Abstracts of Current Literature

The ProGram Manual
Roger Evans, Gerald Gazdar
Cognitive Studies Programme
Arts E, University of Sussex
Brighton, BN1 9QN, U.K.
Cognitive Science Research Paper
CSRP 035, April 1984, 106 pages
(Address requests to Judith Dennison.)

This is the manual for the *ProGram* grammar development system (written in Prolog) intended for use by linguists and computational linguists developing grammars for significant fragments of natural languages. The system incorporates all aspects of the Generalized Phrase Structure Grammar framework, including features, metarules, ID/LP rule format, Kleene star rule feature instantiation principles, schemata, and so on. Among other things, the user can "direct" parsing interactively in order to explore particular analyses.

Natural Language Database Update
Sharon Salveter
Computer Science Department
College of Liberal Arts
Boston University
Boston, MA 02215
Technical Report No. 84/001

Although natural language (NL) querying of data bases (DBs) has been an active research area for many years, and at least one commercial system is available that supports natural language database query, little effort has been expended in support of natural language database *update*. Thus, end-users of data bases must alternate between easy-to-use natural language query systems and harder-to-use format database update systems.

Salveter and Maier have shown that, because NL DB update is a fundamentally different problem than query, it is not possible to naturally extend natural language query systems to also support update. A data base is an attempt to abstract information about the real world. A state of the data base is meant to represent a state of a portion of the real world. To interpret values of the data objects in the data base as statements about the real world, we must be able to connect the values in the data base with various entities and relationships in the data base. A semantic data description indicates a set of real world states (RWS), a database definition gives a set of allowable database states (DBS). The correspondence between the semantic description and the database definition induces connections between database states and real world states.

A Fuzzy-Set-Theoretic Approach to the Compositionality of Meaning: Propositions, Dispositions and Canonical Forms
Lotfi A. Zadeh
Division of Computer Science
ERL, College of Engineering
University of California
Berkeley, CA 94720
Memorandum No. UCB/ERL M83/24

In its traditional interpretation, Frege's principle of compositionality is not sufficiently flexible to have a wide applicability to natural languages. In a fuzzy-set-theoretic setting which is outlined in this paper, Frege's principle is modified and broadened by allowing the meaning of a proposition, p, to be composed not from the meaning of the constituents of p but, more generally, from the meaning of a collection of fuzzy relations which form a so-called *explanatory data base* that is associated with p. More specficially, through the application of test-score semantics, the meaning of p is represented as a procedure which tests, scores and aggregates the elastic constraints which are implicit in p. The employment of fuzzy sets in this semantics allows p to contain fuzzy predicates such as *tall, kind, much richer*, etc.,; fuzzy quantifiers such as *most, several, few, usually*, etc.; and other types of semantic entities which cannot be dealt with within the framework of classical logic.

The approach described in the paper suggests a way of representing the meaning of dispositions, e.g., *Overeating causes obesity, Icy roads are slippery, Young men like young women*, etc. Specifically, by viewing a disposition, d, as a proposition with implicit fuzzy quantifiers, the problem of representing the meaning of d may be decomposed into (a) restoring the suppressed fuzzy quantifiers and/or fuzzifying the nonfuzzy quantifiers in the body of d; and (b) representing the meaning of the resulting dispositional proposition through the use of test-score semantics.

To place in evidence the logical structure of p and, at the same time provide a high-level description o the composition process, p may be expressed in the canonical form "X is F", where $X = (X_1 ..., X_n)$ is an implicit n-ary variable which is constrained by p and F is a fuzzy n-ary relation which may be interpreted as an elastic constraint on X. This canonical form and the meaning-composition process for proposisitions and dispositions are illustrated by several examples among which is the proposition $p \overset{\Delta}{=}$ *Over the past few years Naomi earned far more than most of her close friends.*

Test-Score Semantics for Natural Languages and Meaning Representation via PRUF
Lotfi A. Zadeh
see page 43
In Rieger, B., Ed., *Empircal Semantics I.*
Brockmeyer, Bochum, 1981.

In a sharp departure from the conventional approaches to the problem of meaning representation in natural languages, test-score semantics is based on the premise that almost everything that relates to natural languages is a matter of degree. Thus, in test-score semantics, predicates, propositions and other types of linguistic entities are treated as collections of elastic constraints on a set of objects or relations in a universe of discourse. Viewed in this perspective, the meaning of a linguistic entity may be defined by (a) identifying the constraints which are implicit or explicit in the entity in question; (b) describing the tests that must be performed to ascertain the degree to which each constraint is satisfied; and (c) specifying the manner in which the degrees in question or, equivalently, the partial test scores are to be aggregated to yield an overall test score. In general, the overall test score is a vector whose components are numbers in the unit interval or possibility/probability distributions over this interval.

The first step in the representation of the meaning of a given proposition involves the construction of a relational data base in which the meaning of constituent relations and their attributes is assumed to be known. The choice of the data base affects the explanatory effectiveness of the translation process and is governed by the knowledge profile of the intended user of the translation. The test procedure – which is regarded as the representation of the meaning of the proposition – acts on the data base and returns an overall test score which is interpreted as the compatibility of p with the data base.

Test-score semantics is sufficiently general to allow the translation of almost any proposition in a natural language. However, the price of generality is the difficulty of writing a program which could represent the meaning of a given proposition without recourse to human assistance.

A Computational Approach to Fuzzy Quantifiers in Natural Languages
Lotfi A. Zadeh
see page 43
Comp. & Maths. with Appls.
9(1): 149-184 (1983).

The generic term *fuzzy quantifier* is employed in this paper to denote the collection of quantifiers in natural languages whose representative elements are: *several, most, much, not many, very many, not very many, few, quite a few, large number, small number, close to five, approximately ten, frequently,* etc. In our approach, such quantifiers are treated as fuzzy numbers which may be manipulated through the use of fuzzy arithmetic and, more generally, fuzzy logic.

A concept which plays an essential role in the treatment of fuzzy quantifiers is that of the cardinality of a fuzzy set. Through the use of this concept, the meaning of a proposition containing one or more fuzzy quantifiers may be represented as a system of elastic constraints whose domain is a collection of fuzzy relations in a relational data base. This representation, then, provides a basis for inference from premises which contain fuzzy quantifiers. For example, from the propositions "Most U's are A's" and "Most A's are B's", it follows that "$Most^2$ U's are B's", where $most^2$ is the fuzzy product of the fuzzy proportion *most* with itself.

The computational approach to fuzzy quantifiers which is described in this paper may be viewed as a derivative of fuzzy logic and test-score semantics. In this semantics, the meaning of a semantic entity is represented as a procedure which tests, scores, and aggregates the elastic constraints which are induced by the entity in question.

The Role of Fuzzy Logic in the Management of Uncertainty in Expert Systems
Lotfi A. Zadeh
see page 43
Fuzzy Sets and Systems 11, 1983: 199-227.

Management of uncertainty is an intrinsically important issue in the design of expert systems because much of the information in the knowledge base of a typical expert system is imprecise, incomplete or not totally reliable.

In the existing expert systems, uncertainty is dealt with through a combination of predicate logic and probability-based methods. A serious shortcoming of these methods is that they are not capable of coming to grips with the pervasive fuzziness of information in the knowledge base, and, as a result, are mostly ad hoc in nature. An alternative approach to the management of uncertainty which is suggested in this paper is based on the use of fuzzy logic, which is the logic underlying approximate or, equivalently, fuzzy reasoning. A feature of fuzzy logic which is of particular importance to the management of uncertainty in expert systems is that it provides a systematic framework for dealing with fuzzy quantifiers, e.g., *most, many, few, not very many, almost all, infrequently, about 0.8,* etc. In this way, fuzzy logic subsumes both predicate logic and probability theory, and makes it possible to deal with different types of uncertainty within a single conceptual framework.

In fuzzy logic, the deduction of a conclusion from a set of premises is reduced, in general, to the solution of a nonlinear program through the application of projection and extension principles. This approach to deduction leads to various basic syllogisms which may be used as rules of combination of evidence in expert systems. Among syllogisms of this type which are discussed in this paper are the intersection/project syllogism, the generalized modus ponens, the consequent conjunction syllogism, and the major-premise reversibility rule.

Precisiation of Meaning via Translation into PRUF
Lofti A. Zadeh
see page 43
In Vaina, L. and Hintikka, J., Eds., *Cognitive Constraints on Communication*
D. Reidel, 1984: 373-401.

It is suggested that communication between humans – as well as between humans and machines – may be made more precise by the employment of a meaning representation language PRUF which is based on the concept of a possibility distribution. A brief exposition of PRUF is presented, and its application to precisiation of meaning is illustrated by a number of examples.

A Theory of Commonsense Knowledge
Lotfi A. Zadeh
see page 43
In Skala, H.J.; Termini, S.; and Trillas, E., Eds., *Aspects of Vagueness.*
D. Reidel, 1984: 257-296.

The theory outlined in this paper is based on the idea that what is commonly called *commonsense knowledge* may be viewed as a collection of *dispositions,* that is, propositions with implied fuzzy quantifiers. Typical examples of dispositions are: *Icy roads are slippery, Tall men are not very agile, Overeating causes obesity, Bob loves women, What is rare is expensive,* etc. It is understood that, upon restoration of fuzzy quantifiers, a disposition is converted into a proposition with explicit fuzzy quantifiers, e.g., *Tall men are not very agile → Most tall men are not very agile.*

Since traditional logical systems provide no methods for representing the meaning of propositions containing fuzzy quantifiers, such systems are unsuitable for dealing with commonsense knowledge. It is suggested in this paper that an appropriate computational framework for dealing with commonsense knowledge is provided by *fuzzy logic,* which, as its name implies, is the logic underlying fuzzy (or approxi-

mate) reasoning. Such a framework, with an emphasis on the representation of dispositions, is outlined and illustrated with examples.

The following reports are available from

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

Prices are as indicated; add postage: surface – theses £0.65, others £0.40; air – theses £5.00, others £1.50.

Compound Noun Interpretation Problems
K. Sparck Jones
Technical Report No. 45, July 1983,
16pp., £0.50

This paper was prepared for the SERC/CREST Course on Computer Speech Processing, Cambridge, England, July 1983. It discusses the problems of compound noun interpretation in the context of automatic language processing. Given that compound processing implies identifying the senses of the words involved, determining their bracketing, and establishing their underlying semantic relations, the paper illustrates the need, even in comparatively favourable cases, for inference using pragmatic information. This has consequences for language processor architectures and, even more, for speech processors.

Automatic Summarising of English Texts
J.I. Tait
Technical Report No. 47, December 1982,
137pp., £2.50.

The thesis describes a computer program called Scrabble which can summarise short English texts. It uses large bodies of predictions about the likely contents of texts about particular topics to identify the commonplace material in an input text. Pre-specified summary templates, each associated with a different topic, are used to condense the commonplace material in the input. Filled-in summary templates are then used to form a framework into which unexpected material in the input may be fitted, allowing unexpected material to appear in output summary texts in an essentially unreduced form. The system's summaries are in English.

The program is based on technology not dissimilar to a script applier. However, Scrabble represents a significant advance over previous script-based summarising systems. It is much less likely to produce misleading summaries of an input text than some previous systems, and can operate with less information about the subject domain of the input than others.

These improvements are achieved by the use of three main novel ideas. First, the system incorporates a new method for identifying the topic or topics of an input text. Second, it allows a section of text to have more than one topic at a time, or at least a composite topic which may best be dealt with by the computer program simultaneously applying to the text predictions associated with more than one simple topic. Third, Scrabble incorporates new mechanisms for the incorporation of unexpected material in the input into its output summary texts. The incorporation of such material in the output summary is motivated by the view that it is precisely unexpected material which is likely to form the most salient matter in the input text.

The performance of the system is illustrated by means of a number of example input texts and their corresponding Scrabble summaries.

A Mechanism for the Accumulation and Application of Context in Text Processing
H. Alshawi
Technical Report No. 48, 17pp. £0.50.

The paper describes a mechanism for the representation and application of context information for automatic natural language processing systems. Context information is gathered gradually during the reading of the text, and the mechanism gives a way of combining the effect of several different types of context factors. Context factors can be

managed independently, while still allowing efficient access to entities in focus. The mechanism is claimed to be more general than the global focus mechanism used by Grosz for discourse understanding. Context affects the interpretation process by choosing the results, and restricting the processing, of a number of important language interpretation operations, including lexical disambiguation and reference resolution. The types of context factors that have been implemented in an experimental system are described, and examples of the application of context are given.

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
Reseach Unit for Information Science and Artificial Intelligence
University of Hamburg
Mittelweg 179
D-2000 Hamburg 13, WEST GERMANY

Surface Transformation during the Generation of Written German Sentences
Stephan Busemann
Report ANS-23. In Neumann, B., Ed., *GWAI-83. 7th German Workshop in Artificial Intelligence.*
Dassel, September 1983: 90-99.

This paper gives an overview of the modular and adaptable system SUTRA for surface transformations. The system generates written German sentences from an intermediate structure consisting solely of data based on knowledge which was already required for an earlier stage of the generation process. SUTRA may thus be employed in any generation system for the German language which produces such intermediate structures. SUTRA works in two steps. During the syntactical stage, several processes transform the intermediate structure into a linear string, thereby generating discontinuous verb constituents, arranging sentence components using separate ordering rules, mapping deep cases onto surface cases, and extracting further morpho-syntactical properties from the intermediate structure (e.g., number) and the word lexicon of the generation system (e.g., gender). During the morphological stage, the terminal word forms are generated from the linear string using classification schemas for the different inflectional paradigms of German word stems. SUTRA is currently employed in the generation component of the natural language dialogue system HAM-ANS.

The Real Estate Agent – Modeling Users by Uncertain Reasoning
Katharine Morik,
Claus-Rainer Rollinger
Report No. ANS-24. In Neuman, B., Ed., *GWAI-83:* 158-168.

Two topics are treated in this article. Firstly, a user model patterned after the stereotype approach is presented. This model surpasses Rich's model with respect to its greater flexibility in the construction of user profiles and its treatment of positive and negative arguments. Secondly, an inference machine is presented. This machine treats uncertain knowledge in the form of evidence for and against the accuracy of a proposition. Truth values are replaced by the concept of a two-dimensional evidence space. The consequences of the concept, particularly with regard to verification, are discussed. The connection between these two topics is established by implementation of the user model on the inference machine.

User Modelling and Profiles of Interest in Artificial Intelligence Dialog Systems
Katharina Morik
Report No. ANS-25. To appear in

In this paper the necessity for user modeling is demonstrated for three types of systems: information retrieval systems, tutor systems, and natural language systems. Three facets of user modeling – familiarity with system use, expertise in a domain, and profile of interests – are distinguished. Techniques for building up user profiles and utilizing

Rollinger, C.-R., Ed., *Probleme des (Text-) Verstehens -- Ansätze der Kunstlichen Intelligenz. (Sprache and Information).* Tübingen, 1984

them to govern system behavior are described. With regard to natural language systems, our work on user modeling for HAM-ANS is presented in some detail.

A Study of LISP-Program-Transportation
Thomas Christaller
Memo No. ANS-17, July 1983.

The first part of this study presents the manual transporting of the AI language FUZZY from UCI-LISP to Franz Lisp. The problems encountered in this process and their solution are taken up in the second part, dealing with the evaluation of different transporting procedures for LISP-programs. The best procedures are based on production rule systems. These must, however, offer the capability of interacting with the user in doubtful cases.

Considerations for Automatic Spelling Correction in an AI System
Michael Fliegner
Memo No. GEN-18, September 1983.

This paper investigates the integration of an automatic spelling-correction component in an AI-system. First the basic requirements for such a correction procedure are presented. The major part of the paper consists of a critical examination of different correction methods, which formed the basis for the automatic spelling-correction component integrated in HAM-ANS.

Report on a Lecture and Fact-Finding Tour in the USA in Late Summer 1983
Thomas Christaller, Wolfgang Hoeppner
Memo No. ANS-19, October 1983.

CCA
BBN
MIT
Boston
University of
: Pennsylvania
New Brunswick
Philadelphia
Rutgers University
Washington
AAAI-83

San Francisco
Stanford University
SRI
Hewlett Packard
Berkeley

Dynamic Storage Allocation Strategy for UCI-LISP
Rolf Dannenberg, Bernhard Nebel
Memo No. ANS-20, November 1983.

After criticizing the current strategy of memory reallocation in UCI-LISP, we discuss several alternative possibilities for realizing dynamic strategies under the TOPS10 operating system. The implementation of the chosen method, which is from the users' point of view almost as convenient as memory management in INTERLISP, is described and the user interface is documented.
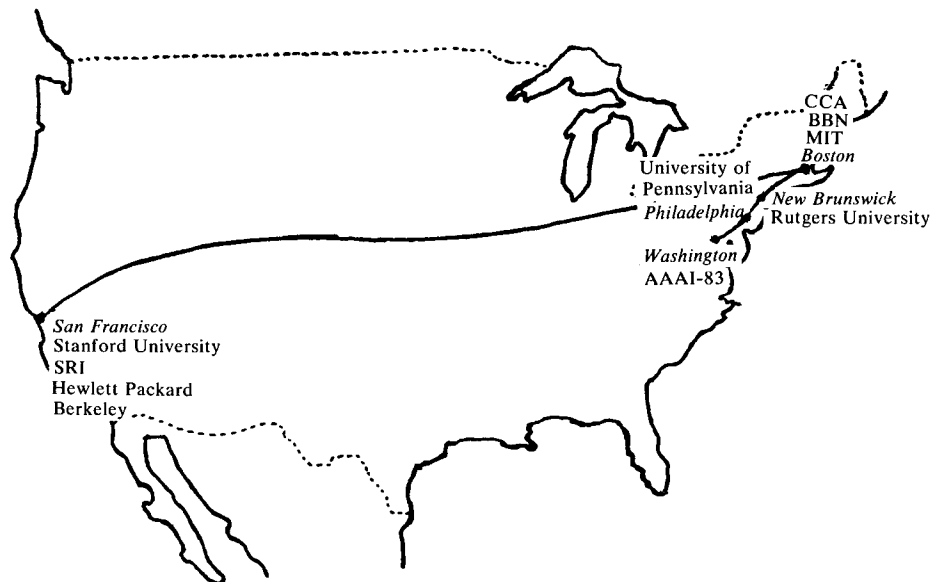
Approaches to a Semantic Representation for Natural Language Access to a Relational Data Base.
Henning Bergmann, Annedore Paeseler
Memo No. ANS-21, December 1983.

Natural language access systems to formatted mass data administered by DBMSs (data base management systems) have been available for a number of years. The inadequacy of these systems is, for the most part, a result of an insufficient representation of knowledge about the interfaced DBMS and the domain of discourse modeled by the stored data. Taking a specific relational data base as a point of departure, we first discuss preliminary considerations for the representation of this knowledge. Next we describe the design of such a representation and its implementation in the frame-oriented AI-langauge FRL. The characteristics of our approach are: organization of knowledge in a generalization hierarchy, the use of deep-case frames and the attachment of procedures for the generation of database queries.

Abstract Control Structures and the Semantics of Quantifers

Steven Cushing
Computer Science Department
St. Anselm College
Manchester, NH 03102
Proc. EACL 1983, pp. 1-8

Intuitively, a *quantifier* is any word or phrase that expresses a meaning that answers one of the questions "How many?" or "How much?" Typical English examples include *all, no, many, few, some but not many, all but at most a very few, wherever, whoever, whoever there is,* and also, it can be argued, *only, also,* and *the.* In this paper we review an empirically motivated analysis of such meanings and draw out its computational significance. For purposes of illustration, we focus our attention on the meanings expressed by the English words *whatever* and *some,* commonly represented, respectively, by the symbols "∀" and "∃", but most of what we say will generalize to the other meanings of this class.

In Section 1, we review the notion of satisfaction in a model, through which logical formulas are customarily imbued implicitly with meaning. In Section 2, we discuss quantifier relativization, a notion that becomes important for meanings other than ∀ and ∃. In Section 3, we use these two notions to characterize quantifier meanings as structured functions of a certain sort. In Section 4, we discuss the computational significance of that analysis. In Section 5, we elaborate on this significance by outlining a notion of abstract control structure that the analysis instantiates.

L'idee de Grammaire avec le Contexte Naturel

Leszek Haduch
Institute of Informatics
Technical University of Lodz
Lodz, ul.Piotrkowska 220, Poland
Proc. EACL 1983, pp. 9-13

Commonly used grammars that describe natural languages (e.g., ATN, Metamorphonic Grammars) can hardly be applied in describing highly inflectional languages. So I propose a grammar called the grammar with natural context, which takes into consideration properties of highly inflectional languages (e.g., Polish) as well as structural languages (e.g., English). I introduce its normal form.

Iterative Operations

Sae Yamada
Notre Dame Seishin University
Ifuku-Cho 2-16-9
700 Okayama, Japan
Proc. EACL 1983, pp. 14-20

We present in this article, as part of an aspectual operation system, a generation system of iterative expressions using a set of operators called iterative operators. In order to execute the iterative operations efficiently, we have classified previously propositions denoting a single occurrence of a single event into three groups. The definition of a single event is given recursively. The classification has been carried out especially in consideration of the durative/non-durative character of the denoted events and also in consideration of existence/non-existence of a culmination point (or a boundary) in the events. The operations concerned with iteration have either the effect of giving a boundary to an event (in the case of a non-bounded event) or of extending an event through repetitions. The operators concerned are: N,F – direct iterative operators; I,G – boundary giving operators; I – extending operator. There are direct and indirect operations: the direct ones change a non-repetitious proposition into a repetitious one directly, whereas the indirect ones change it indirectly. The indirect iteration is indicated with Σ. The scope of each operator is not uniquely definable, though the mutual relation of the operators can be given more or less explicitly.

**Structure of Sentence and Inferencing
in Question Answering**
Eva Hajičová, Petr Sgall
Faculty of Mathematics and Physics
Charles University
Malostranské n. 25
118 00 Praha 1 Czechoslovakia
Proc. EACL 1983, pp. 21-25

In the present paper we characterize in more detail some of the aspects of a question answering system using as its starting point the underlying structure of sentences (which with some approaches can be identified with the level of meaning or of logical form). First of all, the criteria are described that are used to identify the elementary units of underlying structure and the operations conjoining them into complex units (Section 1), then the main types of units and operations resulting from an empirical investigation on the basis of the criteria are registered (Section 2), and finally the rules of inference, accounting for the relevant aspects of the relationship between linguistic and cognitive structures are illustrated (Section 3).

A Phonological Processor for Italian
Rodolfo Delmonte
Centro Linguistico Interfacolta
Università degli Studi di Venezia
Ca' Garzoni-Moro - S. Marco 3417
Proc. EACL 1983, pp. 26-34

A computer program for the automatic translation of any text of Italian into naturally fluent synthetic speech is presented. The program, or Phonological Processor (hence FP) maps into prosodic structures the phonological rules of Italian. Structural information is provided by such hierarchical prosodic constituents as Syllable (S), Metrical Foot (MF), Phonological Word (PW), Intonational Group (IG). Onto these structures, phonological rules are applied such as the "letter-to-sound" rules, automatic word stress rules, internal stress hierarchy rules indicating secondary stress, external sandhi rules, phonological focus assignment rules, logical focus assignment rules. The FP constitutes also a model to simulate the reading process aloud, and the psycholinguistics and cognitive aspects related will be discussed in the computational model of the FP. At present, Logical Focus assignment rules and the computational model are work in progress still to be implemented in the FP. Recorded samples of automatically produced synthetic speech will be presented at the conference to illustrate the functioning of the rules.

**An Expert System for the Production
of Phoneme Strings from Unmarked
English Text using Machine-Induced
Rules**
Alberto Maria Segre
Coordinated Science Laboratory
Bruce Arne Sherwood
Computer-Based Education Research
 Laboratory
Wayne B. Dickerson
English as a Second Language
University of Illinois at Urbana-Champaign
Urbana, IL 61801
Proc. EACL 1983, pp. 35-42

The speech synthesis group at the Computer-Based Education Research Laboratory (CERL) of the University of Illinois at Urbana-Champaign is developing a diphone speech synthesis system based on pitch-adaptive short-time Fourier transforms. This system accepts the phonemic specification of an utterance along with pitch, time and amplitude warping functions in order to produce high quality speech output from stored diphone templates.

This paper describes the operation of a program which operates as a front end for the diphone speech synthesis system. The UTTER (for "Unmarked Text Transcription by Expert Rule") system maps English text onto a phoneme string, which is then used as an input to the diphone speech synthesis system. The program is a two-tiered Expert System which operates first on the word level and then on the (vowel or consonant) cluster level. The system's knowledge about pronunciation is organized in two decision trees automatically generated by an induction algorithm on a dynamically specified "training set" of exam-. ples.

**Vocal Interface for a Man-Machine
Dialog**
Dominique Beroule
LIMSI (CNRS)
B.P. 30
91406 Orsay CEDEX France
Proc. EACL 1983, pp. 43-48

We describe a dialog-handling module used as an interface between a vocal terminal and a task-oriented device (for instance, a robot manipulating blocks). This module has been specially designed to be implanted on a single board using a microprocessor, and inserted into the vocal terminal which already comprises a speech recognition board and a synthesis board. The entire vocal system is at present capable of conducting a real time spoken dialog with its user.

Knowledge Engineering Approach to Morphological Analysis
Harri Jappinen, Aarno Lehtola,
Esa Nelimarkka, Matti Ylilammi
Helsinki University of Technology
Helsinki, Finland
Proc. EACL 1983, pp. 49-51

Finnish is a highly inflectional language. A verb can have over ten thousand different surface forms – nominals slightly fewer. Consequently, a morphological analyzer is an important component of a system aiming at "understanding" Finnish. This paper briefly describes our rule-based heuristic analyzer for Finnish nominal and verb forms. Our tests have shown it to be quite efficient: the analysis of a Finnish word in a running text takes an average of 15 ms of DEC 20 CPU time.

A PROLOG Implementation of Lexical Functional Grammar as a Base for a Natural Language Processing System
Werner Frey, Uwe Reyle
Department of Linguistics
University of Stuttgart
West Germany
Proc. EACL 1983, pp. 52-57

The aim of this paper is to present parts of our system, which is to construct a data base out of a narrative natural language text. We think the parts are of interest on their own. The paper consists of three sections: (I) We give a detailed description of the PROLOG implementation of the parser, which is based on the theory of lexical functional grammar. (II) For the semantic representation of texts we use the Discourse Representation Theory developed by Hans Kamp. (III) Finally we sketch how the parser formalism can be augmented to yield as output discourse representation structures.

Extended Access to the Left Context in an ATN Parser
Irina Prodanof, Giacomo Ferrari
Istituto di Linguistica Computazionale
Via della Faggiola 32
I-56100 Pisa, Italy
Proc. EACL 1983, pp. 58-65

Some Italian sentences related to linguistic phenomena largely known and recently discussed by many computational linguists are discussed in the framework of ATN. They offer certain difficulties which seem to suggest a substantial revision of the ATN formalism. The theoretical assumptions and an experimental implementation of such a revision are presented together with examples. Many related theoretical points, such as some psycholinguistic implications and the relationship between deterministic and non-deterministic hypothesis are also briefly discussed.

An Experiment with Heuristic Parsing of Swedish
Benny Brodda
Institute of Linguistics
University of Stockholm
S-106 91 Stockholm SWEDEN
Proc. EACL 1983, pp. 66-73

Heuristic parsing is the art of doing parsing in a haphazard and seemingly careless manner but in such a way that the outcome is still "good", at least from a statistical point of view, or, hopefully, even from a more absolute point of view. The idea is to find strategic shortcuts derived from guesses about the structure of a sentence based on scanty observations of linguistic units in the sentence. If the guess comes out right, much parsing time can be saved; and if it does not, many subobservations may still be valid for revised guesses. In the (very preliminary) experiment reported here, the main idea is to make use of (combinations of) surface phenomena as much as possible as the base for the prediction of the structure as a whole. In the parser to be developed along the lines sketched in this report, the main stress is put on arriving at independently working, parallel recognition procedures.

The work reported here is aimed both at simulating certain aspects of human language perception and at arriving at effective algorithms for actual parsing of running text. There is, indeed, a great need for fast such algorithms, e.g. for the analysis of the literally millions of words of running text that already today comprise the data bases in various large information retrieval systems, and which can be expected to expand several orders of magnitude both in importance and in size in the foreseeable future.

Towards the Semantics of Sentence Adverbials
Eva Koktová
9. kvétna 1576

In the present paper we argue that the so-called sentence adverbials (typically, adverbs like *probably, admittedly, ...*) should be generated, in the framework of Functional Generative Description, by means of a special deep case – Complementation of Attitude (CA) on grounds of

39001 Tábor, Czechoslovakia
Proc. EACL 1983, pp. 74-80

Dealing with Conjunctions in a Machine Translation Environment
Xiuming Huang
Institute of Linguistics
Chinese Academy of Social Sciences
Beijing, China
Proc. EACL 1983, pp. 81-85

Fallible Rationalism and Machine Translation
Geoffrey Sampson
Linguistics & Modern English Language
University of Lancaster
Lancaster LA1-4YT, G.B.
Proc. EACL 1983, pp. 86-89

The Generation of Term Definitions from an Online Terminological Thesaurus
John McNaught
Centre for Computational Lingustics
UMIST
P.O. Box 88
Manchester, UK
Proc. EACL 1983, pp. 90-95

Relating Syntax and Semantics: The Syntactico-semantic Lexicon of the System VIE-LANG
Ingeborg Steinacker,
Ernst Buchberger
Medical Cybernetics
University of Vienna, Austria
Proc. EACL 1983, pp. 96-100

their special behaviour in the topic-focus articulation (TFA) of a sentence. From the viewpoint of the translation of CA expressions (and also of the multiple occurrence thereof inside a sentence) into a calculus of intensional logic, it should be noted that the TFA properties of CA expressions are directly correlated to the scope properties thereof. Our approach, which is stated in terms of a linguistic theory, serves as a basis for an algorithm of analysis of CA for purposes of a system of man-machine communication without a pre-arranged data base.

A set of rules, named CSDC (Conjunction Scope Determination Constraints), is suggested for attacking the conjunct scope problem, the major issue in the automatic processing of conjunctions which has been raising great difficulty for natural language processing systems. Grammars embodying the CSDC are incorporated into an existing ATN parser, and are tested successfully against a wide group of "and" conjunctive sentences, which are of three types: clausal coordination, phrasal coordination, and gapping. With phrasal coordination, the structure with two NPs coordinated by "and" has been given the most attention.

It is hoped that an ATN parser capable of dealing with a large variety of conjunctions in an efficient way will finally emerge from the presented work.

Approaches to MT have been heavily influenced by changing trends in the philosophy of language and mind. Because of the artificial hiatus which followed the publication of the ALPAC Report, MT research in the 1970s and early 1980s has had to catch up with major developments that have occurred in linguistic and philosophical thinking; currently, MT seems.to be uncritically loyal to a paradigm of thought about language which is rapidly losing most of its adherents in departments of linguistics and philosophy. I argue, both in theoretical terms and by reference to empirical research on a particular translation problem, that the Popperian "falliable rationalist" view of mental processes which is winning acceptance as a more sophisticated alternative to Chomskyan "deterministic rationalism" should lead MT researchers to redefine their goals and to adopt certain currently-neglected techniques in trying to achieve those goals.

A new type of machine dictionary is described, which uses terminological relations to build up a semantic network representing the terms of a particular subject field, through interaction with the user. These relations are then used to dynamically generate outline definitions of terms in online query mode. The definitions produced are precise, consistent and informative, and allow the user to situate a query term in the local conceptual environment. The simple definitions based on terminological relations are supplemented by information contained in facets and modifiers, which allow the user to capture different views of the data.

This paper describes the structure and evaluation of the syntactico-semantic lexicon (SSL) of the German Natural Language Understanding System VIE-LANG. VIE-LANG uses an SI-Net as internal representation. The SSL contains the rules according to which the mapping between net-structures and surface structures of a sentence is carried out. This information is structured in such a way that it can be evaluated from two sides. The parser interprets it as production-rules that control the analysis. Syntactic and semantic features of the input

sentence are evaluated and individuals are created in the semantic net. The generator uses the same rules to express selected net-structures in adequate natural language expressions. It is shown how both processes can make effective use of SSL. The different possibilities for evaluating the SSL are explained and illustrated by examples.

An Island Parsing Interpreter for the Full Augmented Transition Network Formalism
John A. Carroll
University of Cambridge Computer
 Laboratory
Corn Exchange Street
Cambridge CB2 3QG England
Proc. EACL 1983, pp. 101-105

Island parsing is a powerful technique for parsing with Augmented Transition Networks (ATNs) which was developed and successfully applied in the HWIM speech understanding project. The HWIM application grammar did not, however, exploit Woods' original full ATN specification. This paper describes an island parsing interpreter based on HWIM but containing substantial and important extensions to enable it to interpret any grammar which conforms to that full specification of 1970. The most important contributions have been to eliminate the need for prior specification of scope clauses, to provide more power by implementing LIFTR and SENDR actions within the island parsing framework, and to improve the efficiency of the techniques used to merge together partially-built islands within the utterance.

This paper also presents some observations about island parsing, based on the use of the parser described, and some suggestions for future directions for island parsing research.

WEDNESDAY: Parsing Flexible Word Order Languages
Oliviero Stock, Cristiano Castel- franchi, Domenico Parisi
Istituto di Psicologia
del Consiglio Nazionale delle Ricerche
Via dei Monti Tiburtini 509, 00157 Roma
Proc. EACL 1983, pp. 106-110

A parser for "flexible" word order languages must be substantially data driven. In our view syntax has two distinct roles in this connection: (i) to give impulses for assembling cognitive representations, (ii) to structure the space of search for fillers. WEDNESDAY is an interpreter for a language describing the lexicon and operating on natural language sentences. The system operates from left to right, interpreting the various words comprising the sentence one at a time. The basic ideas of the approach are the following:

a) To introduce into the lexicon linguistic knowledge that in other systems is in a centralized module. The lexicon therefore carries not only morphological data and semantic descriptions. Syntactic knowledge, partly of a procedural kind, also is distributed throughout it.

b) To build progressively a cognitive representation of the sentence in the form of a semantic network, in a global space, accessible from all levels of the analysis.

c) To introduce procedures invoked by the words themselves for syntactic memory management. Simply stated, these procedures decide on the opening, closing, and maintaining of search spaces; they use detailed constraints and take into account the active expectations.

WEDNESDAY is implemented in MAGMA-LISP, with a stress on the non-deterministic mechanism.

How to Parse Gaps in Spoken Utterances
G. Goerz, C. Beckstein
Univ. Erlangen-Nuernberg, RRZE
Martensstr. 1
D-8520 Erlangen, W. Germany
Proc. EACL 1983, pp. 111-113

We describe GLP, a chart parser that will be used as a SYNTAX module of the Erlangen Speech Understanding System. GLP realizes an agenda-based multiprocessing scheme, which allows us to apply easily various parsing strategies in a transparent way. We discuss which features have been incorporated into the parser in order to process speech data, in particular the ability to perform direction independent island parsing, to handle gaps in the utterance and its hypothesis scoring scheme.

A Flexible Natural Language Parser Based on a Two-Level Representation of Syntax

In this paper we present a parser which allows us to make explicit the interconnections between syntax and semantics, to analyze the sentences in a quasi-deterministic fashion and, in many cases, to identify

Leonardo Lesmo, Pietro Torasso
Istituto di Scienze dell-Informazione
Università di Torino
C.so Massimo D'Azeglio 42
10125 TORINO - ITALY
Proc. EACL 1983, pp. 114-121

the roles of the various constituents even if the sentence is ill-formed. The main feature of the approach on which the parser is based comprises a two-level representation of the syntactic knowledge: a first set of rules emits hypotheses about the constituents of the sentence and their functional role and another set of rules verifies whether a hypothesis satisfies the constraints about the well-formedness of sentences. However, the application of the second set of rules is delayed until the semantic knowledge confirms the acceptability of the hypothesis. If the semantics reject it, a new hypothesis is obtained by applying a simple and relatively inexpensive "natural" modification; a set of these modifications is predefined and only when none of them is applicable is a real backup performed: in most cases this situation corresponds to a case where people would normally garden path.

An Approach to Natural Language in the SI-Nets Paradigm

Amedeo Cappelli, Lorenzo Moretti
Istituto di Linguistica Computazionale,
 CNR
Via della Faggiola, 32
56100 Pisa - Italy
Proc. EACL 1983, pp. 122-128

This article deals with the interpretation of conceptual operations underlying the communicative use of natural language (NL) within the Structured Inheritance Network (SI-Nets) paradigm. The operations are reduced to functions of a formal language, thus changing the level of abstraction of the operations on SI-Nets In this sense, operations on SI-Nets are not merely isomorphic to single epistemological objects, but can be viewed as a simulation of processes on a different level, that pertaining to the conceptual system of NL. For this purpose, we have designed a version of KL-ONE which represents the epistemological level, while the new experimental language, KL-Conc, represents the conceptual level. KL-Conc would seem to be a more natural and intuitive way of interacting with SI-Nets.

An Experiment on Synthesis of Russian Parametric Constructions

I.S. Kononenko, E.L. Pershina
AI Laboratory, Computing Center
Siberian Branch of the USSSR Ac. Sci.
Novosibirsk 630090, USSR
Proc. EACL 1983, pp. 129-132

The paper describes an experimental model of syntactic structure generation starting from the limited fragment of semantics that deals with the quantitative values of object parameters. To present the input information the basic semantic units of four types are proposed: "object", "parameter", "function" and "constant". For the syntactic structure representation, the system of syntactic components is used that combines the properties of the dependency and constituent systems: the syntactic components corresponding to word forms and exocentric constituents are introduced and two basic subordinate relations ("actant" and "attributive") are claimed to be necessary. Special attention has been devoted to problems of complex correspondence between the semantic units and lexical-syntactic means. In the process of synthesis such sections of the model as the lexicon, the syntactic structure generation rules, the set of syntactic restrictions and morphological operators are utilized to generate the considerable enough subset of Russian parametric constructions.

Learning Translation Skills with a Knowledge-Based Tutor: French-Italian Conjunctions in Context

Stefano A. Cerri
Dipartimento di Informatica
Marie-France Merger
Dipartimento di Lingue e Letterature
 Romanze
Università di Pisa
56100 Pisa, Italy
Proc. EACL 1983, pp. 133-138

This paper describes an "intelligent" tutor of foreign language concepts and skills based upon state-of-the-art research in Intelligent Teaching Systems and Computational Linguistics.
  The tutor is part of a large R&D project in ITS which resulted in a system (called DART) for the design and development of intelligent teaching dialogues on PLATO and in a program (called ELISA) for teaching foreign language conjunctions in context. ELISA was able to teach a few conjunctions in English, Dutch and Italian. The research reported here extends ELISA to a complete set of conjunctions in Italian and French.

Towards Better Understanding Of Anaphora
Barbara Dunin-Keplicz
Institute of Informatics
Warsaw University
P.O. Box 1210
00-901 Warszawa, Poland
Proc. EACL 1983, pp. 139-143

This paper represents a syntactical method of interpreting pronouns in Polish. Using the surface structure of the sentence as well as grammatical and inflexional information accessible during syntactic analysis, an *area of reference* is marked out for each personal and possessive pronoun. This area consists of a few *internal areas* inside the current sentence and an *external area,* i.e., the part of the text preceding it. In order to determine that area of reference, several syntactic sentence-level restrictions on anaphora interpretation are formulated.

Next, when looking at the area of the pronoun's reference, all NPs which number-gender agree with the pronoun can be selected, and this way the set of *surface referents* of each pronoun can be created. It can be used as data for further semantic analysis.

Rules for Pronominalization
Franz Guenthner, Hubert Lehmann
IBM Deutschland GmbH
Heidelberg Science Center
Tiergartenstr. 15
D-6900 Heidelberg, FRG
Proc. EACL 1983, pp. 144-151

Rigorous interpretation of pronouns is possible when syntax, semantics, and pragmatics of a discourse can be reasonably controlled. Interaction with a data base provides such an environment. In the framework of the User Specialty Languages system and Discourse Representation Theory, we formulate strict and preferential rules for pronominalization and outline a procedure to find proper assignments of referents to pronouns.

Local and Global Structures in Discourse Understanding
M. Koit, S. Litvak, H. Oim,
T. Roosmaa, M. Saluveer
Artificial Intelligence Laboratory
Tartu State University
202400 Tartu, Estonian S.S.R, U.S.S.R.
Proc. EACL 1983, pp. 152-154

We are interested in the nature of content structures in terms of which it would be possible to account for reasoning processes in understanding natural language texts. One of the most crucial problems here at the present time is: how and by which mechanisms these reasoning processes are controlled and directed. As the emphasis in the design of discourse understanding systems so far has been on the problems of knowledge organization and representation, we are only beginning to guess what the corresponding processing mechanisms are and how they function, although an increasing number of papers has been devoted to these problems as well. There are studies of the relation of understanding to such types of knowledge processing as problem solving and planning. Various types of content units and structures needed to account for knowledge processing have been proposed in the general context of modeling discourse understanding. We ourselves have discussed an approach to knowledge units and processing mechanisms in questions, as a part of a computer system which understands stories of a certain kind, as well as on a more theoretical level.

Systemic Grammar in Computation: The Nigel Case
Christian M.I.M. Matthiessen
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292
Proc. EACL 1983, pp. 155-164

Computational linguistics needs grammars for several different tasks such as comprehension of text, machine translation, and text generation. Clearly, any approach to grammar has potentially something to offer computational linguistics, say for parsing or text generation (and, by the same token, there is a potential benefit from an application within computational linguistics for each approach). However, it is equally clear that some approaches have much more to offer than others. Here I take a look at Systemic Linguistics in the service of computational linguistics tasks, concentrating on a large computational systemic grammar for text generation (Nigel) that is currently being developed.

Inquiry Semantics: A Functional Semantics of Natural Language Grammar
William C. Mann
USC/Information Sciences Institute
4676 Admiralty Way

Programming a computer to operate to a significant degree as an author is a challenging research task. The creation of fluent multiparagraph text is a complex process because knowledge must be expressed in linguistic forms at several levels of organization, including paragraphs, sentences and words, each of which involves its own kinds of

Marina del Rey, CA 90292
Proc. EACL 1983, pp. 165-174

complexity. Accommodating this natural complexity is a difficult design problem. To solve it we must separate the various relevant kinds of knowledge into nearly independent collections, factoring the problem.

Inquiry semantics is a new factoring of the text generation problems. It is novel in that it provides a distinct semantic for the grammar, independent of world knowledge, discourse knowledge, text plans and the lexicon, but appropriately linked to each. It has been implemented as part of the Nigel text generation grammar of English.

This paper characterizes inquiry semantics, shows how it factors text generation, and describes its exemplification in Nigel. The resulting description of inquiries for English has three dimensions: the varieties of operations on information, the varieties of information operated upon, and the subject matter of the operations. The definition framework for inquiries involves both traditional and nontraditional linguistic abstractions, spanning the knowledge to be represented and the plans required for presenting it.

Natural Language Input for Scene Generation

Giovanni Adorni, Mauro DiManzo
Istituto di Elettrotechnica, U. of Genoa
Viale F.Causa 13
16145 Genoa, Italy
Giacomo Ferrari
Istituto di Linguistica Computazionale,
 CNR
Via della Faggiola
56100 Pisa, Italy
Proc. EACL 1983, pp. 175-182

In this paper a system which understands and conceptualizes scene descriptions in natural language is presented. Specifically, the following components of the system are described: the syntactic analyzer, based on a Procedural Systemic Grammar; the semantic analyzer relying on the Conceptual Dependency Theory; and the dictionary.

A Multilevel Approach to Handle Non-standard Input

Manfred Gehrke
Linguistics and Literature
University of Bielefeld
P.O. Box 8640
D-4800 Bielefeld 1
Proc. EACL 1983, pp. 183-187

In the project "Procedural Dialogue Models" being carried on at the University of Bielefeld, we have developed an incremental multi-level parsing formalism to reconstruct task-oriented dialogues. A major difficulty we have had to overcome is that the dialogues are real ones with numerous ungrammatical utterances. The approach we have devised to cope with this problem is reported here.

Case Role Filling as a Side Effect of Visual Search

Heinz Marburger
Research Unit for Information Science
 and Artificial Intelligence
University of Hamburg
Mittelweg 179
D-2000 Hamburg 13, F.R. Germany
Wolfgang Wahlster
FB10 - Angewandte Mathematik und
 Informatik
University of Saarbrucken
Im Stadtwald
D-6600 Saarbrucken 11, F.R. Germany
Proc. EACL 1983, pp. 188-195

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

gramming 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 Information Retrieval System Dialog
L. Bolc, K. Kochut, A. Lesniewski, T. Strzalkowski
Warsaw University
Institute of Informatics
PKin, pok.850
00-901 Warszawa, Poland
Proc. EACL 1983, pp. 196-203

The presented paper contains a description of an experimental version of the natural language information retrieval system DIALOG. The system is destined for use in the field of medicine. Its main purpose is to ensure access to information to physicians in a conversational manner. The use of the system does not require programming ability from its user.