

## Abstracts of Current Literature

Following are the titles and abstracts of reports and memos recently published by the Research Unit for Information Science and Artificial Intelligence at the University of Hamburg. Single copies of papers still in print are available, free of charge. Please write

Ms. Angela Carstensen  
 Research Unit for Information Science  
 and Artificial Intelligence  
 University of Hamburg  
 Mittelweg 179  
 2000 Hamburg 13  
 West Germany

### Parsing Interactions and a Multi-level Parser Formalism Based on Cascaded ATNs

Th. Christaller, D. Metzger

*Report No. GEN-11, November 1982.* In Sparck Jones, K. and Wilks, Y., Eds., *Automatic Natural Language Parsing*. Horwood, Chichester, 1983.

We first provide a short linguistic motivation for the need for parsing formalisms more general than, for example, ATNs. We then present such a formalism using as an example a cascaded ATN developed according to the object-oriented computation paradigm. Central to this part of our paper is the formulation of a correct backtracking algorithm for CATNs. We then describe some generalisations that proved to be necessary for this formalism to be applicable in the development of dialogue grammars, which was our original concern. These include the possibility of using several types of processes (not necessarily ATNs) at every stage of the cascade; a solution to the problem of linearising tokens between different stages; and an algorithm for specifying when certain stages may be bypassed.

### Knowledge-based NL Access to Different Domains of Discourse with the AI-System HAM-ANS

Th. Christaller, W. v.Hahn, W. Hoepfner, H. Marburger, K. Morik, B. Nebel, W. Wahlster

*Report No. ANS-12.* In Slama, R., Ed., *Workshop Sprachverarbeitung, December 1982*. GMD Bonn, 1982: 100-135.

This report describes the overall goals of the natural language system HAM-ANS (*HAM*burg *App*lication-oriented *N*atural *L*anguage *S*ystem), which is currently being developed at the University of Hamburg. The project's more specific research topics are divided into application demands (e.g., optimization, portability) and basic research (e.g., explanation component, explicit partner model). HAM-ANS encompasses three different application classes: natural language access to a vision system (traffic at a street crossing), to a data base system (fishery data), and for guiding a competi-

tive dialogue with a client (hotel reservation situation). We introduce the system's general architecture and knowledge sources, and describe two processing components in greater detail: the semantic-oriented ATN Parser and the strategy for evaluating the internal representation of a user's question. The final section contains implementation data and a description of modifications of the LISP/FUZZY programming environment. The appendix shows the processing of an example utterance throughout the system and excerpts from its various external knowledge sources.

### An Object-oriented Approach for the Realization of Complex Control Structures

Th. Christaller

*Report No. GEN-13.* In Stoyan, H. and Wedekind, H., Eds., *Objekt-orientierte Software- und Hardwarearchitekturen*. Teubner, Stuttgart, 1983: 300-318.

The use of object-oriented programming languages is often equated to the user of abstract data types. This is only one aspect of object-oriented programming. In this paper we show how to realize simple and complex control structures in a clear and comprehensible way using objects and the computing paradigm based on message passing. The given examples are presented in terms of the flavor system which enables one to program in LISP in an object-oriented style. We describe two different implementation techniques of control structures – by means of the conditional and simple iteration schemes. We then present very general forms of backtracking and co-routines, whose usefulness relies on the fact that the flavor system allows multiple inheritance of information between flavors. This is demonstrated by sketching the implementation of an interpreter for cascaded recursive transition network grammars, in order to show how to combine control structures of different types without one of them predominating.

### Market Inquiry for Natural Language Access System

K. Morik

*Report No. ANS-14, February 1983.*

Application-oriented work on natural language systems (NLS) is not only to be justified by linguistic or software ergonomic considerations, but also by the needs and requirements of the users and customers. While there are several user studies, systematic evaluation studies for particular NLS, and experience with applied NLS, no study has yet been published that determines the demand for NLS and the requirements of potential customers. This report presents in detail the results from a market inquiry in the field of German NLS. The statistical data from the qualitative inquiry and some examples stemming from the quantitative inquiry

are analyzed. It is shown that customers expect NLS to go beyond the limitations of question answering.

### **Over-answering Yes-No Questions: Extended Responses in a NL-Interface to a Vision System**

**W. Wahlster, H. Marburger, A. Jameson, S. Busemann**

*Report No. ANS-15. In Proceedings of the 8th IJCAI. Karlsruhe, 1983: 643-646.*

The paper addresses the problem of over-answering yes-no questions, i.e. of generating extended responses that provide additional information to yes-no questions that pragmatically must be interpreted as wh-questions. Although the general notion of extended responses has already been explored, our paper reports on the first attempt to build a NL system able to elaborate on a response as a result of anticipating obvious follow-up questions, in particular by providing additional case role fillers, by using more specific quantifiers and by generating partial answers to both parts of questions containing coordinating conjunctions. As a further innovation, the system explicitly deals with the informativeness-simplicity trade-off when generating extended responses. We describe both an efficient implementation of the proposed methods, which use message passing as realized by the FLAVOR mechanism and the extensive linguistic knowledge incorporated in the verbalization component. The structure of the implemented NL generation component is illustrated using a detailed example of the system's performance as an interface to an image understanding system.

### **Beyond Domain-Independence: Experience with the Development of a German Language Access System to Highly Diverse Background Systems.**

**W. Hoepfner, T. Chirstaller, H. Marburger, K. Morik, B. Nebel, M. O'Leary, W. Wahlster**

*Report No. ANS-16. In Proceedings of the 8th IJCAI. Karlsruhe, 1983: 588-594.*

For natural language dialog systems, going beyond domain independence means the attempt to create a core system that can serve as a basis for interfaces to various application classes that differ not only with respect to the domain of discourse but also with respect to dialog type, user type, intended system behavior, and background system. In the design and implementation of HAM-ANS, which is presently operational as an interface to an expert system, a vision system and a data base system, we have shown that going beyond domain independence is possible. HAM-ANS is a large natural language dialog system with both considerable depth and breadth, which accepts typed input in colloquial German and produces typed German responses quickly enough to make it practical for real-time interaction. One goal of this paper is to report on the lessons learned during the realization of the

system HAM-ANS. This paper introduces the overall structure of the system's processing units and knowledge sources. In addition we describe some of the innovative features concerning the strategy of semantic interpretation.

### **Explanation Components as Dialog Tools**

**W. Wahlster**

*Report No. ANS-17. In Office Management. Sonderheft, 1983: 45-48.*

This paper discusses the relevance of explanation components for knowledge-based dialog systems and motivates the need for such modules. First, the various tasks of an explanation component are described and a general architecture for such a component is proposed. Then the broad range of applications of explanation components is illustrated by several implemented examples from the current AI literature. In particular, the explanation components of a backgammon program (QBKG), a computer-aided design system (CADHELP) and our natural language system HAM-ANS are examined. As a negative example we discuss the explanation component of a nuclear-power plan consultant and identify some serious flaws in its design. Finally, six guidelines for the design of explanation components summarize the main results of the paper.

### **The Contribution of Artificial Intelligence to the Human Factors of Application Software**

**W. v.Hahn**

*Report No. GEN-18. In Blaser, A. and Zoepfritz, M., Eds., *Enduser Systems and their Human Factors*. Springer, Heidelberg, 1983: 128-138.*

This paper deals with some mechanical, cognitive and social aspects of ergonomics in the following fields of Artificial Intelligence: vision, robotics, theorem proving, speech recognition, natural language processing. A few benefits that are now visible are shown. To evaluate the use of natural language AI-systems, research (1) must consider the natural (language) environment of the problem-solving process with its heterogeneous types of information and (2) must appreciate highly restricted language utterances as linguistically appropriate to specific task environments. The benefits of future natural language systems depend on three conditions.

- dominance of cognitive and communicative abilities over linguistic ones,
- realization of an elementary, dynamic partner model,
- transparency, as provided by an explanation component.

### **Demand and Requirements for Natural Language Systems. Results of an Enquiry.**

**K. Morik**

*Report No. ANS-19.* Extended version of a paper in *Proceedings of the 8th IJCAI*. Karlsruhe, 1983: 647-649.

Application-oriented work on natural language systems (NLS) is not only to be justified by linguistic or software ergonomic considerations, but also by the needs and requirements of users and customers. While several user studies exist, systematic evaluation studies for particular NLS, and experience with applied NLS, no study has yet been published that determines the demand for NLS and the requirements of potential customers. This article presents in detail the results from a market inquiry in the field of German NLS. The statistical data from the qualitative inquiry and examples stemming from the quantitative inquiry are analyzed.

### **The Dialogue System HAM-ANS: what it is based on, how it works and to whom it responds**

**W. Hoepfner, K. Morik**

*Report No. ANS-20.* In *Linguistisch Berichte*. 88, December 1983.

This contribution has two objects. In the first section, the position of language-oriented research in the field of Artificial Intelligence (AI) and its relevance to related disciplines is more closely defined and our theoretical approach in the development of HAM-ANS (Hamburg Application-Oriented Natural Language System) is outlined. In the following two sections, this system is presented with an emphasis on aspects which are of theoretical interest. Although details of implementation are for the most part omitted, it is to be emphasized that all of the processes and knowledge bases described have been fully implemented.

After an introduction to the three dialogue situations investigated in our project, the organization of the system as a whole is elucidated on the basis of examples taken from one of the domains of discourse: A Hotel Reservation Situation. The final section is devoted to a special problem of pragmatics: representation and application of user modelling in an AI system.

### **Case Role Filling as a Side Effect of Visual Search**

**H. Marburger, W. Wahlster**

*Report No. ANS-21.* In *Proceedings of the First EACL Meeting*. Pisa, 1983.

This paper addresses the problem of generating communicatively adequate extended responses in the absence of specific knowledge concerning the intentions of the questioner. We formulate and justify a heuristic for the selection of optional deep case slots not con-

tained in the question as candidates for the additional information contained in an extended response. It is shown that, in a visually present domain of discourse, case role filling for the construction of an extended response can be regarded as a side effect of the visual search necessary to answer a question containing a locomotion verb. The paper describes the various representation constructions used in the German language dialog system HAM-ANS for dealing with the semantics of locomotion verbs and illustrates their use in generating extended responses. In particular, we outline the structure of the geometrical scene description, the representation of events in a logic-oriented semantic representation language, the case-frame lexicon and the representation of the referential semantics based on the Flavor system. The emphasis is on a detailed presentation of the application of object-oriented programming methods for coping with the semantics of locomotion verbs. The process of generating an extended response is illustrated by an extensively annotated trace.

### **Natural Language Data Base Access with HAM-ANS: Syntactic Correspondence, Natural Language Quantification and Semantic Model of the Domain of Discourse**

**H. Marburger, B. Nebel**

*Report No. ANS-22.* In Schmidt, J.W., Ed., *Sprachen für Datenbanken*. Berlin, 1983: 26-41.

This article describes our approach to the solution of three basic problems which arise in interfacing a natural language system to a background system (especially to a relational data base management system): the communication between the systems, the translation process from the internal representational language of the natural language system into the data base query language, and the representation of conceptual knowledge about the domain of discourse.

The coupling of the systems is elucidated with the aid of the ISO reference model of 'open system interconnection'. Syntactic transformation is demonstrated using the treatment of natural-language quantifiers as an example. Finally the structures for representing conceptual knowledge are discussed with reference to a selected part of a data base schema.

### **The Semantics and Pragmatics of Answers: An Annotated Bibliography**

**A. Jameson**

*Memo No. GEN-12*, October 1982.

This bibliography focuses selectively on works within the fields of linguistics, philosophy, and artificial intelligence which, at least in part, treat some aspect of the semantics or pragmatics of answers or question answering. The relevant parts of each of the 75 works listed are annotated. The systematic index which pre-

cedes the annotations is organized around the main categories 'Answer Classifications', 'Pragmatic Analyses', 'Problems of Answer Formulation', 'Answers to Particular Types of Questions', and 'Analytical Frameworks'. The introduction explains the categories used in the index and gives a brief overview of the literature covered.

### Is LISP a 'Slow' Language?

**B. Nebel**

*Memo No. ANS-13, February 1983.*

An experiment is described which compares runtimes of a large collection of PASCAL- and LISP-programs on a DEC-system-10 computer. The most important result of this comparison is that in most cases compiled LISP-programs are as fast as the equivalent PASCAL-programs.. This result gives rise to the conclusion that in a similar setting (machine and quality of the language processors) one is not constrained to the use of conventional programming languages because of runtime-efficiency.

### Evaluative Utterances and the Representation of Valuations

**K. Morik**

*Memo No. ANS-14, March 1983.*

Evaluative utterances and their use in dialogues are investigated. Adequate representations of such utterances are discussed, and two approaches are compared: the goal approach and the feature approach.

### PEVAL: The Interface between the HAM-ANS Core System and PASCAL/R Databases

**M. O'Leary**

*Memo No. ANS-15, April 1983.*

The PEVAL package is a syntax-directed translation system that serves as the interface between the HAM-ANS core system and a PASCAL/R database system. This paper describes the details of the PEVAL package. It will be useful for readers who want to become acquainted with the package as one of the components of the HAM-ANS system, but it is especially intended for programmers who will extend the current version to process more complex input expressions and to generate more varied responses. This report therefore focuses on the functions of the PEVAL package themselves, outlining their algorithms, describing their present abilities, and giving suggestions for extensions of these abilities.

### Dialog Sequences with the System HAM-ANS

**W. Hoepfner, H. Marburger**

*Memo No. ANS-16, May 1983.*

This memo contains examples of dialogues with the German natural language access system HAM-ANS

(Hamburg Application Oriented Natural Language System) in three different applications classes: First the system takes the role of a hotel manager and responds to questions of a client concerning details of a hotel room. In the second dialogue the system accesses data from a vision system, the specific domain of this application being a sequence of images which were recorded at a street intersection. The third dialogue is concerned with natural language access to a large PASCAL/R database containing fishery data. The internal processing of each user utterance is annotated by the system's built-in tracing facilities. In addition, more detailed information about distinguishing features is provided.

□ □ □

Copies of the following technical reports can be obtained by writing

Ms. Hilary Whiteway Cohen  
Department of Computer and Information Science  
The Moore School of Electrical Engineering / D2  
University of Pennsylvania  
Philadelphia, PA 19104

### Automatic Enhancement of a Data Base Knowledge Representation Used for Natural Language Generation

**Kathleen F. McCoy**

*Technical Report No. MS-CIS-81-06.*

The purpose of this paper is to describe the ENHANCE system. ENHANCE automatically generates part of a data base knowledge representation used for natural language generation.

The generation system TEXT answers questions about the structure of a data base. The knowledge representation used by TEXT is a taxonomic description of the data base augmented with descriptive attributes at each node. ENHANCE generates this representation below the actual entity classes in the data base. It uses the logical description of the data base, the actual data base values, and some world knowledge axioms to generate descriptions of sub-classes of the entity classes. These descriptions will enhance the generation system to give richer descriptions of the data base.

This paper describes the knowledge representation used by TEXT and explains how this representation is augmented by the ENHANCE system.

### Invoking a Beginner's Aid Processor by Recognizing JCL Goal

**Jeffrey C. Shrager**

*Technical Report MS-CIS-81-07*

Typical help processors are invoked explicitly by the user or implicitly when an error occurs. Often a beginner will not know that he needs help because the inefficient use of commands will get the job done without raising errors. WIZARD is an expert system that recognizes beginner misbehaviors and can automatically start a help transaction.

The WIZARD processor relies on a special purpose, dynamic, pattern matcher directed by a KL-One based knowledge network. An author studies logs of beginner interactions and develops sequence rules which parse and properly identify misbehaviors. Objects that drive the parser to understand VAX DCL commands are coded into the network and a set of semantic programming utilities is used to perform actual goal recognition.

This thesis deals primarily with the implementation of such a goal recognizing expert invocation system. It is the WIZARD documentation and final working report. I discuss the motivations for the design of the system and detail the knowledge base and heuristics that support goal recognition. Some issues of generality are taken up and potential topics for later research are presented which will extend WIZARD's capabilities.

### **A Skeletal Automaton Approach To The Inference Problem for Context-Free Languages (Thesis)**

**Leona F. Fass**

*Technical Report MS-CIS-81-08.*

The inference problem for context-free languages is approached from an automata-theoretic point of view, through the inference of the skeletal automata described by Levy and Joshi. Techniques for constructing finite state machines from specifications of their behavior are adapted to this class of tree automata which have, as their behavior, sets of skeletons (trees labelled only on their frontier nodes). It is shown that when a bound on the number of states of the automaton is known, the minimal machine may be inferred from a finite sample of its behavior. Since these automata are known to recognize exactly the sets of skeletons of derivation trees of context-free languages, a consequence of this machine inference is grammatical inference. Given a suitably chosen finite sample of the derivation trees of good words in a context-free language, it is shown that a grammar for the language may be obtained from the resultant inferred skeletal automaton. Thus the language will be learned from a finite (structured) sample of its words.

### **Natural Language Research at the University of Pennsylvania - 1981**

**Bonnie Lynn Webber and Tim Finin (eds.)**

*Technical Report MS-CIS-81-10. December 1981*

This report summarizes the natural language processing research done at the University of Pennsylvania during 1981 and discusses the directions we are taking in our current research. Seven natural language projects are described as well as several related ones.

(An abbreviated version of this report appeared in the Sigart special issue on natural language processing, 1982).

### **A General Incremental Evaluator for Attribute Grammars**

**Fahimeh Jalili and Jean Gallier**

*Technical Report MS-CIS-81-11.*

This paper describes a new evaluator for attribute grammars, noncircular or circular. The ordering of attributes and their evaluation is accomplished by a single procedure that simulates a depth-first search of the dependency graph for a given input. The dependency graph is not constructed. Instead the algebraic semantic tree and the dependency graphs of the productions which implicitly represent the dependency graph are used. The incremental version of the evaluator is obtained by a simple modification. It avoids reevaluation of those attributes whose values are not affected by the modifications of the input. The evaluator has time complexity linear in the number of attributes needed to specify the meaning of the input string, and has space complexity linear in the size of the parse tree.

### **An Algorithm for Parsing Nominal Compounds**

**Timothy W. Finin**

*Technical Report MS-CIS-81-12. December 1981*

This note describes an algorithm used to "parse" strings of nominal modifiers. It has been used as a component of a larger system which builds semantic interpretations of nominal compounds such as "aluminum fuel pump repairs" and "navigation equipment failure rates". The algorithm is constrained to examining constituents which lie within a three position window.

### **Taking the Initiative in Natural Language Data Base Interactions: Justifying Why**

**Bonnie L. Webber and Aravind K. Joshi**

*Technical Report MS-CIS-82-01.*

This report describes our work to date on one aspect of what we are calling "Extended Natural Language Data Base Interactions". This is our work on producing justifications for system responses to data base queries. In this report we consider three issues:

- (1) what justification a query calls for
- (2) what justification the system can give
- (3) what constitutes a clear and understandable justification

concentrating on the third. This we discuss in terms of succinct justifications that do not say more than is necessary and well-structured justifications that the user can understand and evaluate.

(A shortened version of this paper appears in the Proc. COLING-82, Prague Czechoslovakia.)

### Translating KL-ONE from Interlisp to FranzLisp

Timothy W. Finin

*Technical Report MS-CIS-82-02.*

This paper describes an effort to translate the Interlisp KL-ONE system into FranzLisp to enable it to be run on a DEC PDP-11/780 VAX computer. Our approach was to first build a general purpose inter-dialect Lisp translation system that is driven by transformation rules. We then developed a set of specific Interlisp to FranzLisp translation rules and an appropriate run-time support system for the resulting FranzLisp version of KL-ONE. The current collection of Interlisp to FranzLisp rules, which is tailored for translating KL-ONE, numbers about forty. The run-time support environment (dubbed InterFranz) contains about 250 functions, mostly macros. In addition, a rudimentary DWIM-like facility has been developed to handle certain classes of expressions which tend to slip through the translation process.

(An abbreviated version of this report appears in Proceedings of the Second KL-ONE Workshop)

### The Interpretation of Nominal Compounds in Discourse

Timothy W. Finin

*Technical Report MS-CIS-82-03.*

In earlier work we described a system which built semantic interpretations of nominal compounds – sequences of two or more nouns (or nominal adjectives) which function as a unit. Our initial system took the representations of the concepts making up the compound and synthesized a number of concepts which represented possible interpretations. Each candidate interpretation was then assigned an “appropriateness” measure based on a variety of factors. One source of knowledge which was not used was the discourse context. Nominal compounding is often used as a naming device to refer to a concept already introduced, either directly or indirectly, into the discourse.

Our current work in this area centers around three questions:

- (1) How can we effectively use discourse context as a primary source of evidence for both hypothesizing candidate interpretations (roughly in order of plausibility) and for choosing a preferred interpretation from among them?
- (2) How much analysis is really required to do an adequate job of understanding nominal compounds used in discourse? We believe that many

compound strings may require only a relatively shallow analysis.

- (3) When is it appropriate to use nominal compounds? We would like to formulate the conditions when a particular nominal compound is likely to be accurately and efficiently interpreted by a hearer.

(An abbreviated version of this paper appears in Penn Linguistics Review, 1982)

### Processing of Sentences with Intra-Sentential Code-Switching

Aravind K. Joshi

*Technical Report MS-CIS-82-04.*

Speakers of certain bilingual communities systematically produce utterances in which they switch from one language to another, possibly several times in the course of an utterance. The two language systems systematically interact with each other in the production (and recognition) of these sentences. We have investigated this phenomenon in a formal or computational framework which consists of two grammatical systems and a mechanism for switching between the two systems. A variety of constraints apparent in these two sentences are then explained in terms of constraints on the switching mechanism. Constraints on the switching closed class items appear pre-eminently in this framework. This investigation has relevance to issues in processing of monolingual utterances.

### Generating Natural Language Text in Response to Questions about Database Structure

Kathleen R. McKeown

*Technical Report MS-CIS-82-05.*

There are two major aspects of computer-based text generation: 1) determining the content and textual shape of what is to be said; and 2) transforming that message into natural language. Emphasis in this research has been on a computational solution to the questions of what to say and how to organize it effectively. A generation method was developed and implemented in a system called TEXT that uses principles of discourse structure, discourse coherency, and relevancy criterion.

The main features of the generation method developed for the TEXT strategic component include 1) selection of relevant information for the answer, 2) the pairing of rhetorical techniques for communication (such as analogy) with discourse purposes (for example, providing definitions) and 3) a focusing mechanism. Rhetorical techniques, which encode aspects of discourse structure, are used to guide the selection of propositions from a relevant knowledge pool. The focusing mechanism aids in the organization of the message by constraining the selection of information to be talked about next to that which ties in with the previous discourse in an appropriate way.

### **The ENHANCE System: Creating Meaningful Sub-Types in a Database Knowledge Representation for Natural Language Generation**

**Kathleen McCoy**

*Technical Report MS-CIS-82-06.*

The knowledge representation is an important factor in natural language generation since it limits the semantic capabilities of the generation system. It is, however, a tedious task to hand code a knowledge representation which reflects both a user's view of a domain and the way that domain is modelled in the database. A system is presented which uses the contents of the database to form part of a database knowledge representation automatically. It augments a database scheme depicting the database structure used for natural language generation. Computational solutions are presented for deriving the information types contained in the schema. Three types of world knowledge axioms are used to ensure that the representation formed is meaningful and contains salient information.

### **Inferring Monitors on Natural Language Queries on Relational Database**

**Sitaramaswamy Lanka**

*Technical Report MS-CIS-82-07.*

In this thesis a Natural Language interface to a database management system has been designed that has the capability to offer to monitor for the relevant future changes of the database and later provide that additional information whenever it is available. This is based on the syntactic structure of the question and also semantic information regarding updatability of the contents of database.

### **User Participation in the Reasoning Processes of Expert Systems**

**Martha Pollack, Julia Hirschberg, and Bonnie Webber**

*Technical Report MS-CIS-82-09*

We argue that expert systems, if they are to satisfy the legitimate needs of their users, must include dialogue capabilities as sophisticated as those proposed in current Natural Language research. Our claims are corroborated by analysis of transcripts of a "naturally-occurring" expert system, a radio talk show in which callers ask an expert for financial advice. We present data demonstrating that user-expert dialogues are best viewed as a negotiation process, and we describe the exchanges that compose the dialogue in terms of the motivations, goals, strategies, and moves of the participants. Included is a catalog of the motivations, goals, and strategies observed in the expert and in the users.

(A shortened version of this paper will appear in the Proceedings of AAAI-82.)

### **Extended Natural Language Data Base Interactions**

**Bonnie Webber, Aravind Joshi, Eric Mays and Kathleen McKeown**

*Technical Report MS-CIS-82-10.*

For many years, the most frequently given reason for developing natural language interaction with computers was that it would increase their accessibility to infrequent (and possibly naive) users. Such users would not have to learn a host of languages and protocols in order to get the information they desired: they could express their needs and requests as they would everyday.

This viewpoint has led to valuable research aimed at removing artificial constraints on a user's freedom of expression, and useful results continue to appear. But just expanding the ways a query can be phrased and still understood does not in itself guarantee fruitful Natural Language interactions. This paper presents two complementary directions for extending Natural Language interfaces to data bases, which can make them more useful systems. These extensions make Natural Language interfaces less simply "windows" through which data can be called into view, and more "articulate experts" on the data base system and what it represents. The research we discuss involves (1) broadening the range of query types that can be handled and (2) extending the range of responses that can be provided.

(This paper will appear in a special issue of the International Journal of Computers and Mathematics edited by Nick Cercone, Simon Fraser University.)

### **Augmenting a Database Knowledge Representation for Natural Language Generation**

**Kathleen Filliben McCoy**

*Technical Report MS-CIS-82-11.*

The knowledge representation is an important factor in natural language generation since it limits the semantic capabilities of the generation system. This paper identifies several information types in a knowledge representation that can be used to generate meaningful responses to questions about database structure. Creating such a knowledge representation, however, is a long tedious process. A system is presented which uses the contents of the database to form part of this knowledge representation automatically. It employs three types of world knowledge axioms to ensure that the representation formed is meaningful and contains salient information.

### **Taking the Initiative in Natural Language Data Base Interactions: Monitoring as Response**

**Eric Mays, Aravind K. Joshi, and Bonnie L. Webber**

*Technical Report MS-CIS-82-13.*

This note is intended as a status report on our ongoing work in the area of natural language interaction with dynamic knowledge bases. Specifically, we discuss that portion of our research involving the offer by the question-answering system to monitor for future changes in the knowledge base.

(This paper appears in the Proceedings of ECAI-82, Orsay, France.)

### **An Expert System That Volunteers Advice**

**Jeff Shrager, Tim Finin**

*Technical Report MS-CIS-82-15.*

This paper describes the design and implementation of an expert system that provides novice users with help in using the VAX/VMS operating system. The most interesting feature of our advisor is that it follows the user's interactions with the system and volunteers its help when it believes that the user would benefit from advice. The user need not ask for help or raise an error condition. The advisor recognizes correct yet inefficient command sequences and helps the beginner become more proficient by indicating how these tasks may be done more efficiently.

(An abbreviated version of this report appears in Proceedings of AAAI-82.)

### **Help and Advice in Task Oriented Systems**

**Timothy W. Finin**

*Technical Report MS-CIS-82-22.*

This note describes current work at the University of Pennsylvania centered around providing intelligent help and advice to users of interactive task oriented systems. This work has focused on three general themes: (1) Help systems should be active rather than passive; (2) help systems should contain explicit models of the user, the task and the system utility being used and (3) the help system should engage in an interactive dialogue with the user in order to identify the information he really needs. An experimental help system, WIZARD, has been implemented for the VAX/VMS operating system to explore some of these issues.

### **Parsing with Logical Variables**

**Timothy W. Finin and Martha S. Palmer**

*Technical Report MS-CIS-82-23.*

Logic based programming systems have enjoyed an increasing popularity in applied AI work in the last few years. One of the contributions to Computational Linguistics made by the Logic Programming Paradigm has been the Definite Clause Grammar. In comparing DCG's with previous parsing mechanisms such as ATN's, certain clear advantages are seen. We feel that the most important of these advantages are due to the use of Logical Variables with Unification as the funda-

mental operation on them. To illustrate the power of the Logical Variable, we have implemented an experimental ATN system which treats ATN registers as Logical Variables and provides a unification operation over them. We would like to simultaneously encourage the use of the powerful mechanisms available in DCG's, and demonstrate that some of these techniques can be captured without reference to a resolution theorem prover.

### **The Text System for Natural Language Generation: An Overview**

**Kathleen R. McKeown**

*Technical Report MS-CIS-82-51.*

Computer-based generation of natural language requires consideration of two different types of problems: 1) determining the content and textual shape of what is to be said, and 2) transforming that message into English. A computational solution to the problems of deciding what to say and how to organize it effectively is proposed that relies on an interaction between structural and semantic processes. Schemas, which encode aspects of discourse structure, are used to guide the generation process. A focusing mechanism monitors the use of the schemas, providing constraints on what can be said at any point. These mechanisms have been implemented as part of a generation method within the context of a natural language database system, addressing the specific problem of responding to questions about database structure.

### **The ENHANCE System: Augmenting a Knowledge Base for Natural Language Generation**

**Kathleen F. McCoy**

A system is described which is able to automatically create a part of a database knowledge representation needed for natural language generation. The knowledge base needed to generate text length responses to questions about a discussion of how part of such a knowledge base can be itself. The contents of the database, however, are not enough to ensure that a meaningful representation is created. Three types of world knowledge axioms are based on database concepts and can therefore be easily provided by the database manager.

### **Reductions in Tree Replacement Systems**

**Jean H. Gallier and Ronald V. Book**

*Technical Report MS-CIS-83-02.*

Certain kinds of Church-Rosser tree-replacement systems called "reduction systems" are studied. Most results presented here generalize some results proved for string rewriting systems. The concept of a tree rewriting system with a "norm" is introduced and studied. A new "tree reducing machine" which reduces



every input tree to an irreducible tree is defined. As a consequence, when the norm under consideration is the size of the tree (the number of nodes), the word problem for finite Church-Rosser reduction systems is decidable in linear-time (in the size of the input tree). Church-Rosser monadic reduction systems generalizing some monadic Thue systems are also studied. It is shown that every congruence class and every finite union of congruence classes defined by such a system is accepted by a deterministic bottom-up tree push-down automation. This new form of a tree automation is briefly studied.

### **SCALAR QUANTITY IMPLICATURE: A Strategy for Processing Scalar Utterances**

**Julia Hirschberg**

*Technical Report MS-CIS-83-10.*

To support cooperative exchanges natural language interfaces must permit a wide range of user behavior, including indirect responses to system queries and indirect questions from the user. In naturally-occurring discourse many such utterances rely for their interpretation upon speaker and hearer recognizing a common scale underlying their contributions. By extending Horn's concept of scalar quantity implicature we may derive the basis for a strategy to facilitate system recognition, interpretation, and response not only to indirect answers and indirect questions but also to a wide variety of other scalar assertions.

### **Understanding Misunderstanding - Recognizing and Responding to User Misunderstandings**

**Ethel Schuster, Tim Finin**

*Technical Report MS-CIS-83-12.*

In designing an interactive system (such as a HELP system or a tutor) that gives advice or instructions to a user, one must cope with a problem that can seriously impair its effectiveness: the user's potential for misunderstanding the system's advice or instructions. Such misunderstandings can occur at many points -- for example, where the system has used an everyday term in a technical sense or a new term that looks or sounds like one the user is already familiar with. If such misunderstandings are not detected and corrected quickly, at best, time will be wasted, at worst, the user will fail to achieve his/her goal. In human interactions the responsibility to recognize and correct misunderstanding is shared by the participants. We are addressing the problem of enabling machines to take up their responsibility in human-machine interactions. In this paper, we categorize causes of misunderstanding in human-computer dialogues and examine some ways by which misunderstandings can be recognized rectified.

### **Custom-Made Responses: Maintaining and Updating The User Model**

**Ethel Schuster**

*Technical Report MS-CIS-83-13.*

Efforts to develop more "user-oriented" help/tutor systems must include development of the system's part in the interaction, if they are to succeed. How the system behaves strongly affects what the user learns, how and whether s/he achieves his/her desired goal. The paper is concerned with part of the system's behavior -- the language it uses in interacting with its users and the effect that that language can have upon them. We claim that by maintaining a model of the user's past experience (and hence his/her current expectations) and by updating this model as the interaction proceeds, the system can tailor its responses to the user's needs and facilitate his/her success in learning what was intended and/or obtaining appropriate help.

### **Incremental Partial Matching of Descriptions In a Knowledge Representation Network**

**Robert Chaefitz**

*Technical Report MS-CIS-83-15.*

This thesis describes a description matcher for the KL-ONE knowledge representation language that is incremental and can perform partial matches. The matcher is incremental in that it can be suspended at any time, producing partial results, and later restarted. The matcher is partial in that it has a formalism for measuring the degree to which two descriptions match. Both features are supported by using four different scalar metrics for the degree of match of two descriptions: a lower bound, an upper bound, a base value, and a current estimate.

### **BUP - A Bottom Up Parser**

**Timothy W. Finin; Bonnie L. Webber**

*Technical Report MS-CIS-83-16.*

BUP is a bottom up syntactic analyser that can be used in several ways: as a simple bottom up parser for context free languages, as a bottom up parser for extended phrase structure grammars (equivalent in power to an ATN), or as a transducer for either. BUP accepts a grammar and lexicon specified as rules. It will then analyse input strings according to those rules, recording its findings in a chart and producing all analyses in parallel. Rules can be displayed, added or deleted incrementally. BUP provides a small set of tools for building structures and specifying tests.

### **Logics and Natural Language**

**Bonnie Lynn Webber**

*Technical Report MS-CIS-83-18.*

Logic is that branch of knowledge concerned with truth and inference – with determining the conditions under which a proposition is true or one proposition may be inferred from others. Such knowledge is essential for communication, since most of our beliefs about the world come not from the world, but rather from what others tell us about it. Hence, it is not surprising that logic underlies a wide range of current research on Natural Language (NL) interaction with machines. However, it is not just first-order logic (FOL) that has been enlisted in this enterprise: other logics have begun to be used as well. This paper discusses three of them: (1) default logic, for computing presuppositions; (2) modal logic, for planning utterances to enlist help; and (3) temporal logic, for offering competent database monitors and correcting certain user misconceptions.

### **CADRS-A System for the Design of Complex Data Structures in Lisp**

**Richard J. Duncan**

*Technical Report MS-CIS-83-19.*

This thesis describes a project to design and implement a data abstraction facility for LISP. The result of this project, the CADRS (Compiler for Access of Data Regardless of Structure) system, supports the definition, creation, and efficient manipulation of hybrid data/procedure structures that closely resemble class objects in languages like Simula and Smalltalk.

The foremost goal of CADRS is to provide a flexible and efficient means of defining data structures and hybrid data/procedure structures in LISP. Another important goal of CADRS is to provide the functionality available in INTERLISP's record package to facilitate a large scale translation effort between INTERLISP and Franz. As such, the CADRS package provides:

- Data structuring for both physical and virtual structures.
- Structural inheritance.
- Procedural Attachment (hybrid data/procedure structures).
- Structure-independent data access.
- INTERLISP compatibility.

As a structure compiler, CADRS uses the following constraints to achieve efficiency: (1) Each data access form produced by the compiler is optimized using a standard set of rules for LISP optimization; (2) the mechanisms which support the more expensive features of CADRS are only evoked when these features are actually used, and when possible only on a case by case basis and (3) package is segmented into compiler and run-time support segments allowing the system to run with minimal space requirements.

### **CADRS Reference Manual**

**Richard J. Duncan**

*Technical Report MS-CIS-83-20.*

This document describes version 1.00 of the CADRS (Compiler for Access of Data Regardless of Structure) package and its implementation under Franz Lisp Opus 37. It is intended as a reference to the features, syntax and functions of CADRS for a reader who is familiar with the LISP language (Franz, Maclisp or INTERLISP).

### **Processing of Sentences with Intra-Sentential Code-Switching**

**Aravind K. Joshi**

*Technical Report MS-CIS-83-23.*

Speakers of certain bilingual communities systematically produce utterances in which they switch from one language to another, possibly several times in the course of an utterance. The two language systems systematically interact with each other in the production (and recognition) of these sentences. We have investigated this phenomenon in a formal or computational framework which consists of two grammatical systems and a mechanism for switching between the two systems. A variety of constraints apparent in these sentences are then explained in terms of constraints on the switching mechanism. Constraints on the switching of closed class items appear pre-eminently in this framework. This investigation has relevance to issues in processing of monolingual utterances.

### **How Much Context-Sensitivity is Required to Provide Structural Descriptions: Tree Adjoining Grammars**

**Aravind K. Joshi**

*Technical Report MS-CIS-83-24.*

Two types of dependencies are especially important: subcategorization and filler-gap dependencies. Moreover, these dependencies can be unbounded. One of the motivations for transformations was to account for unbounded dependencies. The so-called non-transformational grammars account for the unbounded in different ways. In a tree-adjoining grammar (TAG) unboundedness is achieved by factoring the dependencies and recursion in a novel and, we believe, in a linguistically interesting manner. All dependencies are defined on a finite set of basic structures (trees) which are bounded. Unboundedness is then a corollary of a particular composition operation called adjoining. There are thus no unbounded dependencies in a sense. In this paper, we will introduce TAG's, investigate their formal properties, and present some detailed linguistic examples.

## In Response: Next Steps in Natural Language Interaction

**B. L. Webber**

*Technical Report MS-CIS-83-27*

In the area of man-machine interaction, Natural Language has so far primarily been used to simplify people's access to information. The next step beyond simple data access is the kind of cooperative interactive problem-solving that current expert systems aspire to. But support for problem solving (which includes helping the user formulate his/her problems) demands more in the way of interaction than just answering requests for the factual information. In the first part of this paper, we illustrate some of these needed capabilities. In the remainder, we discuss two of them in greater detail: (1) recognizing and responding to user misconceptions and (2) getting from users the information needed to help them solve their problems.

## A Modal Temporal Logic For Reasoning About Change

**Eric Mays**

*Technical Report MS-CIS-83-29.*

We examine several behaviors for query systems that become possible with the ability to represent and reason about change in data bases: queries about alternative histories, and offers of monitors as responses to queries. A modal temporal logic is developed for this purpose. A completion axiom for history is given and modelling strategies are given by example.

## Monitors As Responses to Questions: Determining Competence

**Eric Mays**

*Technical Report MS-CIS-83-30.*

This paper discusses the application of a propositional temporal logic to determining the competence of a monitor offer as an extended response by a question-answering system. Determining monitor competence involves reasoning about the possibility of some future state given a description of the current state and possible transitions.

## Describing Knowledge from Semantic Network

**Hassan Ait-Kaci**

*Technical Report MS-CIS-83-33.*

This document is a study of linguistic means and strategies that a human speaker implicitly uses when providing information that is new to his/her interlocutor. It may lay the ground for the design of an eventual interactive computer system exploiting a knowledge base to brief a user in English. The specific work reported is an experimental investigation of a natural language generation rhetorics in describing knowledge

encoded as a semantic network. The knowledge representation language used is KI-one, a data structure specification language defining conceptual objects organized in a network.

## An Efficient Evaluator for Attribute Grammars with Conditional Rules

**Jean Gallier**

*Technical Report MS-CIS-83-36.*

This paper presents a new efficient evaluator for attribute grammars (circular or noncircular) allowing conditional semantic equations, which generalizes the algorithm presented in Jalili's Ph.D thesis. The essence of the method, investigated and presented for the first time in Jalili (also reported in Jalili and Gallier), is to perform a depth-first search of the (static) reversed dependency graph associated with a parse tree, interleaved with the execution of semantic rules. This algorithm is suggested by the standard iterative method for solving for least fixed points of continuous functionals, and as a consequence, is correct with respect to the fixed point semantics of attribute grammars.

The efficiency of the attribute evaluator is achieved via two mechanisms:

- (1) As in Jalili, the dependency graph is not constructed. Instead, it is implicitly represented by the semantic tree and the dependency graphs of the productions.
- (2) The semantic rules are precompiled as programs written in intermediate code called semantic modules.

The algorithm performs a "call by need" evaluation, and is optimal in the number of attribute instances evaluated. The time and space complexity of the evaluator are linear in the number of nodes in the parse tree for the input string.

## Proofs in Higher-Order Logic

**Dale A. Miller**

*Technical Report MS-CIS-83-37.*

Broadly speaking, this dissertation is an investigation of some of the metatheory for higher-order logic which may be important to builders of automated theorem provers for this logic. In particular, expansion trees are defined as generalizations of Herbrand instances for formulas in a nonextensional form of higher-order logic based on Church's simple theory of types. Such expansion trees can be defined with or without the use of skolem functions. These trees store substitution terms and either critical variables or skolem terms used to instantiate quantifiers in the original formula and those resulting from instantiations. An expansion tree is called an expansion tree proof (ET-proof) if it encodes a tautology, and, in the form not using skolem functions, an imbedding relation among

the critical variables is acyclic. The relative completeness result for expansion tree proofs not using skolem functions, i.e. if  $A$  is provable in higher-order logic then  $A$  has such an expansion tree proof, is based on Andrews' formulation of Takahashi's proof of the cut-elimination theorem for higher-order logic. If the occurrences of skolem functions in place of critical variables is equivalent to the requirement that the imbedding relation is acyclic. This fact not only resolves the open question of what is a sound definition of skolemization in higher-order logic but also provides a direct, syntactic proof of its correctness.

Since subtrees of expansion trees are also expansion trees (or their dual) and expansion trees store substitution terms and critical variables explicitly, ET-proofs can be directly converted into sequential and natural deduction proofs. A naive translation will often produce proofs which contain a lot of redundancies and will often use implicational lines in an awkward fashion. An improved translation process is presented. This process will produce only focused proofs in which much of the redundancy has been eliminated and backchaining on implicational lines was automatically selected if it was applicable. The information necessary to construct focused proofs is provided by a certain connection scheme, called a mating, of the boolean atoms within the tautology encoded by an ET-proof. Such a translation process could be made part of a very generalized explanation facility.

This dissertation was submitted to the Mathematics Department of Carnegie-Mellon University on 27 August 1983 and was written under the direction of Peter Andrews.

### Cooperative Responses to Object-Related Misconceptions: A Thesis Proposal

Kathleen F. McCoy

*Technical Report MS-CIS-83-39.*

Because people's knowledge is often partial and/or faulty, it is inevitable that misconceptions will be revealed in the course of a conversation. If recognized, the other person may say something to correct the misconception, and the conversation continues. Just as this is the case when people interact with each other, so must it be when users interact with a computer system. For example, in interacting with an expert system, a user may reveal misconceptions about objects modelled by the system. By failing to correct such misconceptions, the system may not only confirm the original misconception, but may cause the user to develop further misconceptions. It must therefore be up to the system to recognize and respond to misconceptions in an effective way.

In this paper the space of possible object misconceptions is characterized according to the knowledge base feature involved in the incorrect information.

This characterization is often useful in determining how the user could have arrived at the response. Using such a characterization and a model of what the user knows, the system can reason in a domain independent way about how to best correct the user. Factors which affect the amount of information included in a correction (such as discourse and situational context) are also examined.

□ □ □

The following abstracts are from *Proceedings of the Conference on Applied Natural Language Processing*, February 1983. Copies of the *Proceedings* cost \$15 and are available from

Donald E. Walker, ACL  
Artificial Intelligence Center  
SRI International  
Menlo Park, CA 94025 USA

### Domain-Independent Natural Language Interfaces: Session Introduction

Aravind K. Joshi

Department of Computer and Information Science  
R. 268 Moore School  
University of Pennsylvania  
Philadelphia, PA 19106  
*1983 ANLP Conf. Proc., pp. 1-2.*

In my comments, I will try to summarize briefly the six papers in this session, pointing out some key aspects of each contribution, listing some common themes, and raising some questions, not for the purpose of criticizing one or another contribution, but rather for the purpose of prodding the authors to discuss these questions in their presentations.

### Problems in Natural-Language Interface to DBMS with Examples from EUFID

Marjorie Templeton, John Burger

System Development Corporation  
Santa Monica, CA  
*1983 ANLP Conf. Proc., pp. 3-16.*

For five years the End-User Friendly Interface to Data management (EUFID) project team at System Development Corporation has worked on the design and implementation of a Natural-Language Interface (NLI) system that was to be independent of both the application and the database management system. In this paper we describe application, natural-language and database management problems involved in NLI development, with specific reference to the EUFID system as an example.

### Introducing ASK, A Simple Knowledgeable System

Bozena H. Thompson, Frederick B. Thompson

California Institute of Technology  
Pasadena, CA 91125  
1983 ANLP Conf. Proc., pp. 17-24.

ASK, *A Simple Knowledgeable System*, is a total system for the structuring, manipulation and communication of information. It is a simple system in the sense that its development concentrated on clean engineering solutions to what could be done now with good response times. The user interface is a limited dialect of English. In contrast to expert systems, in which experts build the knowledge base and users make use of this expert knowledge, ASK is aimed at the user who wishes to create, test, modify, extend and make use of his own knowledge base. It is a system for a research team, a management or military staff, or a business office.

This paper is designed to give you a feel for the general performance of the ASK System and overview of its operational capabilities. To this end, the movie you will see will continue throughout the talk. Indeed, the talk itself is a commentary on this background movie. The movie is bona fide and in real time; it is of the ASK System in action. (Many of the illustrations from the movie are reproduced in the written paper.)

### **A Robust Portable Natural Language Data Base Interface**

**Jerrold M. Ginsparg**

Bell Laboratories  
Murray Hill, NJ 07974  
1983 ANLP Conf. Proc., pp. 25-30.

This paper describes a NL data base interface which consists of two parts: a Natural Language Processor (NLP) and a data base application program (DBAP). The NLP is a general purpose language processor which builds a formal representation of the meaning of the English utterances it is given. The DBAP is an algorithm which builds a query in a augmented relational algebra from the output of the NLP. This approach yields an interface which is both extremely robust and portable.

### **IR-NLI: An Expert Natural Language Interface to Online Data Bases**

**Giovanni Guida, Carlo Tasso**

Instituto di Matematica, Informatica e Sistemistica  
Universita di Udine  
Via Antonini, 8 - 33100 Udine, Italy  
1983 ANLP Conf. Proc., pp. 31-38.

Constructing natural language interfaces to computer systems often requires achievement of advanced reasoning and expert capabilities in addition to basic natural language understanding. In this paper the above issues are faced in the frame of an actual application concerning the design of a natural language interface for the access to online information retrieval systems.

After a short discussion of the peculiarities of this application, which requires both natural language understanding and reasoning capabilities, the general architecture and fundamental design criteria of a system presently being developed at the University of Udine are then presented. The system, named IR-NLI, is aimed at allowing non-technical users to directly access through natural language the services offered by online data bases. Attention is later focused on the basic functions of IR-NLI, namely, understanding and dialogue, strategy generation, and reasoning. Knowledge representation methods and algorithms adopted are also illustrated. A short example of interaction with IR-NLI is presented. Perspectives and directions for future research are also discussed.

### **TEAM: A Transportable Natural-Language Interface System**

**Barbara J. Grosz**

Artificial Intelligence Center  
SRI International  
Menlo Park, CA 94025  
1983 ANLP Conf. Proc., pp. 39-45.

A major benefit of using natural language to access the information in a database is that it shifts onto the system the burden of mediating between two views of the data: the way in which the data is stored (the "database view") and the way in which an end-user thinks about it (the "user's view"). Database information is recorded in terms of files, records, and fields, while natural-language expressions refer to the same information in terms of entities and relationships in the world. A major problem in constructing a natural-language interface is determining how to encode and use the information needed to bridge these two views. Current natural-language interface systems require extensive efforts by specialists in natural-language processing to provide them with the information they need to do the bridging. The systems are, in effect, handtailored to provide access to particular databases.

This paper focuses on the problem of constructing *transportable* natural-language interfaces, i.e., systems that can be adapted to provide access to databases for which they were not specifically handtailored. It describes an initial version of a transportable system, called TEAM (for *Transportable English Access Data Manager*). The hypothesis underlying the research described in this paper is that the information required for the adaptation can be obtained through an interactive dialogue with database management personnel who are not familiar with natural-language processing techniques.

## Isolating Domain Dependencies in Natural Language Interfaces

R. Grishman

Department of Computer Science  
New York University  
New York, NY 10012

L. Hirschman

Federal and Special Systems Group  
Burroughs Corporation  
Paoli, PA 19301

C. Friedman

Linguistic String Project, NYU  
1983 ANLP Conf. Proc., pp. 46-53.

Isolating the domain-dependent information within a large natural language system offers the general advantages of modular design and greatly enhances the portability of the system to new domains. We have explored the problem of isolating the domain dependencies within two large natural language systems, one for generating a tabular data base from text ("information formatting"), the other for retrieving information from a data base. We describe the *domain information schema* which is used to capture the domain-specific information, and indicate how this information is used through the two systems.

## Knowledge-Based Approaches: Session Introduction

Ivan A. Sag

Department of Linguistics  
Stanford University  
Stanford, CA 94305

1983 ANLP Conf. Proc., p. 54.

The papers in this session touch on several key issues of our field that make good sense to address. Boguraev and Sparck Jones discuss properties of system design that ensure desirable functioning (e.g., discussions of modularity and portability). Both Finin and Palmer, and Pazzani and Engelman touch on comparisons among existing techniques. Montgomery is concerned with new problems for theories of representation or processing. Lehnert and Shwartz address assessing the limitations of existing systems. In some cases the contribution being made is clear. In other cases, I for one find it difficult to assess whether the systems being described do what they do in a principled way that will generalize to other applications, or in an ad hoc fashion that is devoid of intellectual interest. In the absence of agreed-upon standards for evaluating descriptions of such systems, I can only urge both authors and audience to address the issues I have raised, doubtless for the first time, in the oral presentations and subsequent discussion.

## Distinguishing Fact from Opinion and Events from Meta-events

Christine A. Montgomery

Operating Systems Division  
Logicon, Inc.

Woodland Hills, CA

1983 ANLP Conf. Proc., pp. 55-61.

A major problem in automatically analyzing the text of military messages in order to synthesize data base elements is separating fact from opinion, i.e., identifying factual statements versus evaluative commentary such as degree of belief or confidence in the parameters of a particular event, comparative evaluations of entities and events, predictions of future events, etc. This paper describes a model which accounts for the range of factual to evaluative information in the message traffic, and discusses a means for representing such information in the context of an experimental system for automated data base generation.

## Parsing with Logical Variables

Timothy W. Finin, Martha Stone Palmer

Department of Computer and Information Science  
University of Pennsylvania  
Philadelphia, PA 19104

1983 ANLP Conf. Proc., pp. 62-68.

Logic based programming systems have enjoyed an increasing popularity in applied AI work in the last few years. One of the contributions to Computational Linguistics made by the Logic Programming Paradigm has been the Definite Clause Grammar. In comparing DCG's with previous parsing mechanisms such as ATN's, certain clear advantages are seen. We feel that the most important of these advantages are due to the use of Logical Variables with Unification as the fundamental operation on them. To illustrate the power of the Logical Variable, we have implemented an experimental ATN system which treats ATN registers as Logical Variables and provides a unification operation over them. We would like to simultaneously encourage the use of the powerful mechanisms available in DCG's, and demonstrate that some of these techniques can be captured without reference to a resolution theorem prover.

## EXPLORER: A Natural Language Processing System for Oil Exploration

Wendy G. Lehnert

Department of Computer and Information Science  
Graduate Research Center  
University of Massachusetts  
Amherst, MA 01003

Steven P. Shwartz  
Cognitive Systems Inc.  
234 Church Street  
New Haven, CT 06510

1983 ANLP Conf. Proc., pp. 69-72.

EXPLORER is a non-fragile, "hands-on" language analysis system that allows oil explorationists with no knowledge of computers or computer programming to create customized maps. Users in Tulsa, Denver, and

New Orleans currently have dial-up access to a DEC-20 where EXPLORER is implemented in TLISP. A user converses with EXPLORER about a desired map until both parties have agreed on an adequate and unambiguous set of specifications. Another phone line then carries EXPLORER's output to an IBM 3033 which runs database retrieval routines on commercial well data. When all the information has been secured from well data, a graphics system takes over to perform the actual map generation. EXPLORER is currently undergoing evaluation, and it is targeted for a 1983 installation in all regional offices of a major oil company for restricted use by geologists and geophysicists.

### Knowledge Based Question Answering

Michael J. Pazzani, Carl Engelman

The MITRE Corporation  
Bedford, MA 01730

1983 ANLP Conf. Proc., pp. 73-80.

The natural language database query system incorporated in the KNOBS interactive planning system comprises a dictionary driven parser, APE-II, and script interpreter which yield a conceptual dependency conceptualization as a representation of the meaning of user input. A conceptualization pattern matching production system then determines and executes a procedure for extracting the desired information from the database. In contrast to syntax driven Q-A systems, e.g., those based on ATN parsers, APE-II is driven bottom-up by expectations associated with word meanings. The processing of a query is based on the contents of several knowledge sources including the dictionary entries (partial conceptualizations and their expectations), frames representing conceptual dependency primitives, scripts which contain stereotypical knowledge about planning tasks used to infer states enabling or resulting from actions, and two production system rule bases for the inference of implicit case fillers, and for determining the responsive database search. The goals of this approach, all of which are currently at least partially achieved, include utilizing similar representations for questions with similar meanings but widely varying surface structures, developing a powerful mechanism for the disambiguation of words with multiple meanings and the determination of pronoun referents, answering questions which require inferences to be understood, and interpreting ellipses and ungrammatical utterances.

### How to Drive a Database Front End Using General Semantic Information

K. K. Boguraev, K. Sparck Jones

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

1983 ANLP Conf. Proc., pp. 81-88.

This paper describes a front end for natural language access to databases making extensive use of general, i.e. domain-independent, semantic information for question interpretation. In the interests of portability, initial syntactic and semantic processing of a question is carried out without any reference to the database domain, and domain-dependent operations are confined to subsequent, comparatively straightforward, processing of the initial interpretation. The different modules of the front end are described, and the system's performance is illustrated by examples.

### Handling Ill-Formed Input: Session Introduction

Ralph M. Weischedel

Department of Computer and Information Sciences  
University of Delaware  
Newark, DE 19711

1983 ANLP Conf. Proc., pp. 89-92.

Suppose we call "normative" any system based on a set of constraints (whether pragmatic, semantic, or syntactic). Input that violates the constraints of a system could be termed "ill-formed". Sondheimer and Weischedel (1980) identify two general classes of input appearing ill-formed to a normative system. Input will be called

*absolutely ill-formed*, if native speakers generally agree that it violates one or more linguistic constraints, or

*relatively ill-formed*, if it violates some constraint(s) of the computational system, though native speakers perceive nothing odd about it.

Examples of absolute ill-formedness include misspellings, mistypings, mispunctuations, tense and number errors, word order problems, run-on sentences, sentence fragments, extraneous forms, and meaningless sentences. Examples of relatively ill-formed input include unknown words and requests that are beyond the limits of either the computer system or the natural language interface.

In natural language access (e.g. English access) to information systems, the magnitude of the problem of absolute ill-formedness can be seen in several case studies. If one includes telegraphic and elliptical constructions in the class of absolute ill-formedness, then case studies reported in Thompson (1980) and Eastman and McLean (1981) indicate that as much as 25% of queries to question-answering systems are absolutely ill-formed. On the other hand, no matter how large the dictionary, grammar, and underlying system, there will always be unknown words and phrases (e.g. proper names) and impossible requests (due to user misconceptions of the capabilities of the underlying system).

### The Fitted Parse: 100% Parsing Capability in a Syntactic Grammar of English

Karen Jensen, George E. Heidorn

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

*1983 ANLP Conf. Proc.*, pp. 93-98.

A technique is described for performing *fitted* parsing. After the rules of a more conventional syntactic grammar are unable to produce a parse for an input string, this technique can be used to produce a reasonable approximate parse that can serve as input to the remaining stages of processing. The paper describes how fitted parsing is done in the EPISTLE system and discusses how it can help in dealing with many difficult problems of natural language analysis.

### Utilizing Domain-Specific Information for Processing Compact Text

Elaine Marsh

Linguistic String Project  
New York University  
New York, NY 10012

*1983 ANLP Conf. Proc.*, pp. 99-103.

This paper identifies the types of sentence fragments found in the text of two domains: medical records and Navy equipment status messages. The fragment types are related to full sentence forms on the basis of the elements which were regularly deleted. A breakdown of the fragment types and their distributions in the two domains is presented. An approach to reconstructing the semantic class of deleted elements in the medical records is proposed which is based on the semantic patterns recognized in the domain.

### Scruffy Text Understanding: Design and Implementation of the NOMAD System

Richard H. Granger, Chris J. Staros,  
Gregory B. Taylor, Rika Yoshii

Artificial Intelligence Project  
Department of Information and Computer Science  
University of California  
Irvine, CA 92717

*1983 ANLP Conf. Proc.*, pp. 104-106.

The task of understanding unedited naval ship-to-shore messages is implemented in the presence of a large database of domain specific knowledge. The program uses internal syntactic and semantic expectations to analyze the texts and to correct errors that arise during understanding, such as syntactic errors, missing punctuation, and errors of spelling and usage. The output of the system is a well-formed English translation of the message. This paper describes some of the knowledge mechanisms that have been implemented in the NOMAD system.

### Text Analysis: Session Introduction

Donald E. Walker

Artificial Intelligence Center  
SRI International  
Menlo Park, CA 94025

*1983 ANLP Conf. Proc.*, pp. 107-108.

Text analysis is a promising area for applications of computational linguistics, with regard both to the prospects for technological achievement and to the potential benefits that those achievements can provide. Books, articles, and documents in text form are a major source of information. Being able to use computers more effectively in the analysis of these materials would have significant implications for virtually every intellectual activity. Moreover, developments in this area are essential if mankind is to take advantage of as well as merely to cope with the vast quantity of textual data already stored in computer-readable form, an amount that is increasing rapidly because of the use of computer-directed photocomposition for printing and the practicality of optical character recognition techniques.

### Specialized Information Extraction: Automatic Chemical Reaction Coding from English Descriptions

Larry H. Reeker

Department of Computer Science  
Tulane University  
New Orleans, LA 70118

Elena M. Zamora

P.O. Box 3554  
Gaithersburg, MD 20278

Paul E. Blower

Chemical Abstracts Service  
2540 Olentangy River Road  
P.O. Box 3012

Columbus, OH 43210

*1983 ANLP Conf. Proc.*, pp. 109-116.

In an age of increased attention to the problems of database organization, retrieval problems and query languages, one of the major economic problems of many potential databases remains the entry of the original information into the database. Specialized information extraction (SIE) systems are therefore of potential importance in the entry of information that is already available in certain restricted types of natural language text. This paper contains a discussion of the problems of engineering such systems and a description of a particular SIE system, designed to extract information regarding chemical reactions from experimental sections of papers in the chemical literature and to produce a data structure containing the relevant information.

### Automatic Analysis of Descriptive Texts

James R. Cowie



Computer Centre  
University of Strathclyde  
Royal College, George Street  
Glasgow, G1 1KW, SCOTLAND  
*1983 ANLP Conf. Proc., pp. 117-123.*

This paper describes a system that attempts to interpret descriptive texts without the use of complex grammars. The purpose of the system is to transform the descriptions to a standard form which may be used as the basis of a database system knowledgeable in the subject matter of the text.

The texts currently used are wild plant descriptions taken directly from a popular book on the subject. Properties such as size, shape and colour are abstracted from the descriptions and related to parts of the plant in which we are interested. The resulting output is a standardised hierarchical structure holding only significant features of the description.

The system, implemented in the PROLOG programming language, uses keywords to identify the way segments of the text relate to the object described. Information on words is held in a keyword list of nouns relating to parts of the object described. A dictionary contains the attributes of ordinary words used by the system to analyse the text. The text is divided into segments using information provided by conjunctions and punctuation.

About half the texts processed are correctly analysed at present. Proposals are made for future work to improve this figure. There seems to be no inherent reason why the technique cannot be generalized so that any text of semi-standard descriptions can be automatically converted to a canonical form.

### **"Expertness" from Structure Text? RECONSIDER: A Diagnostic Prompting Program**

Mark S. Tuttle

Section on Medical Information Science, A-16  
University of California  
San Francisco, CA 94143

David D. Sherertz  
Tandem Computers, Inc.  
19333 Vallco Parkway  
Cupertino, CA 95014

Marsden S. Blois  
Section on Medical Information Science, UCSF  
Stuart Nelson

Department of Medicine  
State University of New York  
Stony Brook, NY 11794  
*1983 ANLP Conf. Proc., pp. 124-131.*

RECONSIDER is an interactive diagnostic prompting program which uses simple information retrieval techniques to prompt a physician regarding possible diagnoses, given a list of positive patient findings. Its knowledge base consists of "structured text" definitions of 3262 diseases and a synonym dictionary. Patient findings, and their synonyms, are matched against inverted files of terms from the disease descriptions,

the number and selectivity of the patient findings matching terms in a given disease description determine the disease's "score", and the matched diseases are sorted on this score to form a preliminary differential diagnosis. Definitions of diseases can be referenced for viewing by name, or by their position in a differential. While its first formal evaluation is not yet complete, the performance of RECONSIDER continues to exceed the expectations of user and designer alike.

### **Using Natural Language Descriptions to Improve the Usability of Databases**

Carole D. Hafner, John D. Joyce

Computer Science Department  
General Motors Research Laboratories  
Warren, MI 48090  
*1983 ANLP Conf. Proc., pp. 132-135.*

This paper describes the REGIS extended command language, a relational data language that allows users to name and describe database objects using natural language phrases. REGIS accepts multiple-word phrases as the names of tables and columns (unlike most systems, which restrict these names to a few characters). An extended command parser uses a network-structured dictionary to recognize multi-word names, even if some of the words are missing or out of order, and to prompt the user if an ambiguous name is entered. REGIS also provides facilities for attaching descriptive text to database objects, which can be displayed online or included in printed reports. Initial data from a few databases indicate that users choose to take advantage of the naturalness of multi-word descriptions when this option is available.

### **Natural Language Text Segmentation Techniques Applied to the Automatic Compilation of Printed Subject Indexes and for Online Database Access**

G. Vladutz

Institute for Scientific Information  
3501 Market Street  
Philadelphia, PA 19104  
*1983 ANLP Conf. Proc., pp. 136-142.*

The nature of the problem and earlier approaches to the automatic compilation of printed subject indexes are reviewed and illustrated. A simple method is described for the detection of semantically self-contained word phrase segments in title-like texts. The method is based on a predetermined list of acceptable types of nominative syntactic patterns which can be recognized using a small domain-independent dictionary. The transformation of the detected word phrases into subject index records is described. The records are used for the compilation of Key Word Phrase subject indexes (KWPSI). The method has been successfully tested for the fully automatic production of KWPSI-type indexes to titles of scientific publications. The usage of

KWPSI-type display formats for the enhanced online access to databases is also discussed.

### **Automatic Representation of the Semantic Relationships Corresponding to a French Surface Expression**

**Gian Piero Zarri**

Centre National de la Recherche Scientifique  
Laboratoire d'Informatique pour les Sciences de l'Homme  
54, Boulevard Raspail  
75270 Paris Cedex 06 FRANCE

*1983 ANLP Conf. Proc., pp. 143-147.*

The work presented here is a preliminary study concerning the automatic translation of French natural language statements into the RESEDA semantic metalanguage. The text in natural language is first (pre)processed in order to obtain its syntactic structure. The "semantic parsing" process begins with marking the "triggers", defined as lexical units which call one or more of the predicative patterns allowed for in the metalanguage. The patterns obtained are then merged, and their case slots filled with the elements found in the surface structure according to the predictions associated with the slots.

### **Machine Translation: Session Introduction**

**Gary R. Martins**

Intelligent Software, Inc.  
15450 Cohasset Street  
Van Nuys, CA 91406

*1983 ANLP Conf. Proc., p. 148.*

More than two decades ago, scholars of the stature of the late Y. Bar-Hillel were stating in detail their perceptions of the futility of automated language translation by the methods then under development. Some years (and many "progress reports") later, others of us came to see that general translation, without the benefit of a robust and dominant semantics and pragmatics, was not thinkable. The very slow and scattered development of these latter, in both principle and substance, is the most fundamental impediment to the computer manipulation of language; in those tiny domains where the work has been done, the results are impressive – across a variety of methods.

Why isn't more of this work being done, with more focus? Perhaps both the potential market and the (mostly military) research sponsors are awed by the magnitude of the challenge, and skeptical of the pay-offs. One might envision a world in which, once the skepticism has been answered, such answers would stimulate rather than intimidate research.

Meanwhile, we should be busy with other goals in mind, of which aids-to-translation is an obvious example. Two of the papers before us describe approaches to improving the human translation process. A third valiantly wades into the struggle with computation semantics. The fourth paper offers unsurprising re-

sults from an odd experiment. I should note that all I had available to me were the abstracts.

### **Investigating the Possibility of a Microprocessor-Based Machine Translation System**

**Harold L. Somers**

Centre for Computational Linguistics  
University of Manchester Institute of Science and  
Technology  
PO Box 88  
Manchester M60 1QD, England

*1983 ANLP Conf. Proc., pp. 149-155.*

This paper describes an on-going research project being carried out by staff and students at the Centre for Computational Linguistics to examine the feasibility of Machine Translation (MT) in a microprocessor environment. The system incorporates as far as possible features of large scale MT systems that have proved desirable or effective: it is multilingual, algorithms and data are strictly separated, and the system is highly modular. Problems of terminological polysemy and syntactic complexity are reduced via the notions of controlled vocabulary and restricted syntax. Given these constraints, it seems feasible to achieve translation via an 'interlingua', avoiding any language-pair oriented 'transfer' stage. The paper concentrates on a description of the separate modules in the translation process as they are currently envisaged, and details some of the problems specific to the microprocessor-based approach to MT that have so far come to light.

### **An Application of Montague Grammar to English-Japanese Machine Translation**

**Toyoaki Nishida, Shuji Doshita**

Department of Information Science  
Faculty of Engineering, Kyoto University  
Sakyo-ku, Kyoto 606, JAPAN

*1983 ANLP Conf. Proc., pp. 156-165.*

English-Japanese machine translation requires a large amount of structural transformations in both grammatical and conceptual level. In order to make its control structure clearer and more understandable, this paper proposes a model based on Montague Grammar. Translation process is modeled as a data flow computation process. Formal description tools are developed and a prototype system is constructed. Various problems which arise in this modeling and their solutions are described. Results of experiments are shown and it is discussed how far initial goals are achieved.

### **A Status Report on the LRC Machine Translation System**

**Jonathan Slocum**

Siemens Corporation  
Boca Raton, FL 33431

*1983 ANLP Conf. Proc., pp. 166-173.*

This paper discusses the linguistic and computational techniques employed in the current version of Machine Translation system being developed at the Linguistics Research Center of the University of Texas, under contract to Siemens AG in Munich, West Germany. We pay particular attention to the reasons for our choice of certain techniques over other candidates, based on both objective and subjective criteria. We then report the system's status vis-a-vis its readiness for application in a production environment, as a means of justifying our claims regarding the practical utility of the methods we espouse.

### Computer-Assisted Translation Systems: The Standard Design and a Multi-level Design

Alan K. Melby

Linguistics Department  
Brigham Young University  
Provo, UT 84602

*1983 ANLP Conf. Proc., pp. 174-177.*

The standard design for a computer-assisted translation system consists of data entry of source text, machine translation, and revision of raw machine translation. This paper discusses this standard design and presents an alternative multi-level design consisting of integrated word processing, terminology aids, preprocessing aids and a link to an off-line machine translation system. Advantages of the new design are discussed.

### Speech Interfaces: Session Introduction

Douglas L. Hogan

13625 Middlevale Lane  
Silver Spring, MD 20906

*1983 ANLP Conf. Proc., pp. 178-179.*

The speech interface is the natural one for the human user and is beginning to be used in a limited way in many applications. Some of these applications are experimental; still others have achieved the status of cost-effective utility. A brief summary of the current state-of-the-art of speech input and output is presented. The two papers in the session represent specific examples of current work. Some comments on the need for linguistically oriented development conclude the paper.

### Interactive Natural Language Problem Solving: A Pragmatic Approach

A. Biermann

Department of Computer Science  
Duke University  
Durham, NC

R. Rodman

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

B. Ballard

CS, Duke

G. Bilbro

CS, NCSU

H. Deas, L. Fineman, P. Fink, K. Gilbert, D. Gregory,

F. Heidlage

CS, Duke

*1983 ANLP Conf. Proc., pp. 180-191.*

A class of natural language processors is described which allow a user to display objects of interest on a computer terminal and manipulate them via typed or spoken English sentences.

This paper concerns itself with the implementation of the voice input facility using an automatic speech recognizer, and the touch input facility using a touch sensitive screen. To overcome the high error rates of the speech recognizer under conditions of actual problem solving in natural language, error correction software has been designed and is described here. Also described are problems involving the resolution of voice input with touch input, and the identification of the intended referents of touch input.

To measure system performance we have considered two classes of factors: the various conditions of testing, and the level and quality of training of the system user. In the paper a sequence of five different testing situations is observed, each one resulting in a lowering of system performance by several percentage points below the previous one. A training procedure for potential users is described, and an experiment is discussed which utilizes the training procedure to enable users to solve actual non-trivial problems using natural language voice communication.

### Application of the Liberman-Prince Stress Rules to Computer Synthesized Speech

David L. McPeters

Bell Laboratories  
Indianapolis, IN 46219

Alan L. Tharp

Computer Science Department  
North Carolina State University  
Raleigh, NC 27650

*1983 ANLP Conf. Proc., pp. 192-197.*

Computer synthesized speech is and will continue to be an important feature of many artificially intelligent systems. Although current computer synthesized speech is intelligible, it cannot yet pass a Turing test. One avenue for improving the intelligibility of computer synthesized speech and for making it more human-like is to incorporate stress patterns on words. But to achieve this improvement, a set of stress prediction rules amenable to computer implementation is needed.

This paper evaluates one such theory for predicting stress, that of Liberman and Prince. It first gives an overview of the theory and then discusses modifications which were necessary for computer implementation. It then describes an experiment which was per-

formed to determine the model's strengths and shortcomings. The paper concludes with the results of that study.

□ □ □

The *Conference Proceedings* listed below are available from

Donald E. Walker, ACL  
Artificial Intelligence Center

SRI International  
Menlo Park, CA 94025 USA

*AJCL*

17th Annual Meeting, 1979	\$10.00
18th Annual Meeting, 1980	\$15.00
19th Annual Meeting, 1981	\$15.00
20th Annual Meeting, 1982	\$15.00
21st Annual Meeting, 1983	\$15.00
ACL Europe 1983 Conference	\$15.00



