CURRENT BIBLIOGRAPHY

The selection of material through the current second year of
AJCL's existence remains tentative. A survey of subscriber-
members will be included in the last packet mailed during 1975
to establish patterns of coverage for future years.

Categorization of entries deepens as the field defines itself
and the collection of literature against which new items can·
be matched increases. The advice of members is welcome.

Many summaries are authors' abstracts, sometimes edited for
clarity, brevity, or completeness. Where possible, an infor-
mative summary is provided.

The Linguistic Documentation Centre of the University of
Ottawa provides many entries; by editorial accident, some
of the entries recently received from that source remain to
be included in the next issue. AJCL gratefully acknowledges
the assistance of Brian Harris and his colleagues.

Some entries are reprinted with permission from Computer
Abstracts.

See the following frames for a list of subject headings and
items with extended presentation or review.

SUBJECT HEADINGS

REPRESENTATION AND UNDERSTANDING

Edited by Daniel G. Bobrow and Allan Collins

Linguistics

STRING AND LIST PROCESSING IN SNOBOL4:

TECHNIQUES AND APPLICATIONS

By Ralph E. Griswold
Reviewed by Norman Badler

FORTRAN TECHNIQUES WITH SPECIAL REFERENCE

TO NON-NUMERICAL APPLICATIONS

By A. Colin Day
Reviewed by Richard J. Miller

THEORETICAL ISSUES
IN
NATURAL LANGUAGE PROCESSING

AN INTERDISCIPLINARY WORKSHOP IN

COMPUTATIONAL LINGUISTICS
PSYCHOLOGY
LINGUISTICS
ARTIFICIAL INTELLIGENCE

Cambridge, Massachusetts

June 10-13, 1975

EDITORS:

Professor R. Schank   B. L. Nash-Webber
Department of Computer Science   Bolt Beranek and Newman Inc
Yale University   50 Moulton Street
10 Hillhouse Avenue   *and*   Cambridge, Massachusetts
New Haven, Connecticut 06520   02138

ABSTRACTS FOUND·ELSEWHERE ON THE MICROFICHE

THE PRAGUE BULLETIN OF MATHEMATICAL LINGUISTICS

22

Universita Karlova
Praha 1974

TABLE OF CONTENTS

C O M P U T A T I O N A L   A N A L Y S E S.
OF ASIAN & AFRICAN LANGUAGES

A new journal
Mantaro J. Hashimoto, Editor

No. 1   March, 1975

TABLE OF CONTENTS

Syntax, Semantics, and Speech

William A. Woods
Bolt Beranek and Newman Inc.
Cambridge, Mass   02138

Acquaints speech researchers in the state of the art in the conceptual development of, and the new perspectives they place on, parsing, syntax and semantic interpretation. Includes the Chomsky hierarchy of grammar models, non-determinism in parsing and its implementation in either backtracking or multiple independent alternatives, predictive vs. non-predictive parsing, word lattices and chart parsing, Early's algorithm, transition network grammars, transformational grammars and augmented transition networks, procedural semantics, selectional restrictions and semantic association.

Improving Methodology in Natural Language Processing

William C. Mann
USC Information Sciences Institute
Marina Del Rey, California

Process models are rigorous, process specifications are made very explicit, and complexity is handled by use of computers. A methodology should be reliable, efficient and have integrative power. The distinctive strengths of the current computer oriented methodology are (a) the complexity of data and theory is easy to accommodate, (b) time sequence and dependencies are preserved, and (c) a diversity of hypotheses can be tested. Weaknesses are (a) experiments often take years to perform, (b) the activity is treated as a programming exercise with the status of data and program unclearly defined and (c) in attempting to be general on a particular phenomenon, significant others are missed. As whole systems are produced, they are difficult to disseminate and judge. A system may process its examples, but it is hard to determine if it is ad-hoc and tuned to the examples.

Some methodological issues in natural language understanding research

W. A. Woods
Bolt Beranek and Newman, Inc
Cambridge, Mass

There are two tasks for which methodologies are used, (a) building intelligent machines, and (b) understanding human language performance. Both depend on the development of a 'device-independent' language understanding theory. For theoretical studies, a methodology should be cognitively efficient and should deal effectively with the problem of scale--having a large number of facts embodied in the theory. Studies should be performed in the context of total language understanding; isolation of components limits scope. Intuition on human language performance is a good guide to computational linguistics.

Speech recognition by computer: a bibliography with abstracts

D. W. Grooms
National Technical Information Service
5285 Port Royal Rd.
Springfield, Virginia 22161

Contains 142 abstracts covering recognition, synthesis, and the acoustical, phonological and linguistic processes necessary in conversion of various waveforms. Retrieved using the National Technical Information Service on-line search system.

FUZZY LOGIC FOR HANDWRITTEN NUMERICAL CHARACTER RECOGNITION

P. Siy and C. S. Chen
Akron University

Considers characters as a directed abstract graph, of which the node set consists of tips, corners, and junctions, and the branch set consists of line segments connecting pairs of adjacent nodes. Classification of branch types produces features which are treated as fuzzy variables. A character is represented by a fuzzy function which relates its fuzzy variables, and by the node pair involved in each fuzzy variable. After producing a representation of an unknown character recognition occurs when a previously learned character's representation is isomorphic to the unknown.

A MEANS OF ACHIEVING A HIGH DEGREE OF COMPACTION
ON SCANDIGITIZED PRINTED TEXT

R. N. Ascher and G. Nagy
IBM Corporation

A 16:1 compaction ratio was achieved by storing only the first instance of each pattern class and thereafter substituting this exemplar for every subsequent occurrence of the symbol. Proposed are refinements to yield a 40:1 ratio.

THE MORPHOLOGY OF CHINESE CHARACTERS
A SURVEY OF MODELS AND APPLICATIONS

William Stallings
Center for Naval Analyses
Arlington, Virginia

Various proposals are discussed, principally (1) Rankin, who has a two-level grammar, the first gives the strokes and rules for combination and the second explicates the order, with a recursive definition of subframes. (2) Fujimara has an inventory of strokes and operators. For each stroke 3 functional points are isolated and operators define the linking by reference to these points. Applications include keyboard input, storage and retrieval or characters, and automatic recognition. There are two different approaches. One seeks a logically efficient system; the other one that seems natural to a user of the language.

CHINESE CHARACTER RECOGNITION BY A STOCHASTIC SECTIONALGRAM METHOD

Y-L. Ma
National Taiwan University

An approach to recognition of a block picture by comparing it with stochastic sectionalgrams obtained by grouping many samples. To calculate the risk, the absolute values of the differences between the stroke-occurrence probabilities of corresponding quanta in the two sectionalgrams are summed one of these two sectionalgrams being derived from the input pattern and the other from the prototype pattern. The smaller the sum of these differences is, the more accurate the input pattern recognition.

COMPUTER IDENTIFICATION OF CONSTRAINED HAND PRINTED CHARACTERS WITH A HIGH RECOGNITION RATE

W. C. Lin and T. L. Scully
Case Western Reserve University
Cleveland, Ohio

Hand printed on a standardizing grid made of twenty line segments, yielding twenty features, and input using a television camera, 49 character classes were recognized at a greater than 99.4% rate. Feature values calculated utilizing a Gaussian point-to-line distance concept were used in a weighted minimum distance classifier. All character-dependent data are obtained through training techniques Both statistical linear regression and averaging methods are used to obtain the parameters defining each character class in feature space.

THE PHRASAL LEXICON

Joseph D. Becker

We speak mostly by conjoining remembered phrases. Productive processes have secondary roles of adapting old phrases to new situations and of gap filling.

Programs for Linguistic Statistics
Part 1: Word Roots in Scientific and Technical Russian

[Programme zur Sprachstatistik. Teil 1:
Wortstämme in russischen naturwissenschaftlichen und technischen, Fachsprachen]

S. Halbauer

Angewandte Informatik, 16: 469-470, 1974

Description of a program, written in machine language, that searches for words containing a fixed stem from Russian mathematical texts.

Grammar : Parser   38

Augmented Phrase Structure Grammars

George E. Heidorn
Computer Sciences Dept.
IBM Watson Research Center
Yorktown Heights, NY

Augmented phrase structure grammars consist of phrase structure rules with embedded conditions and structure building actions Data structures are records consisting of attribute-value pairs. Records can be actions, words, verb phrases, etc. There are three kinds of attributes: relations, whose value is a pointer to other records; properties, with values either numbers or character strings; and indicators, whose values have a role similar to linguistic features. Structure building rules have a left part indicating the contiguous segments that must be present for a structure building operation, given in a right part, to apply.

Diagnosis as a Notion of Grammar

Mitchell Marcus
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge

The hypothesis is that every language user knows as part of his recognition grammar, a set of highly specific diagnostics that he uses to decide deterministically what structure to build next at each point in the process of parsing a sentence. This theory rejects 'backup as a standard control mechanism for parsing. A grammar is a set of modules. The parser works on two levels, a group level and a clause level. Group level modules work on a word buffer and build group level structures. Modules have a pattern, a pretest procedure and a body to be executed if the pattern matches and the pretest succeeds. If the parser fails, it keeps the structure constructed to date, and makes whatever substructures it can from the remaining part.

Some Programming Aspects of Computers with Natural Language

William White
National Institutes of Health
Division of Computer Research and Technology
Bethesda, Maryland

A morphological analyzer is written in PL/1 using a recursive macro actuated generator. Called with a word as argument it returns a stem, part of speech, possible transformations, and semantic information.

Topic Analysis

Brian Phillips
Department of Information Engineering
University of Illinois at Chicago Circle

A theory for the structure of discourse is developed. It is shown that propositions of a coherent discourse must be logically connected and exhibit a hierarchic thematic structure that has a single root. An example of a logical connective is 'Cause'; a theme is a generalized pattern that is associated with a single word, e.g., 'poison' is describable as 'Someone ingests something that causes him to become ill'. A theme applies to a discourse if its definiens matches part of the discourse. The topic of a coherent discourse is its matrix theme; an illformed discourse has no topic.

Not all discourse structure is expressed. If omitted, it must be inferrable. The process of inference requires a store of world knowledge - encyclopedic knowledge. An encyclopedia is described that contains all the devices required by the discourse analysis problem. In fact, the encyclopedia is a general model for human cognition and is applicable to many diverse cognitive tasks. The encyclopedia is a directed graph. Categories of nodes and arcs, and of processes, are presented in detail.

On "Fuzzy" Adjectives

Fred J. Damerau
IBM Watson Research Center
Yorktown Heights, N.Y.

Discusses some of the problems that arise when the concept
of a linguistic variable is combined with the concept of a fuzzy
set: the range of the numerical base variable, in ordering usage,
is not fixed for a given linguistic variable. Does not explain the
computation of values of compound expressions from the values of
their components. Not all adjectives can be related to an under-
lying numerical base. Other features involved in a complete anal-
ysis are: average value, typical value, observed value, standard
deviation of values and polarity.

[Distribution limited prior to publication.]

User's Guide to the SOLAR Theoretical Backgrounds File

Timothy Diller and Tom Bye
System Development Corporation
Santa Monica, California   90406

For each analysis in the semantic analysis file the author's
theoretical orientation, his assumptions, and his notational conven-
tions are entered on this file. The data fields are: identifying
number, document source, related sources, words analyzed, conventions,
theoretical basis including - acknowledgements, assumptions, stated
purpose, and limits, a SOLAR critique, and the name of the person
responsible for the entry. This file is available via on-line
queries or in a listing format. The file can be searched using the
identifying number on document source fields. Other fields can be
searched using a string-matching facility.

USER S GUIDE TO THE SOLAR SEMANTIC ANALYSIS FILE

Tom Bye, Timothy Diller, and John Olney
System Development Corporation
Santa Monica, California   90406

Report No. TM-5292/001/00 April 1975

        This file contains formal descriptions of word meanings,
including qualifications, informal explanations, and criticisms of
descriptions. The words used are found in the lexicons of the
Speech Understanding Research groups being sponsored by ARPA. The
semantic analysis produces 23 data fields for each word, of which
the following are searchable: word, domain analysis number, source
part of speech and components Other fields can be searched using
a string matching facility. This file is available via on-line
queries or in a listing format.

USER'S GUIDE TO THE SOLAR BIBLIOGRAPHY FILE

Timothy Diller
System Development Corporation
Santa Monica, California 90406

Report No. TM-5292/000/02 December 1974

        This file provides the citations to the documents refer-
enced in other SOLAR files. Thirty data fields are used, of which
the following are searchable: author, year, index term, document
type, subject ID, document number, and Bell ID. Other fields can
be searched using a string-matching facility. This file available
via on-line queries or in a listing format including an author.
keyword and sequence number index.

PRIMITIVES AND WORDS

Yorick Wilks
Istituto per Gli Studi
Semantici e Cognitivi
Castagnola, Switzerland

If semantic primitives are seen as essentially different
from words, this leads to attempts to justify them directly,
usually psychologically. Otherwise the justification is merely
that they work. Primitives can be taken as a small natural
language, with no essential difference between primitives and
words. But the set of primitives cannot be extended indefinitely,
otherwise the distinction between the representation and the
natural language will be lost. If it is not possible to escape
from natural language into another realm, one cannot separate
semantic representation from reasoning as is attempted. It is
probably more sensible to say that natural language understanding
depends on reasoning rather than vice-versa.

META-COMPILING TEXT GRAMMARS AS A MODEL FOR HUMAN BEHAVIOR

Sheldon Klein
Computer Sciences Department
University of Wisconsin
Madison

A key feature of the system is that the semantic deep struc-
ture of the non-verbal, behavioral, rules may be represented in the
same network as the semantics for natural language grammars, and,
as a consequence, provide non-verbal context for linguistic rules.
The total system has the power of at least the 2nd order predicate
calculus.

The primitive ACTs of conceptual dependency

Roger C. Schank
Yale University
New Haven, Connecticut

Canonical representations of conceptualizations are composed of an ACTOR, an ACTION and a set of ACTION dependent cases. The 12 primitive actions are ATRANS, transfer of possession; PTRANS, transfer of physical location; MTRANS, transfer of information; PROPEL, application of physical force; MBUILD construction of new conceptual information; INGEST, taking in of an object by an animal; GRASP, to grasp; ATTEND, to focus sense organ on an object; SPEAK, to make a noise; MOVE, to move a body part; EXPEL, to push something out of the body; and PLAN, which characterizes the ability to form a course of action that leads to a goal.

Comments on lexical analysis

George A. Miller

An analysis of the verb 'hand' is paraphrased as: 'S had Y prior to some time t at which X used his hand to do something that caused Y to travel to Z, after which Z had Y' The analysis includes a discussion of the subsumed concepts HAPPEN, USE, ACT, CAUSE, ALLOW, BEFORE, TRAVEL, and AT.

A SYSTEM OF SEMANTIC PRIMITIVES

Ray Jackendoff
Department of English
Brandeis University

Primitive functions GO, BE and STAY can be extended from a positional interpretation to possessional and identificational interpretations. Two kinds of cause are distinguished, CAUSATIVE and PERMISSIVE. Inference rules based on the form of semantic representations derive logical entailments. e.g. CAUSE (X,E)-- E.

COMPUTATIONAL UNDERSTANDING

Christopher K. Riesbeck

Comprehension is a memory process; breaking computational understanding into subproblems of parsing and semantic interpretation has hindered progress with much effort wasted on the construction of parsers. A system is described in which a monitor takes words from a sentence one at a time, from left to right. From a lexicon expectations of the word (or its root) are added to a master list of expectations. If an element of the master list evaluates to 'true', programs associated with the element are executed. The final structure built by the triggered expectations is the meaning of the sentence.

DOES A STORY UNDERSTANDER NEED A POINT OF VIEW?

Robert P. Abelson
Yale University
New Haven Connecticut

Reasoning may be propositional or by mental simulation using visual imagery. In the latter situation, do people include acts and objects not present in a given story, but necessary to carry out the simulation. This has not yet been experimentally tested. Experiments have shown that a listener may simulate a story from the point of view of an observer or of a participant in the story. One problem that this raises for AI, if a program can construct an interconnected structure from the text, is the non-uniqueness of this meaning representation. Another problem is that programs should not be designed to preserve all details, but then, what should be forgotten; point of view may be useful here

BRIDGING

Herbert H. Clark
Stanford University
Stanford, California

Listeners draw inferences from what they hear, but different listeners can make different inferences. One kind of inference in comprehension is in the context of given-new information: the speaker tries to construct the given and new information of each utterance, so that the listener is able to compute unique antecedents for the given information, and so that he will not already have the new information attached to the antecedent. Inference mechanisms include direct reference, identity, pronominalization, epithets, set membership, indirect reference by association, indirect reference by characterization, reasons, causes, consequences, and concurrences. Bridging inferences need not be determinate, but in discourse they seemingly are, and further, are the inferences with fewest assumptions. Both backward and forward inferences are possible, but only the former are determinate.

COMPUTERS AND NATURAL LANGUAGE

A. W. Pratt, M. G. Pacak, M. Epstein and G. Dunham
National Institutes of Health
Division of Computer Research and Technology
Bethesda, Maryland

The Systematized Nomenclature of Pathology (SNOP), in use at
NIH, consists of about 15,000 entries in four lists: topography,
morphology, etiology, and function. Only a few binary relations on
terms are needed; e.g., location of morphology, (lesion) at topog-
raphy (body site). Numerous relations on the primary relational
triples evidently have to be defined.

GENERATION AS A SOCIAL ACTION

Bertram C. Bruce
Bolt Beranek & Newman
Cambridge, Mass   02138

Generation is a two stage process. The first formulates a
plan and the second expresses these intentions; there is feedback
between the stages. Intentions can be encoded by (i) establish-
ing presuppositions, (ii) by linguistic conventions, and (iii) by
discourse structure. A Social Action Paradigm is a model of the
flow of social actions.

THE BOUNDARIES OF LANGUAGE GENERATION

Neil M. Goldman
Information Sciences Institute
University of Southern California

In generating natural language from a conceptual structure words and syntactic structure must be deduced from the information content of the message. Words are accounted for by a pattern matching mechanism, a discrimination net. The case framework of verbs is one source of knowledge for choice of syntàctic structure.

SPEAKING WITH MANY TONGUES: SOME PROBLEMS IN MODELING SPEAKERS OF ACTUAL DISCOURSE

John H. Clippinger, Jr.
Teleos
Cambridge, Mass   02138

In therapeutic discourse the subject is not so much generating discourse as regulating it. Statements are made, retracted, qualified and restated. The ERMA model simulates this. It has five stages, represented as CONNIVER contexts. The discourse stream has its source in a special program and then flows back and forth between the contexts before achieving its final expression. Each context determines suitability for expression; whether it should be censored or passed on with suggestions for modification. Concepts are represented by means similar to Minsky's frames.

CONSIDERATIONS FOR COMPUTATIONAL THEORIES OF SPEAKING:
SEVEN THINGS SPEAKERS DO

John H. Clippinger, Jr.
Teleos
Cambridge, Mass   02138

Technological computational linguistics is primarily concerned with software technology whereby computers can use and process natural language. Descriptive computational linguistics uses the computer as a means of developing an accurate and empirically valid model of linguistic and cognitive behaviors of human speakers. There is no inherent representation of intentions in the former, and experience is that it cannot easily be generalized to the latter. One problem of modeling is that important things are often hidden by their familiarity.

CREATIVITY IN VERBALIZATION AS EVIDENCE FOR ANALOGIC KNOWLEDGE

Wallace L. Chafe
Department of Linguistics
University of California
Berkeley

Both propositional and non-propositional knowledge must exist. Interpretive processes during perception individuate and categorize objects. If an object cannot be categorized then the object will be stored with analogic information. During verbalization analogic images will be compared with available category prototypes to decide on the best match for use in the utterance.

REPRESENTATION AND UNDERSTANDING

STUDIES IN COGNITIVE SCIENCE

edited by

Daniel G. Bobrow
Xerox Palo Alto Research Center
Palo Alto, California

Allan Collins
Bolt Beranek and Newman
Cambridge, Massachusetts

Academic Press

1975

Dedicated to the memory of

JAIME CARBONELL, 1928-1973

I. THEORY OF REPRESENTATION

II NEW MEMORY MODELS

The preface is reprinted on the following frames
by permission.

Preface

Jaime Carbonell was our friend and colleague. For many years he worked with us on problems in Artificial Intelligence, especially on the development of an intelligent instructional system. Jaime directed the Artificial Intelligence group at Bolt, Beranek, and Newman (in Cambridge, Massachusetts) until his death in 1973. Some of us who had worked with Jaime decided to hold a conference in his memory, a conference whose guiding principle would be that Jaime would have enjoyed it. This book is the result of that conference.

Jaime Carbonell's important contribution to cognitive science is best summarized in the title of one of his publications: *AI in CAI*. Jaime wanted to put principles of Aritificial Intelligence into Computer-Assisted Instruction (CAI) systems. He dreamed of a system which had a data base of knowledge about a topic matter and general information about language and the principles of tutorial instruction. The system could then pursue a natural tutorial dialog with a student, sometimes following the student's initiative, sometimes taking its own intiative, but always generating its statements and responses in a natural way from its general knowledge. This system contrasts sharply with existing systems for Computer-Assisted Instruction in which a relatively fixed sequence of questions and possible reponses have to be determined for each topic. Jaime did construct working versions of his dream--in a system which he called SCHOLAR. But he died before SCHOLAR reached the full realization of the dream.

It was a pleasure to work with Jaime. His kindness and his enthusiasm were infectious, and the discussions we had with him over the years were a great stimulus to our own thinking. Both as a friend and a colleague we miss him greatly.

Cognitive Science. This book contains studies in a new field we call *cognitive science*. Cognitive science includes elements of psychology, computer science, linguistics, philosophy, and education, but it is more than the intersection of these disciplines. Their integration has produced a new set of tools for dealing with a broad range

i

of questions. In recent years, the interactions among the workers in these fields has led to exciting new developments in our understanding of intelligent systems and the development of a science of cognition. The group of workers has pursued problems that did not appear to be solvable from within any single discipline. It is too early to predict the future course of this new interaction, but the work to date has been stimulating and inspiring. It is our hope that this book can serve as an illustration of the type of problems that can be approached through interdisciplinary cooperation. The participants in this book (and at the conference) represent the fields of Artificial Intelligence, Linguistics, and Psychology, all of whom work on similar problems but with different viewpoints. The book focuses on the common problems, hopefully acting as a way of bringing these issues to the attention of all workers in those fields related to cognitive science.

Subject Matter. The book contains four sections. In the first section, Theory of Representation, general issues involved in building representations of knowledge are explored. Daniel G. Bobrow proposes that solutions to a set of design issues be used as dimensions for comparing different representations, and he examines different forms such solutions might take. William A. Woods explores problems in representing natural-language statements in semantic networks, illustrating difficult theoretical issues by examples. Joseph D. Becker is concerned with the representation one can infer for behavioral systems whose internal workings can not be observed directly, and he considers the interconnection of useful concepts such as hierarchical organization, system goals, and resource conflicts Robert J. Bobrow and John Seely Brown present a model for an expert understander which can take a collection of data describing some situation, synthesize a *contingent knowledge structure* which places the input data in the context of a larger structural organization, and which answers questions about the situation based only on the contingent knowledge structure.

Section two, New Memory Models, discusses the implications of the assumption that input information is always interpreted in terms of large structural units derived

from experience. Daniel G. Bobrow and Donald A. Norman postulate active *schemata* in memory which refer to each other through use of *context-dependent descriptions*, and which respond both to input data and to hypotheses about structure. Benjamin J. Kuipers describes the concept of a *frame* as a structural organizing unit for data elements, and he discusses the use of these units in the context of a recognition system. Terry Winograd explores issues involved in the controversy on representing knowledge in declarative versus procedural form. Winograd uses the concept of a frame as a basis for the synthesis of the declarative and procedural approaches. The frame provides an organizing structure on which to attach both declarative and procedural information.

The third section, Higher Level Structures, focuses on the representation of plans, episodes, and stories within memory. David E. Rumelhart proposes a grammar for well-formed stories. His summarization rules for stories based on this grammar seem to provide reasonable predictions of human behavior. Roger C. Schank postulates that in understanding paragraphs, the reader fills in causal connections between propositions, and that such causally linked chains are the basis for most human memory organization. Robert P Abelson defines a notation in which to describe the intended effects of plans, and to express the conditions necessary for achieving desired states.

The fourth section, Semantic Knowledge in Understander Systems, describes how knowledge has been used in existing systems. John Seely Brown and Richard R. Burton describe a system which uses multiple representations to achieve expertise in teaching a student about debugging electronic circuits. Bonnie Nash-Webber describes the role played by semantics in the understanding of continuous speech in a limited domain of discourse. Allan Collins, Eleanor H. Warnock, Nelleke Aiello, and Mark L. Miller describe a continuation of work on Jaime Carbonell's SCHOLAR system. They examine how humans use strategies to find reasonable answers to questions for which they do not have the knowledge to answer with certainty, and how people can be taught to reason this way.

ORGANIZATION AND INFERENCE IN A FRAME-LIKE SYSTEM OF COMMON SENSE
KNOWLEDGE

Eugene Charniak
Institute for Semantic and Cognitive Studies
Castagnola, Switzerland

Frames are static structures about one stereotyped topic
Each frame has many statements about the topic, each expressed
in a suitable semantic representation. The primary goal in under-
standing is to find instances of frame statements in the discourse
Questions about a source statement can be answered by reference to
the frame of which it is an instance.

COGNITIVE NETWORKS AND ABSTRACT TERMINOLOGY

David G. Hays
Department of Linguistics
State University of New York at Buffalo

By systematic application of a cognitive network or similar
theory of knowledge the internal structure of a (medical) code can
be improved and tools developed for different purposes. Hays's
theory uses paradigmatic, syntagmatic, discursive, attitudinal, and
metalingual (MTL) arcs. The MTL arcs shift level of abstraction;
e.g., anemia is neither a fewness nor an erythrocyte but an abstract
condition. An abstract definition can include several syntagmatic
propositions, linked discursively. A medical term can be linked by
MTL to definitions in different languages (clinical, Pathophysio-
logical, etc.)

STRUCTURAL KNOWLEDGE IN A DOCUMENT INFORMATION CONSULTING SYSTEM

Ronald J. Brachman
Center for Research in Computing Technology
Harvard University
Cambridge, Mass 02138

A data structure schema for creating structured concept nodes in a semantic network is presented, with structuring techniques based on a set of primitive link types including: defined as attribute part, modality, role, structural/condition, value/restriction, subconcept and superconcept. This structure will store descriptions of bibliographic references in a way that will facilitate the important processes of inference, paraphrase and analogy.

THE TROUBLE WITH MEMORY DISTINCTIONS

Allan Collins
Bolt Beranek & Newman
Cambridge, Mass. 02138

Tulving's episodic memory is seen as a record of experiences and their context. However, both episodic and semantic memories must have similar power of representation, so their structures are not distinguishable. Similarly, a lexical memory must have the power to represent propositional information about words. Thus, the fabric of knowledge is merely cut into different shapes.

A FORMALISM FOR RELATING LEXICAL AND PRAGMATIC INFORMATION: ITS RELEVANCE TO RECOGNITION AND GENERATION

Aravind K. Joshi and Stanley J. Rosenschein
The Moore School of Electrical Engineering
University of Pennsylvania
Philadelphia, 19174

        A uniform formal structure for the interpretation of events, initiation of actions, understanding language, and using language is sought. The components of the system are CONTROL --the procedural component; SCHEMATA --a lattice whose points are lexical decompositions; LEXICON --non-definitional information; BELIEFS -- a closed and consistent set of statements in a predicate calculus; and GOALS.

HOW EPISODIC IS SEMANTIC MEMORY?

Andrew Ortony
University of Illinois at Urbana-Champaign

        The distinction between semantic and episodic memory is not so much one between different kinds of memory, but one between different kinds of knowledge. The distinction has been rejected, because it is said that since we know everything from experience, there is no room for the distinction. The error lies in confusing knowledge from knowledge, and knowledge of knowledge. Semantic knowledge is knowledge that has been reorganized around concepts from knowledge originally encoded around events; it is stripped of personal experience. One question raised by the distinction is how does information get into semantic memory, and how and when does it get lost from episodic memory.

BAD-MOUTHING FRAMES

Jerry Feldman
University of Rochester
New York

There is evidence that people use three-dimensional models and that they integrate several views into a single model. This is counter to the claim that we symbolically store a large number of separate views. Another problem is with the assumption of default values for slots in frames. In the extreme, this gives visual perception without vision. The evidence is that people can understand totally unexpected images presented for quite short periods. A third point concerns the relatively static nature of frames. A better model is to construct a goal oriented subsystem making use of context specific knowledge.

SOME THOUGHTS ON SCHEMATA

Wallace L. Chafe
University of California
Berkeley

Stories are broken down into schemata, e.g., plot plus moral. Questions about schemata are: what are the essential ingredients of a schema; are some more abstract than others; and how are they to be discovered--by imagination and intuition?

STEREOTYPES AS AN ACTOR APPROACH TOWARDS SOLVING THE PROBLEM OF PROCEDURAL ATTACHMENT IN FRAME THEORIES

Carl Hewitt

Stereotypes are actor versions of frames. A stereotype has the following parts: a collection of characteristic objects, characteristic relations for these objects and invocable plans for transforming the objects and relations.

MINSKY'S FRAME SYSTEM THEORY

Marvin Minsky
Artificial Intelligence Laboratory
M.I.T.
Cambridge, Mass

Frames are data structures for representing stereotyped situations. Each frame contains information about how to use the frame, what to expect to happen next, and what to do if the expectations are not fulfilled. Lower levels of a frame have terminals that can be filled by specific instances from source statements. Frames are linked together into a frame system and the action to go from one to another indicated. Different frames can share the same terminals. Unfilled slots in instances of frames are filled by default options from the general frame.

Using knowledge to understand

Roger C. Schank
Yale University
New Haven, Connecticut

A SCRIPT is a structure consisting of slots and requirements on what can fill the slots. It is defined as a predetermined causal chain of conceptualizations that describe the normal sequence of things in a familiar situation. A SCRIPT header defines the circumstances under which a SCRIPT is called into play.

Grammatical inference: Introduction and survey. Part 1

K. S. Fu   and   T. L. Booth
Purdue University   Connecticut University

Potential engineering applications. Inference algorithms for finite-state and context-free grammars. Application of some of the algorithms to the inference of pattern grammars in syntactic pattern recognition illustrated by examples.

An Anthropological Linguistic View of Technical Terminology

Paul L. Garvin
Department of Linguistics
State University of New York at Buffalo

The health-care community has a functional dialect, with subdialects for physicians, nurses, etc. Anthropological study of the naming behavior of the community is a suitable preliminary step in thesaurus building. It would determine what are terms to be entered, how they are related, and what theoretical differences require alternative definitions of the same term.

Syntactic Recognition of Imperfectly Specified Patterns

M. G. Thomason and R. C. Gonzalez
Tennessee University

Using for illustration a recognition system for chromosome structures, methods are developed which basically consist of applying error transformations to the productions of context-free grammars in order to generate new context-free grammars capable of describing not only the original error free patterns, but also patterns containing specific types of errors such as deleted, added, and interchanged symbols which often arise in the pattern-scanning process.

METHODOLOGY IN AI AND NATURAL LANGUAGE UNDERSTANDING

Yorick Wilks
Istituto per Gli Studi
Semantici e Cognitivi
Castagnola, Switzerland

Artificial Intelligence has had at least four benefits for the study of natural language: (a) emphasis on complex stored structures, (b) emphasis on the importance of real world knowledge, (c) emphasis on the communicative function of sentences in context, and (d) emphasis on the expression of rules, structure and information within the operational environment. The only test of a natural language system is its success on a task, any demand for more theory must bear this in mind. Neither can recent work in AI be regarded as theoretical; it is the semi-formal expression of intuition. AI is engineering, not a science, and as such there is no boundary to natural language; one counter example does not overthrow a rule system. Further, talk of theory distracts from heuristics.

FORMAL REASONING AND LANGUAGE UNDERSTANDING SYSTEMS

Raymond Reiter
Department of Computer Science
University of British Columbia

There are two mechanisms for formal reasoning: (a) resolution principle, a competence model, by virtue of its completeness, and (b) natural deductive systems, which are attempts to define a performance model for logical reasoning. A system could be designed that interfaces the two systems, each doing what it does best. Natural deductive systems have not considered fuzzy kinds of reasoning. Future questions concern other quantifiers, contexts for representing wanting, needing, etc., and the balance between computation and deduction.

THE COMMONSENCE ALGORITHM AS A BASIS FOR COMPUTER MODELS OF HUMAN MEMORY, INFERENCE, BELIEF AND CONTEXTUAL LANGUAGE COMPREHENSION

Chuck Rieger
Department of Computer Science
University of Maryland

Commonsense algorithms are basic structures for modeling human cognition. The structure is defined by specifying a set of links which build up large structures of nodes of five types: Wants, Actions, States, Statechanges and Tendencies. There are 25 primitive links, e.g., one-shot causality, action concurrency, inducement Various applications are active problem solving, basis for conceptual representation of language, basis of self model, etc.

UNDERSTANDING HUMAN ACTION

Charles F. Schmidt
Rutgers University
New Brunswick, New Jersey

A model of reasoning about human action must include (1) how people arrive at a plan, (2) what can count as a reason for choosing to perform the plan, and (3) discovering plans and motivations from observation or linguistic report of actions. A plan is the internal representation or set of beliefs about how a particular goal may be achieved. The belief by an observer that an actor performed one act to enable a second to be performed can follow neither from deductive nor inductive reasoning. An observer may have other propositions that are reasons for believing or not believing that a plan correctly characterizes the beliefs of the actor. An act name organizes a set of beliefs about how a move of this type might relate to other moves, and the cognitive and motivational states of the actors.

Programming

STRING AND LIST PROCESSING IN SNOBOL4: TECHNIQUES AND APPLICATIONS

Ralph E. Griswold

Prentice-Hall, Inc.
Englewood Cliffs
New Jersey
1975

Reviewed by Norman Badler
Department of Computer and Information Science
The Moore School of Electrical Engineering
University of Pennsylvania

Among popular computer programming languages, SNOBOL4 stands out as the only one offering complex pattern definition and matching capabilities. It also has a flexible function definition facility and programmer-defined data types. While not unique, these two features encourage problem-dependent extensions of the language. . All three aspects of SNOBOL4 form the basic tools in Griswold's new book

Intended as a text for the SNOBOL4 user (it is not an "introductory" text), it presents techniques for the representation and manipulation of data in string, list, or otherwise "structured" form. The text includes many programmed examples, problems with a wide range of difficulty, and answers to many of these problems. The first three chapters develop pattern matching, function definition, and data structures. The last four chapters examine particular application domains: mathematics, cryptography, document preparation, plus a few more specialized problems. Although this may seem to ignore computational linguistics, the greatest immediate benefit for the programmer lies in the first three chapters anyway.

Within Chapter 1, the section on grammars and patterns can be used for the implementation of simple syntactic analysis. For example, there is a straight-forward mapping of a BNF grammar into SNOBOL4 patterns, but there are pitfalls (as well as some more efficient representations in the balance) that the programmer ought to know. These are carefully explained.

A topic that I felt was inadequately covered in Chapter 1 was the definition of the pattern matching mechanism itself. The immediate presentation of examples using pattern matching (page 2) calls for a brief overview of pattern matching syntax and semantics. Surely a programmer would appreciate not having to refer back to his introductory text should some pattern function or construct be hazy in his memory. Even an appendix would be satisfactory. In addition, this would support the section on patterns as procedures by providing the underlying semantics for such "procedures." Further incentive for its inclusion is provided by the excellent review of programmer-defined data types in Chapter 3. Why leave pattern matching to the user's recollection?

The function definition facility discussed in Chapter 2 enables the constructic of generic functions. Since there are no data type declarations for function arguments or parameters, often only one function is required for the execution of related operations on various data types. The proliferation of functions in a complex system might therefore be systematically reduced. The burden falls on the programmer, of course, to sort out the admissible combinations or appropriate actions. An addition function for real and complex numbers is discussed, where the former is a SNOBOL4 primitive and the latter is constructed from programmer-defined data types. Although not in the realm of computational linguistics, it does have a parallel, for example, in a function which inserts data into a semantic network and is expected to handle various chunks of network as well as atomic data. The data type might only be determined during program execution; using a generic function avoids distracting logic within the user's primary function.

The section on functions as generators is a little weak from the point of view of computational linguistic requirements for procedures which generate successive alternatives from a complex structure, for example, sentence parsing or

referent resolution. The use of simple global variables is too limited in these contexts; one often needs to become involved with saving the values of several local variables in special data blocks or stacking the decision points associated with alternatives. The first is a well-known compiler-design technique, while the second involves a backtracking control structure. In fact, an excellent illustration of these ideas would be an implementation of the SNOBOL4 pattern matching system in SNOBOL4.

Chapter 3 is the most useful because it describes how programmer-defined data types can be used to build "structures": stacks, queues, linked lists, binary trees, and trees. The skillful user of such representations will find a reduced role for complicated pattern matching expressions because the implicit structure encoded into a string becomes manifest in the explicit links of the structure. Not only is there often an economic advantage, but the semantics of SNOBOL4 are easier to use than the implicit backtracking semantics of pattern matching. (Griswold himself points this out in the section on patterns as procedures.) The programmer is encouraged to consider economic trade-offs in the implementation of structures. Often over-looked questions are addressed: for example, the relative merits of implementing stacks using strings, arrays, tables, or defined data types. Programs for the use or traversal of structures are also provided.

Although exercise 3.40 requests a representation for directed graphs, neither hint nor answer is provided. The computational linguist having an interest in semantic networks or similar associative structures is thus left to his own expertise. The basic tree representation must be significantly modified to incorporate labelled edges, a means of traversal (search) through the edge set, and, of course, non-tree structures. Griswold apologizes for not covering every application,

but the generality and current popularity of networks for the representation of knowledge calls for expanded treatment of the topic.

Among the applications covered in detail, the ones most relevant to computational linguistics include a random sentence generator (from a grammar), a· macro processor, and (perhaps) a context editor. The input and output of textual material is covered in depth under document preparation (Chapter 6). Since the text does not delve into computational linguistics per se, the reader (or instructor) will often be called upon to map techniques described in the text onto his own problem. I think that a good programmer would be able to perform this transformation since solutions are provided for many of the basic problems in handling input text, setting up data structures, and traversing these structures.

Before you begin programming your next computational linguistics project, a glance through this book may save you considerable programming time and reward you with usable and flexible data structures. Even if you do not program in SNOBOL4, the techniques presented here might guide you to more efficient usage of other languages. On the other hand, it might convince you to try SNOBOL4.

F O R T R A N   T E C H N I Q U E S

WITH SPECIAL REFERENCE TO NON-NUMERICAL APPLICATIONS

A. Colin Day

Cambridge University Press
New York
1972

Reviewed by Richard J. Miller
St. Olaf College

A practical guide for the occasional Fortran IV programmer to
the basic "tricks" and vocabulary used by the systems programmers.
This book ranges over topics from plotting on a line printer to hash-
ing and basic storage structures (stacks, queues, etc.) using a con-
cise, to-the-point writing style. This style reinforces the stated
intention of the book, which is to help a programmer with a problem
by providing descriptions of non-mathematical techniques. The style
and intention do limit the usefulness of this book, as some of the
topics would be well known to advanced programmers and are not cov-
ered in sufficient depth for such a person. It is then the area be-
tween these two extremes to which this book is aimed, and there it
can be of great service.

The only important assumption made of the reader is that he know
the variable types of Fortran (integer, real, Hollerith, etc.) and
their attendant format specifications. A good knowledge of character
formats is especially useful, although the major use for them is in
output statements used in the examples given in the book. It is also
assumed that the reader knows the basic Fortran statements, but this

is simple matter as opposed to the format and variable type problems which confront a Fortran programmer.

The book also includes several exercises at the end of each chapter (answers not supplied unfortunately) and a short but very complete bibliography which includes several sources for each chapter. The book's primary value is as a source for hints to problems encountered during programming, providing an introduction to the more sophisticated literature which can be found by starting with the bibliography. This book is therefore a starting point for picking up a basic vocabulary, techniques, and references for someone who has just completed a programming course or who needs a quick introduction to some technique which he may want to look at later in more detail.

INFORMATION SYSTEMS

VOLUME 1 NUMBER 2

APRIL 1975

Hans-Jochen Schneider, Editor-in-Chief
Institute für Informatik
Universität Stuttgart
Herdweg 51
D-7000 Stuttgart 1/Fedrep. Germany

TABLE OF CONTENTS

ON RETRIEVING INFORMATION FROM VISUAL IMAGES

Stephen Michael Kosslyn
The Johns Hopkins University
Baltimore, MD

A computer graphics metaphor is useful for human visual imagery. Analogous properties are found: as objects become smaller their constituent parts become more difficult to discern perceptually; as more parts are added to an image it becomes more degraded due to capacity limitations; images displaying more identifiable details take longer to construct; images cannot be indefinitely expanded before overflowing; and the existence of decay time for an image which affects the time taken to construct a new image.

REPRESENTATION OF KNOWLEDGE: NON-LINGUISTIC FORMS
DO WE NEED IMAGES AND ANALOGUES?

Zenon W. Pylyshyn
Department of Psychology
University of Western Ontario
London, Canada

Semantic structure is relative to the process that constructs and uses the representation. By positing analogue representations it is suggested that a process does not need to know the rules of transformation, e.g., rotation, but this is impossible unless the analogical modelling medium intrinsically follows the laws of physics, i.e., ascribing these laws to brain tissue.

THE NATURE OF PERCEPTUAL REPRESENTATION: AN EXAMINATION OF THE
ANALOG/PROPOSITIONAL CONTROVERSY

Stephen E. Palmer
Department of Psychology
University of California
Berkeley

Sensory data is considered as having several levels of inter-
pretation. At the sensory end, the representation is analog, and
propositional at the cognitive end. Analog images are incorrectly
seen as having all details of the stimulus whereas quasi-linguistic
representations are only partial. The important issue is not the
partiality but the selection, possibly information that discrimin-
ates the object in context. For structural information there needs
to be a mechanism for both parts and wholes. Parametric informa-
tion can be coded componentially and explicitly, but some seems to
function integrally. It is claimed that structural perception is
qualitative whereas parametric perception is quantitative, but
structural elements may have quantitative aspects--its strength of
association with different groups. Although both structure and par-
ameters are encoded relative to other information, there is evidence
of preferred orientation and perspectives for parameters

AFTERTHOUGHTS ON ANALOGICAL REPRESENTATIONS

Aaron Sloman
Cognitive Studies Programme
School of Social Sciences
University of Sussex
Brighton, England

The distinction between Fregean (symbolic) and analogical
representations is that in the latter both representation and
thing must be complex and there must be correspondence between the
structures, whereas in the former case there is no need for a
correspondence. Attempts to subsume either representation under
the other have not succeeded. There is a mistaken belief that only
proofs in Fregean symbolism are rigorous. Although analogical
representations can sometimes be implemented using Fregean ones,
this does not imply that they are not used.

M E D I C A L   V O C A B U L A R Y

PROCEEDINGS OF THE FIFTH BUFFALO CONFERENCE ON COMPUTERS IN MEDICINE

October 29-31, 1973

Editor-In-Chief:
E. R. Gabrieli

TABLE OF CONTENTS

PROCEEDINGS OF THE FIFTH BUFFALO CONFERENCE ON COMPUTERS IN MEDICINE (CONT'D).

INFORMATION STORAGE AND RETRIEVAL

Gerard Salton, Project Director
Department of Computer Science
Cornell University
Ithaca, New York 14853

TABLE OF CONTENTS

INFORMATION STORAGE AND RETRIEVAL

TABLE OF CONTENTS (Cont'd.)

A THEORY OF TERM IMPORTANCE IN AUTOMATIC TEXT ANALYSIS

G. Salton, C. S. Yang   and   C. T. Yu
Department of Computer Science   Dept. of Computer Science
Cornell University   University of Alberta
Ithaca, NY 14853   Edmonton, Ata, Canada

Discrimination value analysis ranks text words in accordance with how well they are able to discriminate the documents of a collection from each other. The value of a term depends on how much the average separation between individual documents changes when the given term is assigned for content identification. The best words are those which achieve the greatest separation. Effective criteria are given for assigning each term to either single word, phrase or word group categories and for constructing optimal indexing vocabularies. The theory is validated by citing experimental results.

MILITARY APPLICATIONS OF SPEECH UNDERSTANDING SYSTEMS

R. Turn, A. S. Hoffman, T. F. Lippiatt
Rand Corporation
Santa Monica, California

The general military environment. Possible uses: avionics equipment control, field data entry, tactical command systems, and data base management in tactical and administrative systems. New operational capabilities may arise from spoken language translation, biomedical monitoring, and speech-operated writing machines. Applications areas for further research. Methodology for transferring this technology into operational systems.

Automated Review of Professional Services
and the Problems of Medical Records

Paul Y. Ertel
Ohio State University
Columbus

The Medical Advances Institute developed a system to keep records, select cases for review, and information about individuals and categories, for the quality control system now established by law. Over 200 quality criteria packages have been developed. They concern the process and result of medical care. The system screens each case within a day of hospital admission and frequently thereafter. It provides a review of use of facilities and conformity to standards of care.

Content Analysis and Relevance Feedback

A. Wong, R. Peck, and A. van der Meulen
Cornell University
Ithaca, New York

Experimental results indicate that final retrieval system performance, after user feedback is applied using Rocchio's algorithm, is highly dependent on the system performance of the initial indexing process. Therefore every tool which improves the indexing performance as an outcome of the content analysis of natural language is beneficial because initial differences in a system performance are retained after user feedback is applied.

THE BUREAU OF DRUGS, FOOD AND DRUG ADMINISTRATION
SCIENTIFIC INFORMATION SYSTEMS

Alan Gelberg
Bureau of Drugs
Food and Drug Administration
Rockville, Maryland

ASTRO-4 is a file of new drug applications. The Ingredients File lists 36,000 chemicals believed to have biological effects. The National Drug Code is a list of manufacturers and products. A file of Clinical Investigators and a file of Facilities are kept. Also Drug Experience and Adverse Drug Reaction, Poison Control Center file of incidents and a Drug Product Defect file. A dictionary of adverse reaction terms is in progress. Sophisticated hardware, software, and terminological controls are in use or development.

PROPOSED USE IN CANADA OF SNOMED
IN A MEDICAL INFORMATION MANAGEMENT SYSTEM

Roger A. Cote
University of Sherbrooke Faculty of Medicine
Department of Pathology
Sherbrooke, Quebec

SNOP, published in 1965, is not rich enough to code problems, signs, symptoms, disease entities, administrative, diagnostic, and therapeutic procedures. SNOMed is to cover the whole. The code is hierarchical: Topography is organized by system or tract, Morphology by such categories as traumatic, neoplasm, etc., Etiology by categories of organisms and chemicals, Normal function by metabolism, enzyme, etc., Abnormal function correspondingly, and Procedure by medical discipline. Qualifiers such as history of, laboratory diagnosis, etc., are included, and terms can be linked.

Coding Diagnoses of Medical Records: A Challenge

J. van Egmond and R. Wieme
Medische Informatica Gent
Gent; Belgium

The authors' codification system splits compound diagnoses into units, interrelated if relevant by the grammatical operator "complication of". The content of a unit is described with three sets of codes: disturbance, localization, and etiology. Representation is mnemonic for the coder, numeric for the processor.

Retrieval-Oriented Storage of Medical Data: Operational Aspects

Charles W. Conaway and Edward T. O'Neill
School of Information and Library Studies
State University of New York at Buffalo

Records are encoded by a clerk. The system is to give a physician at a terminal the current synopsis of a patient record, the complete record (delay of a few minutes), any facts selected for periodic determination in the pool of Clinical Experience; input is to be interactive, with verification of single statements.

DATA MANAGEMENT SYSTEMS AT THE SOCIAL AND REHABILITATION SERVICES

Webster A. Rogers
Division of Management Systems
Social and Rehabilitation Services
Department of Health, Education and Welfare
Washington, D.C.

To improve the management of Medicaid, which spends (predicted) $9 billion for 27 million persons in 1974, an information system was designed and installed in a pilot state. It maintains data about eligibility of persons, qualification (administrative) of providers, claims, background (e.g. normal prices); it delivers statistical summaries and exception reports for managers in addition to processing claims.

THE CLOWNS MICROWORLD

Robert F. Simmons
Department of Computer Science
University of Texas
Austin

Sentences describing scenes centred around a clown who can balance and move are analyzed by an ATN parser. The parser produces property list semantic structures which are adequate to transmit data to a package that generates the scene on a display screen.

AN HISTORICAL NOTE ON THE USE OF WORD-FREQUENCY CONTIGUITIES
IN CONTENT ANALYSIS

H. P. Iker
Rochester University

Computers and the Humanities, 8: 93-98, 1974

        Discusses the development of this form of content analysis
in information retrieval, the social sciences, and literary analysis.

BIBLIOGRAPHY ON SOCIAL SCIENCE COMPUTING

R. E. Anderson
Minnesota University
Minneapolis

Computer Reviews, 15; 247-261, 1974

        Contains 591 references in the period 1960 to 1973 covering
statistical analysis, simulation, text processing, and laboratory
automation.

Literature

CONTENTS

Humanities : Concordance

INDEX THOMISTICUS

SANCTI THOMAE AQUINATIS OPERUM OMNIUM INDICES ET CONCORDANTIAE

Roberto Busa, S.J.

Friedrich Fromann Verlag / Günther Holzboog KG, Stuttgart, 1974 -
DM 370 per volume, half-leather

A review of volumes 1-10 of the 26-volume Concordantia Prima

Ford Lewis Battles
Pittsburgh Theological Seminary

It is appropriate indeed that, on the seven hundredth anniversary of the death of St. Thomas Aquinas, Father Roberto Busa with his co-workers of the Faculty of Philosophy at the Aloisianum, Gallarate, Italy has begun to publish the long awaited massive concordance to his writings and to texts by other authors long associated with his circle. For many of us who have done lesser work in computerized humanistic studies, rumors and reports of Busa's enterprise aroused our curiosity and, in some cases, led us also to grapple with the manifold problems of producing a concordance by computer.

In studying the specifications and sampling the first ten volumes of *Index Thomisticus* (Sectio II, Concordantia Prima (A-Initor)), this reviewer has been reminded of his own struggle to produce a concordance to the *Institutes of John Calvin* (Pittsburgh, 1972). The *I.T.* provides a hierarchically organized concordance to a literary corpus of 10,600,000 words of Latin Texts; by comparison, the Calvin concordance contains 405,338 words of Latin text in a single sequence.

Thus, the vastly greater literary task of Busa called for a series of basic literary and philological and logical decisions not only to make the enormous work of processing possible, but also to produce a final instrument for the use of scholars that would rationally encompass the vast corpus.

At the outset the character of the Latin language and especially its morphological peculiarities had to be translated into computerizable routines, so that something other than a sea of raw alphabetical sorting would result. Lemmatization by hand sorting after the basic concordancing (feasible for a small corpus), preparation of an interlined ("glossed") lemmatized machine readable text (also suitable for smaller texts), even the elaborated encoding of the text developed by De Latte at the Liege Centre - none of these methods was chosen by Busa and his associates. They turned rather to Forcellini's *Lexicon totius Latinitatis* and encarded the 90,000 Forcellini lemmata (in all possible forms) plus additional ones in the Thomistic corpus to a total of 10,000,000 codes, put this on magnetic tape, and worked out procedures to apply this Latin Machine Dictionary (LEL) to the machine-readable text. This instrument is now available for the use of others working on Latin texts. To anyone knowing the homographs of Latin, the limitations of any mechanical routine are apparent: the *T.*, however, handles these problems in a clear and workable manner.

The size of the literary corpus also called for basic decisions by successively sequestering different fractions of the corpus (in a way that would have doubtless intrigued Thomas himself!), the compilers reduced the mass to manageable proportions. Their decisions

may be set down serially.

(1) LITERARY. First, divide the authentic works of
Thomas (100 + 18?) from those of other authorship (61):
treat each in separate series. Secondly, extract literal
quotations, citations in references to other authors, and
cross-citations to other Aquinian treatises; treat these
separately. This leaves distinct layers of material for
concordancing.

(2) PHILOLOGICAL. First, separate out indeclinables
like prepositions, conjunctions, adverbs, forms of *esse*,
helping verbs, etc., pronouns, numerals, etc., etc ; these
will occupy a particular concordance. Second, put the re-
maining nouns, adjectives and verbs in a primary concordance:
nouns and adjectives arranged alphabetically by termination;
verbs by a standard order.

(3) LOGICAL. To reduce the bulk of the concordance
and to make the vocabulary and context more rationally acces-
sible: First, analyze out frequently used phrases (e.g.,
liberum arbitrium, acceptio personarum, caelum et terra -
there are about 500 of these), concordancing them under one
word only; for example, acceptio personarum would be listed
after all other instances of acceptio. Secondly, distinguish
words which are either proper names or so commonly found
that only a brief context (1½ lines) need be quoted; the
rest can then be set in a context of 2½ lines - and the whole
interfiled in a single series.

These decisions have determined both the cnaracter ana con-
tent of Sectio II, comprising two series (Aquinian and non-Aquin-
ian works) of five concordances each. So much for the Concordance
proper. There remains a further instrument for the use of schol-
ars, Sectio I. In this are included indices of distribution, sum-
maries of the lexicon, and indices of frequency: Through these
lists linguistic and literary studies of all sorts can be made.
By reverse alphabetical ordering of lemmata and forms additional
kinds of analysis are facilitated. Since these volumes have not
yet appeared, they can only be briefly mentioned here.

A massive concordance of this type must carry a concise yet
precise location code for each item. The editors have determined
the proper modern edition to be used, have set a precise order of
works to be followed in the concordancing of each type under its
appropriate lemma, and have summarized this on a separate 4-page
insert, to which the user will doubtless make frequent reference
as he learns to use this grand instrument of research.

A short review cannot do justice to the immense detail and
the intelligence with which this detail has, with human and com-
puter help, been marshalled in the r.r. Father Busa and his
associates are to be commended not only for their achievement,
but also for the example they have set for other laborers.
Future Concordances to comparable literary corpora will obviously
have their own special features; yet Busa's method of attacking
problems - linguistic, philological, logical, quantitative - will
suggest analogous modes of approach to others. While an index to
St. Thomas by no means exhausts even the whole body of Christian
Latinity, it provides a key to the heart of Roman theology during

its period of greatest fruitfulness and to classical and patristic thought that passed through the schoolmen's filter. Space also precludes discussion of the physical aspects of concordancing and printing, for which Father Busa had the assistance of IBM.

In a work of such vast proportions, the care of men cannot obviate error. Some 53 errors have been noted by the compilers in Concordantia Prima. But even the correction of at least one of these contains a minor error: B.012(QDV)23 13.ag8/8 should read B.012(QDV)22.13.ag8/8. See Sectio 2 vol. 1, p. 526, col.2. Also 12.Tabula Syntagmatum, the list of phrases concordanced under only one member of the phrase (pp.xiv-xvi) has at least two errors: bona esteriora should read bona exteriora (p. xiv): drosperitas terrena should read prosperitas terrena (p. xvi) The most useful pentaglot descriptive booklet of 46pp. is somewhat marred in the English version by misprints and verbal infelicities. But these small matters are quite eclipsed by the enormous accomplishment of Father Busa and his co-workers.

POTENTIALITIES OF MACHINE PROCESSING OF LATE MIDDLE HIGH GERMAN TEXTS. REPORT ON A RESEARCH PROJECT

[Möglichkeiten der maschinellen Verarbeitung spatmittelhochdeutscher Texte. Bericht über ein Forschungsunternehmen]

T. Baumgarten
Institute for Communication Research and Phonetics
Bonn University

Computers and the Humanities, 8: 85-91, 1974

Lemmatized and classifying index, verse concordance, rhyming index, reverse morphological index frequency list, computer-readable "MHG Working Dictionary" "Syntactical Rule System" making possible a mechanical text description and expandable to a "descriptive grammar".

ON UNDERSTANDING POETRY

D. L. Waltz
University of Illinois
Urbana

In R. Schank and B.L. Nash-Webber, eds., Theoretical Issues in Natural Language Processing, 1975, 20-23.

The processing of discourse is generally organized around verbs. However the structure at a topical or thematic level may not be so organized, bearing little resemblance to a deep structure. Analogies are used, not only in poetry, to transfer large amounts of information from one domain to another; to enable communications of the otherwise inexplicable; to make distinctions vivid; and to understand new concepts by analogy to old ones.

Thought Clusters in Early Greek Oral Poetry

Cora Angier Sowa and John F. Sowa
21 Palmer Avenue
Croton-on-Hudson, New York 10520

Analyzing associations in a literary text is analogous to the problem of computing term associations in document retrieval. This paper describes how the theory of clumps was used to find clusters of closely associated words in the *Homeric Hymns*. For each cluster, the program printed a mini-concordance to the lines of text containing each word in the cluster. The results showed two types of patterns in the poet's use of words: localized word plays extending over a few lines, and global interactions between a cluster of words and the overall thematic structure of the text.

Computer Aided Instruction
A Bibliography with Abstracts

E. J. Lehmann
National Technical Information Service
Springfield, Virginia

Contains 252 abstracts dating from 1970 to June 1974 covering use in education, computer system requirements, motivation, technical training, learning factors and human factors engineering. Retrieved using the National Technical Information Service on-line search system.