

The 1981 ACM Annual Conference will be held at the Bonaventure Hotel in Los Angeles, California, November 9-11, 1981. [See *AJCL* 6:3-4, pg. 194.] For further information contact:

Mrs. A.C. Toni Shelter
Xerox Corporation, A3-49
701 South Aviation Boulevard
El Segundo, California 90245
(213) 679-4511 x1968

The ACL is sponsoring three sessions on "Computer Modeling of Linguistic Theory" in conjunction with the *Annual Meeting of the Linguistic Society of America* which will be held in New York City at the Grand Hyatt Hotel, December 28-30, 1981. [See *AJCL* 7:1, pg. 49.] For further information contact:

Stan Petrick
IBM T.J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598

or: Terry Langendoen
CUNY Graduate Center
33 West 42nd Street
New York, New York 10036

Evaluation of Natural Language Processors

Harry R. Tennant

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Technical Report T-103, Nov. 1980, 247 pages.

Despite a large amount of research on developing natural language understanding problems, little work has been done on evaluating their performance or potential. The evaluations that have been done have been unsystematic and incomplete. This has led to uncertainty and confusion over the accomplishments of natural language processing research.

The lack of evaluation can be primarily attributed to the difficulty of the problem. The desired behavior of natural language processors has not been clearly specified. Partial progress toward the eventual goals for natural language processors has not been delineated, much less measured.

This thesis attempts to clarify some of the difficulties behind evaluating the performance of natural language processors. It also proposes an evaluation method that is designed to be systematic and thorough. The method relies on considering a natural language processor from three viewpoints in the light of several taxonomies of issues relevant to natural language processing. Finally, an evaluation is described of PLANES, a natural language database query system.

Abstracts of Current Literature

The State-of-the-Art in Natural Language Understanding

David L. Waltz

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-27, Dec. 1980, 35 pages.

Research in computer understanding of natural language has led to the construction of programs which can handle a number of different types of language, including questions about the contents of data bases, stories and news articles, dialogues, and scene descriptions. This research draws on and has in turn had an affect on many other research areas, including software engineering, linguistics, psychology, philosophy, and knowledge representation. This paper provides a brief history and overview of the field, along with examples and explanations of the operation of several natural language understanding programs. The limitations of our current technology are discussed, and assessments are given of the most promising current research directions.

A Program Conversing In Portuguese Providing a Library Service

Helder Coelho

Centro de Informatica
Laboratorio Nacional de Engenharia Civil
101, Av. do Brasil
1799 Lisboa Codex, PORTUGAL

Edinburgh Ph.D. Thesis, Dec. 1979.

TUGA is a program which converses in Portuguese to provide a library service covering the field of Artificial Intelligence. The objective of designing TUGA was the development of a feasible method for consulting and creating data bases in natural Portuguese. The resulting program allows dialogues where the program and its users behave in the way humans normally do in a dialogue setting. The program can answer, and question, in pre-defined scenarios. Users can question, answer and issue commands in a natural and convenient way, without bothering excessively with the form of the dialogues and sentences.

The original contributions of this work are: the treatment of dialogues, the adaptation of Colmerauer's natural language framework to Portuguese, the particular method for evaluating the logical structures involved in Colmerauer's framework, and the library service application itself. The program is implemented

in PROLOG, a simple and powerful programming language essentially identical in syntax and semantics to a subset of predicate calculus in clausal form.

Semantic Grammar and Meaning Representation Language in a Natural Language Question and Answer System

C. Rathke, B. Sonntag, and W. Schopper

Institut für Informatik
Universität Stuttgart
D-7000 Stuttgart 1, WEST GERMANY

Angew. Inf. 22, 4 (April 1980), 155-157 (In German).

We describe a natural language question-answering system. Questions can be put to the system in written natural language about the world of German soccer teams and the people employed by them. A so-called semantic grammar is introduced which is especially designed to cover a large set of possible questions about the world mentioned. An expression of a formal Meaning Representation Language (MRL) is produced by the system. Applied to the database, which is organized in a semantic net, it generates the appropriate answer.

A D-LADDER User's Guide

Daniel Sagalowicz

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Technical Note 224, Aug. 1980.

D-LADDER (DIAMOND-based language Access to Distributed Data with Error Recovery) is a computer system designed to provide answers to questions posed at the terminal in a subset of natural language information. The system accepts natural-language questions about the data. For each question D-LADDER plans a sequence of appropriate queries to the data base management system, determines on which machines the queries are to be processed, establishes links to those machines over the ARPANET, monitors the processing of the queries and recovers from certain errors in execution, and prepares a relevant answer to the original text.

This user's guide is intended for the person who knows how to log in to the host operating system, as well as how to enter and edit a line of text. It does not explain how D-LADDER works, but rather how to use it on a demonstration basis.

Interpreting Discourse: Coherence and the Analysis of Ethnographic Interviews

Michael Agar and Jerry Hobbs

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Technical Note 225, Aug. 1980.

Practitioners of ethnography, in seeking to discover and describe complex patterns of behavior, face a number of serious problems. First, the patterns should be described in as formal a fashion as possible, and yet the formalism that ethnographers have availed themselves of are simply inadequate to the task. Secondly, data such as ethnographic interviews constitutes the most common way of discovering a culture, but there is a dearth of formal methods for going from a text to the cultural presuppositions that underlie it. Thirdly, it is difficult to know in ethnographic interviews how much of what is said is a reflection of the culture, how much is the speaker's personal interpretation, and how much is due to the interview situation itself. Finally, ethnographers face the problem of how to deal in depth with large volumes of ethnographic data.

Artificial intelligence can be viewed in large part as the investigation of complex formalisms. Heretofore, these have been applied primarily in simple domains. In this paper we attempt to use AI formalism as a formal language of description for the complex conversational behavior that occurs in ethnographic interviews. We thus address the first of the ethnographer's problems by exploring the use of formalisms that begin to be adequate to the task. Moreover, work on discourse analysis in the AI framework has sought to characterize the structure of texts in terms of the goals and beliefs of the speaker. It thus suggests methods of using the structure of the text to force the explicitation of the underlying belief system, addressing the second of the ethnographer's problems. Finally, our work has confronted us with the last two problems, and the approach we are taking has suggested tentative ideas for dealing with these problems.

Generalization and Memory in an Integrated Understanding System

Michael Lebowitz

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 186, Oct. 1980.

Generalization and memory are part of natural language understanding. As people read stories describing various situations, they are able to recall similar episodes from memory and use them as a basis to form generalizations about the way such situations normally

occur. This thesis describes an integrated system for language understanding, IPP (Integrated Partial Parser), that encompasses the ability to generalize and record information in long-term memory as well as conceptual analysis.

IPP is a program that learns about the world by reading stories taken from newspapers and the UPI news wire, adding information from these stories to memory, and making generalizations that describe specific situations. It uses the generalizations that it has made to help in understanding future stories.

As it reads stories, IPP adds them to its permanent memory. If it locates similar stories in memory as it does this, then it attempts to make generalizations that describe the similarities among the events. Such generalizations form the basis for organizing events in memory and understanding later stories. IPP also includes a procedure for confirming generalizations as further stories are read.

In order to analyze the text that it reads, IPP makes extensive use of top-down, predictive processing. As it processes a story, IPP accesses memory in an attempt to identify generalizations describing stereotypical situations that can provide predictions to be used in understanding. Such use of memory to provide top-down context results in a robust and efficient understanding system.

Retrieval and Organizational Strategies in Conceptual Memory: A Computer Model

Janet Lynne Kolodner

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 187, Nov. 1980.

People effortlessly recall past events and episodes in their lives many times in the course of a normal day. A reasonable goal in the design of computer programs is to construct a memory with that same capability. To facilitate human-like retrieval of events from a computer memory, we must first specify a reasonable memory organization. We must then design updating and retrieval processes to build up and access that information. This thesis will present such a theory, and will describe a computer program called CYRUS which implements that theory.

CYRUS (Computerized Yale Retrieval and Updating System) stores and retrieves episodes in the lives of Secretaries of State Cyrus Vance and Edmund Muskie. When new events are added to its memory, CYRUS integrates them into memory along with the events it already knows about. CYRUS can then answer questions posed to it in English about the events it stores.

The algorithms and memory organization used in CYRUS have been developed by examining the way people answer questions requiring extensive memory search. Its reconstructive processes include *instantiation strategies*, which construct and elaborate on contexts for search, and *search strategies*, which direct construction.

Reconstructive processes require a vast store of generalized knowledge in order to be applied. Reconstructive retrieval implies a memory organization which organizes both generalized information about different types of events and distinguishing features of particular events. CYRUS' memory is self-organizing. When given a new fact about Vance or Muskie, it integrates the new event into its already-existing memory organization. In the process, it updates its generalized information and indexes the new event in the appropriate places. CYRUS can be seen as both a model of human memory and an intelligent information retrieval system.

Memory, Meaning, and Syntax

Roger C. Schank and Lawrence Birnbaum

Computer Science Department
Yale University
New Haven, Connecticut 06520

Research Report 189, Nov. 1980.

This paper explores the role of syntax in computational theories of natural language. We discuss the *integrated processing hypothesis*, which contends that meaning and world knowledge play a crucial part in language understanding even at the earliest points in the process. The hypothesis implies that syntactic knowledge plays no privileged role in language processing. Computer models of language analysis are discussed in relation to the overall theory.

Word Expert Parsing: A Theory of Distributed Word-Based Natural Language Understanding

Steven Small

Department of Computer Science
University of Maryland
College Park, Maryland 20742

Technical Report TR-954, Sept. 1980.

People have an incredible facility for organizing and selecting word senses in arriving at the intended meaning of sentences in context. The process takes place with such unnoticed and subconscious ease that it has been often overlooked in the study of language. To those building computer programs to understand language, however, this phenomenon represents a central problem. While many syntactic constructions and conceptual relations can be described through systems of rewrite rules, the word sense selection problem remains. The reason for this lies in the incompatibility

of the sense discrimination problem and the rule-based problem-solving method.

The computational theory of Word Expert Parsing approaches natural language understanding as a non-uniform distributed process of interacting words. An expert process for each word actively pursues its intended meaning in the context of other word experts and real-world knowledge. The theory perceives understanding as a behavior of memory interactions, and the computer model emphasizes process rather than output structures.

The Lexical Interaction Language (LIL) formalizes the interactions among individual word experts, and the Sense Discrimination Language (SDL) specifies their actions to determine intended word senses in context. These languages constitute the formal theory of Word Expert Parsing, and permit the representation of all linguistic knowledge in terms of active word-based distributed agents. An existing computer program translates word experts represented in these languages into executable processes that interact to cooperatively analyze sentences.

The thesis makes a number of claims about the processes of natural language comprehension and its computational realization. A formal theory is developed and a computer model constructed to provide evidence to support those claims. Word Expert Parsing explains the understanding of sentences containing highly ambiguous words and complex structures. The distributed word-based approach is advanced as a framework for a full-scale theory of discourse comprehension.

Understanding English Descriptions of Programs

Allan Ramsay

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 140, 1980, 4 pages.

A considerable amount of work has been done on verifying that computer programs fit their specifications. However, providing formal specifications is itself a difficult and tedious task, so that programs are generally only documented incompletely and imprecisely. This paper presents a computer system which accepts English descriptions of procedures and relates them to LISP programs that are supposed to implement them. This system is intended to illustrate how "informal" techniques may be used to provide a rough analysis of a program for which incomplete specifications are provided.

Parsing English Text

Allan Ramsay

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 139, 1980, 5 pages.

This paper presents a technique for parsing English text according to a grammar specified as a set of rewrite rules. The paper describes a compact way of representing such a grammar and presents a program which uses this representation to parse text without backtracking and without repeating work that it has already done.

Using Determinism to Predict Garden Paths

Rob Milne

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 142, 1980, 6 pages.

I am interested in making a psychologically valid model of human natural language understanding, and especially in a processing model for predicting when a sentence will be a garden path. While extending the Marcus deterministic parser to include noun-noun modification, several counter examples to Marcus' garden path prediction were found. In this paper I proposed that when people encounter an ambiguous situation that may lead to a garden path, they use semantics to decide rather than look ahead. I will present an extension to the garden path prediction mechanism of Marcus' parser to account for this and several experiments to test this theory.

Parsing Against Lexical Ambiguity

Rob Milne

Department of Artificial Intelligence
University of Edinburgh
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 144, 1980, 8 pages.

Marcus' original deterministic parsing included almost no part-of-speech ambiguity. In this paper, the addition of part-of-speech ambiguity to a deterministic parser written in PROLOG is described. To handle this ambiguity, it was necessary to add no special mechanisms to the parser. Instead the grammar rules were made to enforce agreement, and reject ungrammatical sentences. The resulting system is very effective and covers many examples of ambiguity.

Predictive Analysis in Sentence Comprehension: A Computer Simulation Model for Surface Structure Parsing

C.P. Whaley

Communications Psychology (Dept. 3Z11)
Bell-Northern Research
P.O. Box 3511
Station C
Ottawa, Ontario K1Y 4H7 CANADA

Int. J. Man-Mach. Studies 13, 3 (Oct. 1980), 259-294.

Through numerous models have been proposed by linguists and computer scientists for the parsing of natural language, Kimball (1973) has outlined one of the few that takes into account the operational limitations of the human perceiver. His *predictive analysis* model is intuitively appealing and is supported by empirical research (e.g. Whaley, 1979). This paper presents the parsing principles suggested by Kimball, and describes a computer simulation which incorporates them. The parsing accuracy of the model is demonstrated and discussed for various sentence types. Finally, extensions to the existing model are considered including the quantification of the model's predictions.

The Use of Artificial Intelligence Techniques in Computer-Assisted Instruction: An Overview

Alice Gable and Carl V. Page

Department of Computer Science
Michigan State University
East Lansing, Michigan 48824

Int. J. Man-Mach. Studies 12, 3 (April 1980), 259-282.

One of the major goals of research in Artificial Intelligence is the representation of knowledge so that a computer can solve problems or communicate in a manner which exhibits "common sense". Few computer programs, including those for education, possess behavior which approaches any facet of the constellation of human skills and knowledge which are imprecisely called "common sense". However, the revolutionary decline in hardware costs now makes it possible to consider economically viable, sophisticated designs for computer-aided instruction systems possessing some of the common sense attributes of a human tutor.

In this survey we examine, in depth, techniques from Artificial Intelligence that can be used to endow a Computer-Aided Instruction system with approximations of some of the desirable qualities of a human tutor. We consider both techniques which have been proved in prototype systems for Computer-Aided Instruction and some techniques which were originally developed for other purposes.

The Structure of The Merriam-Webster Pocket Dictionary

Robert A. Amsler

Department of Computer Sciences
University of Texas
Austin, Texas 78712

Technical Report TR-164, Dec. 1980.

The dissertation has as its purpose the exploration and discussion of the structure of a machine-readable copy of an ordinary pocket dictionary with particular attention to the utility of the information contained therein for future application in computational linguistics, ethnosemantics, and information science. This structure was first explored by hand using two concordance-like printouts of the *Merriam-Webster Pocket Dictionary's* contents prepared from magnetic tapes produced by John Olney at System Development Corporation.

Initially the goal of the analyses was the determination of whether useful semantic information could be derived from dictionary definitions. Once it was determined by several hand analyses that such data did provide a new source of information about the lexicon, the larger goal of assembling a complete taxonomy of the words in the dictionary was conceived and undertaken.

A hand analysis of the rich semantic information contained in the dictionary for verbs defined in terms of "move" is presented in Chapter 2. This componential analysis of a set of definitions revealed the potential of dictionaries for use as the basis of numerous additional studies of high-frequency verbs based upon their usage in defining more specific verbs.

Chapter 3 presents the results of a hand analysis of the taxonomy of "vehicle" terms developed from a large sample of definitions based upon the word "vehicle" and its descendants. This study revealed that indeed the dictionary did contain large, coherent, and computationally useful information in a taxonomic-like organization that could be revealed by connecting together definitions on the basis of their defining terms.

Chapter 4 deals with the steps involved in designing, loading, and selectively accessing large databases containing all the definitions of nouns, verbs, and adjectives contained in the dictionary. Statistical information on the frequencies and nature of the part of speech data contained in the dictionary is given in chapter 4 and appendix 3. A "word sense meaning" representation for uniquely specifying dictionary senses and the problems of semantic ambiguity in lexical usage are also illustrated.

In chapter 5, statistics on the frequency of dictionary defining vocabulary and measures of semantic ambiguity are given. Section 5.2 contains an extensive

discussion of the steps taken to perform semantic disambiguation on dictionary definition terms along with further statistics on the frequencies of such disambiguated words in the dictionary. Appendix 4 gives part of a disambiguation protocol session transcribed from a tape-recording and demonstrating the nature of the human disambiguation task carried out on the dictionary. Finally, chapter 5 concludes with a discussion of the methods used to computationally assemble and enumerate the complete taxonomic structure of the dictionary's noun and verb definitions.

Chapter 6 discusses some of the findings about the nature of the dictionary's taxonomies of nouns and verbs. It discusses the manners in which dictionary taxonomies, ultimately terminate in primitive root concepts, relationships to other taxonomies, case argument relations to verbs, and in partitives and collectives.

Finally, chapter 7 concludes with a discussion of a possible means of automating the analysis procedures of chapter 4 and 5 to perform fully automatic parsing and disambiguation of dictionary entries. This discussion also demonstrates the application of the proposed disambiguation technique to natural language processing and computational linguistics in general.

The Automated Dictionary

Mark S. Fox, Donald J. Bebel, and Alice C. Parker

Carnegie-Mellon University
Schenley Park
Pittsburgh, Pennsylvania 15213

Computer 13, 7 (July 1980), 35-48.

An automated dictionary is a computer-based device that holds all or part of a dictionary and allows access to and the display of entries. The "automation" of a dictionary raises several issues: What information should be stored? How does the automated dictionary's size determine the style of usage? What is the interface between the user and the dictionary? What are the automated dictionary's physical components? Each issue provides a rich set of alternatives from which to choose. Part of the design problem of an automated dictionary is picking feasible points in this multidimensional design space.

The purpose of this article is to present the results of some preliminary studies conducted at the request of the National Institute of Education. We determine the capability of an automated dictionary system built using present technology and estimate present and future costs based on trends in the cost and functionality of digital electronics.

Computer Programs for Detecting and Correcting Spelling Errors

James L. Peterson

Department of Computer Sciences
University of Texas
Austin, Texas 78712

Comm. ACM 23, 12 (Dec. 1980), 676-687.

With the increase in word and text processing computer systems, programs which check and correct spelling will become more and more common. This paper investigates the basic structure of several such existing programs and their approaches to solving the problems which arise when this type of program is created. The basic framework and background necessary to write a spelling checker or corrector are provided.

Recognition of Spoken Spelled Names for Directory Assistance Using Speaker-Independent Templates

A.E. Rosenberg, L.R. Rabiner, and J.G. Wilpon

Bell Telephone Laboratories, Inc.
600 Mountain Avenue
Murray Hill, New Jersey 07974

Bell Syst. Tech. J. 59, 4 (April 1980), 571-592.

In a recent paper, Rosenberg and Schmidt demonstrated the applicability of a speaker-trained, isolated word speech recognizer to the problem of automatic directory assistance. Input to the system was in the form of a string of letters which spelled the last name and initials of an individual for whom a directory listing was required. Rosenberg and Schmidt found that, even though the recognition rate for individual letters was rather low (approximately 80 percent), the rate at which the correct directory listing was found was higher (approximately 95 percent). In this paper, we extend these results to include the case of speaker-independent recognition of letters. We show that overall performance in the speaker-independent mode is comparable to performance in a speaker-dependent mode and examine various factors important for operation in a speaker-independent mode, such as characteristics of the reference templates, choice of decision rule, and threshold parameters. For the most part, the overall system is remarkably robust to the parameters of the recognizer. For the best choice of these parameters, a 95-percent correct string rate is obtained, comparable to the performance in a speaker-dependent mode.

A Minimum-Distance Search Technique and Its Application to Automatic Directory Assistance

B. Aldefeld, S.E. Levinson, and T.G. Szymanski

Bell Telephone Laboratories, Inc.
600 Mountain Hill
Murray Hill, New Jersey 07974

Bell Syst. Tech. J. 59, 8 (Oct. 1980), 1343-1356.

This paper describes a new search procedure and its application to the problem of obtaining telephone directory information from spoken spelled input. The method obtains its speed from using the concept of equivalence classes, with names classified according to their letter-by-letter acoustic similarity. It derives its accuracy from the use of a minimum-distance criterion for selecting answers. The search procedure finds the name with the minimum distance, usually after only a small fraction of the directory file has been examined. Using an acoustic analyzer with an 80 percent correct recognition rate for individual letters, a 98.6 percent correct recognition rate for names was achieved when the method was applied to a directory of 18,000 entries. On the average, only 1.2 percent of the directory had to be examined for each query. With an input recognition rate of 71 percent for letters, the respective figures were 97.2 percent and 2.8 percent.

Synthesis by Rule of Prosodic Features in Word Concatenation Synthesis

S. J. Young

Control Systems Centre
University of Manchester
Institute of Science and Technology
Manchester, M80 1QD, ENGLAND

F. Fallside

Engineering Department
Cambridge University
Corn Exchange Street
Cambridge, CB2 3QG, ENGLAND

Int. J. Man-Mach. Studies 12, 4 (April 1980), 241-258.

The quality of speech obtainable by the technique of Word Concatenation Synthesis depends crucially on the accuracy of the pitch and timing contours which need to be computed for each utterance synthesized. A method for the synthesis-by-rule of these prosodic features is described for utterances for which the only information available is a syntactic phrase marker and the lexical structure of each component word.

Rules and algorithms are presented for the determination of word group boundaries and the placement of stress, accent and nuclei. A timing contour algorithm is described which implements the prominence required for stressed syllables and also generates the stress-timed rhythm of natural English. A corresponding pitch contour algorithm computes appropriate intonation patterns using a minimal set of three tone groups; fall, rise and fall-rise.

The Linguistic Reason Why the Computer Will Never Think

John R. Hammen

P.O. Box 12931
Seattle, Washington 98111

SIGLASH Newsletter 13, 4 (Dec. 1980), 8-16.

This paper draws a comparison between the respective lowest denominators of computer and human language, both of which are the respective mediums of computer programming and human reasoning. The lowest denominator of the former is Boolean Algebra, which is a logic defined by the 3 operators: *And*, *Or*, and *Not*. Underlying human language is the system of logic called semantic structure. The Boolean operators — *And*, *Or*, and *Not* — exist in semantic structure. This paper shows that the 3 determiners: *Definite*, *Indefinite*, and *All-Inclusive*; the 5 moods: *Indicative*, *Imperative*, *Interrogative*, *Conditional*, and *Emphatic*; the 3 tenses: *Past*, *Present*, and *Future*; and the 8 cases: *Agent*, *Possessive*, *Source*, *Direction*, *Instrument*, *Experiencer*, *Location*, and *Object* function just as much like logic operators as do *And*, *Or*, and *Not*. This means that semantic structure is defined by 22 operators (19 more than Boolean Algebra). The author contends that because of this difference in numbers of operators the computer will never be programmed to think in a human fashion.

The Natural Language of Interactive Systems

Henry Ledgard, Andrew Singer, and William Seymour

Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

John A. Whiteside

Digital Equipment Corporation
146 Main Street
Maynard, Massachusetts 01754

Comm. ACM 23, 10 (Oct. 1980), 556-563.

The work reported here stems from our deep belief that improved human engineering can add significantly to the acceptance and use of computer technology. In particular, this report describes an experiment to test the hypothesis that certain features of natural language provide a useful guide for the human engineering of interactive command languages. The goal was to establish that a syntax employing familiar, descriptive, everyday words and well-formed English phrases contributes to a language that can be easily and effectively used. Users with varying degrees of interactive computing experience used two versions of an interactive text editor; one with an English-based command syntax in the sense described above, the other with a more notational syntax. Performance differences strongly favored the English-based editor.

Intensional Concepts in Propositional Semantic Networks

Anthony S. Maida

Rochester Institute of Technology
Rochester, New York

Stuart C. Shapiro

Computer Science Department
State University of New York at Buffalo
4230 Ridge Lea Road
Amherst, New York 14226

Technical Report 171, Feb. 1981.

One of the major characteristics of semantic networks as a representation of knowledge is what we term the *uniqueness principle*: that each concept represented in the network is represented by a unique node. We show that this principle entails that at least some nodes represent intensional, rather than extensional, objects. By taking the position that a semantic network is a model of the conceptual organization of a cognitive agent rather than of the world, we argue further that nodes can represent only intensional objects. To allow for coreferential terms, we include the representation of a proposition that asserts the extensional equivalence of two intensional concepts. This approach is shown to lead to elegant solutions to three apparently unrelated problems in the representation of knowledge. Our discussion brings to bear relevant literature in analytic philosophy and experimental psychology and leads us to suggest two new psychology experiments.

COCCI: A Deductive Semantic Network Program for Solving Microbiology Unknowns

Stuart C. Shapiro

Department of Computer Science
State University of New York at Buffalo
4230 Ridge Lea Road
Amherst, New York 14226

Technical Report 173, March 1981.

"You have been given a culture of one of nine cocci. Identify it." COCCI is a program to solve this problem. To identify the unknown, COCCI requests a human assistant to perform tests and make observations and report the results to it. COCCI consists of 14 specific facts and 17 deduction rules stored in SNePS, a general purpose deductive semantic network processing system, and a small ATN grammar for parsing and for generating English from the semantic network. All of COCCI's reasoning and interaction with humans is driven by the general purpose bi-directional inference sub-system of SNePS. This paper describes COCCI as an illustration of deductive semantic networks in general and SNePS in particular. Specific points covered include structure sharing, procedural attachment, generation grammars and the structure of deduction rules.

Bi-Directional Inference

Joao P. Martins, Donald P. McKay
and Stuart C. Shapiro

Department of Computer Science
State University of New York at Buffalo
4230 Ridge Lea Road
Amherst, New York 74226

Technical Report 174, March 1981.

In this paper we present a brief overview of the SNePS deduction system and show through an example the interaction between forward and backward inference. This interaction — resulting in a class of inference termed bi-directional inference — enables an easy and elegant way of performing resource-limited searches and the possibility of performing resource-limited deduction in a natural and simple way. Furthermore, bi-directional inference focus a system's attention towards the interests of the user and can cut down the fan-out of pure forward or pure backward chaining.

A Belief Revision System Based on Relevance Logic and Heterarchical Contexts

Joao P. Martins and Stuart C. Shapiro

Department of Computer Science
State University of New York at Buffalo
4230 Ridge Lea Road
Amherst, New York 14226

Technical Report 175, March 1981.

This paper describes the underlying theory of a Belief Revision System based on Relevance Logic and Heterarchical Contexts. In our system each statement is indexed by the set of basic (i.e., non-derived) assumptions used in its derivation and by the set of basic assumptions with which it is incompatible. A context is a set of basic assumptions and contains all the statements whose first index is a subset of the context and whose second index is disjoint from the context. This allows straightforward switching between contexts and the possibility of efficiently performing hypothetical reasoning.

Interacting in Natural Language with Artificial Systems: The Donau Project

Giovanni Guida and Marco Somalvico

Milan Polytechnic Artificial Intelligence Project
Istituto di Elettrotecnica ed Elettronica
Politecnico di Milano
Milan, ITALY

Inform. Systems 5, 4 (1980), 333-344.

This paper is intended to propose a new methodological approach to the conception and development of natural language understanding systems. This new contribution is supported by the design, implementation, and experimentation of DONAU: a general purpose domain-oriented natural language understanding

system developed and presently running at the Milan Polytechnic Artificial Intelligence Project. The system is based on a two-level modular architecture intended to overcome the lack of flexibility and generality often pointed out in many existing systems, and to facilitate the exchange of results and actual experiences between different projects. The horizontal level allows an independent and parallel development of the single segments of the system (syntactic analyser, semantic analyser, information extractor, legality controller). The vertical level ensures the possibility of changing (enlarging or redefining) the definition of the semantic domain on which each particular version of the system is oriented and specialized in a simple, incremental, and user-oriented way. In the paper the general architecture of the system and the mode of operation of each segment are illustrated in detail. Linguistic models, knowledge representation, and parsing algorithms are described and illustrated by means of selected examples. Performance evaluations of the system in the application version on data base inquiry are reported and discussed. Promising directions for future research are presented in the conclusions.

Using the Data Base as a Semantic Component to Aid in the Parsing of Natural Language Data Base Queries

Larry R. Harris

Mathematics Department
Dartmouth College
Hanover, New Hampshire 03755

J. Cybern. 10, 1-3 (Jan.-March 1980), 77-96.

Many of the recent advances in artificial intelligence (AI) have been brought about through the use of domain specific knowledge. In this same spirit, this paper presents an approach to understanding natural language data base queries that employs the use of domain specific knowledge to aid the parsing process. The unique aspect of the data base query environment relative to other AI problem domains is that the required body of knowledge already exists in the form of the data base being queried. Thus, there is an important benefit of making use of this knowledge in the form in which it is maintained by the data base management system, since the very difficult problems of gathering and representing this information can be circumvented.

The paper describes the problems encountered in the parsing of data base queries that can be solved by the semantic use of the data base, as well as a precise description of how these problems can be solved by existing data base systems. The proposed use of the data base as a semantic component has been a primary component of the design of a high performance natural language data base query system called ROBOT, that has been successfully installed at several

commercial installations. Since this system is a physical realization of the proposed methodology, a brief description of the system and its experiences in the field are given as evidence of the feasibility of this approach.

Natural Language Access to Medical Text

Donald E. Walker and Jerry R. Hobbs

Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Technical Note 240, March 1981.

This paper describes research on the development of a methodology for representing the information in texts and of procedures for relating the linguistic structure of a request to the corresponding representations. The work is being done in the context of a prototype system that will allow physicians and other health professionals to access information in a computerized textbook of hepatitis through natural language dialogues. The interpretation of natural language queries is derived from DIAMOND/DIAGRAM, a linguistically motivated, domain-independent natural language interface developed at SRI. A text access component is being developed that uses representations of the propositional content of text passages and of the hierarchical structure of the text as a whole to retrieve relevant information.

An Efficient Easily Adaptable System For Interpreting Natural Language Queries

David H.D. Warren

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

Fernando C.N. Pereira

CAAS Studies, Department of Architecture
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 155, 1981, 18 pages.

This paper gives an overall account of a prototype natural language question answering system, called Chat-80. Chat-80 has been designed to be both efficient and easily adaptable to a variety of applications. The system is implemented entirely in PROLOG, a programming language based on logic. With the aid of a logic-based grammar formalism called extraposition grammars, Chat-80 translates English questions into the PROLOG subset of logic. The resulting logical expression is then transformed by a planning algorithm into efficient PROLOG, cf. "query optimisation" in a relational database. Finally the PROLOG form is executed to yield the answer. On a domain of world geography, most questions within the English subset

are answered in well under one second, including relatively complex queries.

A Computer Written Language Lab

Mike Sharples

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 134, 1980, 9 pages.

A group of Edinburgh primary school children are using a prototype computer written language lab, designed to help them understand language and develop their writing style. In a more sophisticated version the language lab can become a resource for teachers, linguists, indeed anyone interested in creating or investigating language.

A Computer Based Language Workshop

Mike Sharples

Department of Artificial Intelligence
University of Edinburgh
Forrest Hill
Edinburgh EH1 2QL SCOTLAND

DAI Research Paper No. 135, 1980, 11 pages.

The paper describes a project to implement and evaluate a computer written language workshop. The pilot phase, now completed, provides pupils with a small range of language manipulation programs. The second phase will offer children a more extensive range of computer based language aids for the exploration of written style.

Reading, Looking and Learning

Christine Urquhart

38 Shipton Road
Sutton Coldfield
W. Midlands B721 NR, ENGLAND

Journal of Inform. Sci. 1 (1980), 333-344.

The survey of aspects of reading research discusses reading styles, 'readability', and the influence of presentation on reading and comprehension. Consideration of these leads to the fundamental problems of how the mind and eye interact in visual processing of text. Some research on the psychology of reading is reviewed, and certain psycholinguistic models of the reading process are discussed. Text structure and content appear to influence comprehension and learning, and relevant research on this, and the interaction of the reader with the text, is outlined. Reading strategies are discussed, and the possibilities of altering reading behaviour by increasing reading speed or improving learning patterns are reviewed. Possible implications for information science of reading research are mentioned.

Expert Systems

Donald Michie

Machine Intelligence Research Unit
University of Edinburgh
Hope Park Square, Meadow Lane
Edinburgh EH8 9NW, SCOTLAND

Computer Journal 23, 4 (Nov. 1980), 369-376.

A central feature of 'expert systems' in chemistry, molecular geology, medicine, plant pathology, robotics, chess and other applications is the man-machine communication of descriptive concepts in the form of patterns. 'Humanisation' of machine-made representations must find a place in future designs.

A Linguistic Approach to Decisionmaking with Fuzzy Sets

Richard M. Tong

Advanced Information and Decision Systems
201 San Antonio Circle
Suite 286
Mountain View, California 94040

Piero P. Bonissone

Corporate Research and Development Department
General Electric Company
P.O. Box 43
Building 37-579
Schenectady, New York 12301

IEEE Trans. Sys. Man Cyb. 10, 11 (Nov. 1980), 716-722.

A technique for making linguistic decisions is presented. Fuzzy sets are assumed to be an appropriate way of dealing with uncertainty, and it is therefore concluded that decisions taken on the basis of such information must themselves be fuzzy. It is inappropriate then to present the decision in numerical form; a statement in natural language is much better. For brevity only a single-stage multiattribute decision problem is considered. Solutions to such problems are shown using ideas in linguistic approximation and truth qualification. An extensive example illuminates the basic ideas and techniques.

Natural Language Programming and Natural Programming Languages

Larry H. Reeker

Department of Computer Science
University of Queensland
St. Lucia, Brisbane
Queensland, AUSTRALIA 4067

Aust. Comput. J. 12, 3 (Aug. 1980), 89-92.

The idea of "natural language programming" has a good deal of superficial attractiveness, but it also raises a number of questions. The most immediate of these is what one means by "natural language programming", since the use of unfettered, everyday language is not feasible, for a number of reasons. Two approaches to moving toward the goal of more natural programming languages using natural language are

discussed, and the second, or "bottom up" approach is advocated.

Natural Language Programming: Styles, Strategies and Contrasts

Lance A. Miller

Computer Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

IBM Systems Journal 20, 2 (1981), 184-215.

College students who were not familiar with computers were asked to produce written natural language procedural instructions as directions for others to follow. These directions were solutions for six file-manipulation problems that also could reasonably be solved by writing computer programs. The written texts were examined from five points of view: solution correctness, preferences of expression, contextual referencing, word usage, and formal programming languages. The results provide insight both on the manner in which people express computer-like procedures "naturally" and on what features programming languages should include if they are to be made more "natural-like."

Probabilistic Languages: A Review and Some Open Questions

C.S. Wetherell

Bell Telephone Laboratories, Inc.
600 Mountain Avenue
Murray Hill, New Jersey 07974

Computing Surveys, 12, 4 (Dec. 1980), 361-379.

Context-free languages are commonly used to describe the structure of programming languages. However many interesting problems involve not just a language's structure but also the actual usage of the language. Adding a notion of probability to ordinary grammars gives rise to probabilistic context-free grammars. Interesting in their own right because of some pretty theorems, probabilistic context-free languages can be applied to the analysis of programming languages, automatic parsers, and error correctors. A complete outline of the theory is presented with examples. Some open questions are posed.

Compiler Testing Using a Sentence Generator

A. Celentano, S. Crespi Reghizzi, P. Della Vigna and
C. Ghezzi

Istituto di Elettrotecnica ed Elettronica
Politecnico di Milano
Piazza L. da Vinci 32
I-20133 Milano ITALY

G. Granata and F. Savoretti

Ing. C. Olivetti & C. S.p.A.
Via Jervis 77
Ivrea ITALY

Softw. Pract. Exper. 10, 11 (Nov. 1980), 897-918.

A system for assisting in the testing phase of compilers is described. The definition of the language to be compiled drives an automatic sentence generator. The language is described by an extended BNF grammar which can be augmented by actions to ensure contextual congruence, e.g. between definition and use of identifiers. For deep control of the structure of the produced sample the grammar can be described by step-wise refinements: the generator is iteratively applied to each level of refinement, producing at last compilable, complete programs. The implementation is described and some experimental results are reported concerning PLZ, MINIPL and some other languages.

Soft Display Key for Kanji Input

Jouko J. Seppänen

Helsinki University of Technology
Computing Centre
021 Espoo 15, FINLAND

Research Report 17, 1980.

The concept of a soft display key as applied to input of large character sets or vocabularies such as Kanji, the ancient Chinese ideographic script, is discussed. The Japanese orthography and the necessity of using Kanji characters in data terminals are explained. Problems arising from the number and complexity of Kanji symbols for the manufacture and use of keyboard devices are stated. A review is made of devices and methods presently used or suggested. The feasibility of the soft display key is then demonstrated. Some requirements for the design and implementation of a soft display keyboard for Kanji are considered. In conclusion, implications to man/computer interface design, human factors engineering and hardware unification and standardization are stated.