

## ABSTRACTS OF CURRENT LITERATURE

### **Automatic Text Generation: Application to the French Stock Market Report Sublanguage**

*Chantal Contant*

Département de linguistique et philologie

Université de Montréal

Montréal, Québec, Canada H3C 3J7

*M.A. Thesis, December 1985, 154 pages*

The field of natural language text generation has evolved over the last few years to the point where automatic systems can now produce linguistically well-formed texts in a well-defined technical sublanguage.

The starting point for this M.A. thesis is a system that automatically generates English stock market reports from numerical data from the New York Stock Exchange. This earlier system, named ANA, was developed by Karen Kukich at the University of Pittsburgh. ANA is made up of four sequential modules. The second of these deals with the semantics of stock market reports (WHAT to say), and the fourth with their linguistic form (HOW to say it). The objective of the present research was to create a French linguistic module capable of being inserted into the existing ANA system to automatically produce French stock market reports. The resulting module accounts for the semantics, syntax, morphology, lexicology, and rhetoric of these texts.

The thesis describes the structure of the ANA, the original system, as well as the development and functioning of FRANA, the computer system to be inserted into ANA as a French linguistic module. The descriptions of ANA and FRANA follow two chapters devoted to aspects of the problem of automatic generation, and to a linguistic study of a corpus for the stock market sublanguage.

This thesis is in French; only the chapter on FRANA (Chapter Four) is available in English on request (60 pages), along with an appendix showing output samples.

### **Selected Dissertation Abstracts**

Compiled by:

Susanne M. Humphrey, National Library of Medicine, Bethesda, MD 20209

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 initial DAI subject category and second by author. Citations denoted by an MAI reference do not yet have abstracts in the database 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 by University Microfilms International), and may not be reproduced without their prior permission.

### **Improvement of Automatic Indexing through Recognition of Semantically Equivalent, Syntactically Different Phrases**

*Olorunfemi Stephen Aladesulu*

Document indexing is the intellectual process of identifying the major topics discussed in the documents of some particular domain to facilitate subsequent retrieval of documents dealing with a specific topic in that domain. By and large, this operation is still performed manually in most operational document retrieval systems, even though a considerable

The Ohio State University Ph.D. 1985,  
192 pages  
Computer Science  
*University Microfilms International*  
ADG85-26134

amount of research effort has been expended on ways of automating this activity.

This dissertation is concerned with methods of improving automatic indexing operations to make them more closely approach the intellectual level of human indexing. Towards this end, a semantic recognition algorithm (SRA) is described and implemented, which has merit for both the concept selection and concept representation processes of frequency-based automatic indexing.

Additionally, an evaluation method based on a comparative measure of confidence (CMC) function is also described. The CMC function provides a method of measuring the comparative effectiveness of different vocabulary control techniques on the frequency-based concept selection processes used in the research.

Specifically, the SRA provides for recognition of syntactic synonymy found in semantically equivalent phrases containing the same significant words, but in inverted order, with or without function words interposed (e.g., management system, system for management). The algorithm also makes use of stemming algorithms to recognize derivational and inflectional synonymy as well (e.g., management system, management systems).

As one would intuitively expect, comparative evaluations using the CMC on several frequency-based automatic indexing techniques show that, on the average, techniques using the SRA perform better than those using stemming algorithms alone, followed by techniques using no vocabulary control at all. Suggestions for further improvement of the SRA are also described.

#### **Pragmatic Modeling in Information System Interfaces**

*Mary Sandra Carberry*  
University of Delaware Ph.D. 1985,  
364 pages  
Computer Science  
*University Microfilms International*  
ADG85-25275

Natural language interfaces currently treat each query as an isolated request for information, with little use of the dialogue context within which the utterance occurs. This thesis investigates how natural language systems can assimilate an on-going dialogue and use the resulting knowledge to increase their robustness.

The thesis first presents a strategy for inferring a model of the task-related plan motivating an information-seeker's queries. Focusing heuristics are used to relate each new utterance to the existing plan context and construct an enlarged context model. It then develops a pragmatics-based approach to handling two classes of problematic utterances: utterances involving pragmatic overshoot and intersentential elliptical fragments. The framework for handling pragmatic overshoot rephrases the pragmatically ill-formed query based on the speaker's perceived intentions in making the utterance. The ellipsis interpretation strategy identifies the discourse goal the speaker is pursuing with the utterance and interprets it relative to the speaker's inferred task-related plan.

The results of this research indicate that natural language interfaces must place greater emphasis upon established dialogue context, both the inferred task-related plan and anticipated conversational goals, in understanding utterances.

#### **AUGUST: An Expert System for Translating a Conventional File System into an Entity-Relationship Model**

*Kathi Hogshead Davis*  
Illinois Institute of Technology Ph.D. 1985,  
291 pages  
Computer Science  
*University Microfilms International*  
ADG85-25931

The problem addressed by this dissertation is that of translating a conventional file system into a commercial database management system (DBMS). Currently, the translation process consists of starting from scratch with a new database conceptual model and then translating it into a physical DBMS.

The methodology proposed in this dissertation takes the record layouts (COBOL) of the conventional file system (CFS) and creates a current physical model (CPM). The CPM is then translated into a current logical model (CLM). The CLM represents the conceptual characteristics of the current environment using the Entity-Relationship (ER) model.

The CLM is then used to create a new logical environment (NLM). The NLM differs from the CLM in two ways: one, the problems of the CFS have been corrected; and, two, any new structures that are needed in the new environment have been added.

The NLM is used to create the physical description for the specific DBMS of the new environment. This is the new physical model (NPM).

The data models of AUGUST are defined in four areas: (1) the data structures, (2) the instances of the data structures, (3) the processing constraints under insertion and deletion, and (4) the navigational paths. As each model (CFS, CPM, etc.) is translated into the succeeding model, all four aspects are included in the process. The equivalence of the CFS, CPM, and CLM is also addressed, with the result that the models are indeed equivalent.

AUGUST is a prototype expert system for doing the conversion process described. The database designer interfaces with AUGUST to supply any information that cannot be obtained from the CFS's record layouts.

The first stage, CFS to CPM, of AUGUST called AUGCPM was implemented and tested on a CORONA microcomputer using micro PROLOG. The testing showed that the concepts behind AUGUST lend themselves very nicely to the expert system approach. The problems encountered were mostly due to the inefficiency of micro PROLOG and can be corrected by moving to a larger machine and the PROLOG language.

#### **A Foundational Approach to Conjecture and Knowledge in Knowledge Bases**

*James Patrick DelGrande*

University of Toronto (Canada) Ph.D. 1985,

1 page

Computer Science

*This item is not available from University*

*Microfilms International.*

*ADG05-56965*

A foundational investigation of the notion of hypothesis in knowledge representation schemes is presented. The problem has two distinct but related components. The first concerns the ultimately non-deductive problem of forming and maintaining a set of hypotheses, or a theory, based on a stream of ground atomic formulae. The second concerns deductively reasoning with a theory, together with arbitrary known and hypothesized sentences.

For the first part, theory formation, a language, HL (and from it an algebra and logic), is derived for forming hypotheses framed in set-theoretic terms. Two soundness and completeness results for the logic are presented. The first explicitly links the logic to the algebra; the second treats the logic as a three-valued system. Through the formal systems, the set of potential hypotheses is precisely specified, and a procedure is derived for restoring the consistency of a set of hypotheses after conflicting evidence is encountered. For the second part, reasoning with a theory, an existent first-order language that can represent and reason about what it knows is extended to one that can reason with knowledge and hypothesis. The original proof-theoretic and semantic results for the language are extended appropriately.

The relation between these two parts, for introducing and reasoning with hypotheses, is also explored. The theory formation process is considered as a source of hypothetical sentences for the deductive (reasoning) component, and the deductive component is considered as a source of a priori knowledge and hypothesis for the theory formation process.

#### **The Design of Graphical User Interface**

*Mark William Green*

University of Toronto (Canada) Ph.D. 1985,

1 page

Computer Science

*This item is not available from University*

*Microfilms International.*

*ADG05-56976*

A new approach to the design of graphical user interfaces is presented. This approach is based on a design methodology and a set of design tools supporting this methodology. The design methodology is motivated by research in both ergonomics and software engineering. It consists of seven orthogonal design tasks. Each of these design tasks has one well defined goal and a document that records the results of the task. The division of the design process into tasks allows the designer to concentrate on one aspect of the design at a time and gives the design manager the ability to establish milestones. A unique feature of this design methodology is its concentration on design evaluation and correctness.

A consistent set of tools has been developed to support this design methodology. The most important tools are a user modeling language for describing the user's view of the problem and the user interface, and a specification language for user interfaces. The user modeling language is based on the concepts of objects and operators. The specification language is based on state machines. The combination of these two languages covers all four levels in the language model of user interfaces proposed by Foley. Techniques have been developed for showing the compatibility of different descriptions of the user interface and for calculating important ergonomic measures given a description of the user interface.

**A Text Generation Module for a Decision Support System**

*Ping-Yang (Frank) Li*

Illinois Institute of Technology Ph.D. 1985,  
141 pages  
Computer Science  
*University Microfilms International*  
ADG85-25933

Our text generator is one module of a decision support system designed to assist physicians in the management of stroke, which is being developed jointly by the Neurology Department at Michael Reese Hospital and the Computer Science Department at Illinois Institute of Technology, using data from the Michael Reese Stroke Data Base. The text generator produces multi-paragraph reports on stroke cases stored in the Stroke Data Base or on cases being processed by the Decision Support System. Analysis of human-generated case reports using Sager's Linguistic String Parser led to a characterization of the stroke sublanguage in terms of four components: a Text Grammar for stroke case reports, a set of Stroke Information Formats, a Relational Lexicon for the stroke sublanguage, and a Linguistic String Grammar for this sublanguage. Our first reports were highly constrained; the program chooses between alternative formulas depending on the symptoms input by the physician and deductions from that information. We are now producing freer text by using reverse transformations from our LSP grammar to combine fragments into sentences. Further interest lies in discovering how to generate good paragraphs, using the Text Grammar, the Stroke Information Formats, the Relational Lexicon, and the Linguistic String Grammar as tools.

**Decision Processes in Understanding English Discourse Anaphora**

*Paul C. Lustgarten*

The University of Wisconsin - Madison Ph.D.  
1985, 187 pages  
Computer Science  
*University Microfilms International*  
ADG85-16775

This report presents a theory of the decision processes used to resolve natural language discourse anaphora. The anaphora problem is how to determine which of the entities mentioned in a discourse is the one meant by some anaphoric phrase.

Under the proposed theory, anaphora resolution has two stages: evaluation of the plausibility of each potential interpretation, followed by integration of these plausibilities into a composite rating. The potential interpretations are generated using a mental model of the discourse, where that discourse model contains an explicit representation of the items that have been introduced into the discourse. Furthermore, these interpretations are evaluated independently and in parallel. The second stage combines these plausibilities according to a relative judgment rule to obtain the relative plausibility of a given interpretation. The final interpretation of the anaphor is based directly on the relative plausibility. A distinctive feature of this theory is that the parameters vary over a continuous range of values, in contrast to the discrete, binary-valued parameters typically found in propositional information processing models.

Three experiments were performed to test the theory. In two, subjects were shown two-sentence texts involving two objects and a (more-or-less ambiguous) anaphoric reference to one of the objects. In each trial, they used a continuous scale to indicate their judgment of the relative plausibility of the two interpretations of the anaphor. The third experiment was similar in form, but tested the application of the model to the complementary task of deciding whether an indefinite noun phrase introduced a new entity into the discourse. The proposed model performed very well in the first two experiments, but only moderately well in the third.

The present theory is significant in several respects. First, it contributes to the development of a psychologically real and computationally viable model of discourse anaphora resolution. It also presents a significant decomposition of the problem, by identifying the evaluation of candidate interpretations as independent components, integrated according to a known rule. Finally, this study extends to "higher-level" language understanding processes a general decision model already successfully applied to several other aspects of language processing.

**An Empirical Study of Robust Natural Language Processing**

*Amir M. Razi*

University of Delaware Ph.D. 1985,

249 pages

Computer Science

*University Microfilms International*

ADG85-25293

Case studies in the literature have shown that as much as 25% of queries to question-answering systems are ill-formed in the sense that they violate the constraints of the grammar.

If natural language processing systems are ever to achieve natural and cooperative behavior, they must be able to process input that is ill-formed lexically, syntactically, semantically, or pragmatically. Systems must be able to partially understand, or at least give specific appropriate error messages, when input does not correspond to their model of language.

The approach of Sondheimer and Weischedel is to relate the processing of ill-formed input to the rules of well-formed input processing via meta-rules. The left-hand-side of a meta-rule diagnoses a problem with the input that violates a rule of the grammar. The right-hand-side rewrites the violated rule by relaxing the violated constraint. Normal processing can be resumed when some constraint is relaxed. Part of this thesis is the implementation of this metaprocessing in both an ATN interpreter and an ATN compiler.

Since processing of ill-formed input requires relaxation of the very constraints that limit the search for interpretations, one faces a potentially large number of partial interpretations to be examined. In order to trim the search space, a number of heuristics are suggested for localizing the problem. In order to determine the trade-offs between accuracy of the heuristics and added computation time to obtain that accuracy, an experiment is designed. The parameters used for this study include, among others, the amount of time required to localize the problem, the number of alternatives considered, and the percent of time that a heuristic might miss the correct cause of the problem. The experiment also suggests an ordering of the meta-rules and tuning up of some meta-rules to enhance their accuracy in diagnosing the failed constraints.

The structure of the grammar on which the experiment is run makes the implementation of some heuristics very difficult. It is my conclusion that a grammar with an underlying context-free structure is more suitable for ill-formed input processing.

**A Computer Model of Case-Based Reasoning in Problem Solving: An Investigation in the Domain of Dispute Mediation**

*Robert Lee Simpson, Jr.*

Georgia Institute of Technology Ph.D. 1985,

393 pages

Computer Science

*University Microfilms International*

ADG85-25622

Rather than approach each problem as a unique event, people often try to solve problems by recalling similar previous experiences as guides to problem solving. This analogical process, which we call case-based reasoning, seems to provide an explanation for the change in problem solving behavior of people over time. This research presents a computer process model of problem solving based on the use of case-based reasoning. The necessary reasoning processes, operational measures of similarity, and memory structures needed for effective storage and retrieval are presented via the specifications for an advisory system called the MEDIATOR, which offers advice on resolving common sense disputes. In this context, issues associated with enabling machines to dynamically adapt their reasoning and automatically recover from failure are discussed. The model of case-based problem solving which has been developed seems to offer promise as an integrated solution for some issues in problem solving, analogical reasoning, and machine learning.

### Determination of the Formal Vocabulary of Physicians through Analysis of Medical Literature

*Jack Edward Peterson*

Georgia State University - College of Education Ph.D. 1984, 273 pages  
Health Sciences, Education. Language, Linguistics. Education, Higher  
*University Microfilms International*  
ADG85-25638

**Purpose.** The purpose of this study was to determine the formal medical vocabulary in use by physicians and whether it changed significantly over time. A secondary purpose was to determine the constituent morphemes of the terminology and whether they changed significantly over time.

**Research Methodology.** A 100,000-word sample of the medical terminology contained in ten 1982 issues of the *Journal of the American Medical Association (JAMA)* was analyzed, resulting in a list of medical terms used, ordered from the most-frequently used to the least-frequently used. Each term was then reduced to its constituent morphemes, which were then sorted in order of decreasing frequency.

This process was then replicated for a comparable number of issues of *JAMA* for ten years previously (1972).

**Results.** The primary result of this research is a frequency-sorted listing of the medical vocabulary used by physicians in ten issues of the 1982 *JAMA*. A secondary result was a similar listing of terms from the 1972 *JAMA*.

The morphemes constituting these terms were analyzed similarly, resulting in two listings of morphemes sorted in decreasing frequency.

No significant differences were noted between the two sets of chronologically displaced samples, either at the term or morpheme levels. A correlation coefficient was computed for each of the two comparisons. It was greater than 0.9 in both cases, indicating little change in either terms or morphemes over the ten-year period.

**Implications.** The word- and morpheme-frequency lists resulting from this research may be helpful in designing the content of medical terminology texts. Terms or morphemes used most frequently may be taught first, and level of instruction may be adjusted by altering the cutoff point to suit the objectives of the text.

In addition, the techniques developed during this research may be helpful in analysis of the terminology of other disciplines for educational purposes.

### A Compositional Semantics for Aktionsarten and NP reference in English

*Erhard Walter Hinrichs*

The Ohio State University Ph.D. 1985,  
345 pages  
Language, Linguistics  
*University Microfilms International*  
ADG85-26186

Written in the framework of Montague Grammar, this dissertation offers a model-theoretic semantics for Aktionsarten in English that formally explicates the structural analogies between the count/mass distinction and the distinction between telic and atelic events. The semantics of NP reference incorporates the lattice-theoretic approach to mass terms developed by Link (1983) and the ontology of kinds, objects and stages proposed by Carlson (1977). A parallel ontology of event types, individual events, and event stages is developed for the domain of Aktionsarten. The two ontologies are connected by taking event stages and stages of individuals to be spatio-temporal locations, which we identify as the realizations (in the sense of Carlson 1977) of objects, kinds, event type, and individual events. The set of locations (used in the sense of Barwise & Perry 1983) is defined as a complete join-semilattice, which makes it possible to generalize Link's account of the homogeneous reference property of mass terms to the semantics of atelic events. The semantic treatment for mass terms and bare plurals is based on the distinction between singular and plural individuals, which (following Link) leads us to adopt the structure of a complete atomic Boolean algebra for the domain of objects.

The distinction between temporally homogeneous and heterogeneous predicates provides the basis for a compositional semantics of the aspectual classes of states, activities, accomplishments and achievements recognized by Vendler (1967). Since the treatment of Aktionsarten in the fragment of English included in the dissertation is fully compositional, the semantic interaction between homogeneous/heterogeneous predicates and homogeneous/heterogeneous noun phrase arguments can be properly accounted for. The fragment provides a new approach to the semantics of

**Negation in English: An Essay in Game-Theoretical Semantics***Michael Robert Hand*

The Florida State University Ph.D. 1985,  
170 pages  
Philosophy  
*University Microfilms International*  
ADG85-24606

**Learning by Reasoning from Multiple Analogies***Mark Howard Burstein*

Yale University Ph.D. 1985,  
225 pages  
Computer Science  
*University Microfilms International*  
ADG86-00973

**A Model and Notation for Specifying User Interfaces***Uli Han-Hsiang Chi*

University of Washington Ph.D. 1985,  
231 pages  
Computer Science

motion verbs and their directional and temporal modifiers. The distinction between object accomplishments such as *build a house* and event accomplishments such as *play a sonata*, which was first observed by Dowty (1979), is formally treated and defended as an important subclassification of telic events.

This essay treats semantical negation from the standpoint of game-theoretical semantics. It consists of an exposition and critique of the one previous treatment of negation from this viewpoint, a formulation of negation-formation rules with special attention to the interaction of quantifier scopes and the scope of negation, and applications of this treatment to a number of problems: negation of sentences containing definite descriptions, sentences having contextually determined restrictions on quantifier domains, reciprocal quantification, negative polarity, negative raising, term negation of quantifier phrases, no, the noneffectiveness of negation in general, quantificational only, and tag questions.

When students are first being introduced to a new subject by a teacher or textbook, basic concepts are often illustrated by analogies to things that they are more familiar with. Although this is apparently a very powerful form of instruction, the method by which students use these analogies when learning has not been extensively studied by the artificial intelligence and psychology communities.

This thesis presents a cognitive process model of learning from analogies suggested by a tutor or textbook. The model was based on examples from recorded protocols of several students who were tutored on the programming language BASIC. The protocols revealed several important attributes of analogically-based learning. First, the analogies suggested by the tutor were used primarily to generate hypothetical explanations of observed interactions with the computer, and to form expectations about possible computer responses to newly issued commands. When additional examples were given, these students incrementally extended their analogical models to account for the new situations. This behavior is modeled by a process that finds and maps stereotypical causal representations from a familiar domain, given an analogy to that domain.

The protocols also showed analogical hypothesis formation to be an inherently error-prone process. The conceptual models these students developed were repeatedly tested and corrected. Subjects sometimes used alternate analogies to give "second-guess" answers when they encountered a problem or were told they were wrong. Their ability to use and integrate information from several analogies when learning helped them both because it enabled them to generate alternative hypotheses, and because many situations cannot be completely characterized by any single analogy.

To illustrate the theory of analogical learning developed from these observations, a computer program, CARL, is presented that learns to use variables in BASIC assignment statements. While learning about variables, CARL generates many of the same erroneous hypotheses seen in the recorded protocols. CARL exhibits this behavior by incrementally mapping stored representations of causal and planning knowledge about "familiar" domains presented in analogies. CARL's learning process produces a single target model with some aspects preserved from each of several analogies.

A new model and notation for describing and specifying user interfaces is proposed and evaluated. The model presents an object-oriented view of user interfaces that explicitly provides for concurrent interaction with multiple users and/or applications. The notation is developed based on the dual approaches of algebraic specifications (for describing the objects of the interface) and flow expressions (for describing interaction between

*University Microfilms International*  
ADG85-29893

**Intention-Based Diagnosis of Errors in Novice Programs**

*William Lewis Johnson*

Yale University Ph.D. 1985,

304 pages

Computer Science

*University Microfilms International*

ADG86-01069

interface objects, users, and applications). Using the notation, a range of generic user interface abstractions are specified, including logical input and output devices, editing-based objects, and time-constrained interaction. Combining these notions, three existing commercial user interfaces that involve text, graphical animation, and concurrency are specified, demonstrating the practical utility of the model and notation. Finally, the notation is applied as a design tool and as a basis for rapid prototyping of a new user interface design for a simplified multi-user computer conferencing system.

This thesis investigates the process whereby faults are diagnosed and corrected in human-designed artifacts in general, and in computer programs in particular. Established automatic diagnostic techniques are useful in cases where the artifact design is known to be correct, but they are inadequate when the design itself may be faulty. Instead, it is necessary for the diagnostician to identify the intentions underlying the design of the artifact and reason about these intentions in order to identify and correct faults. In other words, the diagnostician must understand the artifact in order to correct it. Such understanding is doubly necessary if the designer is a student, and the diagnostician is a teacher who is trying to find out why the student is having difficulties. Intention-based error diagnosis has been implemented in a program called PROUST, which identifies non-syntactic bugs in programs written by novice Pascal programmers. Empirical studies of PROUST's performance show that it achieves high performance in finding bugs in non-trivial student programs.

**A Self-Organizing Retrieval System for Graphs**

*Robert Arlen Levinson*

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

101 pages

Computer Science

*University Microfilms International*

ADG85-27601

The theory, design, and implementation of a graph-based, self-organizing database retrieval system. The system is designed to support the expert problem solving tasks of recall, design, and discovery. The fundamental design principle is the production of a partial ordering by the relation subgraph-of. This relation is considered to be equivalent to more-general-than. This document discusses this design from three different levels: an abstract level in which the nodes in the partial ordering are concepts, the implementation level described above (the nodes are graphs), or an application level in which the nodes are domain specific objects such as molecules or reactions.

The primary problem domain explored is organic chemistry. A large data base of organic reactions and starting materials can be queried to extract reactions or molecules that match, either exactly or approximately, desired structures. The system may also suggest precursors to a desired target molecule. The queries are answered by exploiting a set of concepts that are common subgraphs of molecule or reaction graphs. Concepts serve multiple purposes: They constrain the search involved in the matching process so that the time required to answer a query grows sub-linearly in the size of the data base. Concepts define the notion of "similarity" that is crucial if approximate match is desired. They also may be useful generalizations of reactions or molecular structures. The concepts can be "discovered" (i.e., constructed) by the system itself using largely syntactic criteria based on the topology of the database. A variety of performance tests are performed, including a comparison of the system's precursor recommendation capability with graduate students in organic chemistry.

The system is also applied to the retrieval and generalization of chess positions.

**Editing as a Paradigm for User Interaction**

*Jeffrey Alan Scofield*

University of Washington Ph.D. 1985,

173 pages

The most difficult part of the programming task is often the creation of the component for interacting with the person who will use the program. This component is called the user interface. This dissertation describes the design of a framework for high-quality, highly-interactive interfaces for all

Computer Science  
*University Microfilms International*  
 ADG85-29946

parts of a system. The uniformity, extensibility, and flexibility of this framework guarantee the existence of nearly identical user interfaces for all objects, automatically provide a means of interacting with new objects, and allow new interfaces to be built easily, when necessary, from a set of standard components.

The user interfaces of this system are based on the natural and responsive principles of the screen-oriented editor, generalized to allow interactions with objects of all types. Editors use structural information to control the editing process, which allows them to guarantee that edited objects are well formed, while assisting the user with structural details.

The editors are designed as interfaces to an object-oriented system. Furthermore, editors themselves are built using a set of cooperating objects and types (corresponding to Smalltalk classes). In particular, the syntactic descriptions of valid object structures are implemented as direct extensions to the type system.

The dissertation includes a detailed design and examples of interaction with many types of objects. The design has been tested by means of a prototype implementation in Lisp, running on a standard timesharing system, from which the examples are drawn. The dissertation concludes with a discussion of the experience derived from this implementation, followed by suggestions for improvements and further research.

**Evidential Reasoning in Semantic Networks: A Formal Theory and its Parallel Implementation**

*Lokendra Shastri*

The University of Rochester Ph.D. 1985,  
 258 pages  
 Computer Science  
*University Microfilms International*  
 ADG85-28562

The problem of representing and utilizing a large body of knowledge is fundamental to artificial intelligence. This thesis focuses on two important issues related to this problem. (1) An agent cannot maintain complete knowledge about any but the most trivial environment, and therefore, he must be capable of reasoning with incomplete and uncertain information. (2) An agent must act in real-time. Human agents take a few hundred milliseconds to perform a broad range of intelligent tasks, and agents endowed with artificial intelligence should perform similar tasks in comparable time.

It is argued that the best way to cope with partial and incomplete information is to adopt an evidential form of reasoning wherein inference does not involve establishing the truth of a proposition but instead involves finding the most likely hypothesis from among a set of alternatives.

It is also argued that in order to satisfy the real-time constraint, we must identify the kinds of inference that need to be performed very fast, and provide a computational account of how this limited class of inference may be performed in an acceptable time frame. This latter requirement prompts us to consider massively parallel models of computation, in particular models that do not require an interpreter.

Inheritance and categorization within a conceptual hierarchy are identified as two operations that humans perform very fast. It is suggested that these operations are important because they seem to lie at the core of intelligent behavior and are precursors to more complex reasoning.

The above concerns and proposed solutions lead to an evidential framework for representing conceptual knowledge, wherein the principle of maximum entropy is applied to deal with uncertainty and incompleteness. It is demonstrated that the proposed framework offers a uniform treatment of inheritance and categorization, and solves an interesting class of inheritance and categorization problems, including those that involve exceptions, multiple hierarchies, and conflicting information. The proposed framework can be encoded as an interpreter-free, massively parallel (connectionist) network, that can solve the inheritance and categorization problems in time proportional to the depth of the conceptual hierarchy.

**An Extension of a First-Order Language and its Applications**

In the currently known many-sorted logic, a variable ranging over a sort that does not exist in the sort structure cannot be introduced. To avoid

**Dong-Guk Shin**

The University of Michigan Ph.D. 1985,  
244 pages  
Computer Science  
*University Microfilms International*  
ADG86-00552

this problem, extended calculi are proposed that are obtained by embedding a new kind of syntactic object called aggregate variables in a one-sorted language and in a many-sorted language. The resulting languages are called, respectively, a one-sorted language with aggregate variables, denoted by  $L(\sigma)^1$  and a many-sorted language with aggregate variables, denoted by  $L(\sigma)$ . Theoretic foundation for the extended predicate calculi is provided, and their practical usage in real applications is demonstrated:  $L(\sigma)$ , in the distributed database design area and  $L(\sigma)^1$ , in the automatic theorem proving area.

As an application in the distributed design area,  $L(\sigma)$  is used for the representation of the user queries and the knowledge about data. A knowledge-based approach is proposed with which the user reference clusters (URCs) to a data base are estimated. The URCs can then be used in partitioning a relational data base horizontally during distributed database design. Using the knowledge about the data, the user queries are converted to equivalent queries by a proposed inference procedure. The URCs that were estimated from these revised queries are more precise than those that can be estimated from the original user queries. The types of knowledge to be used are discussed. An example illustrates the way inference is carried out, and the correctness of the inference is also discussed.

As an application of  $L(\sigma)^1$  in the automatic theorem proving area, a type of problem is first identified which may occur when a resolution scheme is applied to many-sorted theory. It is shown that any many-sorted theory can be converted into an equivalent theory in  $L(\sigma)^1$ . Aggregate variables allow the introduction of range-restricted variables dynamically in the structure which is expanded by definitions. This allows the introduction of a new resolution scheme named Unification over the Weakest Range (or UWR-resolution). The completeness of UWR-resolution is shown and the efficiency of UWR-resolution is discussed.

**A Systems Model of Cognition for Improving Human Factors of Computing Environments**

**James Loftin Snell**

State University of New York at Binghamton  
Ph.D. 1986, 225 pages  
Computer Science  
*University Microfilms International*  
ADG85-27918

The literature on computer-human interaction is strong on empirical studies of particular problems and phenomena, and on practical design guidelines, but offers little in the way of overall theoretical framework for either. Meanwhile, in cognitive psychology and artificial intelligence, various theories of human cognition are currently being debated, but have not been systematically related to issues of computer-human interaction. In this dissertation, a systems model of human cognition is developed that encompasses the relevant areas, based partly on models of Norman, Minsky, and Lowen. On the basis of this cognitive system model, a model of computer-human interaction is then developed, and from it are derived design principles with direct application to improving human factors of computing environments.

**First-Order Unification in Equational Theories and its Application to Logic Programming**

**Jia-Huai You**

The University of Utah Ph.D. 1985,  
161 pages  
Computer Science  
*University Microfilms International*  
ADG85-27948

As an effective mechanism primarily used in theorem-proving, unification has been extensively studied by several researchers since the invention of the resolution principle. This thesis studies first-order unification in equational theories, called E-unification, paying particular attention to complete unification algorithms for classes of equational theories. It also investigates how results and notions of E-unification can be applied to logic programming systems that support equality handling and functional notation.

First-order unification in equational theories is proved to be complete, i.e., it is shown that there exists an E-unification algorithm which is complete for all equational theories. This result is primarily of theoretical interest, since such an algorithm tends to be too inefficient to be of practical use. An E-unification algorithm using the narrowing process is then shown to be complete for the class of theories that can be described by a

closed linear term rewriting system with the nonrepetition property. This result sets the stage for investigation of applications to logic programming.

An extended equational programming paradigm, referred to as equational logic programming, is proposed, which uses narrowing and deletion upon successful unification as the inference rules, and enjoys the semantic simplicity of the classical theory of equality. It is shown that the kind of programming features that were initially possessed solely by Prolog can also be provided in this extended equational programming paradigm.

Finally, a computational model integrating functional programming and logic programming is described. An important notion in the model is called semantic unification, which is a special case of E-unification for ground terms. The model supports computations with infinite data structures, and provides the user the flexibility of choosing between a backtracking free computation framework or a conventional logic computation framework, that is, a nondeterministic one involving backtracking. The model can be extended to include the notion of equality when complete E-unification algorithms are used. An interpreter for the proposed computational model is presented, written in DEC-20 Prolog.

### **Text Relations and Prose Comprehension**

*Connie Kendall Varnhagen*

University of California, Santa Barbara Ph.D.

1985, 194 pages

Education, Psychology

*University Microfilms International*

ADG85-28619

This thesis compares the ways in which three prose analysis systems describe the underlying relations in a text and predict recall of expository prose. Each of these approaches is based on different predictions about how a particular text is initially processed, represented in memory, and later recalled. The systems contrasted were Meyer's (1975) content structure analysis, Kintsch's (Kintsch & van Dijk, 1978) argument repetition analysis, and a taxonomy of logical relations developed in this thesis. The classification system for the logical relations is based on philosophical definitions of causation, the logical relations connecting information in the text in some meaningful way, associating similar information, identifying causes or precursor conditions and events, resultant solutions, etc.

Seventh- and ninth-grade students and adults read two expository passages and wrote their recall immediately following reading and after a one-week delay. Although each of the three prose analysis systems was independently successful in terms of adequately and durably predicting recall, they did not combine equally to predict recall of the two texts by the different age groups. In comparison with Meyer's and Kintsch's systems of prose analysis, the system of logical relations provided a more adequate account of adults' recall and, in the absence of explicit, top-level, globally cohesive information, the logical relations system provides a better representation of adolescents' recall of expository prose. Theoretical and practical implications of the logical relations system of prose analysis are discussed.

### **Statistical Models of Text: A System Theory Approach**

*Ye-Sho Chen*

Purdue University Ph.D. 1985, 198 pages

Engineering, Industrial

*University Microfilms International*

ADG85-29267

A study is made of statistical models of text, including the laws proposed by Lotka, Bradford, and Zipf, and the models proposed by Markov, Mandelbrot, and Simon-Yule. A system theory approach is developed and applied first to show the equivalence of the three laws; secondly, to propose a multivariate representation of text which exhibits three important empirical properties: marginal skewness, type-token relationship, and exponential gaps; and thirdly, to compare four leading models with respect to the multivariate representation and select the Simon-Yule model as an appropriate statistical model of text. A modification of the selected model, which gives better performance than the original one, is discussed. Further modification relating the Simon-Yule model to computational models of text generation in artificial intelligence is suggested for future research.

### **A Theory of Constituent Structure and Constituent Utterances**

Participants in a discourse regularly use utterances that are smaller than grammatically complete sentences, and researchers in generative grammar

**Ellen Lynn Barton**

Northwestern University Ph.D. 1985,  
394 pages  
Language, Linguistics  
*University Microfilms International*  
ADG86-00849

have suggested that deletion rules account for the structures that underlie these fragment utterances. In contrast to a deletion analysis, I propose that some fragments are actually constituent structures, which derive from major categories as initial nodes. I present a theory of constituent structures and constituent utterances by developing a competence model of the grammar of constituent structures and a pragmatic model of the interpretation and acceptability or unacceptability of constituent utterances in context.

I motivate a constituent structure analysis by demonstrating the empirical inadequacy of any deletion analysis. I then develop a competence model of constituent structures, which describes well-formed constituent structures at each of the levels of representation in a Revised Extended Standard Theory grammar. My central claim in terms of Government-Binding theory is that S and X''' are contrasting initial nodes: S as an initial node generates a sentence structure; X''' as an initial node generates a constituent structure.

I also develop a pragmatic model of constituent utterances, which describes the interpretation and the acceptability or unacceptability of constituent utterances in context. The pragmatic model consists of two levels of representation: the representation of linguistic context and the representation of conversational context. Each level has principles that describe the structure of context; each principle has associated rules that describe the operation of inference that fills in the structure of the context. Each principle also has associated conditions of acceptability and unacceptability. The output of the pragmatic model of constituent utterances is a representation of meaning and acceptability in context.

**Germar Verbal Prefixes and Modern  
Generative Theories of Word Structure**

**Karl Steik Brehmer**

Princeton University Ph.D. 1985, 214 pages  
Language, Linguistics  
*University Microfilms International*  
ADG85-29030

The title of this dissertation contains two separate elements – German prefixes and morphological theory. Although German prefixes provide the main source of data throughout, it is not so much about German prefixes as it is about morphological theory in light of them. The thesis is organized as follows.

In chapter 1 several recently proposed theories of morphology are discussed. Out of this discussion arises a general framework, which forms the basis for the analyses of the ensuing chapters. It is argued that the morphology is an interpretive rather than a generative component, and that it is highly restricted in its scope. Of great importance is the distinction between word structure and word formation. Only the former is taken to fall within the scope of the grammar.

Chapter 2 deals with non-affixational derivation – specifically, morphological conversion. A derivational analysis is rejected in favor of a redundancy analysis. It is seen that the notion derivation, taken in the usual sense of the word, is not relevant for the grammatical morphology. It must be replaced by the concept of relatedness. Arguments based on the evidence of the behavior of German prefixes in connection with morphological conversion are presented which require that the morphology be viewed as an interpretive component.

The internal structure of German words containing verbal prefixes is the topic of chapter 3. Discussed are: the status of the so-called separable prefixes; the need for tertiary-branching structures; prefixes as head affixes, and prefixes as non-head affixes. It is shown, among other things, that the right-hand head rule must be rejected, and that the notion affix is significant for grammatical theory.

The fourth and final chapter deals with questions of argument structure. A recently proposed theory which virtually eliminates these matters from the derivational morphology is rejected. Specifically, the assignment of case in complex (prefixed) forms is shown to require morphologically stip-

**The Linguistics and Pedagogy of Complex Nominals and the Complex Noun Phrase in Computer Manuals: An EST Perspective**

*Luis Guillermo LaTorre*

Purdue University Ph.D. 1985, 213 pages

Language, Linguistics

*University Microfilms International*

ADG85-29298

ulated argument frames. Further, questions concerning the inheritance of argument structure in complex verbs are discussed.

This dissertation analyzes the most relevant properties of complex nominals (CNs) from the points of view of three descriptive and theoretical models. The rationale is that these syntactic structures require attention because their productivity renders them invaluable in designating new concepts and processes in the rapidly expanding field of computer technology. At the same time, they are so peculiar to English that a description of their properties may prepare the ground for pedagogic approaches that help reduce the time non-English-speaking technologists need to become proficient in reading technical texts in English.

The Introduction establishes the limits of the dissertation by relating it to the notions of descriptive grammar (linguistics), register (pragmatics), and ESP/EST (pedagogy).

Chapter I explores the issue of research in ESP/EST and of how new teaching approaches may have led to a neglect of the properties of technical English. Data is provided on how this neglect has affected research into specialized lexis, particularly that of computer technology. A case is made for the importance of CNs in technical vocabulary.

Chapter II reviews the main approaches to CN description: traditional, structuralist, and generative. Important concepts in each school are selected and applied to CNs from a corpus obtained from computer manuals. A few gaps in current descriptions and theories are tentatively suggested.

Chapter III concentrates on the surface properties of CNs from the corpus, with indications about the most frequently-occurring types. Further resort is made to generative processes. The chapter closes with a proposed pedagogic model of the complex noun phrase in computer texts, to be contrasted with a generalized model derived from current surface descriptions.

Chapter IV takes a lead from Noam Chomsky in focusing on the EST learner facing a specialized text. In spite of the complexity of technical vocabulary in English, it is suggested that technical training, even in the mother tongue, produces in EST learners a readiness to decode specialized texts in English. Another model is presented that describes the various factors that may be part of such readiness and that may facilitate decoding. Teachers are urged to become aware of such factors.

**Pronouns and Prepositional Phrases**

*Greer Watson*

Yale University Ph.D. 1985, 526 pages

Language, Linguistics

*University Microfilms International*

ADG86-00998

Reinhart's (1976) widely accepted rule for pronoun interpretation permits coreference freely in either direction provided the pronoun does not c-command its antecedent (determined by its bounding node in S-structure): the prepositional phrase (PP) is specifically included as a bounding node. In this thesis I present theoretical arguments and experimental data to support my view that coreference is not determined simply on S-structure, that the PP cannot always be a bounding node, and that linear order is sometimes relevant. Methodologically, a contribution is made by developing a new experimental test that, through the effect on coreference of variation in the discourse context, distinguishes sentences where coreference is ungrammatical from sentences where it is grammatical but disfavoured.

I demonstrate that constraints on the possibility of coreference of reflexive and non-reflexive pronouns are not applied on S-structure: the reflexive rule must apply before either PP-preposing or heavy-NP shift, while non-reflexives are interpreted on a structure after sentential PP-preposing but before verb-phrasal PP-preposing or heavy-NP shift. Rules preposing sentential and verb-phrasal PPs are thus distinguished not only structurally but in rule order.

Most sentential PPs are bounding nodes, but some verb-phrasal PPs are certainly non-bounding – variation in part lexically determined. I suggest, however, that verb-phrasal PPs may be divided into two types: a group of “core” PPs, possibly the same as subcategorized PPs, and additional PPs optionally generated after these. I demonstrate that within the “core” only the last PP can be a bounding node.

I also investigate factors affecting the variability of the probability of coreference (as determined experimentally); these include the subordination of the pronoun or its antecedent, and the direction of coreference in the surface structure of the sentence.

Word order does, therefore, play a crucial indirect role in the determination of the possibility of coreference, and a direct role in determining its probability.

**Foundations for a Method for Knowledge Analysis**

*Peter Joseph Lazzara*

The University of Tennessee Ph.D. 1985,  
182 pages

Psychology, Experimental

*University Microfilms International*

*ADG86-00039*

This thesis develops and justifies a method for acquiring expert knowledge and representing this knowledge as sets of practical arguments. The methodology involves an iterative analysis, wherein prior interview analysis cycles provide data for and guide subsequent interview analysis cycles. Argument structures derived from self report protocols are analyzed and hidden premises of those arguments are derived from those argument analyses. Subsequent analyses are structured by treating the hidden premises of the practical arguments as hypotheses to be confirmed. Over the course of repeated cycles the knowledge base and rule structure expands thereby increasing the constraint on the interpretation of data from subsequent interview analysis cycles.

Data was collected in two interviews with a clinical psychologist. The two interviews were subjected to a complete analysis. The second interview was structured by hypotheses formulated on the basis of data derived from the first interview. A selection of argument structures, complete with hidden premises, were then represented in the formalism of the predicate calculus.

Upon analysis of the two interview transcripts and the rule structures generated by the application of the methodology to them it has been shown that the method successfully locates hidden premises in such a way that these premises can be used to generate, in a subsequent interview, a more detailed and constrained representation of the rule structures of an expert.