

Abstracts of Current Literature

GLISP User's Manual

Gordon S. Novak Jr.

Computer Science Department
Stanford University
Stanford, California 94305

Report No. HPP-82-1, January 1982, 36 pages.

GLISP is a high-level, LISP-based language which is compiled into LISP. GLISP provides a powerful abstract datatype facility, allowing description and use of both LISP objects and objects in AI representation languages. GLISP language features include PASCAL-like control structures, infix expressions with operators which facilitate list manipulation, and reference to objects in PASCAL-like or English-like syntax. English-like definite reference to features of objects which are in the current computational context is allowed; definite references are understood and compiled relative to a knowledge base of object descriptions. Object-centered programming is supported; GLISP can substantially improve runtime performance of object-centered programs by optimized compilation of references to objects. This manual describes the GLISP language and use of GLISP within INTERLISP.

How to Solve It with PROLOG

Helder Coelho, Jose C. Cotta, and Luis M. Pereira

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

Research Report, 2nd edition, 1980, 215 pages.

Our purpose is to present the outstanding features of PROLOG through a collection of small problems and exercises, divided in general application areas such as deductive reasoning over data bases, natural language, symbolic calculus, etc.

Logic Programming Bibliography

Helder Coelho

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

Research Report, 2nd edition, October 1981, 41 pages.

This is a provisional version of an evolving bibliography on logic programming and PROLOG, covering field work carried out in all known groups spread over the whole world. It is the first draft, and consequently incomplete. The purpose behind its construction is very simple. Every researcher needs to have available an up-to-date and complete list of references. The motivation to work out this list came from the bibliog-

raphy for my previous report "How to Solve it With PROLOG," and from the Hungarian list delivered during the Logic Programming Workshop at Debrecen, July 1980.

A New Point of View on Children's Stories

Bertram Bruce

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

Reading Education Report No. 25, July 1981, 45 pages.

Recent work on text analysis at the Center for the Study of Reading and elsewhere has produced surprising results regarding the texts that children read in school. These results support the hypothesis that part of the difficulty children encounter in making the transition from beginning to skilled reading lies in an abrupt shift in text characteristics between lower and upper elementary school. Moreover, a comparison between school texts and popular trade books shows that the school texts may provide inadequate preparation for the texts that skilled readers need to master. Thus, characteristics of the texts that children are expected to read may hinder rather than help in the attainment of educational goals.

Why Readability Formulas Fail

Bertram Bruce, Andee Rubin, and Kathleen Starr

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

Reading Education Report No. 28, Aug. 1981, 13 pages.

Being able to measure the readability of a text with a simple formula is an attractive prospect, and many groups have been using readability formulas in a variety of situations where estimates of text complexity are thought to be necessary. The most obvious and explicit use of readability formulas is by educational publishers designing basal and remedial reading texts; some states, in fact, will consider using a basal series only if it fits certain readability formula criteria. Increasingly, public documents such as insurance policies, tax forms, contracts, and jury instructions must meet criteria stated in terms of readability formulas. Unfortunately, readability formulas just don't fulfill their promise. We attempt here to categorize and summarize some of the problems with readability formulas and their use.

Stories Within Stories

Bertram Bruce

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

Reading Education Report No. 29, Aug. 1981, 15 pages.

What appears to be a single story is often a complex set of stories within stories, each with its distinct author and reader. Examples of such stories within stories are presented. Results of analyses of basal readers and trade books in terms of embedded stories are also discussed. These suggest that a greater variety of stories could and should be made available to children.

**Conceptual Readability:
New Ways to Look at Text**

Andee Rubin (Editor)

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

Reading Education Report No. 31, Sept. 1981, 61 pages.

The papers in this collection describe a notion of "conceptual readability" which contrasts our approach to text with that assumed by standard readability formulas. Traditionally, the readability level of a text has been calculated by considering text characteristics such as the number of words per sentence and the degree of familiarity of individual words. Our approach focuses instead on the concepts communicated by the text: how arguments are presented, what place examples play in an exposition, how characters' interactions are developed and described.

In this report, we first demonstrate how certain uses of traditional readability formulas may actually lead to more difficult texts. Next, we discuss two alternative text analysis methods which are sensitive to structural, semantic and discourse characteristics. Finally, we suggest an educational method which encourages children to focus on the conceptual level of text in their early reading and writing experiences.

**Text Readability:
Proceedings of the March 1980 Conference**

Alice Davison, Richard Lutz, and Ann Roalef (Editors)

Center for the Study of Reading
University of Illinois
51 Gerty Drive
Champaign, Illinois 61820

Technical Report No. 213, August 1981, 145 pages.

The papers which make up this technical report are summaries of oral presentations on readability and readability formulas delivered in March 1980 at the Center for the Study of Reading. The papers included here represent as closely as possible the content and organization of the oral presentations, in a more readable format than a verbatim transcript. The purpose of the conference was to raise a number of issues for discussion and to present a spectrum of ideas and viewpoints from which readability formulas could be judged or criticized. We have not tried to make the papers exhaustive summaries of all that has been done on a certain subject or to represent only the most

"correct" and orthodox positions on any topic. We feel that the views expressed here, while diverse and in some cases programmatic, will be useful in provoking discussion and reexamining assumptions about readability formulas, perhaps in defining research which might lead to a better understanding of what makes things difficult to read.

**Learning the Rules of the Game:
Four Views of the Relation Between
Social Interaction and Syntax Acquisition**

Marilyn Shatz

Center for the Study of Reading
University of Illinois
51 Gerty Drive
Champaign, Illinois 61820

Technical Report No. 214, September 1981, 38 pages.

That language is a phenomenon belonging primarily to the domain of social activities is hardly an arguable point. While one can list nonsocial uses of language as well as types of social interaction that are not linguistic, the fact remains that the overlap between language use and social interaction, though imperfect, is still considerable. Moreover, some minimal amount of social interaction seems to be necessary for language acquisition to take place. The obvious kinship between language and social interaction suggests the possibility of a relationship between knowledge in the social sphere and the learning of linguistic forms. In this paper four different kinds of relationships between social interaction and syntax acquisition are outlined and evaluated. The positions range from a strong one, deriving syntax acquisition directly from interactionally provided social knowledge, to a weak one, where the relatively autonomous process of syntax acquisition can be facilitated by the efficient distribution of processing resources.

A Social Interaction Model of Reading

Bertram Bruce

Bolt Beranek and Newman Inc.
10 Moulton Street
Cambridge, Massachusetts 02238

Technical Report No. 218, September 1981, 83 pages.

An author and a reader are engaged in a social interaction which depends on their goals and their beliefs about the world and each other. One aspect of this interaction is the creation of another level of social interaction involving an "implied author" and an "implied reader." The newly created characters may, in their turn, create another level of social interaction involving, for example, a "narrator" and a "narratee." Each level so created permits the creation of an addi-

tional level. A model for the levels of social interaction in reading is discussed in the paper. The model provides a framework for examining devices such as author commentary, irony, stories within stories, first person narration, and point of view. Examples such as *The Tale of Benjamin Bunny* and *The Turn of the Screw* are discussed.

Translating English Into Logical Form

Stanley J. Rosenschein and Stuart M. Shieber

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

Proc. 20th Annual ACL Meeting, June 1982, 1-8.

A scheme for syntax-directed translation that mirrors compositional model-theoretic semantics is discussed. The scheme is the basis for an English translation system called PATR and was used to specify a semantically interesting fragment of English, including such constructs as tense, aspect, modals, and various lexically controlled verb complement structures. PATR was embedded in a question-answering system that replied appropriately to questions requiring the computation of logical entailments.

Linguistic and Computational Semantics

Brian Cantwell Smith

XEROX Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California 94304

Proc. 20th Annual ACL Meeting, June 1982, 9-15.

We argue that because the very concept of computation rests on notions of interpretation, the semantics of natural languages and the semantics of computational formalisms are in the deepest sense the same subject. The attempt to use computational formalisms in aid of an explanation of natural language semantics, therefore, is an enterprise that must be undertaken with particular care. We describe a framework for semantical analysis that we have used in the computational realm, and suggest that it may serve to underwrite computationally-oriented linguistic semantics as well. The major feature of this framework is the explicit recognition of both the declarative and the procedural import of meaningful expressions; we argue that whereas these two viewpoints have traditionally been taken as alternative, any comprehensive semantical theory must account for how both aspects of an expression contribute to its overall significance.

The Representation of Inconsistent Information in a Dynamic Model-Theoretic Semantics

Douglas B. Moran

Department of Computer Science
Oregon State University
Corvallis, Oregon 97331

Proc. 20th Annual ACL Meeting, June 1982, 16-18.

Model-theoretic semantics provides a computationally attractive means of representing the semantics of natural language. However, the models used in this formalism are static and are usually infinite. Dynamic models are incomplete models that include only the information needed for an application and to which information can be added. Dynamic models are basically approximations of larger conventional models, but differ in several interesting ways.

The difference discussed here is the possibility of inconsistent information being included in the model. If a computation causes the model to expand, the result of that computation may be different than the result of performing that same computation with respect to the newly expanded model (i.e. the result is inconsistent with the information currently in the dynamic model). Mechanisms are introduced to eliminate these local (temporary) inconsistencies, but the most natural mechanism can introduce permanent inconsistencies in the information contained in the dynamic model. These inconsistencies are similar to those that people have in their knowledge and beliefs. The mechanism presented is shown to be related to both the intensional isomorphism and impossible worlds approaches to this problem.

What's in a Semantic Network?

James F. Allen and Alan M. Frisch

Computer Science Department
The University of Rochester
Rochester, New York 14627

Proc. 20th Annual ACL Meeting, June 1982, 19-27.

Ever since Woods's "What's in a Link" paper, there has been a growing concern for formalization in the study of knowledge representation. Several arguments have been made that frame representation languages and semantic-network languages are syntactic variants of the first-order predicate calculus (FOPC). The typical argument proceeds by showing how any given frame or network representation can be mapped to a logically isomorphic FOPC representation. For the past two years we have been studying the formalization of knowledge retrievers as well as the representation languages that they operate on. This paper presents a representation language in the notation of FOPC whose form facilitates the design of a semantic-network-like retriever.

Dependencies of Discourse Structure on the Modality of Communication: Telephone vs. Teletype

Philip R. Cohen

Department of Computer Science
Oregon State University
Corvallis, Oregon 97331

Scott Fertig

Bolt Beranek and Newman, Inc.
10 Moulton Street
Cambridge, Massachusetts 02239

Kathy Starr

Bolt Beranek and Newman, Inc.
10 Moulton Street
Cambridge, Massachusetts 02239

Proc. 20th Annual ACL Meeting, June 1982, 28-35.

A desirable long-range goal in building future speech understanding systems would be to accept the kind of language people spontaneously produce. We show that people do not speak to one another in the same way they converse in typewritten language. Spoken language is finer-grained and more indirect. The differences are striking and pervasive. Current techniques for engaging in typewritten dialogue will need to be extended to accommodate the structure of spoken language.

Towards a Theory of Comprehension of Declarative Contexts

Fernando Gomez

Department of Computer Science
University of Central Florida
Orlando, Florida 32816

Proc. 20th Annual ACL Meeting, June 1982, 36-43.

An outline of a theory of comprehension of declarative contexts is presented. The main aspect of the theory being developed is based on Kant's distinction between concepts as rules (we have called them conceptual specialists) and concepts as an abstract representation (schemata, frames). Comprehension is viewed as a process dependent on the conceptual specialists (they contain the inferential knowledge), the schemata or frames (they contain the declarative knowledge), and a parser. The function of the parser is to produce a segmentation of the sentences in a case frame structure, thus determining the meaning of prepositions, polysemous verbs, noun groups, etc. The function of this parser is not to produce an output to be interpreted by semantic routines or an interpreter, but to start the parsing process and proceed until a concept relevant to the theme of the text is recognized. Then the concept takes control of the comprehension process *overriding* the lower level linguistic process. Hence comprehension is viewed as a process in which high level sources of knowledge (concepts) *override* lower level linguistic processes.

Natural Language Database Updates

Sharon C. Salveter and David Maier

Computer Science Department
SUNY Stony Brook
Stony Brook, New York 11794

Proc. 20th Annual ACL Meeting, June 1982, 67-73.

Although a great deal of research effort has been expended in support of natural language (NL) database querying, little effort has gone to NL database *update*. One reason for this state of affairs is that in NL querying, one can tie nouns and stative verbs in the query to database objects (relation names, attributes, and domain values). In many cases this correspondence seems sufficient to interpret NL queries. NL update seems to require database counterparts for active verbs, such as "hire," "schedule," and "enroll," rather than for stative entities. There seem to be no natural candidates to fill this role.

We suggest a database counterpart for active verbs, which we call *verbgraphs*. The verbgraphs may be used to support NL update. A verbgraph is a structure for representing the various database changes that a given verb might describe. In addition to describing the variants of a verb, they may be used to disambiguate the update command. Other possible uses of verbgraphs include specification of defaults, prompting of the user to guide but not dictate user interaction and enforcing a variety of types of integrity constraints.

Processing English with a Generalized Phrase Structure Grammar

J.M. Gawron, J. King, J. Lamping, E. Loebner,
E.A. Paulson, G.K. Pullum, I.A. Sag, and T. Wasow

Computer Research Center
Hewlett Packard Company
1501 Page Mill Road
Palo Alto, California 94304

Proc. 20th Annual ACL Meeting, June 1982, 74-81.

This paper describes a natural language processing system implemented at Hewlett-Packard's Computer Research Center. The system's main components are: a Generalized Phrase Structure Grammar (GPSG); a top-down parser; a logic transducer that outputs a first-order logical representation; and a *disambiguator* that uses sortal information to convert *normal-form* first-order logical expressions into the query language for HIRE, a relational database hosted in the SPHERE system. We argue that theoretical developments in GPSG syntax and in Montague semantics have specific advantages to bring to this domain of computational linguistics. The syntax and semantics of the system are totally domain-independent, and thus, in principle, highly portable. We discuss the prospects for extending domain-independence to the lexical semantics as well, and thus to the logical semantic representations.

Experience with an Easily Computed Metric for Ranking Alternative Parses

George E. Heidorn

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

Proc. 20th Annual ACL Meeting, June 1982, 82-84.

This brief paper, which is itself an extended abstract for a forthcoming paper, describes a metric that can be easily computed during either bottom-up or top-down construction of a parse tree for ranking the desirability of alternative parses. In its simplest form, the metric tends to prefer trees in which constituents are pushed as far down as possible, but by appropriate modification of a constant in the formula other behavior can be obtained also. This paper includes an introduction to the EPISTLE system being developed at IBM Research and a discussion of the results of using this metric with that system.

An Improved Heuristic for Ellipsis Processing

Ralph M. Weischedel

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

Norman K. Sondheimer

Software Research
Sperry Univac MS 2G3
Blue Bell, Pennsylvania 19424

Proc. 20th Annual ACL Meeting, June 1982, 85-88.

Robust response to ellipsis (fragmentary sentences) is essential to acceptable natural language interfaces. For instance, an experiment with the REL English query system showed 10% elliptical input. This paper presents a method of automatically interpreting ellipsis based on dialogue context. Our method expands on previous work by allowing for expansion ellipsis and by allowing for all combinations of statement following question, question following statement, question following question, etc.

Planning Natural Language Referring Expressions

Douglas E. Appelt

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

Proc. 20th Annual ACL Meeting, June 1982, 108-112.

This paper describes how a language-planning system can produce natural-language referring expressions that satisfy multiple goals. It describes a formal representation for reasoning about several agents' mutual knowledge using possible-worlds semantics and the general organization of a system that uses the formalism to reason about plans combining physical and linguistic actions at different levels of abstraction. It

discusses the planning of concept activation actions that are realized by definite referring expressions in the planned utterances, and shows how it is possible to integrate physical actions for communicating intentions with linguistic actions, resulting in plans that include pointing as one of the communicative actions available to the speaker.

The TEXT System for Natural Language Generation: An Overview

Kathleen R. McKeown

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

Proc. 20th Annual ACL Meeting, June 1982, 113-120.

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

Augmenting a Database Knowledge Representation for Natural Language Generation

Kathleen F. McCoy

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

Proc. 20th Annual ACL Meeting, June 1982, 121-128.

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

Saliency: The Key to the Selection Problem in Natural Language Generation

E. Jeffrey Conklin and David. D. McDonald

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

Proc. 20th Annual ACL Meeting, June 1982, 129-135.

We argue that in domains where a strong notion of saliency can be defined, it can be used to provide: 1) an elegant solution to the selection problem, i.e. the problem of how to decide whether a given fact should or should not be mentioned in the text; and 2) a simple and direct control framework for the entire deep generation process, coordinating proposing, planning, and realization. (Deep generation involves reasoning about conceptual and rhetorical facts, as opposed to the narrowly linguistic reasoning that takes place during realization). We report on an empirical study of saliency in pictures of natural scenes, and its use in a computer program that generates descriptive paragraphs comparable to those produced by people.

A Knowledge Engineering Approach to Natural Language Understanding

Stuart C. Shapiro and Jeannette G. Neal

Department of Computer Science
State University of New York at Buffalo
Amherst, New York 14226

Proc. 20th Annual ACL Meeting, June 1982, 136-144.

This paper describes the results of a preliminary study of a knowledge engineering approach to natural language understanding. A computer system is being developed to handle the acquisition, representation, and use of linguistic knowledge. The computer system is rule-based and utilizes a semantic network for

knowledge storage and representation. In order to facilitate the interaction between user and system, input of linguistic knowledge and computer responses are in natural language. Knowledge of various types can be entered and utilized: syntactic and semantic; assertions and rules. The inference tracing facility is also being developed as a part of the rule-based system with output in natural language. A detailed example is presented to illustrate the current capabilities and features of the system.

A Model of Early Syntactic Development

Pat Langley

The Robotics Institute
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

Proc. 20th Annual ACL Meeting, June 1982, 145-151.

AMBER is a model of first language acquisition that improves its performance through a process of error recovery. The model is implemented as an adaptive production system that introduces new condition-action rules on the basis of experience. AMBER starts with the ability to say only one word at a time, but adds rules for ordering goals and producing grammatical morphemes, based on comparisons between predicted and observed sentences. The morpheme rules may be overly general and lead to errors of commission; such errors evoke a discrimination process, producing more conservative rules with additional conditions. The system's performance improves gradually, since rules must be relearned many times before they are used. AMBER's learning mechanisms account for some of the major developments observed in children's early speech.

