American Journal of Computational Linguistics Microfiche 22

NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS
VOLUME 12 - NUMBER 4 SEPTEMBER 1975

This issue of the Journal contains no bibliography beyond announcements of some new books; a retrospective list from the Fondazione Dalle Mole, and two reviews. The Editor has been away from his office and the former Editorial Assistant is now employed full time elsewhere. Prospects for the next issue look good.

AMERICAN JOURNAL OF COMPUTATIONAL LINGUISTICS is published by the Center for Applied Linguistics for the Association for Computational Linguistics.

EDITOR: *David G. Hays, Professor of Linguistics and of Computer Science, State University of New York, Buffalo*

EDITORIAL SECRETARY: *Jacquin Brendle*

EDITORIAL ADDRESS: *Twin Willows, Wanakah, New York 14075*

MANAGING EDITOR: *A. Hood Roberts, Deputy Director, Center for Applied Linguistics*

PRODUCTION AND SUBSCRIPTION ADDRESS: *1611 North Kent Street Arlington, Virginia 22209*

TABLE OF CONTENTS

A C L Nominating Committee 3
13th Annual Meeting: Abstracts of Papers 4
Fabens, William, 5
Hobbs, Jerry R., 6
Miller, Perry L., 7
Martin, William A., 8
Burger, J. F. et al., 9
Medewa, P., et al., 10
McDonald, David, 11
Knaus, Rodger, 12
Rhyne, J. R., 13
Shapiro, Stuart C., 14
Slocum, Jonathan, 15
Meehan, Jim, 16
Bates, Madeline, 17
Paxton & Robinson, 18
Robinson, J. J., 19
Hendrix, Gary G. 20
Sondheimer, Norman K., 21
Cercone, Nick, 22
Deutsch, Barbara G., 23
Bruce, Bertram, 24
Phillips, Brian, 25
Cullingford, R. E., 26
Badler, Norman, 27
Kegl and Chinchor, 28
Rosenschein, Stan, 29
Klappholz and Lockman, 30
Beckles, Carrington, Warner, 31
Brill and Oshika, 32
Salton, G., 33
Anderson, Bross, Sager, 34
A S I S Annual Meeting 35
A C M Annual Conference 39
SOLAR Project Terminates--ARPA support ends 42
Energy Information Tools: IIA-NFAIS Workshop 42
MT: Chinese Journals offered on subscription from Hong Kong 43
Librarian of Congress: ASIS telegram to President Ford . . . 44
Bibliography on Semantics and Cognition at Castagnola 46
Recent Publications 48
Conceptual Information Processing, 48
Automatic Translation at Grenoble, 49
Frequency and Distribution (French novels), 51
A L L C Bulletin, 52
Word Order and Word Order Change, reviewed by James M. Dunn 53
Informal Speech, reviewed by John B. Carroll 78
Personal Notes 92

NOMINATING COMMITTEE

A slate of officers for 1976 will be presented at the annual meeting in Boston by

William A. Woods, Jr., Chairman

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

Robert Simmons

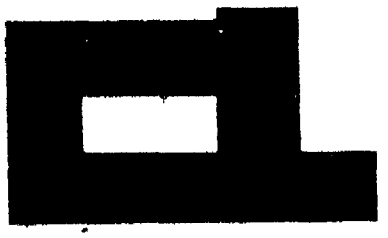
University of Texas

Robert Barnes

Lehigh University

Members of the Association can submit suggestions to the chairman or any member of the Committee.

American Journal of Computational Linguistics *Microfiche 22 : 4*



THIRTEENTH ANNUAL MEETING

ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

ABSTRACTS OF PAPERS ACCEPTED

PROGRAM COMMITTEE:

TIM DILLER, CHAIRMAN
Sperry-Univac
JON ALLEN
Massachusetts Institute of Technology
JOYCE FRIEDMAN
University of Michigan
BONNIE NASH-WEBBER
Bolt Beranek and Newman, Inc.
CHUCK RIEGER
University of Maryland

TIME AND PLACE: Sheraton Boston Hotel
October 30 - November 1, 1975

SESSIONS:

1. LANGUAGE UNDERSTANDING SYSTEMS
2. LANGUAGE GENERATION SYSTEMS
3. PARSING, SYNTAX, AND SEMANTICS
4. MODELING DISCOURSE AND WORLD KNOWLEDGE
5. TEXT ANALYSIS

BUSINESS MEETING AND ELECTION OF OFFICERS

ABSTRACTS APPEAR ON THE FOLLOWING FRAMES

PEDAGLOT AND UNDERSTANDING NATURAL LANGUAGE PROCESSING

WILLIAM FABENS
Rutgers University

PEDAGLOT is a language processor that contains a meta-parser. This parser is programmable not only in its syntax, semantics, and their related activities, but also with respect to the modes of operation. The idea of parsing modes differs from other ways of programming parsers in that it is related to the behaviors desired from a parser rather than to its lower level functions. The dozen modes described are roughly independent of one another.

The main activities of parser form a dialectic process of prediction, discovery and construction, all embedded in a higher order control structure. The modes indicate the behavior of each such activity.

This way of programming a parser is seen as a step toward representing theories of natural language in a unifiable way, in the sense that various mode settings might yield parsers, generators, language inferrers, etc. The matter of unification is found to be non-trivial (i.e. a generator is not just an inverse parser), yet the notion of modes makes it at least tractable. Applications of PEDAGLOT to various current theories are discussed

A SYSTEM FOR GENERAL SEMANTIC ANALYSIS AND ITS USE IN DRAWING MAPS FROM DIRECTIONS

JERRY R. HOBBS
The City College of CUNY

Our data base is a set of facts involving spatial terms in English; our system takes as input directions of how to get from one place to another and outputs a map. As input we use the output of the Linguistic String Project's transformational program.

The problem of semantic analysis is to find quickly out of a potentially enormous collection the appropriate inferences. The key to selection is Joos' Semantic Axiom Number One (restated) "The important facts in a text will be repeated, explicitly or implicitly." Those inferences which should be drawn are those which are keyed by more than one element in the text.

Semantic interpretations include interpretation of higher predicates, finding antecedents of definite noun phrases, and identifying intersentence relations

We divide the sets of inferences into clusters, this organization of the data base is task dependent, in our application, the top-level cluster concerns the one-dimensional aspects of objects and actions. When a fact in a cluster is accessed it becomes the top-level cluster

We tag our inferences always, normally, or sometimes

The task component makes arbitrary decisions required by the map but not given in the text. A geometry is imposed on the topological natural language information and the map is drawn.

AN ADAPTIVE NATURAL LANGUAGE SYSTEM THAT LISTENS, ASKS, AND LEARNS

PERRY L. MILLER
Massachusetts Institute of Technology

When a user interacts with a natural language system, he may well use words and expressions which were not anticipated by the system designers. This paper describes a system which can play TIC-TAC-TOE, and discuss the game while it is in progress. If the system encounters new words, new expressions, or inadvertent ungrammaticalities, it attempts to understand what was meant, through contextual inference, and by asking intelligent clarifying questions of the user. The system then records the meaning of any new words or expressions, thus augmenting its linguistic knowledge in the course of user interaction.

CONCEPTUAL GRAMMAR

WILLIAM A. MARTIN
Massachusetts Institute of Technology

In OWL, an implementation of conceptual grammar, the two types of data items are symbols and concepts and the two basic data composition operations are specialization and restriction

A symbol is an alphanumeric string headed by " Symbols correspond to words, suffixes, prefixes, and word stems in English and the programmer can introduce them at will.

OWL conce ts correspond to the meanings of English words and phrases. They are constructed using the specialization operation, comparable to CONS in LISP· (A B) is the specialization of A, a concept, by B, a concept or symbol OWL forms a branching tree under specialization, with SOMETHING at the top.

Concepts are given properties by restriction, which puts a concept on the reference list of another concept (compare property lists and S-expressions in LISP) A/B is the restriction of A by B.

The categories in the specialization tree are semantic, but we use them also for the purposes usually assigned to syntactic categories.

A predication is a double specification of a model such as present tense or can. Examples are

The pool is full of water. ((PRES-TNS (BE (FULL WATER))) POOL/THE)
The cookie can be in the jar ((CAN (BE (IN JAR/THE))) COOKIE/THE)
Bob is the father of Sam ((PRES-TNS (BE (FATHER SAM)/THE)) BOB)
Bob hits the ball. ((PRES-TNS (HIT BALL/THE)) BOB)
Bob is hitting the ball ((PRES-TNS (BE (-ING (HIT BALL/THE))))BOB)

Starting from this base we will discuss a number of issues such as nominalization, incorporation, and deep vs surface cases.

SEMANTIC-BASED PARSING AND A NATURAL-LANGUAGE INTERFACE FOR INTERACTIVE DATA MANAGEMENT

JOHN F BURGER, ANTONIO LEAL, AND ARIE SHOSHANI
System Development Corporation

This paper describes the current state of work-in-progress on a system having both applied and theoretical relevance for computational linguistics. At the applications level, the program we are developing is an interface that will give a natural-language communications facility to users of existing data management systems. The interface can be adjoined to a wide variety of data management systems with relatively little effort. Advantages and disadvantages of this approach to reducing the "DMS communications gap" are discussed.

The theoretical part of the work shows that useful information in a natural-language expression (i.e., its "meaning") can be obtained by a parser controlled by a grammar that uses no description of syntax whatever. The construction of the parsing tree is controlled primarily by semantics in the form of a concept network consisting of an abstraction of the "micro-world" of the data management system's data organizing methods, its functional capabilities, and the semantic relations of the data base content material. Discussion includes a possible method whereby such a parser might be extended to be used with more general "worlds" if "frames" could be used to temporarily restrict world-view information.

PHILIQA I: MULTILEVEL SEMANTICS IN QUESTION ANSWERING

P. MEDEWA, W. J. BRONNENBERG, H. C. HUNT, S. P. J. LANDSBERGEN,
R. J. H. SCHA, W. J. SCHOENMAKERS, E. P. C. VAN UTTEREN
Philips Research Laboratories, Eindhoven

This paper outlines a recently implemented question answering system, called PHLIQA 1, which answers English questions about a data base.

Unlike other existing systems, that directly translate a syntactic deep structure into a program to be executed, PHLIQA 1 leads a question through some intermediate stages of semantic analysis thereby providing for a perspicuous treatment of some hitherto largely neglected semantic phenomena. In every stage, a question is represented as an expression in a formal language. The paper sketches the syntax and semantics of the languages designed for this purpose, and points out the distinctions between the semantic representations used at the most important levels of semantic analysis, viz.:

- one at which the constants and language constructs are derived from those of English,
- one that may be characterized as representing the system's assumptions about the "structure of the world", reflecting neither the peculiarities of English nor those of the data base,
- one at which the constants match the ones in the data base and the available arithmetical and logical functions, while the language constructs are well-suited to algorithmic interpretation.

PHILQA 1 was designed in such a way that it can be described abstractly, independent of the implementation, as a series of sets of transformation rules, called "convertors" A convertor translates expressions of one level of semantic interpretation into expressions of the next level. The paper gives a sketch of each of these convertors, and shows how they translate an example question step by step into the data base language

A FRAMEWORK FOR WRITING GENERATION GRAMMARS FOR INTERACTIVE COMPUTER PROGRAMS

DAVID McDONALD
Massachusetts Institute of Technology

Interactive programs which wish to employ fluent natural language will be communicating specific messages to their users in specific pragmatic and discourse situations. Accordingly, a generation grammar must be so structured that the particulars of the message and the context themselves direct what linguistic devices--words, phrases, syntax--are used in translating the message to English.

A program s message describes what objects and relations are to be mentioned and what the program's intentions are (e g to make a prediction, to answer a query, etc.). In the framework described, a grammar consists of a body of procedures which perform the translation in two phases. The first phase uses procedures associated with the intentions to construct a plan for the utterance as a whole--a surface level model incorporating the message's elements largely unanalyzed, but in their intended positions (subject, main verb, etc.).

During the second phase, the plan is refined, from left to right, following its constituent structure. Each message element is analyzed and translated in turn as it is encountered in the plan. The translation is done by discourse-sensitive procedures taken from a lexicon of the possible elements which the program might use. By working from left to right, possible combinatorial interactions between descriptive processes are avoided Also, a convenient formalism for encoding surface structure dependant ~~phenomena~~ such as pronominalization and quantifier scope becomes possible.

The framework is described with examples drawn from the imple mentation in progress for an appointment scheduling program. Particular attention is given to how the characteristics of natural language dictate or influence the framework's design.

INCREMENTAL SENTENCE PROCESSING

RODGER KNAUS
University of California, Irvine

Human short term memory is bounded while the speaking rate is nearly constant. Therefore a listener must use a parsing algorithm for which there is a uniform bound over all sentences for the time needed to add a new word onto existing partially completed parses A context free parsing algorithm, the incremental parser, is presented which has a bounded parsing time per word on a class of context free sentences containing the syntactic phrase markers of natural language sentences. Psycholinguistic evidence supports the breadth first strategy used by the incremental parser

The incremental parser is extended to a class of generation grammars with function-like rules more suitable for natural language sentence generation than context free grammars. Instead of grammar nonterminals each generation rule expands a syntactic form consisting of a grammatical category C (noun phrase, sentence, etc.) followed by a list of rule arguments, including a semantic network node N and an association list about the purpose of the sentence. The generation rule expands the syntactic form into a list of words and syntactic subforms each of which generates a syntactic substructure of C from a semantic substructure of N

A heuristic function FIND is defined which when given a syntactic form F and a list L of syntactic forms generated by F finds a set of association lists A such that with the variable bindings of A, evaluating F produces L By a bottom-up application of FIND to the syntactic structures built by a regular expression incremental parser, the generation grammar parser finds the semantic nodes and sentence purpose alists which might have produced a given sentence from a generation grammar.

A LEXICAL PROCESS MODEL OF NOMINAL COMPOUNDING IN ENGLISH

J. R. RHYNE
University of Houston

"Lexical processes" are rather idiosyncratic. Some examples are nominalization, lexical incorporation, nominal compounding, and lexical substitution. These processes are partly syntactic, partly semantic, and partly controlled by information associated with each lexical item. These processes have been largely ignored by computational linguists; recent models with a substantial knowledge base suggest that modeling lexical processes will not be quite as large a task as might have previously been thought.

I have constructed a model which accepts relative clauses and produces nominal compounds. The syntax is simple, if compounding changes a relative clause into a noun-noun pair, it has no semantic aspect. However, only a few relative clauses can be changed into compounds; this process must be controlled by the lexical items

The computer model uses lexical rules associated with items in the lexicon. These rules consist of left part and right part structures representing relative clauses and nominal compounds in a case structure system.

One form of English compound is made by deleting the main verb and several noun phrases and placing the noun which remains in front of the head noun of the original relative clause This process can be allowed only when the deleted information is "lexically recoverable".

The model has been used to generate several hundred nominal compounds and is very efficient, even with a lexicon containing nearly a hundred rules and capable of generating several thousand different nominal compounds. The kinds of rules used can account for other lexical processes. The same rules can be used for recognition of nominal compounds; this has not been done and would probably require a substantial knowledge base to disambiguate some compounds.

GENERATION AS PARSING FROM A NETWORK INTO A LINEAR STRING

STUART C. SHAPIRO
Indiana University

In this paper, we discuss the approach we are taking to the generation of English surface strings from a semantic network The semantic network representation does not contain surface features of the original sentences such as tense or voice Instead of tense, temporal information is stored relative to a growing time line. Voice is considered to be information about the original speech act rather than essential information conveyed by the sentence. As a result, the generated sentences are not necessarily the same as the original sentence.

We view generation as the creation of a linear surface string that describes a node of the semantic network The form of the surface string is controlled by a recursive augmented transition network grammar which is capable of examining the form and content of the semantic network connected to the semantic node being described. A single node of the grammar network may result in different forms of surface strings depending on the semantic node it is given, and a single semantic node may be described by different surface strings depending on the grammar node it is given to For example, a semantic node may be described as an independent sentence in one instance, as a relative clause in another, and as a nominalized sentence in another.

Our approach is to generate surface strings left to right Rather than start with several deep phrase markers which must be connected appropriately and transformed, we are starting with a network where deep structures are already properly connected These can be examined by the grammar network which can then build the final surface string directly.

SPEECH GENERATION FROM SEMANTIC NETS

JONATHAN SLOCUM
Stanford Research Institute

Natural language output can be generated from semantic nets by processing rules that are associated with verbs which correspond to concepts in the net A rule is essentially a sequence of case names The set of rules is being derived from a study of the surface syntax of some 3000 English verbs The active forms of the verbs have been classified according to subject, object(s), and complements--ignoring adverbials of manner, time, distance, etc., which are inserted by heuristic rules Passives are similarly derived. These major argument patterns are in the process of being converted semi-automatically into sequences of case names which will be used as templates in the generation of text The text will be in a form that can be entered into a speech synthesis program Some initial experiments with a VOTRAX speech synthesizer are being conducted.

USING PLANNING STRUCTURES TO GENERATE STORIES

JIM MEEHAN
Yale University

This paper discusses a computer program which makes up stories from its knowledge of the world, including the characters planning structures (goals and plans for achieving those goals) Goals can be constant (such as preserving one s health) or recurring (eating whenever you're hungry), and can produce subgoals which are more immediate in nature (eating a hamburger now) A plan for achieving a goal can include several subplans, each with a set of applicability tests and enabling preconditions. There must also be a decision algorithm to use in deciding which of several subplans to try first. The criteria for a subplan can relate to the character who is doing the planning, to the other characters in the story, or to general information about the world After the initial state of the world is established, the program can change details, present obstacles to goals, and introduce unusual events to make the story interesting. A causal chain of states and acts, representing the story in Conceptual Dependency notation, is generated by programs which model combinations of planning structures. The structures are combined by nesting (when a precondition for a particular subplan requires other plans), in series (when several plans are required), or in parallel (when goals are concurrent). The characters in the present system are talking bears, birds, and other animals. This simplifies somewhat the problem of handling enormous quantities of world knowledge, but without changing the basic problems in generating coherent, interesting stories.

SYNTACTIC PROCESSING IN THE BBN SPEECH UNDERSTANDING SYSTEM

MADELINE BATES
Bolt Beranek and Newman Inc.

The syntactic analysis system presented here is composed of two parts, a modified augmented transition network grammar and a parser which is designed for a speech understanding environment.

The parser operates on partial utterances called theories A theory may be thought of as a set of words which are hypothesized to be in the utterance. The parser processes the words in a theory by building partial syntactic paths using the words of the theory. These paths do not depend on left context, which will be missing if there are gaps in the theory. Syntactic constituents are built where possible and, whenever a constituent is built, the parser can interface with the semantic component of the total speech understanding system for guidance and verification.

The parser tries to predict words and/or syntactic categories to fill or reduce gaps in the theory, particularly small function words which are difficult to detect reliably on acoustic grounds alone. The parser does not follow all possible parse paths, but attempts to select the most likely ones for extension. It uses a judicious mixture of top down, bottom up, depth first, and breadth first parsing strategies to take advantage of local, reliable information It saves all the information gained while following alternative parse paths, so that several parse paths which share a common part, even if the paths are in different theories, can share that portion without reparsing. This is true even if the parse paths split before and or after the common part and even if the common section analyzes only part of a syntactic constituent

SYSTEM INTEGRATION AND CONTROL FOR SPEECH UNDERSTANDING

WILLIAM H. PAXTON AND ANN E. ROBINSON
Stanford Research Institute

Acoustics, syntax, semantics, discourse, and pragmatics play roles in speech understanding and can be integrated into a system that allows the interactions to be easily visible.

The language definition is the focal point for the integration. Basic to the language definition are phrases built from individual words and from other phrases. Integration occurs at the level of each individual phrase. For each phrase type, a statement in the language definition specifies (1) which kinds of knowledge to use and (2) how much weight to give to each source in computing the likelihood that an instantiation of the phrase type is a correct interpretation.

The executive uses a complex heuristic control strategy to control sources of knowledge, establishing priorities for alternative tasks. The processing of an utterance is factored into tasks that make incremental changes to a global data structure and spawn other tasks. Priorities reflect both the expected values of interpretations and the relation of the task to the executive s current focus of activity. The expected values take into account the context established by prior tasks and are based on phrase scores that combine non-Boolean evaluation factors from a variety of knowledge sources. The focus mechanism allows tentatively accepted phrases to inhibit the search for others that would replace them. The tasks and the global data structure are structured in a way that brings together related activities to eliminate duplication of effort and makes it possible to coordinate processing driven by acoustic data with processing driven by goals based on predictions made by higher level linguistic components.

A TUNEABLE PERFORMANCE GRAMMAR

JANE J ROBINSON
Stanford Research Institute

A performance grammar (PG) aims to define the form and meaning of intelligible speech uttered during the course of spontaneous dialog. Its definitions are tuneable to particular utterances in particular dialogs. That is, given the problem of determining the applicability of a definition to the understanding of some portion of an utterance, the definition itself specifies how the attributes of that portion and the properties of the discourse and the speaker affect the likelihood that it should be applied.

This paper presents a tuneable PG being developed for a computer-based speech understanding system. Two different discourse contexts from which its definitions are derived are compared and contrasted, and ways of tuning the word and phrase definitions to them by means of 'factor' statements are described. Sequences of definitions involved in parsing and interpreting utterances are examined in detail, emphasizing the interaction of selected factors for evaluating the likelihood of their application. The selected factors are called 'syntactic', but the attributes that are evaluated may be semantic, pragmatic, acoustic-phonetic, or discourse based. It is shown that superficial syntactic factors are useful for disconfirming a wrong parsing path or confirming a correct one, in ways that reduce the need to call on the semantic, discourse, and acoustic components for in-depth evaluations. It is also shown that factors evaluating number agreement, which is traditionally a syntactic matter, need to refer to semantic attributes. This demonstration points to the conclusion that integrating information from different kinds of analyses or sources of knowledge is well-motivated on both linguistic and heuristic grounds.

SEMANTIC PROCESSING FOR SPEECH UNDERSTANDING

GARY G. HENDRIX
Stanford Research Institute

The semantic component of the speech understanding system being developed jointly by SRI and SDC performs two functions. it rules out those word combinations that are not meaningful, and it produces a semantic interpretation for those combinations that are. The semantic system described in this paper consists of a semantic model embodied in a network and a number of routines that interact with it. The model may be characterized as a description of objects, actions, and relations in the world. The semantic network encoding the model is partitioned into a set of hierarchically ordered sub-nets. This partitioning facilitates the categorization of objects, the encoding of quantification, and the maintenance of multiple interpretation hypotheses during parsing.

Interacting with the semantic network is a set of routines, associated with a set of language definition rules, that combine utterance components into larger phrases. In the course of their operations, these routines consult network descriptions of prototype situations and events and reference data describing how surface cases may be mapped into network representations. The output from these routines is a semantic network fragment consisting of several sub-nets. At the utterance level, the composite of these sub-nets is a complete network description of the semantics of the utterance, while the hierarchical ordering of the sub-nets reflects the syntactic composition of the input

SPS: A FORMALISM FOR SEMANTIC INTERPRETATION AND ITS USE IN PROCESSING PREPOSITIONS THAT REFERENCE SPACE

NORMAN K. SONDHEIMER
The Ohio State University

This paper presents a formalism, called SPS, for writing semantic processors for natural language understanding systems SPS is intended for use in turning underlying syntactic structures in the form of constituent structure trees into underlying semantic structures in the form of nets composed of PLANNER-like assertions The formalism is based on Woods-style "pattern→action" rules The pattern element specifies tree fragments and various types of selectional restrictions. On the action side a variety of devices, including the use of registers, allow common reference to entities in the assertions produced. The registers used for reference can also be used to specify selectional restrictions across rules and for establishing default conditions for handling semantic ellipsis. Finally, SPS provides a control structure for the ordering of the application of the rules that interpret constituents and to control, in part, where the tree fragments are matched.

The power of SPS is seen in its unique ability to allow for the development of Case structures, especially the structures connected with the English prepositions that reference location, orientation and motion in space. These forms have always been troublesome for Case systems. Particularly difficult are the facts that 1) more than one of these prepositions can appear in a sentence in the same role, 2) their appearance can correspond to the need for multiple predicator semantic structures, and 3) they exhibit complex distributional and semantic relations among themselves and with respect to other sentential elements. SPS can allow for each of these phenomena.

An interpretation system for SPS has been implemented in LISP 1.6.

THE NATURE AND COMPUTATIONAL USE OF A MEANING REPRESENTATION FOR WORD CONCEPTS

NICK CERCONE
University of Alberta

The proposition-based semantic network notation of Schubert is especially well suited for including pragmatic and semantic information as part of the meaning representation of individual word concepts. Reperesentations are networks based on propositions that consist of an n-ary predicate with a finite number of arguments. Terms are used to represent a given word concept can also be represented by semantic networks; there is no insistence that a given set of "primitives" comprise the meaning of a word.

Implication templates stored with the network help identify arguments that we expect to find in the surface utterance, they make the most commonly used inferences part of the meaning representation of a word concept.

Whenever the meaning representations of several word concepts are similar, we can extract the similarity and use that part of the network as a higher or more general level concept. More efficient memory storage utilization is obtained. Implication templates that involve only the extracted propositions are stored with the higher level concepts.

The use of networks has several consequences. A natural hierarchy of levels of analysis is suggested, the terms used in the network representation for another term are themselves represented and explicated through networks. Heuristic algorithms and other organizations can be superimposed to take advantage of special situations. The representation is suggestive of the meaning of a term independently of the routines used to process it. Binary decomposition trees are neither suggestive of the type of processing required nor of how they are internally consistent.

ESTABLISHING CONTEXT IN TASK-ORIENTED DIALOGS

BARBARA G. DEUTSCH
Stanford Research Institute

Task-oriented dialogs comprise conversation directed toward the completion of some task. For these dialogs, context is supplied both by the surrounding task (the task context) and by the surrounding dialog (the dialog context). This paper describes the discourse component of a speech understanding system for task-oriented dialogs. This component evaluates a proposed interpretation for an utterance in terms of how well it fits the context surrounding the utterance. In particular, the discourse component identifies the referents of noun phrases and fills in missing information in elliptical expressions. Task context is supplied by a model of the subtasks constituting a task and their relationships to one another. The dialog context is supplied by a history of the preceding utterances In building a representation of the dialog context, the discourse processor takes advantage of the fact that task-oriented dialogs have a structure that closely parallels the structure of the task A semantic network is partitioned into focus spaces with each focus space containing those concepts pertinent to the dialog relating to a subtask. The focus spaces are linked to their corresponding subtasks, they are ordered in a hierarchy determined by the relations of the subtasks to one another. Hence, this mechanism both supplies a dialog context and coordinates it with a task context.

DISCOURSE MODELS AND LANGUAGE COMPREHENSION

BERTRAM C. BRUCE
Bolt Beranek and Newman Inc.

Work of Goffman, Sacks, Labov, Schegloff, Searle, Schmidt and others has demonstrated the prevalence of higher order structures in communication. These structures are not just undifferentiated means for organizing discourse, but rather, essential carriers of information. As such, they must be included in any complete theory of language understanding.

This paper compares two approaches to modeling discourse. The first centers on the concept of a "discourse grammar" which defines the set of likely (i.e. easily understood) discourse structures. Participants in a discourse are then, in effect, following a path through the discourse grammar. A major advantage of the grammar approach is that it provides a relatively easy way to give an economical characterization of a wide range of discourse types. On the other hand, it sometimes makes faulty predictions from which it is difficult to recover.

The second approach is a "demand processing" model in which utterances create demands on both the speaker and the hearer. Responses to these demands are based on their relative "importance" the length of time they have been around, and conditions attached to each demand The flow of responses provides another level of explanation for discourse structure

These two approaches are discussed in terms of flexibility and efficiency. Finally, the paper considers their role in a more complete theory of discourse understanding.

JUDGING THE COHERENCY OF DISCOURSE (AND SOME OBSERVATIONS ABOUT FRAMES SCRIPTS)

BRIAN PHILLIPS
University of Illinois at Chicago Circle

The surface form of any discourse is logically incomplete The prime task for a reader is to fill in the gaps, if he is unable to do this, the discourse is incoherent.

Chomsky notes certain syntactically recoverable items, but the structure of discourse cannot be described by an extension of syntactic devices, the most perspicuous description can be made in terms of cognitive concepts. In such a model we can give canonical representation to facts which enable discourse with gaps to be understood. Omissions will be inferred from world knowledge

The hypothesis is that a coherent discourse must be (a) connected and (b) have a single topic My system has been applied to a conceptual analysis of some half-dozen examples of stories of accidental drowning written by students The examples conform to a single abstract prescribed topical pattern that can only be recognized after inference of logical connections omitted by writers.

A conceptual encyclopedia in the system has frame/script-like structures. There are several possible objections to the systems of Minsky and Schank.

1 A lack of flexibility. The structures do not seem well suited to handling novel scenes.

2 ' There should be some bottom-up means of selecting a suitable frame which may then be used predictively.

3. Frame recognition must be recursive

In the present system, 1 is overcome by having groups of concepts of various degrees of abstraction Abstract structures are located during analysis, this improves on 2. The relations between a complex concept and constituent concepts are explicit and can be used as in 3.

AN APPROACH TO THE ORGANIZATION OF MUNDANE WORLD KNOWLEDGE: THE GENERATION AND MANAGEMENT OF SCRIPTS

R. E. CULLINGFORD
Yale University

In understanding stories or natural language discourse, human hearers draw upon an enormous base of shared world knowledge about specific situations to help establish the needed background or context. Much of this knowledge appears to be episodic in nature, distilled from many experiences in common situations. Thus, for example, ordinary members of middle-class American culture have in common much information about mundane activities like going to birthday parties, restaurants, or supermarkets, simply because these things are done in much the same way all over the country. This paper presents an approach to the representation and management of this type of low-level world knowledge, based upon the concept of a situational script (Schank and Abelson). The application of scripts in story understanding will be illustrated via a working computer model called SAM (Script Applier Mechanism).

As implemented in SAM, a situation script is a data base comprised of interlinked causal chains describing the widely understood paths and turning points encountered in stories about mundane activities like eating in a restaurant or riding a subway The process of story understanding begins with the construction of a 'trace' through this structure which contains the explicit input, other events not mentioned but commonly known to have happened, the more important enabling or resultative inferences to be drawn from the events, and the causal links connecting them. This trace or scenario is then examined by programs which generate summary, paraphrase or question-answering output. In complicated stories requiring several scripts, SAM handles the invocation and disabling of sequential, parallel and nested scripts.

THE CONCEPTUAL DESCRIPTION OF PHYSICAL ACTIVITIES

NORMAN BADLER
University of Pennsylvania

In proposing conceptual theories of language it is easy to overlook a primary motivation for language: The description of sensory information. The visual system provides much of this input and we easily generate descriptions of the activities we perceive happening around us.

In the first part of this paper we outline a representation for objects and events which enables us to describe certain classes of changes in object attributes and relations in conceptual terms. The visual aspects of this problem, as well as the linguistic, are described in the the thesis from which this paper is condensed. The methodology for the description of the motion of rigid or jointed objects in a simulated man-made environment utilizes simple "demon" procedures with a straightforward control structure to watch for semantically-significant situations or changes at several different levels. The organization of these levels is hierarchic, depending on spatial trajectory and rotations, observer movement, and spatial context at the low levels, directional prepositions and adverbs (adverbials) at the intermediate level, and motion verbs at the highest level.

In the second part of this paper we describe the event building algorithm which essentially fills in "deep" cases of a generic motion verb, such as Schank's "PTRANS" or Miller"s "TRAVELS". In this sense the resultant descriptions are compatible with current paradigms of natural language understanding systems (for example, Rumelhart et al.). Moreover, the event construction semantics are based on physical activities so that conceptual events can be related to actual occurrences. Examples are given for descriptions obtained for a few scenes.

A FRAME ANALYSIS OF AMERICAN SIGN LANGUAGE

JUDY KEGL
Massachusetts Institute of Technology

NANCY CHINCHOR
University of Massachusetts

Due to the fact that frames were first used explicitly in terms of visual imagery, the clearest examples of frames are those dealing with visual information. In order to see how a frame analysis of language would work, we thought it would be instructive to examine American Sign Language. a multidimensional spatial language. Our major aim is to use the computational formalism of frames in order to reach a more sophisticated understanding of American Sign Language. In so doing, we have been forced into certain definitions of computational concepts which we hope will be helpful in he general computational work on natural language.

Definitions of such computational concepts as perspective, scenario. frame, and prototype are given in the spirit of Minsky and, more primarily, Winograd. These definitions lead to a better understanding of how signs are abstracted from the perception of an event in order to communicate that event. From the linguistic standpoint, we focus mainly on the concept of verb and the processes of indexing and referring to objects in ASL. A by-product of these findings is a further clarification of the distinction between mime, statement, and discourse in ASL. The data which we will present as an illustration for our work is a telling by a native speaker of ASL of the story of "The Three Little Pigs" which we have on voice-over videotape.

HOW DOES A SYSTEM KNOW WHEN TO STOP INFERENCING?

STAN ROSENSCHEIN
University of Pennsylvania

Natural language processing systems that are sensitive to semantics and pragmatics generally draw 'inferences'; the presence of certain 'thoughts' or 'beliefs' triggers the retrieval and/or construction (usually pattern-matched) of other related thoughts or beliefs (the inferences). The problem we attack is this How can this process be controlled?

Various approaches are possible. One might use external criteria, such as ordering the inferences associated with a given antecedent, attaching 'probabilities' to the inferences, having the control program set an arbitrary limit on the length of inference chains, etc. The drawback of this approach is its arbitrariness. Alternatively, one could choose a strongly goal-oriented approach, however it is not clear how this approach might be reconciled with the data-driven (bottom-up) nature of free inferencing.

We have been looking at inferencing as an operation depending on the whole set of beliefs. The purpose of the operation is to find the least extension of the set which causes the beliefs to cohere, that is, to satisfy some pre-defined pattern. The set of patterns has to have some well-defined structure, it is then possible to define an inference operation in which there is no sharp distinction between antecedent and consequent, a pattern is a collection of subpatterns each of which may serve as an antecedent on one occasion and as a consequent on another

We have viewed the problem of restraining the process of inference as essentially one of making precise the idea of minimal unifying structure for a set of beliefs in such a way that internal (rather than external) criteria are established for inference cut-off. The ideas will be illustrated by various examples of inferring event descriptions, including descriptions of speech acts.

CROSS-SENTENTIAL REFERENCE RESOLUTION

DAVID KLAPPHOLZ AND ABE LOCKMAN
Columbia University

The problem of cross-sentential reference resolution involves the determination of the normally selected referents of pronouns across sentence boundaries, both in the absence and presence of "referent-forcing" context (sentences 1 and 2 vs. 3 and 4)

1. Yesterday a group of boys ran after a pack of dogs, the largest one broke a leg.
2. Yesterday John chased Bill half a block, he was soon out of breath.
3. The wild dogs outside our village all seem to suffer from a bone-weakening disease. Yesterday a group of boys ran after a pack of dogs, the largest one broke a leg
4. My friend Bill has a severe case of asthma. Yesterday John chased Bill half a block, he was soon out of breath

Note that in the former cases the preferred referents seem to be the surface subjects of the first sentences of each pair.

A further ramification of the problem is the determination of the relationship which the reference bears to the referent when the former is other than a pronoun (sentences 5 and 6)

5. John went for a long walk yesterday, the park was all abloom.
6. I met a lovely family yesterday, the father is a computer scientist.

Here the problem is to determine that "the park" is an area through which John took a long walk, and that "the father" is the father of the lovely family.

An attempt is made to formalize the notion of cross-sentential "focus", this notion is incorporated into an attempt to devise a general algorithm for establishing cross-sentential referents and their relationships to their references in the context of a primitive-based, inference driven model of natural language conversation. The question of properly directing an inference mechanism through a large base of world knowledge in the solution of the reference problem is discussed, and partial solutions presented

DEVELOPING A COMPUTER SYSTEM FOR HANDLING INHERENTLY VARIABLE LINGUISTIC DATA

DAVID BECKLES, LAWRENCE CARRINGTON, AND GEMMA WARNER
The University of the West Indies

Linguistic communication in Trinidad and Tobago is characterized by use of varieties of English and varieties of Creole English in a sociolinguistic complex that appears similar to what has been described in Jamaica and Guyana as a post-creole dialect continuum. A host of pedagogical problems result from the absence of adequate description of the language system and the mismatch between the socio-linguistic facts and instructional methodology The tape-recorded speech of a sample of children (aged 5-11+) is being analyzed to determine

(a) the structure of their language,
(b) the correlation of socio-linguistic factors with structures,
(c) their progress in the acquisition of English

Given the inherently variable nature and volume of the data, manual counting of features and correlation with factors is not feasible. This paper is concerned with the development of a computer system for handling such data Because of the difficulty of performing linguistic analyses by computer, the system is designed to deal with manually codified data, the results of such coding being among other things derivational trees with associated grammatical and semantic information Since the communication complex does not have readily identifiable norms, the analytical method and matching computer system have to effect recognition of stable sub-systems (regardless of which set of external criteria constitute the determinants of these sub-systems) as well as state the evolution of the children's language. The computer system takes as input the derivational trees with associated grammatical information and semantics and classifies them in a fashion that allows the output stated above.

A NATURAL LANGUAGE PROCESSING PACKAGE

DAVID BRILL AND BEATRICE T. OSHIKA
Speech Communications Research Laboratory

A set of SAIL programs has been implemented for analyzing large bodies of natural language data in which associations may exist between strings and sets of strings. (A file containing parallel orthographic, phonemic, and phonetic transcriptions of a discourse would be an example of this type of data) These programs include facilities for compiling information such as frequency of occurrence of strings (e.g. word frequencies) or substrings (e.g. consonant cluster frequencies), and describing relationships among strings (e.g various phonological realizations of a given word)

Also, an associative data base may be interactively accessed on the basis of keys corresponding to the different types of data elements, and a pattern matcher allows retrieval of incompletely specified elements For example, a pattern specifying the sequence

voiceless stop - /i/ - voiceless stop

can be used to retrieve the strings

'keep' /kip/
'peeking' /pikiŋ/
'repeated' /ripitəd/

from orthographic and phonemic transcriptions

Applications of this natural language processing package will be demonstrated. These include

a) analysis of systematic phonological variation which serves as the basis for specifying and testing phonological rules,

b) interactive testing of word recognition error rates associated with indeterminacy in the phonological or orthographic string,

c) analysis of phonotactic patterns which can be used as the basis for specifying and testing syllabification algorithms,

d) comparison across languages or dialects, to discover systematic sound correspondences, and to aid in the study of historical reconstruction or dialect relationships

ON THE ROLE OF WORDS AND PHRASES IN AUTOMATIC TEXT ANALYSIS

G. SALTON
Cornell University

Automatic indexing normally consists in assigning to documents either single terms, or more specific entities such as phrases, or more general entities such as term classes. Discrimination value analysis assigns an appropriate role in the indexing operation to the single terms, term phrases, and thesaurus categories. To enhance precision it is useful to form phrases from high-frequency single term components. To improve recall, low-frequency terms should be grouped into affinity classes, assigned as content identifiers instead of the single terms.

Collections in different subject areas are used in experiments to characterize the type of phrase and word class most effective for content representation.

The following typical conclusions can be reached:

a) the addition of phrases improves performance considerably;

b) use of phrases is better with corresponding deletion of single terms in practically all cases;

c) the use of both high-frequency and medium-frequency phrases is generally more effective than the use of either phrase-type alone;

d) the most effective thesaurus categories are those which include a large number of low-frequency terms;

e) the least effective classes either consist of only one or two terms, or else they include terms with unequal frequency characteristics permitting the high-frequency terms to overcome the others.

The discrimination value theory is developed and appropriate experimental output is supplied.

GRAMMATICAL COMPRESSION IN NOTES AND RECORDS: ANALYSIS AND COMPUTATION

BARBARA B. ANDERSON
University of New Brunswick

IRWIN D. J. BROSS
Roswell Park Memorial Institute

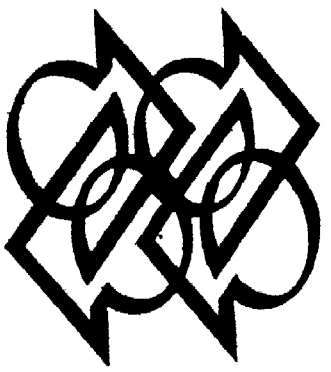
NAOMI SAGER
New York University

All languages have mechanisms of compression. One situation where these are used is when people are making notes. Usually, such note taking occurs within a practical context where the objects and meanings are known; the linguistic forms are degenerate but the message is unambiguous. This paper describes the linguistic mechanisms of compression which achieve this result, as they appeared in a study of the notes used in medical records for collaborative study of breast cancer. The syntactic devices were found to be mainly deletions of fixed types. the deletion of words having a special status in the grammar of the whole language (e.g. the verb be); and the deletion in particular positions of words having a special status in the particular subject matter (e.g. in the medical sublanguage, the word physician). A linguistic description of the forms with deletion was made and sublanguage word classes were defined. To test the description, a subcorpus of the medical records (357 sentences and sentence fragments on Xray findings) was parsed by an existing computer parsing system, using an English grammar to which a small component covering the deletion-forms was added. The paper concludes with a discussion of the modifications required in the computer grammar to parse the deletion forms and a summary of the parsing results.

AMERICAN SOCIETY for INFORMATION SCIENCE

38th ANNUAL MEETING

SHERATON-BOSTON
OCTOBER 26-30, 1975



SUNDAY, OCTOBER 26, 9 AM TO 5 PM: WORKSHOPS

(Separate registration required)

A. MANAGEMENT OF THE LIBRARY IN TRANSITION: COSTING ANALYSIS AND MANAGEMENT TECHNIQUES FOR THE ADMINISTRATOR OF AUTOMATED SYSTEMS (BEGINS OCTOBER 25)

B. AUTOMATED TEXT PROCESSING AND PHOTOCOMPOSITION

C. MICROGRAPHICS IN INFORMATION SYSTEMS: NEW APPLICATIONS

10-4:30 PSYCHOLOGICAL RESEARCH ON USER ON-LINE INTERACTION

10-5 THE FUTURE OF SCIENTIFIC COMMUNICATION--WORK IN PROGRESS

1:30-4:30 NUMERICAL DATA

1:30-4:30 ORGANIZATION AND DISTRIBUTION OF PRIMARY SOCIAL SCIENCE DATA IN A COMPUTERIZED SETTING

MONDAY, OCTOBER 27

9:15 OPENING SESSION Ruth Tighe, Chairman

9:45 1965-1975--A DECADE OF INNOVATION Carlos Cuadra, Chmn

1:15 INTERDISCIPLINARY ISSUES CONCERNING THE USER/COMPUTER INTERFACE

1:15 STATE LEGISLATIVE INFORMATION SYSTEMS

1:15 COPYRIGHT 1975: A SEMI-STRUCTURED FREE-FOR-ALL

3:30 IMPROVEMENT OF METHODS FOR FORECASTING INFORMATION REQUIREMENTS AND SERVICES

3:30 TRENDS IN CLASSIFICATION

3:30 CONTRIBUTIONS TO INFORMATION IN THE BEHAVIORAL AND SOCIAL SCIENCES

5:30 OPPORTUNITIES IN INFORMATION SCIENCE TODAY

TUESDAY, OCTOBER 28

9:15 KEYNOTE SESSION Daniel Bell and panel

1:15 PLANS AND FUNDING FOR NATIONWIDE INFORMATION PROGRAMS

1:15 SPECIAL PROBLEMS IN STORING AND RETRIEVING HUMANITIES MATERIALS

1:15 NATIONAL INSTRUCTIONAL MATERIALS INFORMATION SYSTEM

1:15 INFORMATION ANALYSIS CENTERS

3:30 FEDERAL LEGISLATIVE INFORMATION SYSTEMS

3:30 COMPARISON OF SYSTEM REQUIREMENTS FOR ON-LINE AND BATCH RETRIEVAL

3:30 EDUCATION INFORMATION RESOURCES--AFTER ERIC, WHAT?

3:30 MICROFORM CATALOGS--NOW MORE THAN EVER

8:00 REPORT OF ASIS LONG-RANGE PLANNING COMMISSION

WEDNESDAY OCTOBER 29

9:15 MICROPROCESSORS AND RECENT TECHNOLOGICAL INNOVATIONS

9:15 A LOOK AT SATELLITE-MEDIATED MEDICAL COMMUNICATION EXPERIMENTS

9:15 TOWARD A UNIFIED THEORY OF INFORMATION

9:15 INTERACTIVE SYSTEM EVALUATION: METHODOLOGY AND RESULTS

1:15 INTERACTIVE SYSTEM EVALUATION· PSYCHOLOGY RESEARCH METHODS AND RESULTS

1:15 LIBRARY NETWORKS: ORGANIZATION AND GOVERNANCE

1:15 REPORTS OF NATIONAL AND INTERNATIONAL EDUCATION PROGRAM STUDIES

1:15 REPORT ON ASIS PROJECT TO INVESTIGATE THE PLANNING REQUIREMENTS OF THE SCIENTIFIC AND TECHNICAL INFORMATION COMMUNITY

1:30 ERIC DATA BASE USERS CONFERENCE

3:30 BUSINESS MEETING Dale Baker, President

7:00 AWARDS BANQUET Jules Bergman, ABC News

THURSDAY, OCTOBER 30

9:15 TELEPROCESSING AND INFORMATION NETWORKS

9:15 INTERACTION BETWEEN PUBLIC AND PRIVATE SECTORS IN INFORMATION POLICY FORMATION

10:00 OBTAINING QUICK, INEXPENSIVE HARD COPY FROM MICROFILM AND COMPUTER SYSTEMS

10:00 CLINICAL DATA, MEDICAL RECORDS, AND THE NEW TECHNOLOGY

1:15 PRICING--A RATIONAL MECHANISM FOR ALLOCATING INFORMATION

1:15 TRENDS AND CURRENT RESEARCH IN AUTOMATED LANGUAGE PROCESSING

1:15 FEEDBACK AND CONTROL: IS THE USE OF HUMAN BEINGS MORE HUMAN AFTER 25 YEARS?

3:30 DOCTORAL FORUM

3:30 AVAILABILITY AND ACCESSIBILITY PROBLEMS FACING THE INFORMATION USER

3:30 THE INFORMATION REVOLUTION IN MEDICINE

FEES AT CONFERENCE

ASIS MEMBER, FULL CONFERENCE	$60.00
NON MEMBER, FULL CONFERENCE	$90.00
STUDENT, FULL TIME, FULL CONFERENCE	$10.00
NON MEMBER, ONE DAY	$25.00

REGISTRATION BEGINS AT 2 PM ON SATURDAY, OCTOBER 25

ASSOCIATION FOR COMPUTIG MACHINERY

ANNUAL CONFERENCE 1975

RADISSON HOTEL, MINNEAPOLIS

OCTOBER 20-22

CONFERENCE SESSIONS

MONDAY 20 OCTOBER 9:00

Opening session Alan F. Westin, The next decade of the computer revolution· Privacy, partic pation, and power

MONDAY 20 OCTOBER 1:15

SIGCAS-1 Panel Information and public policy

SIGBDP-1 Panel Researcher/user dialogue on file design

SIGOPS Panel Analysis of memory management and operating systems

SICSOFT-1 Panel Software management

MONDAY 20 OCTOBER 3:30

Panel. Computer user and vendor legal issues

SIGMICRO-1 Tutorial Microprocessors

SIGPLAN-2 Panel Programming and its implication on programming languages

MONDAY 20 OCTOBER 8:00

Panel Dollars and sense of D P

SIGMICRO-3 Panel Microprocessors--chips to working systems

SIGPLAN-1 Debate Should high level languages be used to write systems software?

TUESDAY 21 OCTOBER 8:30

SIGGRAPH Panel Evaluating computer graphics systems organization

SIGBDP-2 Panel GUIDE efforts in data base management

SIGMICRO-2 Panel Microprocessors and their architectural implications

SIGPLAN-3 Papers: New ideas in programming language theory
SIGSAM-1 Tutorial: Presentation of symbol manipulation systems
SIGCUE-3 Panel: The role of instructional simulations

TUESDAY 21 OCTOBER 10:15

SIGCAS-3 Panel: Computers in the electoral process
SIGBDP-3 Panel: Performance measurement and data base design
SIGPLAN-4 Papers: New methods and techniques in programming languages
SIGSAM-2 Papers: Symbolic and algebraic manipulation
SIGCUE-4 Panel: Computer programming in mathematics

TUESDAY 21 OCTOBER 1:15

SIGCAS-2 Panel: Computers and public policy
SIGBDP-5 Panel: Human resource requirements for business application development
SIGCOMM-2 Panel: Providing user services in a computer network environment
SICSOFT-2 Panel: Programming environments
SIGMAP-2 Panel: Math programming data structures
SIGCUE-1 (CUE/CSE/CAS) Panel: Computers in teaching environment
STUDENTS: Special tutorial-seminar: Computer science and the future

TUESDAY 21 OCTOBER 3:30

SIGDOC Tutorial and Panel: Experience with HIPO
SIGBDP-4 Panel: EPOS systems in the distributing industry
SIGIR-1 Panel: Information networks
SIGMINI Panel: The challenge of minicomputer software
SIGMAP-1 Panel: Computational practice in mixed-integer programming
SIGCUE-2 (CUE/CAS) Panel: Certification of CS teachers for secondary schools

TUESDAY 21 OCTOBER 7:

SIGSAM-3 Demonstration: Symbol manipulation systems

WEDNESDAY 22 OCTOBER 8:30

SIGARCH Papers

SIGBDP-6 Panel: Innovations on computer charging mechanisms

SIGCOMM-1 Panel: Applications of distributed computing

SIGSIM Tutorial: Simulation: State of the art

SIGMOD Panel: Data base administration

SIGCUE-5 Panel: Computers and society

WEDNESDAY 22 OCTOBER 10:15

PANEL: Summary of CODASYL report on selection and acquisition of data base management systems

SIGBDP-7 Panel: Developing user-oriented business systems

SIGDA Panel: Human factors engineering issues in design automation systems

SICSOFT-3 Panel: Software product assurance

SIGNUM Papers and Talks: Applications of numerical analysis

SIGCUE-6 Panel: Tutorial computing: the state of the art

WEDNESDAY 22 OCTOBER 1:15

SIGART-1 Papers

SIGBDP-8 Panel: Performance measurement

SIGIR-2 (IR/ARCH) Panel: Non-numeric processing and computer architecture

SIGBIO Panel: Computer systems and the quality of health care

SIGMETRICS Papers: Computer system performance analysis

SIGCUE-7 Panel: Computer managed instruction and guidance

WEDNESDAY 22 OCTOBER 3:30

SIGART-2 Panel: Artificial intelligence and perception

SIGBDP-9 Panel: Structure of future systems

PANEL: ACM self assessment

PANEL: Computers outside the fishbowl

SIGCUE-8 Panel: Time-sharing instructional systems in Minnesota

SOLAR PROJECT TERMINATES

ARPA support for the semantic bibliography and lexicography project at System Development Corporation is ending. Online access to the five SOLAR files through ARPANET ends September 15. Arrangements for printed distribution of data are being explored. Expressions of interest in receiving such listings can be addressed to Dr. Tim Diller, 3244 Butler Avenue, Los Angeles, California 90066.

ENERGY INFORMATION TOOLS

The Information Industry Association and the National Federation of Abstracting and Indexing Services will hold a Workshop November 10-11 at the Quality Inn Capitol Hill, Washington.

Lectures, discussions, and workshops with access to online services will treat the tools presently available. The fee is $55. For information, call Paul Zurkowski, IAI, 4720 Montgomery Lane, Bethesda, Maryland 20014; 301-654-4150.

COMPUTER TRANSLATION OF CHINESE JOURNALS

Shiu-Chang Loh, Professor of Computer Science, began research on machine translation at the Chinese University of Hong Kong in 1969. The system presently in use requires pre-editing by high-school graduates with two weeks of special training; the output is direct from a lineprinter.

The project is accepting subscriptions to machine-translated editions of Acta Mathematica Sinica and Mathematics in Practice and Theory, both quarterlies, and to the bimonthly Acta Physica Sinica. The price for the quarterlies is $40 surface, $48 air; for the bimonthly, $60 surface and $72 air.

Professor Loh is considering publication of translated editions of several other Chinese journals and designing a program for English-to-Chinese translation.

asis
AMERICAN SOCIETY FOR INFORMATION SCIENCE
1155 SIXTEENTH STREET N W WASHINGTON D C 20036 • Telephone 202/659-3644

L I B R A R I A N O F C O N G R E S S

ON MAY 23, A S I S SENT THE FOLLOWING TELEGRAM TO THE PRESIDENT OF THE UNITED STATES:

This is in regard to the position of Librarian of Congress as related to today's information age.

The management of the Library of Congress has become an extremely complex task. With more than 4,000 employees and an annual budget of nearly $100 million, the administrative skills required to direct the operation and growth of our de facto national library are comparable to those required to manage a company that is roughly the size of the New York Times, or about twice that of the Washington Post.

When the Library of Congress was created in 1800, America was just entering the industrial age, and information was relatively unimportant. Less than half of the American people could read. The few libraries that existed were owned and used primarily by the aristocracy.

Today, virtually every community in the United States has a public library. News is transmitted instantly via satellite. Computer-based data bases have become the new storehouses of information, and information itself has become the key to power.

As the importance of information has changed, the functions and responsibilities of the Library of Congress have changed as well. Originally established as a facility for Congress, the

Library has accepted many additional responsibilities including

1. It serves many common services in the areas of technical processing and reference for libraries throughout the country.
2. It provides many user services that supplement local efforts of libraries.
3. It performs many services in the areas of technical processing and reference for other libraries.
4. It serves as the hub of the recorded knowledge system of society.

Many professionals who are familiar with the present functions and responsibilities of the Library of Congress believe that the organization's next head must be more than a scholar. He or she must be a skilled and accomplished administrator; familiar with the implications of major national issues in information science, able to evaluate, and take advantage of, emerging technologies; effectively deal with political and personnel problems; and able to satisfy the information requirements of educators, researchers, professionals, legislators, local libraries, and the public at large.

Members of the American Society for Information Science believe that, while academic scholarship should be considered for our next Librarian of Congress, it certainly is not totally sufficient.

Sincerely,

JOSHUA I. SMITH
Executive Director
A S I S

B I B L I O G R A P H Y :

WORKING PAPERS PUBLISHED IN 1974 - 1975

FONDAZIONE DALLE MOLLE
ISTITUTO PER GLI STUDI SEMANTICI E COGNITIVI
VILLA HELENEUM, CASTAGNOLA 6976, SWITZERLAND

Note: A = hard copies available
B = microfiches available
S = Computer Science Department, Stanford
* = in preparation

1. Causality and reasoning. Roger C. Schank. A, B.
2. Computer generation of natural language from a deep conceptual base. Neil Murray Goldman. A, S.
3. Is there a semantic memory? Roger C. Schank. A, B, *.
4. Computational understanding: Analysis of sentences and context. Christopher Riesbeck. A, B.
5. "He will make you take it back": A study in the pragmatics of language. Eugene Charniak. A, B.
6. Understanding paragraphs. Roger C. Schank. A, B.
7. Selezione di parole per l'estrazione di unità foniche atte alla sintesi delle lingue tedesca e italiana. G. B. Debiasi and A. M. Mioni. A, B.
8. Fonetica e fonologia autonoma della lingua Hindi. Romeo Galassi. A, B.
9. Ottimizzazione delle caratteristiche delle unità normalizzate per la sintesi del parlato mediante mini-computer. Mildonian Offeli. A, B.
10. Ueber die Struktur des semantischen Langzeitgedächtnisses. Manfred Wettler. A, B.

11. "Information Storage and Retrieval" modello di un sistema integrato. B. Treusch, F. Lestuzzi, S. Rova. A, B.
11. Konzepttheorie--Ein praktischer Beitrag zur Textverarbeitung und Textrezeption. Wolfgang Samlowski. A.
13. A partial taxonomy of knowledge about actions. Eugene Charniak. A.
14. Organization and inference in a frame-like system of common sense knowledge. Eugene Charniak. A.
15. Linguistischer Thesaurus. B. Treusch. A*.
16. Semantics, preference and inference--A full description of a system and a program. Margaret King and Yorick Wilks. A*
17. Seven theses on artificial intelligence and natural language Yorick Wilks. A*.

Proceedings of the Tutorial on Computational Semantics given by the members of the Institute of Semantic and Cognitive Studies.

CONCEPTUAL INFORMATION PROCESSING

ROGER C. SCHANK

With contributions by

NEIL M. GOLDMAN, CHARLES J. RIEGER III, & CHRISTOPHER K. RIESBECK

Fundamental Studies in Computer Science, Volume 3

NORTH-HOLLAND PUBLISHING COMPANY
1975

1. MARGIE
2. The conceptual approach to language processing
3. Conceptual dependency theory
4. The conceptual analyzer (Riesbeck)
5. Conceptual memory and inference (Rieger)
6. Conceptual generation (Goldman)

viii + 384 pages

ISBN 0-7204-2507-7

Price: $27.50

AUTOMATIC TRANSLATION AT GRENOBLE

(La Traduction Automatique a Grenoble)

BERNARD VAUQUOIS

Documents de Linguistique, 24

Dunod
1975

Foreword 9
FIRST PART: The CETA MT Experiment (1961-71) 13
CHAPTER I: THE STATE OF MT IN 1961 14
1. The birth and spectacular growth of MT 14
2. Diverse approaches to MT in 1961 16
3. Design of the Grenoble project 31
CHAPTER II: MORPHOLOGICAL ANALYSIS 35
1. Purpose of morphological analysis 35
2. Organization of the morphological model 37
3. Limits and peculiarities of the morphological model . . 49
4. Sample of the results of morphological analysis . . 50
CHAPTER III: THE MODEL OF SYNTACTIC ANALYSIS 52
1. Purpose of syntactic analysis 52
2. Syntactic analyzers 61
3. The CETA model of syntactic analysis 82
4. Examples 97
CHAPTER IV: THE PIVOT LANGUAGE 105
1. Choice of a level of transfer 105
2. Definition of the pivot language 106
3. Informatic methods 112
4. Linguistic information 116
5. Examples of representation in pivot language . . . 118

CHAPTER V: THE PROCESS OF GENERATION 123
1. Generation of superficial syntactic structure . . . 123
2. Morphological generation 128
3. Examples. 128
SECOND PART: The work of GETA since 1972 141
CHAPTER VI: REFLECTIONS ON AUTOMATIC TRANSLATION . . 142
1. MT at the start of the '70s 142
2. Critique of the CETA MT system 143
3. Perspectives on MT 148
CHAPTER VII: THE WORK OF GETA SINCE 1972 157
1. The new conceptions of informatic systems 157
2. The ATEF system 160
3. The CETA system 163
4. Algogrammars and interactive methods 168
5. The LEIBNIZ group 171
Bibliography 173

ISBN 2-85344-024-9 (Association Jean-Favard)
ISBN 2-04-009956-5 (Editions Dunod)

FREQUENCY AND DISTRIBUTION

Frequence et distribution du vocabulaire dans un choix de romans français

GUNNEL ENGWALL

Språkförlaget Skriptor AB
Stockholm 1974

1. Introduction 9
2. Constitution of the corpus 14
3. Units to be examined in a frequency study . . 24
4. Results on the level of occurrences and types 30
5. Methods for classifying words by order of importance 43
6. Definition of the inflected form and lemma . . 56
7. Results on the level of inflected forms and lemmas 69
8. Comparisons among five earlier studies and the corpus of tex novels 85
9. Conclusion . 101
Appendices 108
Abbreviations and symbols 176
Index 178
Bibliography 181

ISBN 91-7282-059-4

ASSOCIATION FOR
LITERARY AND
LINGUISTIC
COMPUTING

BULLETIN

Volume 3 Number 2

Summer Term 1975

EDITOR:

Mrs. Joan M. Smith
6 Sevenoaks Avenue
Heaton Moor, Stockport
Cheshire, England

CONTENTS

GUEST EDITORIAL . . . . . *J. R. Allen* 95

THE ACTIVITIES OF THE INSTITUT FÜR DEUTSCHE SPRACHE'(IDS) MANNHEIM IN THE FIELD OF LITERARY AND LINGUISTIC COMPUTING . . . . . *Godelieve Berry-Rogghe* 97

CARLYLE AND THE MACHINE: A Quantitative Analysis of Syntax in Prose Style . . . . . *R. L. Oakman* 100

HOW TO USE COCOA TO PRODUCE INDEXES (TO BOTH BOOK AND SUB-ROUTINE LIBRARIES) . . . . . *Kathleen M. Crennell* 115

COMPUTATIONAL LINGUISTICS or WHAT'S IN A NAME? *W. Martin* 124

THE BILINGUAL ANALYTICAL LITERARY AND LINGUISTIC CONCORDANCE - BALCON . . *Susan M. Hockey* and *V. Shibayev* 133

TOWARDS AN ALGORITHMIC METHODOLOGY OF LEMMATIZATION *M. L. Hann* 140

THE BODLEIAN PROJECT: A COMPUTERIZED INDEX TO THE VISUAL CONTENT OF MANUSCRIPT ILLUMINATIONS *Alice F. Worsley* 151

A COMPUTER-ASSISTED PROJECT AT THE HEBREW UNIVERSITY, ISRAEL: THE COMPILATION OF AN INDEX TO THE WRITINGS OF AGNON *U. Sahm* 156

THE USE OF A COMPUTER IN DEVISING A BEGINNERS' LATIN COURSE . . . . . *C. W. E. Peckett* 158

TEACHING COMPUTATIONAL LINGUISTICS: A continuation of the Discussion . . . . . *Margaret King* 161

LITERARY STATISTICS VI: ON THE FUTURE OF LITERARY STATISTICS . . . . . *N. D. Thomson* 166

COMPTE RENDU D'UNE TABLE-RONDE DU C.N.R.S. SUR LE THEME: 'PROCEDURES D'ANALYSE ET METHODES DE VALIDATION DANS L'ETUDE DES DONNEES TEXTUELLES' . . . . . *J. Virbel* 172