

Abstracts of Current Literature

Interestingness: Controlling Inferences

Roger C. Schank

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Artificial Intelligence 12, 3 (Nov. 1979), 273-297.

One of the central issues in natural language understanding for the last few years has been the problem of making inferences and controlling those inferences. In this paper I will discuss an overall method for controlling inferences that has effects on parsing, script application and the tracking of goals and plans.

Reminding and Memory Organization: An Introduction to MOPs

Roger C. Schank

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 170, Dec. 1979.

A key question for researchers in Cognitive Science is the problem of how human memory is organized. The solution to this problem is fundamental to Cognitive Scientists regardless of whether their basic orientation is towards Artificial Intelligence or Cognitive Psychology. The theories that we test by means of psychological experiments ought to have some ramifications on how we build computer models of the processes tested by those experiments, and the computer models that we build ought to supply testable hypotheses for psychologists. This paper discusses some issues in memory organization that are fundamental to an understanding system, and thus to Cognitive Science.

Knowledge-Based Parsing

Anatole V. Gershman

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 156, April 1979.

A model for knowledge-based natural language analysis is described. The model is applied to parsing English into Conceptual Dependency representations. The model processes sentences from left to right, one word at a time, using linguistic and non-linguistic knowledge to find the meaning of the input. It operates in three modes: structure-driven, position-driven, and situation-driven. The first two modes are expectation-based. In structure-driven mode concepts underlying new input are expected to fill slots in the previously built conceptual structures. Noun groups are handled in position-driven mode which uses

position-based pooling of expectations. When the first two modes fail to account for a new input, the parser goes into the third, situation-driven mode which tries to handle a situation by applying a series of appropriate experts.

Four general kinds of knowledge are identified as necessary for language understanding: lexical knowledge, world knowledge, linguistic knowledge, and contextual knowledge. The concrete knowledge structures and representational mechanisms necessary for understanding are examined using examples of English constructions involving the preposition "by". These include passives, instrumental constructions, constructions introducing physical settings, role-frame characterizations, and authorship expressions. It is shown that the parsing framework outlined in this thesis is adequate for the integration of the first three kinds of knowledge but difficulties with the integration of contextual knowledge still remain.

Skimming Stories in Real Time: An Experiment in Integrated Understanding

Gerald Francis DeJong II

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 158, May 1979.

This dissertation describes a new method of automated text analysis. FRUMP (Fast Reading Understanding and Memory Program) is a working natural language processing system that has been implemented to demonstrate the viability of this new approach. The system skims news stories directly from the United Press International news wire and produces a summary of what it understands. FRUMP is able to correctly process news articles it has never before seen.

The process of interpreting input text words can be greatly simplified if it is viewed as one component of a highly integrated understanding process. In FRUMP the text analyzer is embedded in a predictive understander. This embedding is the key to FRUMP's robustness. FRUMP's integration makes all of the world knowledge and top-down predictions of the understander available to the text analyzer. FRUMP uses a data construct called a *sketchy script* to store its world knowledge. There is one sketchy script for each real world "situation" FRUMP knows about. The system uses this knowledge to make predictions about what might happen next in a given situation. FRUMP continually jumps to conclusions about what the text means and generates predictions about what might occur next. The text analysis process then is reduced to finding a reading of the text that satisfies these predictions. The process of looking for readings in the text is much simpler than the process of generating a conceptual structure from an arbitrary input. Thus

there is no need in FRUMP for an extremely powerful English parser.

Given a new input FRUMP must be able to decide which of its sketchy scripts contains the knowledge needed to process the input. This is the process of "script selection" which is a major problem for an approach such as FRUMP's. A workable solution to the script selection problem must be computationally manageable. The process must not be significantly slowed down by the addition of more world knowledge in the form of more sketchy scripts. That is, the computational complexity of script selection must not depend significantly on the number of scripts in the system. Furthermore, the script selection process must often be completely bottom-up; most news stories cannot be anticipated before they are seen. Yet during this process, the FRUMP text analyzer must still be supplied with adequate top-down guidance. This problem is addressed and a general solution for FRUMP's purposes is given.

A Process Model of Language Acquisition

Mallory Gordon Richard Selfridge

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 172, Jan. 1980.

How do children acquire language? This thesis approaches this problem by studying how children between the ages of 1 and 2 years learn to understand commands containing action words, object words, and relation words. It presents observations of one child made during language development and presents rules which account for this development. A computer program model of this child's development, incorporating the observations and rules, was written and tested in learning a subset of English and Japanese.

There are four components to the problem of learning to understand commands involving action, object, and relation words. The first is that of knowledge which the child brought to language learning. The second is a characterization of his developed comprehension abilities at age 2. The third is that of his experiences with language -- the various situations in which people said things to him. The fourth is the rules which can account for his learning the comprehension skills based on his initial knowledge and his experience with language.

This thesis presents a characterization of this child's knowledge prior to learning language at age 1, describes his comprehension skills at age 2, examines his experiences with language between these times, and proposes a set of rules which can account for the development of these skills from his prior knowledge

and his language experiences. It also presents a computer program which embodies this model.

The results of this work suggest that language learning is dependent upon situational understanding and world knowledge to a greater degree than realized. They suggest that notions of adult language understanding, in particular the nature and role of language structure in understanding, should perhaps be modified in light of what appears both to be easily learnable and also is sufficient to account for the observations.

Affect Units and Narrative Summarization

Wendy G. Lehnert

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 179, May 1980.

The analysis of narrative text involves various levels of description. On the lowest level are word meanings and syntactic structure within single sentences. On a higher level there are problems of generating inferences and integrating information into memory. At the highest level is the notion of a macro-structure or narrative plot. The identification of high-level narrative structures is central to the problem of narrative summarization. But the intuitive notion of a plot is useless for a process model of summarization unless we can specify the hierarchical representations that allow us to analyze input and produce plots as output.

A representational strategy has been developed for high-level structural analysis in conjunction with the BORIS system (a narrative text understanding system). The structures produced effectively encode plot lines in terms of connected graph structures where graph nodes correspond to specific affect units. An affect unit is an abstraction of affective causality which is recognized in a bottom-up manner at the time of understanding. Simple manipulations of these graph structures yield the conceptual basis for narrative summaries, but the actual process of summarization depends on certain connectivity properties in the graph structure. Summarization techniques for the general case are presented, and a specific algorithm for one class of graphs is proposed.

VEGE: Variable Processing in a Natural Language Query System

Wendy G. Lehnert and William M. Bain

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Research Report 183, April 1980.

VEGE (VEgetable Gardening Expert) is a natural language query system designed to answer questions about vegetable gardening by employing a strategy of variable-depth sentence analysis. Each time an input question is encountered, VEGE can rely on one of three analysis strategies varying from superficial key-concept recognition to fully predictive conceptual analysis. The system determines which analysis is sufficient as it moves through the question in a one-pass left-to-right parsing procedure. A taxonomy of question types then determines what retrieval heuristics to execute in searching for an appropriate answer to the question. By expending no more processing effort than is absolutely necessary, VEGE implements a question analysis technique that is both semantically efficient and relatively robust in the range of sentence constructions it can handle. The processing techniques outlined in this report appear to be specific to the task orientation of an expert query system, but independent of vegetable gardening as a knowledge domain.

Inferring Conceptual Graphs

Sharon C. Salveter

Department of Computer Science
State University of New York
Stony Brook, New York 11794

Cognitive Science 3, 2 (April-June 1979), 141-166.

This paper investigates the mechanisms a program may use to learn conceptual structures that represent natural language meaning. A computer program named Moran is described that infers conceptual structures from pictorial input data. Moran is presented with "snapshots" of an environment and an English sentence describing the action that takes place between the snapshots. The learning task is to associate each root verb with a conceptual structure that represents the types of objects that participate in the action and the changes the objects undergo during the action. Four learning mechanisms are shown to be adequate to accomplish this learning task. The learning mechanisms are described along with the conditions under which each is invoked and the effect each has on existing memory structures. The conceptual structure Moran inferred for one root verb is shown.

Elements of a Plan-Based Theory of Speech Acts

Philip R. Cohen

Bolt Beranek and Newman, Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

C. Raymond Perrault

Department of Computer Science
University of Toronto
Toronto, Ontario M5S 1A7 CANADA

Cognitive Science 3, 3 (July-Sept. 1979), 177-212.

This paper explores the truism that people think about what they say. It proposes that, to satisfy their own goals, people often plan their speech acts to affect their listeners' beliefs, goals, and emotional states. Such language use can be modelled by viewing speech acts as operators in a planning system, thus allowing both physical and speech acts to be integrated into plans.

Methodological issues of how speech acts should be defined in a plan-based theory are illustrated by defining operators for requesting and informing. Plans containing those operators are presented and comparisons are drawn with Searle's formulation. The operators are shown to be inadequate since they cannot be composed to form questions (requests to inform) and multiparty requests (requests to request). By refining the operator definitions and by identifying some of the side effects of requesting, compositional adequacy is achieved. The solution leads to a metatheoretical principle for modelling speech acts as planning operators.

An Evaluation of Story Grammars

John B. Black

Department of Psychology
Yale University
Box 11A Yale Station
New Haven, Connecticut 06520

Robert Wilensky

Computer Science Division
Department of EECS
University of California, Berkeley
Berkeley, California 94720

Cognitive Science 3, 3 (July-Sept. 1979), 213-230.

We evaluate the "story grammar" approach to story understanding from three perspectives. We first examine the formal properties of the grammars and find only one to be formally adequate. We next evaluate the grammars empirically by asking whether they generate all simple stories and whether they generate only stories. We find many stories that they do not generate and one major class of nonstory that they do generate. We also evaluate the grammars' potential as comprehension models and find that they would add nothing to semantic models that focus on the story content. Hence we advocate a story content oriented approach to studying story understanding instead of the structural story grammar approach.

Prediction and Substantiation: A New Approach to Natural Language Processing

Gerald DeJong

Department of Computer Science
Yale University
New Haven, Connecticut 06520

Cognitive Science 3, 3 (July-Sept. 1979), 251-273.

This paper describes a new approach to natural language processing which results in a very robust and efficient system. The approach taken is to integrate the parser with the rest of the system. This enables the parser to benefit from predictions that the rest of the system makes in the course of its processing. These predictions can be invaluable as guides to the parser in such difficult problem areas as resolving referents and selecting meanings of ambiguous words. A program, called FRUMP for Fast Reading Understanding and Memory Program, employs this approach to parsing. FRUMP skims articles rather than reading them for detail. The program works on the relatively unconstrained domain of news articles. It routinely understands stories it has never before seen. The program's success is largely due to its radically different approach to parsing.

Points: A Theory of Story Content

Robert Wilensky

Computer Science Division
Department of EECS
University of California, Berkeley
Berkeley, California 94720

Memorandum No. UCB/ERL M80/17, April 1980.

Attempts to produce computer story understanding systems have generated a number of interesting ideas, particularly in the areas of knowledge representation and organization. However, many basic questions still remain largely unaddressed. In particular, the idea of what actually constitutes a story has never been clearly delineated.

A theory of story *points* has been developed that attempts to characterize those texts that describe situations that constitute stories. Unlike previous attempts at such a characterization, points are based not on the structure or form of a text, but on its content. Points describe those situations that generate reader interest and therefore give a text some poignancy.

The theory of stories proposed here is intimately connected with basic issues of language understanding, language generation, cognition and memory. In addition to characterizing stories, knowledge about poignancy is as necessary for the construction of intelligent story understanding programs as are theories of inference and knowledge representation. A story understanding system (PAM - Plan Applier Mechanism) has been given some of this knowledge about points, and as a result, its language processing capabilities have been extended to facilitate summarization and intelligent *forgetting*.

Meta-Planning: Representing and Using Knowledge About Planning in Problem Solving and Natural Language Understanding

Robert Wilensky

Computer Science Division
Department of EECS
University of California, Berkeley
Berkeley, California 94720

Memorandum No. UCB/ERL M80/33, Aug. 1980.

This paper is concerned with those elements of planning knowledge that are common to both understanding someone else's plan and creating a plan for one's own use. This planning knowledge can be divided into two bodies: knowledge about the world, and knowledge about the planning process itself. Our interest here is primarily with the latter corpus. The central thesis is that much of the knowledge about the planning process itself can be formulated in terms of higher-level goals and plans called *meta-goals* and *meta-plans*. These entities can then be used by the same understanding and planning mechanisms that process ordinary goals and plans; however, the meta-planning knowledge now enables these mechanisms to handle much more complicated situations, and in a quite uniform manner.

Systems based on meta-planning would have a number of advantages over existing problem solving and understanding systems. The same knowledge could be shared by both a planner and understander, and both would be able to handle complex situations elegantly. In addition, in planning, the use of meta-planning has several advantages over more traditional methods involving constraints or critics. Meta-planning allows the full power of a problem solver to be applied to situations that are generally amenable only to special purpose processing. In addition, meta-planning facilitates the representation of some situations that are difficult to express otherwise. We have begun to introduce meta-planning knowledge into two systems: PAM, a story understanding program, and PANDORA, a problem solving and planning system.

PHRAN: A Knowledge-Based Approach to Natural Language Analysis

Robert Wilensky and Yigal Arens

Computer Science Division
Department of EECS
University of California, Berkeley
Berkeley, California 94720

Memorandum No. UCB/ERL M80/34, Aug. 1980.

We have developed an approach to natural language processing in which the natural language processor is viewed as a knowledge-based system whose knowledge is about the meanings of the utterances of its language. The approach is oriented around the phrase rather than the word as the basic unit. We

believe that this paradigm for language processing not only extends the capabilities of natural language systems, but handles those tasks that previous systems could perform in a more systematic and extensible manner.

We have constructed a natural language analysis program called PHRAN (PHRasal ANalyzer) based in this approach. PHRAN reads texts in English and produces structures representing their meanings. PHRAN's knowledge of English is not confined to the word level. Instead, the system has knowledge about language constructs of varying levels of specificity, from canned literal phrases to general verb-oriented patterns. Associated with each language pattern is a corresponding meaning component. As PHRAN reads a sentence, it searches its data base for the language patterns that best interpret the incoming text. Then the associated meaning components of those patterns are used to create a meaning representation for that utterance.

This model has a number of advantages over existing systems, including the ability to understand a wider variety of language utterances, increased processing speed in some cases, a clear separation of control structure from data structure, a knowledge base that could be shared by a language production mechanism, greater ease of extensibility, and the ability to store some useful forms of knowledge that cannot readily be added to other systems.

Computational Interpretation of English Spatial Prepositions

Lois Carolyn Boggess

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Technical Report T-75, Feb. 1979.

It seems clear to anyone who pays attention to the use of prepositions in language that any one preposition, when used to describe the spatial relationship between different objects can produce strikingly different mental models for different objects. The mental model produced by the description "a bowl on a table" seems to be somewhat different from that produced by "a poster on a wall" which in turn is somewhat different from "a shelf on a wall" which again is different from "a fly on a ceiling".

It is the contention of this paper that the preposition *in conjunction with* a small set of features of the objects (mostly perceptual features) can account for such variations in spatial relations. The thesis discusses a means of taking English-language descriptions involving prepositions and their semantic subjects and objects and deriving a three-dimensional model of the spatial relationships of the subject and object.

The relationship of some of the spatial prepositions to a coordinate system is explored, as well as canonical definitions for prepositions based on analyses of descriptions using "neutral" subjects and/or objects ("whatchamacallit", "you-know-what", and so on).

Examples are taken from a simple program which accompanies the theory. The program is supplied with approximate descriptions of the shapes of a variety of objects. Each preposition in the program has *one* definition (e.g., there is only one procedure for *on*, rather than several--ON1, ON2, ON3, and so on); in general the definition is made up of several components, each of which is responsive to a perceptual characteristic of the semantic subject or object.

The program takes extended descriptions involving many objects, each of which is incorporated into the overall model. Once an object has been described, it is possible to interrogate the model about the relation of that object to any other in the model, *without* recourse to inference rules of the following kind: "if A is on B and B is in C then A is (probably) in C."

Experience with the Evaluation of Natural Language Question Answerers

Harry Tennant

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-18, Feb. 1979.

Research in natural language processing could be facilitated by thorough and critical evaluations of natural language systems. Two measurements, conceptual and linguistic completeness, are defined and discussed in this paper. Testing done on two natural language question answerers demonstrated that the conceptual coverage of such systems should be extended to better satisfy the needs and expectations of users. Three heuristics are presented that describe how conceptual coverage of question answerers should be extended.

JETS: Achieving Completeness through Coverage and Closure

Tim Finin, Brad Goodman, and Harry Tennant

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-19, Feb. 1979.

Work in progress on JETS, the successor to PLANES, is described. JETS is a natural language question answering system that is intended to interface users to a large relational data base. The architecture is designed to extend the conceptual coverage of JETS to better meet the conversational and data base usage requirements of users. The implementation of JETS is designed to gain a high degree of closure over concept

manipulation, contributing to a solution to the problems of perspicuity and scale. Specific examples are given on concept manipulation through the implied relationships of modification and of an approach to problem-solving through the use of frames.

Visual Analog Representation for Natural Language Understanding

David L. Waltz and Lois Boggess

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-20, Feb. 1979.

In order for a natural language system to truly "know what it is talking about," it must have a connection to the real world correlates of language. For language describing physical objects and their relations in a scene, a visual analog representation of the scene can provide a useful target structure to be shared by a language understanding system and a computer vision system. This paper discusses the generation of visual analog representations from input English sentences. It also describes the operation of a LISP program which generates such a representation from simple English sentences describing a scene. A sequence of sentences can result in a fairly elaborate model. The program can then answer questions about relationships between the objects, even though the relationships between the objects may not have been explicit in the original scene description. Results suggest that the direct testing of visual analog representations may be an important way to bypass long chains of reasoning and to thus avoid the combinatorial problems inherent in such reasoning methods.

Problem Solving in a Natural Language Environment

Bradley A. Goodman

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-22, July 1979.

The kinds of requests that can be currently handled by natural language data base systems are constrained mainly to simple queries to retrieve information from the data base. The requests must be completely specified by the user (though certain information can be assumed from past context). This paper is a proposal for a Ph.D. thesis that explores requests of a more complicated nature. The goal is to take vague and complex requests from users and turn them into well-defined problems. Missing information will be filled in through world knowledge or from the current dialog context. The transformation of the request into a well-defined problem and the generation of a plan to

solve the problem will be guided by a set of problem solving frames.

Relating Images, Concepts and Words

David L. Waltz

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-23, May 1979.

Examination of verbal descriptions of objects suggests that we use hierarchical structures for shape description; the highest levels of the hierarchy provide a general object framework or breakdown into component parts and a description of each part by analogy to a well-understood set of shapes called *prototypes*. Lower levels of the hierarchy provide refinement of the analogies and ways in which shapes deviate from the prototypes. The set of prototypes on which the analogies are based contains many common objects, especially natural objects and the parts of the human body, plus certain shapes with special symmetry properties. It is argued that no single 3-D representation scheme is natural for all members of this set of prototypes, and that since unfamiliar objects are described with respect to the basic set of shapes, these objects will have varying shape representation schemes also.

Generating and Understanding Scene Descriptions

David L. Waltz

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-24, March 1980.

This paper explores design issues for a system which has both vision and language, in particular, a system which addresses both the problem of selecting appropriate words and sentences to describe a particular perceptual event, and the related problem of making appropriate inferences about a natural language description of a perceptual event. It argues that perception is basically a *description-building* process, and that the understanding of scene descriptions is ultimately based on our ability to first use scene descriptions to drive processes of "*picture-building*", and then to drive processes of "*event-simulation*" which cause the "pictures" we build to mimic the dynamics of the world.

Syntactic Analysis in JETS

Harry Tennant

Coordinated Science Laboratory
University of Illinois
Urbana, Illinois 61801

Working Paper WP-26, May 1980.

Our experience with PLANES has lent evidence to the opinion that syntactic rules ought to be included in a natural language processing system. They were not explicitly included in PLANES, with desirable and undesirable results. The desirable result was that PLANES was quite capable at interpreting elliptical and non-grammatical utterances, which occur fairly frequently in user interactions. The undesirable result was that PLANES could not interpret many grammatical utterances whose interpretation would have been greatly facilitated by using syntactic information. This report is the result of an investigation and experiments into using an extension of the LUNAR grammar for JETS, the successor to PLANES. The goal is to gain the regularities of syntactic analysis, but still allow for elliptical and "non-grammatical" utterances.

The Semantic Interpretation of Noun-Noun Modification

Timothy W. Finin

Department of Computer and Information Science
Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Illinois CSL Working Paper WP-21, May 1979.

This paper is a proposal for a Ph.D. thesis concerned with the semantic interpretation of noun-noun modification. The study of this topic is being coordinated with the development of the JETS natural language question answering system. One of the goals of this research is a computer program which will interpret instances of noun-noun modification within the domain of discourse of JETS. Related work on noun-noun modification in the disciplines of Linguistics and Artificial Intelligence is described and contrasted to the proposed research. The basic approach being taken is described and the scope of the work is outlined.

The Semantic Interpretation of Nominal Compounds

Timothy W. Finin

Department of Computer and Information Science
Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Illinois CSL Working Paper WP-25, April 1980.

This paper briefly introduces an approach to the problem of building semantic interpretations of *nominal compounds*, i.e. sequences of two or more nouns related through modification. Examples of the kinds of nominal compounds dealt with are: "engine repairs", "aircraft flight arrival", "aluminum water pump", and "noun noun modification".

The Semantic Interpretation of Nominal Compounds

Timothy W. Finin

Department of Computer and Information Science
Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Illinois CSL Technical Report T-96, March 1980.

This report deals with one aspect of enabling machines to communicate with people in a natural language. The particular problem which is the focus of this work is the interpretation of *nominal compounds*, i.e. sequences of two or more nouns related through modification. Examples of the kinds of nominal compounds dealt with are: "engine repairs", "aircraft flight arrival", "aluminum water pump", and "noun noun modification".

The interpretation of nominal compounds can be divided into three intertwined subproblems: *lexical interpretation* (mapping words into concepts), *modifier parsing* (discovering the structure of strings with more than two nominals) and *concept modification* (assigning an interpretation to the modification of one concept by another). This last problem is the central one. The essential feature of this form of modification is that the underlying semantic relationship which exists between the two concepts is not explicit. Moreover, a large number of relationships might, in principal, exist between the two concepts. The selection of the most appropriate one depends on a host of semantic, pragmatic and contextual factors.

As a part of this research a computer program was written to build an appropriate semantic interpretation when given a string of nouns. This program has been designed as one component of the natural language question answering system JETS. The interpretation is done by a set of semantic interpretation rules. Some of the rules are very specific, capturing the meaning of idioms and canned-phrases. Other rules are very general, representing fundamental case-like relationships which can hold between concepts. A strong attempt has been made to handle as much as possible with the more general, highly productive interpretation rules.

The approach has been built around a frame-based representational system which represents concepts and the relationships between them. The concepts are organized into an abstraction hierarchy which supports inheritance of attributes. The same representational system is used to encode the semantic interpretation rules. An important part of the system is the *concept matcher* which, given two concepts, determines whether the first describes the second and, if it does, how well.

Cooperative Responses from a Portable Natural Language Data Base Query System

S. Jerrold Kaplan

Department of Computer Science
Stanford University
Stanford, California 94305

Univ. of Penn. Ph.D. Thesis, 1979.

This work describes the design and implementation of a Natural Language (NL) Data Base (DB) query system (named CO-OP) that provides cooperative responses to NL requests for data retrieval and operates with a CODASYL DB management system. Cooperative responses to questions address a questioner's goals and intentions beyond a literal, direct response while observing the conventions normally associated with discourse. For example, if a course CIS500 was not given in the Spring, '77 semester, the indirect response R1 to question Q is more cooperative than the direct, correct response R2.

Q: How many students failed CIS500
in Spring, '77?

R1: CIS500 was not given in Spring '77.

R2: Zero.

Providing cooperative responses in a computational setting requires a systematic approach to *pragmatic* issues - those aspects of the meaning of utterances that arise from the fact or context of use. A limited theory of cooperative behavior in question-answering is presented, and applied in a query system that is capable of various types of direct and indirect responses.

This work also explores problems in the representation of the knowledge required to produce such responses. The implementation of CO-OP illustrates that a DB, a DB schema and a suitably encoded lexicon are sufficient sources of domain specific knowledge to provide appropriate responses to a habitable class of simple NL questions. As a result, the system achieves a high degree of portability to new DB domains. Transcripts of the program's behavior on two radically different DB's are presented and analyzed.

In addition, issues in the production of useful paraphrases, effective error handling, transparency of DB update, and efficient path finding in DB schemas are discussed.

Premier Bilan D'Une Experience de Compréhension du Langage Naturel

D. Coulon, D. Kayser, J. Fuss, F. Jakob, M. Monfils

Laboratoire de Recherche en Informatique
Bâtiment 490
Université de Paris - Sud
F-91405 Orsay, FRANCE

Research Report No. 40, 1979.

After a brief presentation of our hypotheses on the understanding process, we describe how they are implemented in the experimental system of our mechanisms and the problem unsolved yet.

Acquisition de Connaissances a Partir de Textes: Une Expérimentation en Documentation Automatique

J.C. Bassano

Laboratoire de Recherche en Informatique
Bâtiment 490
Université de Paris - Sud
F-91405 Orsay, FRANCE

Research Report No. 41, 1979.

The proposed syntactic interference matching method induces abstractions in a powerful prototype information retrieval system implemented with data-base software. During the retrieval process the initial term assignment of the incoming information request might be automatically improved by adding synonymous or related terms and new term combinations or phrases in the query. Heuristics enable us to keep the search space to a manageable size and provide the search with direction by indicating which of the unused instances are to be used in a cycle of construction.

Les Grammaires Sémantiques, Outil Puissant Pour Interroger les Bases de Données en Langage Naturel

A. Bonnet

Laboratoire de Recherche en Informatique
Bâtiment 490
Université de Paris - Sud
F-91405 Orsay, FRANCE

Research Report No. 52, 1979.

Data bases usually provide a framework in which the relations between objects are sufficiently precise that there is no need for a complete analysis (i.e. syntactic) in order to interpret the requests of the user. Semantic grammars take into account the structures of the questions without explicitly using conventional syntactic categories such as nouns, verbs Instead, the categories intervening in the rules emphasize the domain of application. The use of semantics in the grammar makes the parsing process deterministic and therefore very efficient. Furthermore, such grammars can be rapidly implemented. This is because writing the formal clause resulting from the request (syntactic pattern/response expression correspondence) is fairly straightforward, even for those queries which have a relatively complex structure.

Über Dialogkohärenz in Natürlichsprachlichen AI-Systemen.

Walther v. Hahn

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

HAM-RPM Report No. 8, Oct. 1979.

It is pointed out that dialogue coherence involves the intentions of the speakers, not just particular features of the text. Aspects of coherence involving three areas of research are then discussed using the dialogue system HAM-RPM as an example: The cognitive capabilities of the natural language system (in particular, linguistic and metalinguistic capabilities, knowledge, and adaptive strategies), the coherence-supporting behaviour of the natural language partner (in particular, cooperation, metacommunication, and goal-directed action) and the text indices (in particular, anaphora and lexical contingency). The interdependencies among these three areas can be seen clearly in natural language AI systems.

Analysis by Means of Sentence Patterns: Examples for Which- and How Many- Questions

Wolfgang Hoepfner

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

Memo 10, Dec. 1979, (In German).

This memo gives a short introduction into the component called 'Analysis by means of sentence patterns' in HAM-RPM. The main part contains examples of WHICH- and HOW MANY- questions. Running the system in the 'protocol mode' the internal processing of these questions is described in detail.

Dialogue Sequences with the System HAM-RPM in Protocol Mode

Wolfgang Hoepfner and Wolfgang Wahlster

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

Memo 11, Jan. 1980, (In German).

This memo describes the highly flexible tracing and backtracing mechanisms for internal processing phases ('protocol mode') of the system HAM-RPM and summarizes its new linguistic, communicative and cognitive capabilities. A dialogue sequence about a living-room scene illustrates both the tracing facilities and the new capabilities.

Composite Objects: Representation and Inferences

Wolfgang Hoepfner

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

Report 15, March 1980, (In German).

This report addresses problems concerning the conceptual representation of composite physical objects. To define part-whole relations in the framework of an associative semantic network a simple set of primitive arcs and simple inference rules are introduced. Intensional and extensional arcs provide the natural language dialogue system HAM-RPM with general and specific knowledge about objects, respectively. The main disadvantage of these simple constructs appears to be the unrestricted transitivity of the part-whole relation. The approach taken in HAM-RPM is discussed together with proposals concerning related problems. Examples illustrate the way the system uses its knowledge when answering HOW MANY- questions. Finally extensions to the existing representation structures concerning composite physical objects are proposed.

Towards a Computational Model for the Semantics of Why-questions

Wolfgang Wahlster

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

HAM-RPM Report No. 16, June 1980.

This paper discusses aspects of a computational model for the semantics of why-questions which are relevant to the implementation of an explanation component in a natural language dialogue system. After a brief survey of all of the explanation components which have been implemented to date, some of the distinguishing features of the explanation component designed and implemented by the author are listed. In the first part of the paper the major types of signals which, like the word 'why', can be used to set the explanation component into action are listed, and some ways of recognizing them are considered. In addition to these linguistic signals, communicative and cognitive conditions which can have the same effect are discussed. In the second part the various schemata for argumentative dialogue sequences which can be handled by the explanation component in question are examined. Particular attention is paid to the problems arising in connection with the iteration of why-questions and the verbalization of multiple justifications. Finally schemata for metacommunicative why-

questions and for why-questions asked by the user are investigated.

The Natural Language System HAM-RPM as a Hotel Manager:

Some Representational Prerequisites

Anthony Jameson, Wolfgang Hoepfner, and Wolfgang Wahlster

Universität Hamburg
Germanisches Seminar
Von-Melle-Park 6
D-2000 Hamburg 13 WEST GERMANY

HAM-RPM Report No. 17, Aug. 1980.

The dialogue system HAM-RPM simulates a person conversing about one of several interchangeable scenes in colloquial German. Attention is currently being focused on a hotel reservation situation in which the system is required to construct and make use of a model of its dialogue partner's goals and beliefs. This situation calls for a semantic representation language which is declarative and which permits the representation, among other things, of various numerical and natural language quantifiers, conjoined noun phrases and belief modifiers. This paper describes this language and sketches the processes in the course of which expressions of the language are analysed, constructed, transformed, or evaluated. The system's behavior is described in detail to illustrate the relationships among these processes.

SWYSS - A Natural Language Question-Answering System for Scene-Analysis

Peter Schefe and Bernd Pretschner

Fachbereich Informatik
Universität Hamburg
Schlüterstr. 70
D-2000 Hamburg 13 WEST GERMANY

Technical Report, 1979.

The adequate representation of a universe of discourse in a computational system may be tested by introducing a sensory channel to the "real world". SWYSS ("Say-what-you-see-system") is a natural language system connected to a TV-camera gathering snapshots analysed by a scene analysis component. Since there is a continuum of object shapes and other property values, the uncertainty of definitions becomes a main problem of communication. It is described how representations of natural language inputs and pictorial input are computed and tied together. At present, the system is focused on the processing of natural language queries containing vague descriptors and imprecise quantifiers. The answers of SWYSS are in elaborate German as well as the questions.

On Foundations of Reasoning with Uncertain Facts and Vague Concepts

Peter Schefe

Fachbereich Informatik
Universität Hamburg
Schlüterstr. 70
D-2000 Hamburg 13 WEST GERMANY

Report No. 56, 1979.

"Fuzzy sets theory" and "fuzzy logic" based on the former have become of rapidly increasing interest. The foundations, however, are still disputed. Especially the definitions of some set operations and logical connectives appear to be somewhat arbitrary. The relations of "fuzziness" to "probability" and "possibility" are not yet clear. This paper contains an outline of a probabilistic foundation of multi-valued ("fuzzy") reasoning. The fundamental concept is "agreement probability". It is shown that some undesirable consequences of "fuzzy logic", e.g., that tautologies of propositional calculus are not preserved, can be avoided. A proposal for alternative definitions of "degree of membership" and operations on membership-graded sets are given.

"Fuzziness" is interpreted as a subjectivistic concept, i.e., subjective uncertainty pertaining to the truth of a proposition. An important consequence thereof is that, from a graded agreement associated with a conjecture, an agreement degree pertaining to its negation cannot be computed.

According to this foundation Shortliffe's model of medical diagnosis is reviewed as an application paradigm. There is no fundamental disagreement with Shortliffe's interpretation. However, Zadeh's "linguistic modelling" is shown to be inadequate. "Fuzziness" of linguistic concepts is interpreted as uncertainty of the applicability of a predicate in a given situation. This leads to the conclusion that the definition of derived concepts, especially, of hedged expressions referring to continuous scales cannot be modelled using Zadeh's fuzzy set operations. Experimental findings of Hersh and Caramazza support an alternative interpretation. Particularly, Zadeh's conjecture that truth values can be equated with membership degrees is shown to be inadequate. Alternative interpretations of linguistic phenomena considered and of the sorites paradox are given. Especially the meta-linguistic character of the phenomena is emphasized. It is argued that "vagueness" and "uncertainty" should be clearly distinguished, as well as "possibility" and "applicability". Suggestions are made how underlying measuring scales and orderings of objects are used in reasoning processes involving vague concepts.

On Representing Uncertainty in AI-Systems**Peter Scheffé**

Fachbereich Informatik
 Universität Hamburg
 Schlüterstr. 70
 D-2000 Hamburg 13 WEST GERMANY

Note 69, 1979.

AI systems dealing with natural environments and texts have to cope with "fuzziness", i.e., factual uncertainty and uncertainty of meaning. A formal model of "fuzzy reasoning" is proposed. It is based on the notion of "agreement probability" of a proposition. Combining and inference rules are defined similar to, but definitely deviating from, those in probability theory. Epistemologically, "agreement probability" is interpreted as subjective certainty or uncertainty of belief in the truth (or falsehood) of a proposition.

According to this, some theories and theoretical assumptions underlying AI systems are reviewed. Especially, MYCIN's inference rules conceived of on a purely intuitive base are shown to be valid in general. However, systems of language and story understanding do not cope with uncertain inferences adequately as yet. Suggestions are made how Schank's "conceptual inference" and some aspects of Wilks' "preference semantics" can be reinterpreted and provided with a more coherent model of inference making under uncertainty.

Arten von Wissen und Inferenzen in Natürlichsprachlichen Systemen**Peter Scheffé**

Fachbereich Informatik
 Universität Hamburg
 Schlüterstr. 70
 D-2000 Hamburg 13 WEST GERMANY

Note 73, 1979.

Since "intelligent" information systems using natural language interfaces have to be considered formal systems, philosophically, there is no point to ascribe to these systems natural capabilities corresponding to human understanding, intelligence, and knowledge. Of course, the implementability of a formal system does not support the contrary. On the other hand, the possible usefulness of such systems should not be doubted in general. There are problems of formal representations of knowledge and belief pertaining to an adequate epistemological taxonomy. Different kinds of knowledge and belief require different formal systems for representation and inference making. Definiteness - undefiniteness, preciseness - unpreciseness, and completeness - incompleteness are the main dimensions pertaining to the epistemological state of the knowledge base. Models available so far are discussed. Beyond that, knowledge may be given in vague terms, such that there is a tolerance space for a set of com-

patible assertions. The notions of "definitional uncertainty" and "prototype" can be accounted for by a probabilistic model of "tolerant" knowledge. For all kinds of knowledge addressed, conditions of inference making are given, and implementability is demonstrated using the programming languages FUZZY and AMORD.

Sprachliche Bildinterpretation für ein Frage-Antwort-System**Karl-Jürgen Hanßmann**

Fachbereich Informatik
 Universität Hamburg
 Schlüterstr. 70
 D-2000 Hamburg 13 WEST GERMANY

Note 74, 1980.

A component of a system (SWYSS - 'Say what you see system') is described that generates a linguistic representation of a two-dimensional scene from the output of a scene analysis component which recognizes objects by use of prototypes. Most of the linguistically possible spatial relations holding among the objects are computed. These computations are triggered by natural-language queries, using a knowledge base, except for relations pertaining to gravitation. Most of the relations are vague, i.e., their applicability can be graded. For these relations, a numerical degree of applicability is computed which can be reflected in natural language expressions. The procedures for computing the spatial relations are implemented in FUZZY and LISP.

KNET: An Extended SI-Net Formalism for Knowledge Representation Systems**Michael W. Freeman**

ADO/FSSG
 Burroughs Corporation
 P.O. Box 517
 Paoli, Pennsylvania 19301

Technical Report TR 80-1, Jan. 1980.

We discuss in this paper four areas where we feel additional epistemological primitives should be introduced into Brachman's Structured Inheritance Network formalism. These concern respectively: definitional vs. descriptive concepts, generic vs. individuated description, coreference vs. denotation, and the representation of dynamic attributes as event logs and monitors.

Representing Type Hierarchies**Freeman Rawson**

IBM General Systems Division
 Boca Raton, Florida 33432

IBM SRI Technical Report TR 73-006, Jan. 1980.

Symbol mapping is the problem of associating properties with an object based upon typed informa-

tion about the object so that the properties can be inherited from general class to specific instances. The reference window problem is the problem of finding an individual object given a description of the object. Both of these problems cannot be solved efficiently using the standard methods of procedural inference found in PLANNER and CONNIVER. However, by using conceptual graph theory and organizing the data base using type information, both of these problems can be handled with reasonable efficiency.

The conceptual graph search algorithms and the methods available in PLANNER and CONNIVER are of equal power in the sense that they both derive the same inferences. Using some methods from concrete complexity theory, the efficiency of both the PLANNER approach and the conceptual graph approach are analyzed, and formulas are derived for the average number of nodes searched.

An English Parser

Jonathan L. Handel

IBM Systems Research Institute
205 East 42nd Street
New York, New York 10017

Technical Report TR 73-005, Jan. 1980.

ENGLISH is a purely syntactic parser for English sentences. It was originally designed as an illustrative program for a course in linguistics at the IBM Systems Research Institute. The program has the following features: high speed, generation of all parses in the grammar, interactive execution, and good coverage with a small grammar. This report is a user's guide and reference source for the ENGLISH program.

An Ideational Parser

Alan L. Tharp

Computer Science Department
North Carolina State University
Raleigh, North Carolina 27650

Jeffrey F. Eastman

Hewlett-Packard Corporation
Loveland, Colorado

SIGLASH Newsletter 12, 1 (March 1979), 17-36.

Recent parsers have extended the variety and difficulty of natural language constructs recognizable by a computer. However, one shortcoming of existing parsers is that as the complexity of the sentences recognized increases, the computational complexity increases quadratically. One reason for this computational explosion is that much if not most of the control information is embedded in the parser.

A different approach, referred to as an ideational parser, is proposed. Although the design for this parser was motivated by the manner in which people might use language, it is not necessarily intended to be

a model for human cognition. The control information, for the most part, is removed to the lexicon, and the words are considered *operators* in the information of a mental picture rather than as *operands* to the parser. The parser mechanism is detailed with an example parse and a typical conversation is given in the appendix.

Two primary benefits of the ideational parser are 1) an improved, simpler control mechanism and 2) the ability to acquire new knowledge which is automatically stored in the same format as the given knowledge.

Automatic Resolution of Linguistic Ambiguities

B.K. Boguraev

Computer Laboratory
University of Cambridge
Corn Exchange Street
Cambridge, CB2 3QG ENGLAND

Technical Report No. 11, 1979.

The thesis describes the design, implementation and testing of a natural language analysis system capable of performing the task of generating paraphrases in a highly ambiguous environment. The emphasis is on incorporating strong semantic judgement in an augmented transition network grammar: the system provides a framework for examining the relationship between syntax and semantics in the process of text analysis, especially while treating the related phenomena of lexical and structural ambiguity. Word-sense selection is based on global analysis of context within a semantically well-formed unit, with primary emphasis on the verb choice. In building structures representing text meaning, the analyser relies not on screening through many alternative structures - intermediate syntactic, or partial semantic - but on dynamically constructing only the valid ones. The two tasks of sense selection and structure building are procedurally linked by the application of semantic routines derived from Y. Wilks' preference semantics, which are invoked at certain well chosen points of the syntactic constituent analysis - this delimits the scope of their action and provides context for a particular disambiguation technique. The hierarchical process of sentence analysis is reflected in the hierarchical organization of application of these semantic routines - this allows the efficient coordination of various disambiguation techniques, and the reduction of syntactic backtracking, non-determinism in the grammar, and semantic parallelism. The final result of the analysis process is a dependency structure providing a meaning representation of the input text with labelled components centered on the main verb element, each characterized in terms of semantic primitives and expressing both the meaning of a constituent and its function in the overall textual unit. The representation serves as an input to the generator, organized around the same underlying

principle as the analyser - the verb is central to the clause. Currently the generator works in paraphrase mode, but is specially designed so that with minimum effort and virtually no change in the program control structure and code it could be switched over to perform translation.

The thesis discusses the rationale for the approach adopted, comparing it with others, describes the system and its machine implementation, and presents experimental results.

Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with ATN's

Fernando C. N. Pereira and David H. D. Warren

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

Artificial Intelligence 13, 3 (May 1980), 231-278.

A clear and powerful formalism for describing languages, both natural and artificial, follows from a method for expressing grammars in logic due to Colmerauer and Kowalski. This formalism, which is a natural extension of context-free grammars, we call "definite clause grammars" (DCG's).

A DCG provides not only a description of a language, but also an effective means for analysing strings of that language, since the DCG, as it stands, is an executable program of the programming language PROLOG. Using a standard PROLOG compiler, the DCG can be compiled into efficient code, making it feasible to implement practical language analysers directly as DCG's.

This paper compares DCG's with the successful and widely used augmented transition network (ATN) formalism, and indicates how ATN's can be translated into DCG's. It is argued that DCG's can be at least as efficient as ATN's, whilst the DCG formalism is clearer, more concise and in practice more powerful.

Toward A PROLOG Text Grammar

Georgette Silva and Don Dwiggin

Operating Systems, Inc.
21031 Ventura Blvd., Suite 1200
Woodland Hills, California 91364

SIGART Newsletter 73 (Oct. 1980), 20-25.

This paper focuses on the syntactic and semantic processing components -- the Text Grammar -- of an application-oriented experimental natural language understanding system, called MATRES, currently under development at Operating Systems Inc. The primary application for this system is the automated generation and update of data bases from natural language text. The current version of the Text Grammar

is implemented in the programming language PROLOG, which provides a unified and powerful environment capable of representing the structures and processes required for the analysis of natural language text.

Our methodological approach combines concepts from Text Grammar theory with Artificial Intelligence techniques for discourse analysis. We subscribe to the hypothesis that as people read and understand text, they construct a multi-level mental representation of its content, with the most concrete level at the bottom of the conceptual structure. Thus, at the lowest level of this mental structure are the sentences and phrases of the text, while the representation becomes more concise and abstract at higher levels. This concept of understanding text is embodied in our PROLOG text grammar.

Our approach is centered around the notion of event, and utilizes two major knowledge sources: (1) a model of the sublanguage that characterizes the domain of application and (2), a model of the entities and relations characteristic of this domain. These knowledge sources are used by our text grammar to derive content representations approximating a human's understanding of a text.

Logic Programming and Compiler Writing

David Warren

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

Softw. Pract. Exper. 10, 2 (Feb. 1980), 97-125.

The concept of "logic programming," and its practical application in the programming language PROLOG, are explained from first principles. The ideas are illustrated by describing in detail one sizable PROLOG program which implements a simple compiler. The advantages and practicability of using PROLOG for "real" compiler implementation are discussed.

Translation of Phrase Structured Programming Languages

William Buttelmann

Computer and Information Science
Ohio State University
Columbus, Ohio 43210

Technical Report AFOSR-TR-80-0088, Dec. 1979.

An integrated formal theory of phrase structure linguistic descriptions has been developed, based on classical phrase linguistic theory and on a new formal theory of phrase structure semantics developed by this research project. The theory provides for both context-sensitive syntax and semantic structure, interpretations, the meaning of certain linguistic entities

such as morphemes, phrases, sentences, and statements, and specifies the nature of and relationship between syntactic and semantic ambiguity. The theoretical model of syntax and semantics was developed in order to study computational aspects of semantic interpreters and language translators. Based on the model, a theory of interpretation and translation was constructed, and a number of results in these two areas have been developed. It was shown how linguistic properties can be treated either as context-sensitive or context-free and either as semantic or syntactic information. This theory establishes the basis for further research into the computational complexity of semantic interpreters and especially into the problem of minimizing the complexity of interpreters by proper trade-offs between syntax and semantics and/or 'context-sensitive'-ness and 'context-free'-ness.

A System for Natural Language Computation

**Alan W. Biermann, Bruce W. Ballard,
and Anne M. Holler**

Department of Computer Science
Duke University
Durham, North Carolina 27706

SIGLASH Newsletter 12, 1 (March 1979), 6-16.

This paper briefly describes a natural language programming system called NLC which allows users to type commands in English and watch the resulting actions carried out on a computer display screen. The data entities appear on the screen at all times, so that the user can observe the effect of each input before proceeding to the next one. Should some input ever be incorrectly processed, a backup facility allows requests to be clarified after restoring the previous context and data values. The user may identify a sequence of natural language commands as a procedure which, when given a name, acts as a new imperative verb and may be used to define further customizations.

The design of NLC stresses reliability and depth of linguistic coverage, so that the facilities provided can be employed arbitrarily. For the purpose of testing a prototype version of the system, the domain of matrix elements and scalar variables was selected. For instance, in a recent test of the system, a paid subject was asked to use NLC to compute the final averages for a hypothetical class of six students using a specified formula. The subject requested of the system that a matrix be displayed and that the columns be labeled. He then labeled the rows of the matrix with the students' names and entered their grades.

Statistical Techniques for Free-Text Processing

John M. Morris

PAR Corporation
228 Liberty Plaza
Rome, New York 13440

SIGLASH Newsletter 13, 2 (June 1980), 14-32.

Over the past eight years we have developed statistical methods for characterizing, classifying, and retrieving brief natural-language messages. Our goal was to provide a tool for people who had to deal with enormous numbers of heterogeneous documents, using ill-defined criteria of relevance and interest. Initially, we worked with a large, general-purpose system, the On-Line Pattern Analysis and Recognition System (OLPARS). More recently, we have developed a system called Message Extraction Through Estimation of Relevance (METER). Our present aim is the construction of a Testbed system for further studies and the development of more complex systems.

Anaphora in Natural Language Understanding: A Survey

Graeme Hirst

Department of Computer Science
Box 1910, Brown University
Providence, Rhode Island 02912

UBC Technical Report 79-2, May 1979.

A problem that all computer-based natural language understanding systems encounter is that of linguistic reference, and in particular anaphora (abbreviated reference). This report is an extensive review of the reference and anaphor problem, and the approaches to it that natural language understanding systems have taken, from early systems such as STUDENT through to current discourse-oriented ones such as PAL.

The problem is first examined in detail, and examples are given of many different types of anaphor, some of which have been overlooked by previous authors. The approaches taken in traditional systems are then described and abstracted and it is shown why they were inadequate, and why discourse theme and anaphoric focus need to be taken into account. The strengths and weaknesses of current anaphora theories and approaches are evaluated. The report closes with a list of some remaining research problems.

The report has been written so as to be as comprehensible as possible to both AI workers who know no linguistics, and linguists who have not studied artificial intelligence.

An Artificial Intelligence Approach to Natural Language Teaching

Larry H. Reeker

University of Queensland
St. Lucia, Queensland, AUSTRALIA

Aust. Comput. Sci. Commun. 2, 1 (Jan. 1980), 97-106.

This paper develops the rationale and outlines the structure of a program, LANGLAB, for computer-aided instruction in a second language. Although the first stage of the program is based on conventional teaching principles, the second stage uses an "intelligent" tutor to guide the student into more advanced structures of the language. Essentially, the program mimics situations that appear to be important in a first language acquisition situation, providing a rich, structured environment without overly restricting the student's ability to move forward into new material.

Dylan Thomas, the Craftsman: Computer Analysis of the Composition of a Poem

Mary Dee Harris Fosberg

Central State University
Edmond, Oklahoma

ALLC Bulletin 7, 3 (1979), 295-300.

Over 400 pages of manuscript exist of Dylan Thomas's "Poem on His Birthday," published in its final form with 108 lines. These manuscripts provide evidence of the poet's work while composing the poem. Computer analysis of this material first required designing an encoding scheme to capture the two dimensionality of the manuscript page and translate it into a linear string of characters. When the information from the pages had been transcribed into computer readable form, several collations were made using the published version of the poem as copy text. These collations produced a history of the composition process. Thomas's use of Roget's Thesaurus was analyzed to reveal some of his methods of choosing words. The data from the manuscripts has been organized into a database on disk for convenient access to the poetry to facilitate further research. Later study will include statistical analysis of metric and sound patterns to learn more about Thomas's method of composing the poem.

Tolvukonnun a Tidni Orda og Stafa i Islenskum Texta

Baldur Jonsson, Bjorn Ellertsson,
and Sven P. Sigurosson

Raunvisindastofnun Haskolans
Science Institute
University of Iceland
Dunhaga 3
Reykjavik ICELAND

Report RH-80-12, Oct. 1980.

In 1972, on the initiative of the first author, the Science Institute and the Institute of Nordic Linguistics (both at the University of Iceland) began working together on a project in linguistic computing. The main purpose was to establish the feasibility of carrying out such a project in Iceland. It was decided to try to make a frequency study at the graphic level of the novel *Hreiðio* by Olafur Johann Sigurosson, published in 1972, and to aim at producing the following: (1) Word-frequency lists in alphabetic order, reverse order, and order of decreasing frequency. (2) A complete concordance to the text. (3) A report whose objective was in turn threefold: a) to account for the execution of the project, b) to include results in tabular form on frequencies of letters, word-structures, and related matters, c) to serve as an introductory guide to anybody wishing to carry out a similar study.

The word-frequency lists have been published in ten copies (Baldur Jonsson 1975). The concordance was line-printed in three copies in 1976 and later on made available for public use with the addition of a preface and a title page (Baldur Jonsson 1978). The publication of the present report has - for various personal reasons - been considerably delayed. In this summary we mainly cover the second objective of the report.

Chapter 1 describes the general progress of the project. Chapter 2 refers to previously published frequency counts of Icelandic texts. Chapter 3 accounts for some linguistic concepts and terms used in the succeeding chapters. Chapter 4 describes the key-punching rules used in this project. Chapter 5 contains a description of the programs used and some efficiency considerations. Chapter 6 contains samples from the published word-frequency lists. The main tabular information is contained in Chapter 7. In Chapter 8 the concordance is discussed and samples from it are shown.