3) no DSS exposure. Measures of performance from several stages of the decision making process are analyzed.

The subject of this work is the implementation of an intuitionistic theory of types as the basis for the PRL proof development system. Several aspects of the implementation are discussed. First, the problem of organizing the proof theory to support refinement, a style of top-down proof construction, and extraction, a means of obtaining programs from constructive proofs, is discussed. The expressive power of type theory precludes the incorporation of ordinary well-formedness constraints; in this setting such constraints must be proved. A version of type theory is presented which alleviates much of the burden of demonstrating well-formedness. Several key proof-theoretic facts, including a relative consistency theorem, are obtained.

Then the problem of providing automated assistance for proofs of type membership and equality in a type is considered. Both of these relations are undecidable, so any assistance that can be provided must be incomplete. Therefore the strategy employed is to construct heuristic decision methods for those cases that arise often in the course of proof construction. The semantics of type theory is such that a given term may inhabit any number of unrelated types; furthermore, two terms may be equal in one type and unequal in another. This leads to the consideration of a logic of typed terms, a subsystem of type theory that formalizes the notion of attaching a type expression to term occurrences.

The decision methods for membership and equality are based on annotation, the attachment of a type expression to a term occurrence. The method is similar to that used by Hindley and Milner in their type inference algorithm for the  $\lambda$ -calculus. An extension of Robinson's unification algorithm, called constrained unification, is used to propagate contextual constraints. This algorithm incorporates a limited theory of equality including the computation rules for type theory. The annotation algorithm is based on several heuristics for type assignment, and is designed to be a tunable framework for the construction of automated proof assistance for the PRL system. An equality decision method based on annotation and unification is also presented.

MINDS (Multiple Intelligent Node Document Servers) is a distributed system of knowledge-based query engines for efficiently retrieving multimedia documents in an office environment of distributed workstations. By

University of South Carolina Ph.D. 1985,  
147 pages  
Computer Science  
*University Microfilms International*  
ADG85-18042. 8512

### **A Knowledge-Based Approach to Natural Language Understanding**

*Jeannette Grace Neal*

State University of New York at Buffalo  
Ph.D. 1985, 234 pages  
Computer Science  
*University Microfilms International*  
ADG85-18765. 8601

**Representing Constructive Theories in High-Level Programming Languages**  
*Ryan Dale Stansifer*

learning document distribution patterns, as well as user interests and preferences during system usage, it customizes document retrievals for each user.

The nodes cooperate by sharing knowledge (documents and their properties), metaknowledge (distribution of documents from individual user viewpoints) and query-processing tasks. An architecture for the intelligent node has been presented. The expert query handler uses the metaknowledge stored at the node to plan task and query decompositions while the expert document manager performs routine document handling activities and uses domain-level heuristics for acquiring and updating metaknowledge.

A two-layer learning system has been implemented for studying plausible heuristics. The metaknowledge base used by the query engine is learned at the lower level with the help of heuristics for assigning credit and recommending adjustments; these heuristics are incrementally refined at the upper level. Insights gained from these simulations will enable system designers to integrate and refine heuristics in other domains.

An extremely significant feature of any Natural Language (NL) is that it is its own meta-language. One can use a NL to talk about the NL itself and to give instruction in the use and understanding of the same NL. In this thesis we present a language processing expert system that we have implemented in the role of an educable cognitive agent whose domain of expertise is language understanding and whose discourse domain includes its own language knowledge. We present a representation of language processing knowledge and a core of knowledge, including a Kernel Language, which forms the knowledge base for this AI System. Since linguistic knowledge is part of its domain of discourse, the System can be instructed in the processing and understanding of ever more sophisticated language, with instruction initially given in the predefined Kernel Language. As the System's language knowledge is expanded beyond the primitive Kernel Language, instruction of the System is expressed in an increasingly sophisticated subset of the language being taught. Thus the System's language is used as a meta-language for the self-same language.

Our NLU System is implemented in the form of a general purpose inference system which reasons according to the rules of its knowledge base. This knowledge base comprises the System's task domain knowledge and includes, but is not restricted to, its language processing knowledge.

In this thesis we discuss two experiments that we conducted. In the first experiment, our approach was to teach the System to treat linguistic knowledge in a manner that is commonly used for general knowledge (e.g. property-value pairs) and to use its acquired natural language subset as a meta-language for the same language. In the second experiment, we taught the System to process language according to a subset of a Lexical-Functional Grammar. One of our original objectives was to design a system that was as theory-independent as possible. The purpose of this second experiment was to test, at least to some extent, whether we had achieved this objective.

We also discuss the parsing and interpretation strategies of the System. Parsing is performed according to a combined bottom-up top-down strategy with a focusing context resulting from the bi-directional inference subsystem. Parsing and interpretation take place in an integrated manner in our System, governed by the language definition input to the System by a teacher-user.

This thesis concerns certain constructive theories we call programming logics. A programming logic is a strongly typed, functional programming language. Its semantics can be defined by a set of rewrite rules. We

Cornell University Ph.D. 1985, 225 pages  
 Computer Science  
*University Microfilms International*  
 ADG85-16902. 8601

consider only programming logics in which predicate logic can be embedded, thus justifying the use of the term "logic". Furthermore, we single out those programming logics for which proofs of existential formulas have a canonical form explicitly revealing the witness. Thus, only constructive theories can be represented in these programming logics, and proofs in programming logics are executable.

Three programming logics are described in this thesis. The first is based on  $HA^\omega$  (Heyting arithmetic of the omega order). The second is based on a Martin-Lof style type theory. The third is based on intuitionistic set theory. These programming logics have been implemented in the programming language ML (described in detail in this thesis). The core of each implementation is the same, and this logic engine, written in ML, is described in an appendix.

Programming logics would be woefully inadequate as a basis for automatic deduction without provision for reasoning at a higher plane. We show how to implement proof strategies for programming logics. As an example, we show how PROLOG, or more precisely, linear input resolution could be implemented as a proof strategy for a programming logic. Finally, we demonstrate how certain elements of classical logic can be used in proof development and then eliminated in these cases.

The experience gained in implementing these programming logics and described in this thesis contributes to the design of theories in which proofs are to be executable. The techniques of implementation demonstrated in this thesis can be used to build prototypes of a wide variety of theories. The ease in experimenting with new theories and the clarity with which the underlying mechanisms can be discussed are due to the representation of these theories at the level of abstract syntax and the directness by which the representation has been implemented in ML.

**An Efficient Context-Free Parsing  
 Algorithm for Natural Languages and Its  
 Applications**

*Masaru Tomita*

Carnegie-Mellon University Ph.D. 1985,  
 221 pages  
 Computer Science  
*University Microfilms International*  
 ADG85-17539. 8512

This thesis introduces an efficient context-free parsing algorithm and emphasizes its practical value in natural language processing. In the theoretical worst case analysis, the parsing algorithm occasionally takes more than  $O(n^3)$  time with kinds of context-free grammars which are very unlikely to appear in natural languages. As far as practical natural language processing is concerned, on the other hand, the parsing algorithm seems more efficient than any existing algorithms including Earley's algorithm. Experiments with several English grammars and sample sentences show that our algorithm is 5 to 10 times faster than Earley's standard algorithm.

The parsing algorithm can be viewed as an extended LR parsing algorithm which embodies the concept of a "graph-structured stack". Unlike the standard LR, the algorithm is capable of handling arbitrary non-cyclic context-free grammars including ambiguous grammars, with little loss of LR efficiency. In particular, if its grammar is "close" to LR, most of the LR parsing efficiency can be preserved. Natural language grammars are, fortunately, considerably "close" to LR, compared with other general context-free grammars.

The algorithm is an all-path parsing algorithm; it produces all possible parse trees (a parse forest) in an efficient representation called a "shared-packed forest". This thesis also shows that Earley's forest representation has a defect and his representation cannot be used in natural language processing.

The last chapters of the thesis suggest practical applications of the algorithm. A concept of left-to-right on-line parsing is introduced, taking advantage of the fact that our algorithm parses a sentence strictly from left to right. Several benefits of on-line parsing are described, and its application to user-friendly natural language interface is discussed. This thesis also proposes a technique to disambiguate a sentence out of the shared-

**Unifying Representation and Generalization: Understanding Hierarchically Structured Objects**

*Kenneth Hal Wasserman*

Columbia University Ph.D. 1985, 245 pages  
Computer Science

*University Microfilms International*

ADG85-23256. 8602

packed forest representation by asking the user questions interactively. Finally, a personal/interactive machine translation system is suggested.

Hierarchies are pervasive. They are used to organize and describe many artificial and natural phenomena. In general, humans are very good at understanding them. It therefore seems reasonable to give computers the same ability if they are to be "intelligent".

The integration of representation and generalization is necessary in order to understand hierarchically structured objects. In this thesis we address the issues involved and present a scheme, MERGE, designed to be used in computer systems that understand and automatically classify instances of hierarchies in a given domain.

The MERGE scheme uses a form of dynamic generalization-based memory in order to achieve this integration. Representations of individual hierarchies are stored in terms of how they vary from previously created generalized concepts. Memory is continually reorganized as new data becomes available to a MERGE-based system so that it accurately reflects the known information. The overall effect of this scheme is that representations of individual hierarchies are enhanced by the use of information in the knowledge base. These representations are in turn used to enhance the knowledge base by permitting more and better generalizations to be made.

We have developed two MERGE-based computer systems that intelligently understand hierarchies. CORPORATE-RESEARCHER is a program that learns about upper-level corporate management hierarchies when it is fed representations of corporate charts. RESEARCHER is a larger, natural language processing program that reads and understands patent abstracts about physical objects. Both programs serve as intelligent information systems that automatically classify representations of instance hierarchies.

**Mathematical Foundations of Manufacturing Science: Theory and Implications**

*Steven Hyung Kim*

Massachusetts Institute of Technology  
Ph.D. 1985

Engineering, System Science

*This item is not available from University*

*Microfilms International.*

ADG85-56516. 8602.

A mathematical architecture for manufacturing science has been constructed by building on the foundations of mathematics. The contributions in this thesis may be outlined more specifically as follows: (1) Conceptual. Frameworks are presented for a systematic study of manufacturing systems as well as for a mathematical architecture for manufacturing theory. (2) Technical. This category consists of three types of results: (a) Algebraic structures are shown to be appropriate mathematical structures for the analysis phase. (b) Symbolic logic is used to formalize the synthesis phase. (c) Both the analysis and synthesis phases may be grounded on the foundations of mathematics, namely set theory and logic.

Starting from set-theoretic concepts, the notion of an algebraic structure is formalized, then used as a uniform base for specialized structures such as lattices, graphs and groups. These constructs are in turn used to develop an infrastructure for the study of manufacturing systems. Algebraic and set-theoretic structures are shown to be appropriate bases for system representations such as matrices, graphs, and state spaces. Moreover they provide (1) a precise symbolism for specifying qualitative systems concepts such as linkages and hierarchies, (2) a uniform framework for more specialized theories such as automata theory and control theory, and (3) a base on which to build quantitative theories.

In addition symbolic logic is used to formalize the Design Axioms, a set of decision principles which were previously available only informally. The implications of such formalization are carried forward to two levels, in terms of theoretical as well as operational consequences. The results are as follows: (1) Theoretical. The relationships among the Axioms and their corollaries can be studied more rigorously. An unexpected result is that the propositions which had previously been considered to be corollaries of the Function and Information Axioms may, in fact, be divided into two categories consisting of direct and indirect consequences. The class of

indirect consequences may be further partitioned into those which follow from the Axioms plus some weak assumptions, versus those that require strong assumptions. (2) Practical. The long-term goal of axiomatics research is to establish concrete decision rules and techniques to enable computer systems to design manufacturing and engineering systems. Stating the Axioms in symbolic logic makes it clear how they may be written as clauses in a logical programming language such as PROLOG. Generalized structures are presented for encoding procedures and data in PROLOG, then illustrated through incorporation into a program called the Computerized Axiomatic System (CAS). The operating modes of CAS are discussed, as is the overall architecture for a full-fledged expert system.

### The Interpretation of Functional Relations

*David Loring Farwell*

University of Illinois at Urbana-Champaign

Ph.D. 1985, 385 pages

Language, Linguistics

*University Microfilms International*

ADG85-21761. 8601

This thesis deals with the representation of the functional relations such as agent, instrument, source and so on that are assigned to the various participants in a situation. It can be viewed as an investigation of that conceptual knowledge which defines what is and what is not a conceivable situation and how such knowledge is applied in the interpretation of natural language utterances.

The main body of the study concerns two interrelated topics. First, it contains a discussion of those functional relations that are the most likely to have universal application in the conceptual representation of situations. This discussion includes a concise definition of each of the functional relations proposed as well as a number of examples demonstrating the range of situations that it is intended to cover. Second, there is a description of the way in which this class of knowledge contributes to the interpretation of natural language utterances. The approach requires that a distinction be made between the literal interpretation of an utterance and its ultimate interpretation. Functional structure directly determines the former interpretation. However, the literal interpretation may be incomplete or ill-formed and, as a result, further processing on the basis of domain specific kinds of knowledge is required.

The general approach differs from others in that functional structure is viewed as but one of various levels or components of conceptual knowledge that effect both the form of linguistic expressions as well as the complexity of their associated conceptual representations. This enriched representation allows for directed inferencing with respect to particular domains of knowledge.

### Theme and Case as Determinants of the Domain of Movement

*Eileen M. Fitzpatrick*

New York University Ph.D. 1985, 214 pages

Language, Linguistics

*University Microfilms International*

ADG85-22033. 8602

The movement domain in generative grammar is currently structurally defined, either by bounding (Chomsky (1981)) or by path-by-government (Cattell (1976); Kayne (1981)).

However, structural definitions incorrectly predict that (1) and (2) are, domain-wise, identical: (1) \*In whom (,S) does (,NP) your trust e command her respect? (2) In whom (,S) does (,NP) your trust e remain unshaken?

The Lexical Domain Hypothesis (LDH) defines a domain thematically. Each head that chooses a thematic subject establishes a new domain. Thus, the LDH correctly predicts that *remain unshaken* establishes no domain since it chooses no thematic subject. Therefore, *trust*, which chooses a subject, functions as the only thematic head, thereby permitting extraction. In (1), since *command her respect* chooses a thematic subject, *command* functions as a thematic head. Thus, extraction from the subject results in movement into a distinct domain, thereby blocking extraction.

Membership of a phrase in a domain is indicated in the surface string through the Case and selectional properties of the phrase. For example, the dependents of a phrase share in a domain larger than the phrase if the left edge of the phrase is either Case-marked or selected by an item

external to the phrase. Thus, extraction from the complement in (3) is licit since the verb *consider* Case-marks the complement's left edge, *them*, externally: (3) What did he consider them e? In contrast, phrase-internal determination of the phrase's left edge marks the head of the phrase as the thematic head of a distinct domain. Thus, extraction from the complement in (4) is illicit since the complement's left edge, *their*, is Case-marked and selected internal to the complement: (4) \*What did he consider their e?

The LDH provides an explanation both for the data handled by the structural approaches and for the data that exhibit a misalignment between structure and thematic relations. In addition, unlike the structural approaches, it provides an account of how the child intuits the domain definition; namely, from the Case and selectional cues in the spoken surface string.

### **Complex Predicates and Lexical Operations in Japanese**

*Akira Ishikawa*

Stanford University Ph.D. 1985, 321 pages  
Language, Linguistics  
*University Microfilms International*  
ADG85-22161. 8602

This thesis presents a lexicalist analysis of Japanese complex predicates. It discusses a way of organizing the lexicon so that complex predicates can be formed in the lexicon rather than in syntax. The analysis is shown to provide enough levels of representation to take care of grammatical phenomena involved in complex predicates, thus overcoming the deficiencies of previous lexicalist analyses.

Chapter One introduces the reader to the framework employed in the thesis: Lexical Functional Grammar.

Chapter Two discusses three problems that the traditional transformational analysis of complex predicates cannot solve, showing that the analysis is untenable.

Chapter Three develops the basic mechanisms for word-formation used in this thesis. Two word-formation rules are introduced to take care of suffixation in Japanese, and it is argued that predicates enter into word-formation in their root form.

Chapter Four deals with concrete problems concerning complex predicates in Japanese. Causative-type complex predicates are first taken up as representative cases. A lexical operation called Object Function Sharing is shown to accompany complex predicate formation in Japanese. The problems of constituent order and of order among suffixes are also considered with regard to complex predicates.

Chapter Five is concerned with case-marking in Japanese. Case-features are introduced to represent case-markers as bundles of case-features and to specify the case-marking possibilities of arguments of predicates.

Chapter Six summarizes the whole thesis.

### **The Understanding of Anaphora in Written Text by Fourth Grade English Dominant, Spanish Dominant, and English-Spanish Bilingual Pupils**

*Jorge E. Schneider*

Hofstra University Ph.D. 1985, 278 pages  
Language, Linguistics  
*University Microfilms International*  
ADG85-23278. 8602

The major purpose of this study was to investigate the understanding of anaphora in written text by fourth grade English dominant pupils reading English, Spanish dominant pupils reading Spanish, and English-Spanish bilingual pupils reading English and Spanish. A secondary purpose was to identify different types of anaphora that present the most difficulty to the English dominant pupils in English, the Spanish dominant pupils in Spanish, and the bilingual pupils in English and Spanish.

A test in English and a separate test in Spanish were developed in which ten types of anaphora were emphasized. The English dominant pupils took the test in English, the Spanish dominant pupils took the test in Spanish, and the bilingual pupils took the tests in English and Spanish. The subjects were forty English dominant fourth graders from an elementary school in Lawrence, New York, and twenty-three Spanish dominant, as well as twenty-three bilingual fourth graders from a school in a similar neighboring school district.

The differences among groups and the differences among the various types of anaphora were tested with one-way analysis of variance tests. Pertinent findings included: (1) Significant differences were found in the

performance of English dominant and Spanish dominant pupils and in the performance of the bilingual group in English and Spanish. (2) No significant differences were found between English dominant and bilingual pupils in English, or between Spanish dominant and bilingual pupils in Spanish. (3) Significant differences were found among the various types of anaphora in both English and Spanish in the degree of difficulty they presented to the readers.

Conclusions from the study included: (1) It appears that anaphora is a syntactical structure which is developed sooner in English than in Spanish. (2) Anaphora is a difficult structure for fourth graders in both English and Spanish. (3) Bilingual pupils seem to develop English and Spanish language strategies at the same rate as their monolingual counterparts. (4) Bilingual pupils seem to apply strategies learned in a particular language without transferring these skills from one language to the other.

### **The Semantics and Pragmatics of Preposing**

*Gregory Louis Ward*

University of Pennsylvania Ph.D. 1985,  
313 pages

Language, Linguistics

*University Microfilms International*

ADG85-23465. 8602

The extrasentential competence which underlies a speaker's knowledge about the appropriateness of some syntactically well-formed sentence in a particular context has come to be called pragmatic competence. This research examines that aspect of pragmatic competence involved in English preposing constructions (e.g., That part we haven't finished yet). An analysis of 915 tokens of naturally occurring data reveals that preposing performs two simultaneous functions in discourse. First, the referent of the preposed constituent marks the BACKWARD LOOKING CENTER (BLC) of an utterance (cf. Grosz, Joshi, and Weinstein 1983). A BLC is a discourse entity which is related to the set of previously evoked discourse entities, i.e. the set of FORWARD LOOKING CENTERS (FLC), via a salient SCALAR relationship (cf. Hirschberg 1985). Second, preposing constructions are "presuppositional" (cf. Jackendoff 1972) in that they mark an OPEN PROPOSITION (OP) as salient in the discourse (cf. Prince 1981). A preposed sentence consists of two parts: the OP and the FOCUS. The OP contains one or more unbound variables which are instantiated with the "new information" or FOCUS (cf. Wilson and Sperber 1979) of the utterance. The FOCUS (or FOCI) of an OP is representable as a value on some SCALE, and is realized prosodically with an accented syllable. Based on whether or not the FOCUS is preposed, two types of preposing are distinguished: FOCUS PREPOSING and TOPIC. The BLC of FOCUS PREPOSING contains the FOCUS of the utterance, in which case it bears the single accented syllable of the utterance. TOPIC, on the other hand, involves a non-FOCUS BLC, and bears multiple accented syllables: one on or within the preposed constituent and one on the constituent containing the (non-preposed) FOCUS. A taxonomy of preposing is presented, based on the type of scalar relation that holds between the BLC and the FLC, the discourse status of the OP, and the semantic type of information which instantiates the variable of the OP. It is shown how preposing, together with intonation, specifically affects utterance interpretation.

### **Textual Analysis and the Assignment of Index Entries for Social Science and Humanities Monographs**

*Michael W. Grunberger*

Rutgers University, the State U. of New  
Jersey (New Brunswick) Ph.D. 1985,  
136 pages

Library Science

*University Microfilms International*

ADG85-20363. 8601

This dissertation examines, indirectly, the process by which an indexer selects an index entry by focusing upon the results of indexing: the index entry and its relation to indexed text. A sample of index entries from social science and humanities monographs is investigated in order to examine the assumptions which support the naturalist and the formalist theories of indexing. In an effort to discover relationships between the location and frequency of textual references and the generation of corresponding index entries, two related hypotheses are tested: (1) that term location within paragraphs will correlate with indexability, and (2) that term frequency on cited pages will correlate with indexability. Based upon the literature, it

was expected that there would be a direct relationship among index entries and term location and frequency.

The data, after analysis by descriptive and inferential methods, did not support either of the two hypotheses. The researcher concluded that: (1) Indexed terms are evenly distributed within the paragraph of social science and humanities monographs and do not significantly cluster in discrete segments of paragraphs, (2) Not-indexed terms are evenly distributed within the paragraph structure of social science and humanities monographs and do not cluster in a significant manner in a discrete segment of a paragraph, (3) There is a statistically significant, but trivial, difference between the location of indexed terms in posted locations and indexed terms in not-posted locations, (4) Indexed term frequencies on posted pages are concentrated in the low and medium frequency ranges, (5) Indexed term frequencies on non-posted pages are concentrated in the low frequency range, and (6) There is a statistically significant, but trivial, difference between the frequency of indexed terms on posted pages and their frequency on not-posted pages.

These conclusions suggest that the assumptions of the naturalist approach to indexing, i.e., that term location and frequency are related to indexability, may not apply to full-text index units. It is suggested that future research in indexing take into account the findings of computational linguistics, artificial intelligence, and cognitive theory in order to better understand the complex process by which a human indexer determines indexability.

### **A Formal Semantics for Some Discourse Anaphora**

*Jeffrey C. King*

University of California, San Diego, Ph.D.

1985, 194 pages

Philosophy

*University Microfilms International*

*ADG85-17906. 8512*

The dissertation is an attempt to provide a formal semantics for occurrences of (singular) anaphoric pronouns and definite descriptions whose quantifier antecedents occur in sentences other than those in which the anaphoric pronouns and descriptions themselves occur, (henceforth "q-terms"). The predominant view of anaphoric pronouns whose quantifier antecedents occur in the same sentence as they do is that they function as bound variables (though this view is subject to certain well-known difficulties). Chapter 1 of this dissertation is constituted by a series of arguments against a bound variable treatment of q-terms and the observation that the semantic behavior of q-terms is similar to that of certain singular terms in English arguments. Given the similarity of semantic behavior between the latter and q-terms, it seems plausible to suppose that a theory of the semantic behavior of these singular terms in English arguments would provide a model for the eventual production of a semantic theory of q-terms. In Chapter 2, a formal semantics for these singular terms in arguments is produced. In Chapter 3, a formal semantics for q-terms is produced, on the model of the semantics in Chapter 2. Finally, in Chapter 4 it is shown that the formal semantics of Chapter 3 can be extended to handle pronouns in discourses containing verbs of propositional attitudes such as "wants", "dreams" etc. It is also shown that some occurrences of sentences containing q-terms have truth conditions not expressible by any quantified first order sentence. One must use finite partially ordered quantifiers to express the truth conditions of such occurrences of sentences.

The dissertation contains two appendices: the first is a technical discussion of the formal semantics of Hans Kamp ("A Theory of Truth and Semantic Representation") which shows that his theory cannot accommodate the linguistic data mine is designed to handle. The second examines the views of Gareth Evans, Charles Chastain and Keith Donnellan on certain anaphoric pronouns.