⋐the FINITE STRING ⋙

NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS
VOLUME 11 - NUMBER 1   SEPTEMBER 1974

NSF SPONSORSHIP FOR AJCL

The Office of Science Information Service of the National Science Foundation has awarded $34,700 for publication of the American Journal of Computational Linguistics. Principal investigator is A. Hood Roberts, managing editor of the new journal; the Center for Applied Linguistics administers the grant, which expires on September 30, 1976.

The aim of the publication is to report current research. promptly and to announce meetings, conferences, courses, publications, and other activities of interest and importance to computational linguists.

AJCL is the official journal of the Association for Computational Linguistics, which published Mechanical Translation until 1968 and The Finite String until 1973. ACL's Executive Committee in 1968 decided to publish a new journal under a three-point policy.

> ACL appoints the editor and editorial board, thus controlling policy.

> A large working editorial board and adequate clerical support assures timeliness. AJCL attempts to review submissions within a month.

> Microfiche publication permits distribution of programs, grammars, dictionaries, illustrations, etc.

The grant provides for extensive mailing of announcements and sample issues in order to develop a full subscription list as quickly as possible.
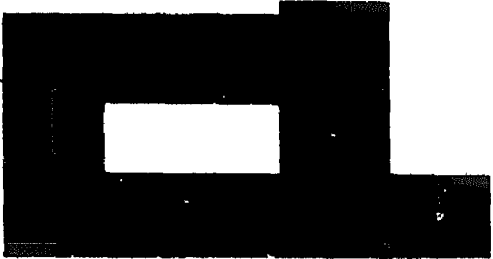
The subscriber who does not
already own a reader for
microfiches may wish to
consult this document.

MICROFICHE
VIEWING
EQUIPMENT
GUIDE

Ronald F. Borden
Defense Documentation Center

Surveys 81 models of microfiche readers
and reader-printers available in the
United States in 1971. 156 pages.

Order XC-734 400 $3.00
National Technical Information Service
Springfield, Virginia

ASSOCIATION
FOR
COMPUTATIONAL
LINGUISTICS

OFFICERS 1975

PRESIDENT: ARAVIND K. JOSHI
University of Pennsylvania

VICE-PRESIDENT: STANLEY PETRICK
IBM Research

SECRETARY-TREASURER: A. HOOD ROBERTS
Center for Applied Linguistics

EXECUTIVE COMMITTEE: BONNIE NASH-WEBBER   *1975-1977*
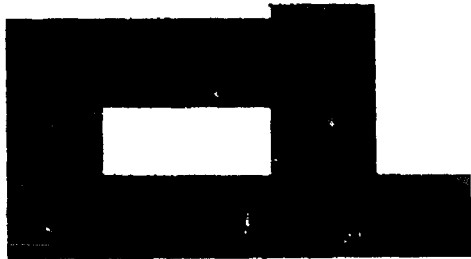Bolt Beranek and Newman

CHARLES J. RIEGER   *1975-1976*
University of Maryland

GARY MARTINS   *1973-1975*
Stonehenge Systems Laboratory

NOMINATING COMMITTEE: WILLIAM A. WOODS   *1975-1977*
Bolt Beranek and Newman

ROBERT BARNES   *1974-1976*
Lehigh University

ROBERT SIMMONS   *1973-1975*
University of Texas

ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

PROGRAM

1974

Twelfth Annual Meeting
Amherst, Massachusetts
July 26-27, 1974

SYNTAX AND SEMANTICS

Automatic creation of an ATN grammar from a transformational grammar. *Perry Miller, Massachusetts Institute of Technology*

String transformations in the REQUEST system. *Warren J. Plath, IBM Research*

A computational treatment of coordinate conjunctions. *Carol Raze, New York University.*

Toward formal solutions to philosophical problems. *James Dunn, Princeton University.*

An investigation of algorithmic translation procedures from standard semantic feature representation to predicate logic. *Robert M. Harnish and Michael Houghtaling, University of Arizona.*

Lana's progress. *Ernst von Glasersfeld, University of Georgia.*

NATURAL LANGUAGE SYSTEMS

Computer as model and metaphor. *Stephan Isard, University of Edinburgh.*

SOPHIE: A knowledge based CAI system using an analogical inference mechanism. *Richard R. Burton, Bolt Beranek and Newman Inc.*

Linguistics and graphics: a compatible system. *Christina Davis and Jeffrey Eastman, North Carolina State University.*

CLET: A computer program that learns arithmetic from an elementary textbook. *Nagib Badre, IBM.*

Construct. *Robert Smith, F. L. Rawson, and Nancy Smith, Stanford University.*

The believer system. *Geoffrey Brown, Rutgers University.*

Junction grammar as a base for automatic language processing. *Eldon Lytle, Brigham Young University.*

More

DISCOURSE AND COMPREHENSION

Simulating processes of verbalization and translation. *Wallace Chafe, University of California at Berkeley*

Episode understanding and belief guided parsing. *Bertram Bruce and C. F. Schmidt, Rutgers University*

Computer understanding of metaphorical phrases. *Sylvia Russell*

Understanding by conceptual inference. *Charles Rieger, University of Maryland*

Paraphrasing paragraphs. *Roger Schank, Institute for Semantics-Cognition, Switzerland*

Cybernetic model of conscious behavior. *Answim Vinje-Morpurgo*

SPEECH PRODUCTION AND UNDERSTANDING

Simple digital speech synthesis. *William Fisher and A. M. Engebretson, Central Institute for the Deaf, St. Louis, Missouri*

Fundamental frequency contours of auxiliary phrases in English. *Jonathan Allen and D. O'Shaughnessy, Massachusetts Institute of Technology*

Non-deterministic phonetic transcription of speech. *Richard Schwartz, Bolt Beranek and Newman Inc.*

Computer testing of fast speech rules. *Douglas B. Moran, University of Michigan*

The role of lexical semantics in automated speech understanding. *Tim Diller, System Development Corporation*

An experiment in the use of iconic language to present graphic structures via the auditory channel. *John B. Eulenberg and Morteza Amir Rahimi, Michigan State University*

Phonological rules for a text-to-speech system. *Sharon Hunnicutt, Massachusetts Institute of Technology*

ASSOCIATION FOR LITERARY AND LINGUISTIC COMPUTING

The inaugural general meeting of the Association was held at King's College, University of London, on April 30, 1973 under the chairmanship of R. A. Wisbey. Its program includes seminars, general meetings, and the publication of a Bulletin three times a year.

Almost 70 persons from 11 countries attended the first international meeting on December 14, 1973. Reports were given by

      Bernard Quemada, Besancon - history of French vocabulary

      W. Martin, Louvain - analysis of Dutch texts

      Winfried Lenders, Bonn - medieval German texts

      Wilhelm Ott, Tubingen - textual editing techniques

      Felicien de Tollenaere, Leiden - the Dutch thesaurus

      I. T. Piirainen, Jyväskylä University - standardization of Finnish

      H. Schanze, Aachen - modern German texts

ALLC is developing an organization with regional branches and national representatives. The U.S. representatives are J. Raben and W. Hirschmann.

The current issue of the Bulletin contains articles on computing and phonology (J. Berger), the processing of late middle high German (T. Bungarten) and early middle Dutch (F. de Tollenaere and W. Pijnenburg) texts, and the activities of groups in Germany (W. Lenders) and Belgium (W. Martin).

Membership for 1974 is $9.50, from Mrs. Joan M. Smith, 6 Sevenoaks Avenue, Heaton Moor, Stockport, Cheshire SK4 4AW, England.

Computer at MIT Can Read

By DAVID HASKELL

CAMBRIDGE, Mass. (UPI)— A computer that reads, and talks, once the figment of writers of science fiction, is now a reality.

Engineers at Massachusetts Institute of Technology (MIT) have developed a computer which can pronounce any word in the English language, or any string of words.

Under the guidance of Associate Prof, Jonathan Allen, the MIT engineers have programmed the computer so that it can pronounce any English word correctly, even if it has never encountered it before.

THE COMPUTER, much in the manner of humans, figures out the word by applying literally thousands of learned rules of pronunciation. And the voice which comes out of the loudspeaker atop the computer is completely nonhuman in origin, constructed by a model of the human vocal tract programmed into the computer.

It is a far cry from the back-talking, troublesome "Hal" of the movie "Space Odyssey: 2001,'" but scientists have a history of eventually catching up with the imaginations of science fiction writers.

The MIT. talking computer speaks in flat, featureless tones, but the engineers are working on this aspect and are teaching it how to pause and inflect at the proper place in a sentence, and even to change the pronunciation of words according to how they are used in a sentence.

ALLEN SAID the reading machine project involved two primary problems—building a machine to scan and recognize printed matter and to transform it into computer language, and building a computer to transform the scanned text into understandable speech.

Both were solved by drawing on earlier efforts in MIT's Research Laboratory of Electrics (RLE). This text-to-speech project began as part of an overall effort at RLE to build a machine to read to the blind, but one practical near future application might be, Allen said, using a phone—in a library, for example—to dial up the computer and having the desired information automatically read to the caller.

ALLEN SAID he and his colleagues avoided force-feeding the computer so that it would memorize all the words in the English language.

"We could have attempted to feed all the words in the English language into the computer's memory, and instructed the computer to match each word in a text with a pronunciation," he said. But, he added, "this would have been unwise because the number of English words is enormous — several-hundred thousand—and because new words are constantly being invented.

"IT IS MUCH BETTER to proceed from a basic understanding of the general linguistic rules of pronunciation," he said. "The basic properties of English would be applicable for a much longer period than mere word lists."

Allen said a much simpler method was to use "morphs," which make up all words. He said there are only about 11,000 morphs. The computer, he said, can understand about 10 times that number of words by using morphs.

To that Allen and his colleagues added some 400 letter-to-sound rules and came up with a computer which would pronounce individual syllables strung together. And by drawing on rules of linguistics developed by many researchers, the engineers taught the computer how to combine the syllables into correctly pronounced words.

COMPUTER-ASSISTED LEXICOGRAPHY

A PRELIMINARY BIBLIOGRAPHY

Richard W. Bailey
Department of English Language and Literature
The University of Michigan

A volume on Computer Uses for Language Research, edited by Sally Yeates Sedelow and Walter A. Sedelow, Jr., is in preparation for the series on Current Trends in the Language Sciences, published under the general editorship of Thomas A. Sebeok.

Richard W. Bailey is writing on historical dictionaries and John Olney on synchronic dictionaries. Among the topics they propose to treat are

>   Techniques of data selection and preparation
>   Lemmatization
>   Compilation of semantic information
>   Man-machine interaction in preparing explanatory material
>   Innovative approaches to publication

The present bibliography is a part of Bailey's preparation for the assignment. Both Bailey and Olney request reprints, citations, and suggestions.

Computer-Assisted Lexicography: A Preliminary Bibliography

Richard W. Bailey
Department of English
The University of Michigan
Ann Arbor, Michigan 48104
June 1974

Aitken, A.J. "Historical Dictionaries and the Computer," in The Computer in Literary and Linguistic Research, ed. R.A. Wisbey (Cambridge, Cambridge University Press, 1971), pp. 3-17.

Aitken, A.J. "The Literary Uses of Computers," Times Literary Supplement (21 April 1972):456.

Aitken, A.J., R.W. Bailey, and N. Hamilton-Smith, eds. The Computer and Literary Studies. Edinburgh: Edinburgh University Press, 1973.

Aitken, A.J., and Paul Bratley. "An Archive of Older Scottish Texts for Scanning by Computer," Studies in Scottish Literature 4(1966):45-47; reprinted in English Studies 48(1967):61-62.

Alford, M.H.T. "The Computer and Lexicography," Bulletin of the Association for Literary and Linguistic Computing 1,iii(1973):8-9.

Alinei, M.L. "Lexical, Grammatical and Statistical Indexing of Italian Texts with the Help of Punched Card Machines at the University of Utrecht," Levende Talen (1963):166-73.

Alinei, Mario. Spogli elettronici dell'Italiano delle origini e del duecento. Bológna: Societa Editrice il Mulino, 1971.

Allen, Sture. "Vocabulary Data Processing," in The Nordic Languages and Modern Linguistics, ed. H. Benediktsson (Reykjavik: Visindafelag Islendinga, 1970).

Allen, Sture. "Travaux du centre de recherches pour le suedois contemporain de Goteborg," in Zampolli, pp. 3-18.

Anderson, Jean Wood. Computers and Natural Language Processing, with Special Reference to Old Spanish Concordance and Dictionary Compilation. Unpublished M.S. thesis, U. of Essex, 1973. 134 pp.

Bailey, Richard W. "Research Dictionaries," American Speech 44(1969):166-72.

Bailey, Richard W. "Reflections on Technology in Lexicography," in McDavid & Duckert, pp. 293-97.

Bailey, Richard W. and Jay L. Robinson, "The University of Michigan Early Modern English Dictionary Project," Shakespeare Research and Opportunities 4(1968-69):120-21.

Bailey, Richard W. and Jay L. Robinson. "MEMEM: A New Approach to Lexicography," Source 4(1974):2-6.

Bailey, Richard W. and Jay L. Robinson. "Computers and Dictionaries," in Cameron, Frank, and Leyerle, pp. 94-102.

Bailey, Richard W. and Jay L. Robinson. "The Computer in Lexicography," in Lexicography and Linguistic Geography: Festgabe für Hans Kurath, eds. Harald Scholler and John Reidy (Wiesbaden: Franz Steiner Verlag, 1973), pp. 37-45.

Ben-Hayyim, Ze'ev. "A Hebrew Dictionary on Historical Principles," Ariel: A Review of the Arts and Sciences in Israel 13(1966):14-20.

Branford, William, et al. Towards a Dictionary of South African English on Historical Principles. Grahamstown: Institute for the Study of English in Africa, 1971.

Bromwich, J.I. "A Semasiological Dictionary of the English Language," in Aitken, Bailey, and Hamilton-Smith, pp. 15-24.

Brunner, Theodore F., ed. Thesaurus Linguae Graecae Newsletter. Irvine: University of California, 1973- .

Busa, Roberto, S.J. "Un lexique Latin electronique," in Stindlova, pp. 251-71.

Cameron, Angus, Roberta Frank, and John Leyerle, eds. Computers and Old English Concordances. Toronto: The University of Toronto Press, 1970.

Danielsson, Bror, ed. Studies in Early Modern English Newsletter (On proposals for a Dictionary of Early Modern English Pronunciation). Stockholm: Engelska Institutionen, 1968- .

Dimitrescu, Florica. "Projet d'un dictionnaire de la langue roumaine du XVIe siecle," in Zamplli, pp. 41-48.

Duro, Aldo. "Les nouvelles methodes du dictionnaire historique de la langue italienne," Cahiers de Lexicologie 8,i(1966):95-111.

Duro, Aldo. "Elaborations electroniques de textes effectuees par L'Academia della Crusca, pour la preparation du dictionnaire historique de la langue italienne," in Zampolli, pp. 53-75.

Duro, Aldo. "L'emploi des moyens electroniques pour la constitution du fichier lexicographique general par l'Academia della Crusca," in Stindlova, pp. 201-20.

Engels, Leopold Karel. "ITL: Instituut voor Toegepaste Linguistiek, Katholieke Universiteit van Leuven," in Zampolli, pp. 77-80.

Finkenstaedt, Thomas, Ernst Leisi, Dieter Wolff. A Chronological English Dictionary. Heidelberg, 1970.

Finkenstaedt, Thomas, and Dieter Wolff. Ordered Profusion: Studies in Dictionaries and the English Lexicon. Heidelberg, 1973.

Frank, Roberta, and Angus Cameron, eds. A Plan for the Dictionary of Old English. Toronto: The University of Toronto Press, 1973.

Frumkina, R.M. "Automatatizacija issledovatel;skich rabot v leksikologii i lekisikografii" (The Automation of Lexicological and Lexicographical Research), Voprosy Jazykoznanija 13,ii(1964):114-19/

Grimes, Joseph E. "Computing in Lexicography," Linguistic Reporter 12,v-vi(1970): 1-4.

Hellberg, Staffan. "Computerized Lemmatization without the Use of a Dictionary: A Case Study from Swedish Lexicology," Computers and the Humanities 6(1972): 209-212.

[Imbs, Paul.] Centre de Recherche pour un Tresor de la Langue Francaise. Paris: Bureau des relations exterieures et de l'information du CNRS, 1967. 47 pp.

Imbs, Paul, ed. _Tresor de la langue francaise: Dictionnaire de la langue du XIXe et du XXe siecles (1789-1960)_. I (A-Affiner). Paris: Editions du Centre National de la Recherche Scientifique, 1971.

Imbs, Paul, and Robert Martin, with Roland Vienney. _Dictionnaire des frequences: Vocabulaire litteraire des XIXe et XXe siecles_. Paris: Didier, 1971. 4 vols.

Josselson, Harry H. "Lexicography and the Computer," in _To Honour Roman Jakobson_ (The Hague: Mouton, 1967), vol. 2, pp. -046-59.

Josselson, Harry H. "The Matrix as a Concept for Structuring the Lexicon," in _Proceedings of the Conference on Linguistics, The University of Iowa, October 1970_, eds. Robert H. Oehmke and Robert S. Wachal (Iowa City, 1970), pp. 118-58.

Kasten, Lloyd. "Creation of an Old Spanish Dictionary through Computerized Techniques," (project proposal).

Klappenbach, R. "L'emploi des cartes perforees dans la dictionnaire de l'allemand contemporain," _Cahiers de Lexicologie_ 2(1960):87-97.

Klappenbach, R. _Wörterbuch der deutschen Gegenwartssprache: I (A-annahem)_. Berlin, 1961.

Lehmann, W.P. "On the Design of a Central Archive for Lexicography in English," in McDavid & Duckert, pp. 312-17.

Leyerle, John. "The Dictionary of Old English: A Progress Report," _Computers and the Humanities_ 5(1971):279-83.

Martin, Robert. "Travaux du centre de recherche pour un Tresor de la langue francaise (T.L.G. de Nancy)," in Zampolli, pp. 103-10.

Martin, Willy. _Analyse van een vocabularium met behulp van een computer_. Brussels: Aimav., 1970.

Mater, Erich. "Menschliches Denken und maschinelles Systematisieren in der Lexikographie," in _Probleme der Lexikographie_, ed. Kaspar Riemschneider (Berlin: Akademie der Wissenschaften zu Berlin, Institut für Orientsor- schung, Veröffentlichen, no. 73, 1970), pp. 23-29.

Mater, Erich. "A Methodological Problem of Mechanising Lexicography," in Stindlova, pp. 271-74.

McDavid, Raven I., Jr., and Audrey R. Duckert, eds. _Lexicography in English_. New York: Annals of the New York Academy of Sciences, vol. 211, 1973.

Neuhaus, H. Joachim. "Towards a Diachronic Analysis of Vocabulary," _Cahiers de Lexicologie_ 9(1971):113-26.

Neuhaus, H. Joachim. "Towards a Diachronic Analysis of Vocabulary," _Cahiers de Lexicologie_ 18(1971):29-42.

Neuhaus, H. Joachim. "Analysis and Retrieval of Conventional Dictionaries," in Zampolli, pp 115-19.

Nitti, John J. "A Description of the Initiation of Computer-Aided Procedures for the Old Spanish Dictionary Project at the University of Wisconsin-Madison," (mimeographed, 1974).

Olney, John, ed. _Newsletter of the Semantic Foundations Project_ (formerly SDC Lexicography Project). Santa Monica, 1972- .

Olney, John, and Donald Ramsey. "From Machine-Readable Dictionaries to a Lexicon Tester: Progress, Plans, and an Offer" (project proposal, 1971).

Olney, John, Carter Revard, and Paul Ziff. "Processor for Machine-Usable Version of Webster's at Systems Development Corporation," The Finite String 4,iii (1967):1-2.

Papp, Ferenc. "O plane masinnoi obrabotki etimologiceskogo slovarja vengerskogo jazyka" (On a Plan for the Mechanical Processing of the Etymological Dictionary of Hungarian), Sovetskoje Finno-ugrovedenie 4,iii(1968):207-14.

Quemada, Bernard. "La mechanisation des inventaires lexicologique," Cahiers de Lexicologie 11(1967):89-111; 12(1968):71-84.

Quemada, Bernard, ed. Cahiers de Lexicologie (Journal). Paris; Didier-Larousse, 1959- .

Quemada, Bernard. "Lexicology and Lexicography," in Current Trends in Linguistics, ed. Thomas A. Sebeok (The Hague. Mouton, 1972), vol. 9, pp. 395-476.

Quemada, Bernard. "Bilan des applications de l'informatique aux etudes lexicologiques," META 18,i-ii:87-102.

Quemada, Bernard. "Les travaux du centre d'etude du francais moderne et contemporain," in Zampolli, pp. 121-32.

Read, Allen Walker. "Approaches to Lexicography and Semantics," in Current Trends in Linguistics, ed. Thomas A. Sebeok (The Hague: Mouton, 1973), vol. 10, pt. 1, pp. 145-205. (See esp. "Technological Developments," pp. 182-87.)

Revard, Carter. "On the Computability of Certain Monsters in Noah's Ark," Computer Studies in the Humanities and Verbal Behavior 2(1969):82-90.

Robinson, Jay L., and Richard W. Bailey. "Computer-produced Microfilm in Lexicography: Toward a Dictionary of Early Modern English," in Aitken, Bailey, and Hamilton-Smith, pp. 3-14.

Sherman, Donald. "A Common Structure for Lexicographic Data," to appear in the Proceedings of the 1973 International Conference on Computers and the Humanities, ed. L. Mitchell (Edinburgh and Minneapolis).

Sherman, Donald. "A New Computer Format for Webster's Seventh New Collegiate Dictionary," Computers and the Humanities (forthcoming).

Sherman, Donald. "A Computer Version of the 'Dictionary of Early Modern English Pronunciation," (project proposal).

Smith, Raoul N. "Interactive Lexicon Updating," Computers and the Humanities 6(1972):137-45.

Spevack, Martin. "Shakespeare's English: The Core Vocabulary," Review of National Literatures 3,ii(1972):106-22.

Spevack, Martin. "Concordances: Old and New," Computer Studies in the Humanities and Verbal Behavior 4(1973):17-19.

Spevack, Martin, H. Joachim Neuhaus, and Thomas Finkenstaedt. "A Shakespeare Dictionary (SHAD)." Münster and Augsburg: mimeo, December 1973.

Stachowitz, Rolf A. "The Construction and Application of a Computerized Dictionary," unpublished paper presented to the Conference on Lexicography, Ohio State University, 1970.

Stickel, G., and M. Gräfe. "Automatische Textzerlegung und Herstellung von Zettelregistern für das Goethe-Wörterbuch," Sprache im technischen Zeitalter 19(1966):247-57.

Stindlova, Jitka, ed. Les Machines dans la Linguistique. Prague: Academia; and The Hague: Mouton, 1968.

Stindlova, Jitka. "Le dictionnaire de la langue tcheque litteraire et l'inscription de ses entrees et leur caracteristiques sur les cartes et la bande perforees pour les machines a traiter les informations," in Stindlova, pp. 235-43.

Streeter, Mary L. "DOC 1971: A Chinese Dialect Dictionary on Computer," Computers and the Humanities 6(1972):259-70.

de Tollenaere, F. Nieuwe Wegen in de Lexicologie. Amsterdam: North-Holland, 1963.

de Tollenaere, F., et al., eds. Informatie Nederlandse Lexikologie (Newsletter). Leiden, 1970- .

de Tollenaere, F. "La table ronde des dictionnaires historiques (Florence, 3-5 mai 1971)," Cahiers de Lexicologie 19(1971):116-28.

de Tollenaere, F. "Encoding Techniques in Dutch Historical Lexicography," Computers and the Humanities 6(1972):147-52.

de Tollenaere, F., and W. Pihnenburg. "Processing a Corpus of Early Middle Dutch Texts," Bulletin of the Association for Literary and Linguistic Computing 2,1(1974):16-23.

de Tollenaere, F. "The Problem of the Context in Computer-Aided Lexicography," in Aitken, Bailey, and Hamilton-Smith, pp. 25-35.

de Tollenaere, F. "Travaux de l'Institut de lexicologie neerlandaise," in Zampolli, pp. 29-39.

Urdang, Laurence. "The Systems Designs and Devices Used to Process and Random House Dictionary of the English Language," Computers and the Humanities 1(1966):31-33.

Urdang, Laurence. "Technological Potentials," in McDavid & Duckert, pp. 282-86.

Van Buren, P. "Preliminary Aspects of Mechanisation in Lexis," Cahiers de Lexicologie 11(1967):89-111; 12(1968):71-85.

Venezky, Richard L. "Storage, Retrieval, and Editing of Information for a Dictionary," American Documentation 19(1969):71-79.=

Venezky, Richard, Nathan Relles, and Jack Bennett. "Man-Machine Interaction: The Design of a Command Language for Non-programmers," Lexico Project Papers (Dept. of Computer Sciences, U. of Wisconsin, Madison (1(1974).

Venezky, Richard L. "Computational Aids to Dictionary Cimpilation," in Frank & Cameron, pp. 307-27.

Wahrig. G. <u>Neue Wege in der Wörterbucharbeit.</u> Hamburg, 1967.

White, Stephen V.F., ed. <u>Calculi</u> (Newsletter). Hanover, N.H.

Wang, William S.-Y. "Project DOC: Its Methodological Basis," <u>Journal of the American Oriental Society</u> 90,i(1970):57-66.

Wilks, Yorick Alexander. <u>Grammar, Meaning and the Machine Analysis of Language.</u> London: Routledge & Kegan Paul, 1972.

Wisbey, Roy A. "The Analysis of Middle High German Texts by Computer: Some Lexicographical Aspects," <u>Transactions of the Philological Society, 1963,</u> 28-48.

Wisbey, Roy A. "Ein computerlesbares Textarchiv des Frühmittelhochdeutschen," <u>Jahrbuch für internationale Germanistik</u> 1(1969):37-46.

Wölck, Wolfgang. "A Computerized Dictionary of Andean Languages," <u>Language Sciences</u> 8(1969):1-9.

Zampolli, Antonio, ed. <u>Linguistica matematica e calcolatori: Atti del convegno e della prima scuola internazionale, Pisa, 1970.</u> Firenze: Olschki (Accademia toscana di scienze e lettere, studi 28), 1973.

Zampolli, Antonio. "L'automatisation de la recherche lexicologique: etat actuel et tendences nouvelles," META 18,i-ii(1973):103-36.

Zampolli, Antonio. "La section linguistique du CNUCE," in Zampolli, pp. 133-99.

Zgusta, Ladislav. <u>Manual of Lexicography.</u> Prague: Academia, and The Hague: Mouton, 1971. (See esp. "The Use of Automatic Data-Processing Machines," pp. 353-57.)

CURRENT BIBLIOGRAPHY

Both the selection of material for this issue and the choice of subject categories are tentative. The opinions of readers will influence both in the future.

Completeness of coverage, especially for reports circulated privately, depends on the cooperation of authors. Summaries or articles to be summarized should be sent to the editorial office Twin Willows, Wanakah, New York 14075.

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

The Linguistic Documentation Centre of the University of Ottawa provides a substantial number of entries; AJCL gratefully acknowledges the assistance of Brian Harris and R. Laskowski.

See the following frame for a list of subject headings with frame numbers.

SUBJECT HEADINGS

THEORETICAL LINGUISTICS

A new journal

Contents of Volume 1, Number 1/2:

A. Kasher, Mood Implicatures: A Logical Way of Doing Generative Pragmatics

H. Lieb, Grammars as Theories: The Case for Axiomatic Grammar

S. Soames, Rule Orderings, Obligatory Transformations, and Derivational Constraints

D. Gabbay, J. M. E. Moravcsik, Branching Quantifiers, English, and Montague-Grammar

J. Hoepelman, Tense Logic and the Semantics of Russian Aspects

L. Karttunen, Presupposition and Linguistic Context

M. Dascal, A. Margalit, A New 'Revolution' in Linguistics? 'Text Grammar' vs. 'Sentence Grammar'

NATURAL LANGUAGE PROCESSING

Randall Rustin, Editor
Courant Institute of Mathematical Sciences
New York University

Algorithmics Press
New York
1973

TABLE OF CONTENTS

AUTOMATIC TEXT PROCESSING

(Avtomaticeskaya obrabotka tekstov)

Faculty of Mathematics and Philosophy

Charles University

Prague
1973

Table of Contents

E=English;   G=German;   R=Russian

*Texts and abstracts of papers presented at an East European
conference on computational linguistics*

Recent Romanian investigations in the field of Mathematical and computational linguistics

Solomon Marcus
Facultatea de Filologul
Bucharest

Automatic Text Processihg, 15 - 4̇2

(1)   A new concept of domination covers several earlier. theories (Dobrusin, Franclova, Kunze, and Tribulec) as special cases.

(2)   Marcus and Veronica du Feu are analyzing English distributional classes; examples are given.

(3)   Marcus and E. Celan are studying medical diagnosis as a linguistic problem, assuming strings over a vocabulary of symptoms.

(4)   Calude is using Marcus's distributional theory to study FORTRAN-IV and ASSEMBLER-360.

Consideration for a future analysis

György Szépe
Hungarian Academy of Sciences
Budapest

Automatic Text Processing, 43 - 44

Remark on the problem of language analysis in connection with machine-translation. The relevance for language analysis of the generative-semantics approach and Prague's multilevel functional approach; language variation; natural language message comprehension; disambiguation; semiotic aspects of language analysis.

LIST: LIBRARY AND INFORMATION SERVICES TODAY

An International Registry of Research and Innovation
Volume 4, 1974

Paul Wasserman, Editor

College of Library and Information Services

University of Maryland

Gale Research Company

Book Tower, Detroit, Michigan 48226

A survey of 1362 research projects involving about 3000 persons. Project descriptions are grouped by topic.

Table of Contents

COMPUTER EDUCATION FOR THE HUMANITIES: MULTIPLE POSSIBILITIES AT THE UNIVERSITY OF SOUTH CAROLINA

R. L. Oakman
Shared Educational Computer System, Inc.
Poughkeepsie, N.Y.

Mimeographed 1973

Two sequences of courses at the University of South Carolina offer liberal arts students an introduction to computers and to the relationship between technology and the humanities. These provide training in rigorous thinking and new means of probing the record of human experience. The lower level sequence consists of a general computer appreciation course and an introduction to non-numerical processes. In the latter, students learn about the fundamental concepts of computers. [EDRS: ED083818 $.65-MF, 3.29-HC]

SPECTRAL ANALYSIS OF SPEECH BY LINEAR PREDICTION

John Makhoul
Bolt Beranek and Newman, Inc.
Cambridge, Massachusetts

The autocorrelation method of linear prediction is formulated in the time, autocorrelation, and spectral domains. The analysis is shown to be that of approximating the short-time signal power spectrum by an all-pole spectrum. The method is compared with other methods of spectral analysis such as analysis-by-synthesis and cepstral smoothing. It is shown that this method can be regarded as another method of analysis-by-synthesis where a number of poles is specified, with the advantages of non-iterative computation and an error measure which leads to a better spectral envelope fit for an all-pole spectrum. Compared to spectral analysis by cepstral smoothing in conjunction with the chirp z transform (CZT) this method is expected to give a better spectral envelope fit (for an all-pole spectrum) and to be less sensitive to the effects of high pitch on the spectrum. The normalized minimum error is defined and its possible usefulness as a voicing detector is discussed.

A DESCRIPTIVE TECHNIQUE FOR AUTOMATIC SPEECH RECOGNITION

Renato De Mori
Centro di Eleborazione Numerale dei Segnali
Turin, Italy

The parameters used are the gravity centers of the zero-crossing interval distributions obtained at the output of two filters. Local aspect descriptions are qualitative descriptions of the stationary and the nonstationary segments of the speech waveform, and the values of the most important attributes (e.g., duration). They are composed into global aspect descriptions, which are analyzed by a set of acceptors, each one having to recognize just one word. A recognition rate of 98 percent for the ten spoken digits for four male speakers has been reached with an acceptable computation time.

Real-time Pitch Extraction by Adaptive Prediction
of the Speech Waveform

Joseph N. Maksym
Department of Electrical Engineering
Carleton University, Ottawa

With the exception of relatively sophisticated methods such
as cepstrum analysis, the problem of reliable pitch-period extrac-
tion has remained largely unsolved. This paper examines the feas-
ibility of pitch-period extraction by means of the nonstationary
error process resulting from adaptive-predictive quantization of
speech. A real-time hardware system that may be realized at low
cost is described.

Application of a Digital Inverse Filter
for Automatic Formant and F_0 Analysis

John D. Markel
Speech Communications Laboratory
Santa Barbara, California

A new algorithm based upon a digital inverse filter for-
mulation determines VU. (VU = 0 during unvoiced speech and VU = 1
during voiced speech), F_0, the fundamental frequency, and F_i,
i = 1,2,3, the first three formant frequencies, as a function of
time. Formant trajectory estimates are obtained for all speech
sounds that satisfy VU = 1. The central element in the analysis is
the digital inverse filter. Based upon the first M + 1 terms of
the input autocorrelation sequence, coefficients of an Mth degree,
all-zero digital filter are calculated. The formant trajectory es-
timates for each frame are based solely upon the locations of the
local minima of the corresponding spectrum of the resultant inverse
filter. The VU decision is determined by the amplitude of the
largest peak of the normalized autocorrelation sequence of the out-
put of the inverse filter (excluding the origin). If VU = 1, then
F_0 is defined as the reciprocal of the peak location.

DISCRETE-WORD RECOGNITION

UTILIZING A WORD DICTIONARY AND PHONOLOGICAL RULES

Shuichi Itahashi, Shozo Makino, and Ken'iti Kido
Research Institute of Electrical Communication
Tohoku University
Sendai, Japan

A discrete-word recognition system utilizing a word dic-
tionary and phonological rules is described. In this system,
nine distinctive features are extracted from a discrete-word
input. Segmentation is performed using these features. Segmen-
tation errors are corrected by applying a phoneme connecting rule.
The input word is transformed into an input feature matrix. The
comparison of this matrix with the standard derived from the dic-
tionary is performed in the feature (matrix) space. Another me-
thod of segmentation is also described in which segmentation is
performed using a duration dictionary. The effectiveness of
utilizing a word dictionary and phonological rules in automatic
discrete-word recognition is discussed.

A MODEL AND A SYSTEM FOR MACHINE RECOGNITION OF SPEECH

D. Raj Reddy, Lee D. Erman, and Richard B. Neely
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, Pennsylvania

This paper presents a model for machine recognition of
connected speech and the details of a specific implementation of
the model, the HEARSAY system. The model consists of a small set
of cooperating independent parallel processes that are capable of
helping in the decoding of a spoken utterance either individually
or collectively. The processes use the "hypothesize-and-test"
paradigm. The structure of HEARSAY is illustrated by considering
its operation in a particular task situation: voice-chess. The
task is to recognize a spoken move in a given board position.
Procedures for determination of parameters, segmentation, and
phonetic descriptions are outlined. The use of semantic, syn-
tactic, lexical, and phonological sources of knowledge in the
generation and verification of hypotheses is described. Prelim-
inary results of recognition of some utterances are given.

SPEECH PROCESSING WITH WALSH-HADAMARD TRANSFORMS

F. Ying Y. Shum, A. Ronald Elliott and W. Owen Brown
Department of Electrical Engineering   Bell-Northern Research Ltd.
McMaster University   Ottawa
Hamilton

        High-speed algorithms to compute the discrete Hadamard and
Walsh transforms of speech waveforms have been developed. Intel-
ligible speech has been reconstructed from dominant Hadamard or
Walsh coefficients on a medium sized computer in a non-real-time
mode. Degradation of some phonemes was noted at low bit rates of
reconstruction, but the reconstruction could be improved by vary-
ing the position of the sampling window. A digital processor,
which allows real-time analysis of speech to be conducted on the
system, is described.

SPEAKER VERIFICATION BY COMPUTER

USING SPEECH INTENSITY FOR TEMPORAL REGISTRATION

Robert C. Lummis
Bell Laboratories
Piscataway, N.J.

        Voice pitch, low-frequency intensity, and the three lowest
formant frequencies, all as functions of time, are used to repre-
sent an individual utterance. Verification consists of computing
these features for a test utterance and comparing them with stored
reference versions for the claimed identity. Before the test-versus-
reference comparison is effected, the time dimension of the test
utterance is warped to optimally register its intensity pattern onto
the reference intensity pattern. Performance of the system is mea-
sured on a speaker population of moderate size. A variety of com-
parison formulas and various subsets of the five speech features
are evaluated. The system responds either "accept" or "reject" to
every utterance; "no decision" is not allowed. Automatic verifica-
tion based solely upon voice pitch and intensity, both of which can
be computed rapidly, yields average error rates below 1 percent.

A VOCAL DATA MANAGEMENT SYSTEM

Jeffrey Barnett
System Development Corporation
Santa Monica, California

This paper describes an implementation strategy for a
vocal data management system (VDMS) being developed by the voice
input/output project at the System Development Corporation. VDMS
will accept connected speech of a language describable by 25-50
phrase equations and having a vocabulary of approximately 1000
words formed from about 100 data records.
The strategy is based on the concept of predictive lin-
guistic constraints (PLC). The present concepts of fixed direc-
tionality in parsing are replaced by a more generalized approach.
To facilitate this flexibility, the system comprises a set of
near-independent coroutines that are interconnected by a software
busing structure. The VDMS acoustic processors verify the pre-
dictions. Very loose matching criteria are used for locating the
predicted words. Special attention is given to word segments
that are experimentally determined to be most invariant.

SEGMENT CLASSIFICATION IN CONTINUOUS SPEECH

Kung-Pu Li, George W. Hughes, and Thomas B. Snow
School of Electrical Engineering
Purdue University
West Lafayette, Indiana

A primary-recognition computer program provides segmenta-
tion, acoustical parameters, and phonetic features of continuous
speech, together with classification of some vowel and consonant
segments. Based on fundamental frequency, level, and duration
information provided by the primary recognition program from short-
term spectra, a procedure to mark stressed and reduced vowels is
proposed. Listener judgments of stress and vowel reduction can
be correlated with the physical parameters, but talker differences
are apparent. It is clear that feature extraction at the segment
level and at the suprasegmental level are mutually interactive.

AN APPROACH TO SYNTACTIC RECOGNITION WITHOUT PHONEMICS

Wayne A. Lea
School of Electrical Engineering
Purdue University
Lafayette, Indiana

Linguistic and perceptual arguments suggest that syntactic hypothesis be formed before phonemic segments are identified. Prosodic features can provide some cues to constituent structure. A decrease in voice fundamental frequency (F_0) usually occurs at the end of each major syntactic constituent, and an increase near the beginning of the following. A program based on this regularity correctly detected over 80 percent of all syntactically predicted boundaries. Some minor boundaries were also detected by the fall-rise patterns in F_0. False boundary detections resulted from F_0 variations at boundaries between vowels and consonants, but most such false alarms could be eliminated by setting a minimum percent variation in F_0. Sentence boundaries were accompanied by large F_0 increases and substantial pauses. Categories affect boundary detection results; NP-V sequences are rarely detected.

RECOVERING PARENTHESES FROM SPOKEN ALGEBRAIC EXPRESSIONS

Michael H. O'Malley, Dean R. Kloker, and Benay Dara-Abrams
Phonetics Laboratory
University of Michigan
Ann Arbor

Subjects were very consistent in their placement of junctures when reading algebraic expressions slowly. Furthermore, there was an almost perfect correlation between measured silence and perceived juncture. Rules were developed for inserting parentheses based on the location and measured duration of silence intervals in an utterance. Listeners inserted parentheses and the consistency of their answers was measured by a chi-square test. Where listeners agreed, the rules agreed with the listeners from 91 to 95 percent of the time. Mathematically experienced and naive listeners performed similarly. The acoustic cues that indicate syntactic structure in this restricted domain may be more general.

Listener Performance in Speaker Verification Tasks

Aaron E. Rosenberg
Bell Laboratories
Murray Hill, N.J.

        The ability of listeners to perform some speaker verifi-
cation tasks has been measured experimentally and compared with
the performance of an automatic system for speaker verification.
A test presentation in the subjective experiments consists of a
pair of utterances. One of these is drawn from the recordings of
a group of speakers designated customers while the second utter-
ance is either a distinct recording from the same customer or the
recording of an impostor. Listeners must respond whether the
utterances are from the same or different speakers. The impostor
classes that have been considered are casual impostors making no
attempt to mimic customers, trained professional mimics, and an
identical twin of a customer. Listener performance is specified
by the two types of error that can be committed.

Design and Simulation of a Speech Analysis-Synthesis System Based on Short-Time Fourier Analysis

Ronald W. Schafer and Lawrence R. Rabiner
Bell Laboratories
Murray Hill, N.J.

        This paper discusses the theoretical basis for represen-
tation of a speech signal by its short-time Fourier transform.
The results of the theoretical studies were used to design a
speech analysis-synthesis system which was simulated on a general-
purpose laboratory digital computer system. The simulation uses
the fast Fourier transform in the analysis stage and specially de-
signed finite duration impulse response filters in the synthesis
stage. The results of both the theoretical and computational
studies lead to an understanding of the effect of several design
parameters and elucidate the design tradeoffs necessary to achieve
moderate information rate reductions.

Application of Sequential Decoding for Converting Phonetic to Graphic Representation, in Automatic Recognition of Continuous Speech (ARCS)

C. C. Tappert, N. Rex Dixon, and Arthur S. Rabinowitz
IBM Corporation
Research Triangle Park, N.C.

Following segmentation and phonetic classification in automatic recognition of continuous speech (ARCS), it is necessary to provide methods for linguistic decoding. In this work a graph search procedure, based on the Fano algorithm, is used to convert machine-contaminated phonetic descriptions of speaker performance into standard orthography. The information utilized by the decoder consists of a syntax, a lexicon containing transcription variation for each word, and performance-based statistics from acoustic analysis. The latter contain information related to automatic segmentation and classification accuracy and certainty (anchor-point) data. A distinction is made between speaker-and-machine-dependent corruption of phonetic input strings. Preliminary results are presented and discussed, together with some considerations for evaluation.

Evaluation of Various Parameter Sets in Spoken Digits Recognition

Akira Ichikawa, Yasuaki Nakano, and Kazuo Nakata
Central Research Laboratory
Hitachi Limited
Tokyo

A spectrum envelope, cepstrum, autocorrelation function, linear predictive coefficients, and partial autocorrelation coefficients (PACs) are evaluated experimentally for spoken digit recognition by simple pattern matching with nonlinear adjustment of the time axis. The spectrum envelope and cepstrum score 100 percent for ten spoken digits of a single-male speaker. PACs are easy to extract and have theoretical orthogonality but suffer from computation errors in fixed-point arithmetic with a short accumulator. Two effective means to reduce errors are variable use of the PAC dimensions controlled by computation accuracy, and smoothing along the time axis.

Computer recognition of the continuant phonemes in connected English speech

Russell J. Niederjohn   Ian B. Thomas
Department of Electrical Eng.   Department of Electrical Eng.
Marquette University   University of Massachusetts
Milwaukee   Amherst

A method of phoneme recognition of connected speech is described. Input to the system is assumed to consist of the 24 continuant phonemes in connected English speech. The system first categorizes each successive 20-ms segment of the input speech utterance as either voiced fricative, voiced nonfricative, unvoiced fricative or no-speech, utilizing a measure of the relative energy balance between low and high frequencies. Next the recognition of each 20-ms segment is performed from a distribution of axis-crossing intervals of speech prefiltered to emphasize each formant frequency range. Segmentation is performed from the results of the recognition of each 20-ms segment and from changes in categorization. Finally, the results of the recognition of each 20-ms segment between each pair of segmentation boundaries are combined and the phonemic sound occurring most frequently is printed out. The system has been trained for a single male speaker. Preliminary results for this speaker and for four 3-4-s sentences indicate: a correct categorization decision for about 97 percent of the input 20-ms segments, a correct recognition for about 78 percent of the input 20-ms segments, and an overall correct phoneme recognition for about 87 percent of the input phonemes.

Delta Modulation of Pitch, Formant, and Amplitude Signals for the Synthesis of Voiced Speech

Nuggehally S. Jayant
Bell Laboratories
Murray Hill, N.J.

A computer simulation of delta modulation (DM) as an alternative to pulse-code modulation (PCM) for a voiced-speech synthesizer. Quantized signals representing the time variations of pitch period, amplitude, and the first three formant frequencies all band limited to 16 Hz, were available in a 1500-b/s PCM format. Each was over sampled at 100 Hz for delta encoding, resulting in a representation at 500 b/s. Low-pass filtered versions of the DM signals were used to synthesize the utterance with a quality very close to the original 1500-b/s system Both "linear' and "adaptive" delta modulators were considered; in the latter case, the step size is adapted continuously to the changing slope statistics of an input signal. When additional band limiting was applied to the original control signals, resulting in a 700-b/s representation, adaptive DM at 250 b/s was sufficient to encode the information without further degradation.

SPEECH IN DIGITAL COMMUNICATION SYSTEMS

Ernst H. Rothauser
IBM Zurich Research Laboratory
Rüschlikon, Switzerland

The advent of integrated digital communication systems will
initiate a new period in speech communications research and devel-
opment. The new and still adaptable systems environment together
with the cost performance advantages of modern digital technolo-
gies will open up the way to practical utilization for many speech
technology concepts and services that are not feasible within the
framework of the present "classical" telephone networks.

Economy spelling computerized

Kenneth H. Ives
Director of Research and Statistics
United Charities of Chicago

Wi the growth v data processing, transmitting n storage systems, th need fr simplification v English spelling has become mor acute. Several approaches t ths. r presented. In combination they cn effect a saving v over 10% v letrz, yet b intelligible (whn read aloud) wiout retranslation. F adopted fr standard use, ths wd save $2 per person per year in printing costs, n mor in typing n ryting.

A system for generating Urdu - Farsi - Arabic script

S. S. Hyder
Department of Information
University of Montreal

The author has designed auziliary hardware for attachment to teletypes enabling them to be used for transmitting and receiving the Arabic alphabet with correct automatic selection of initial, medial and final shapes of letters.

Unique decipherability of codes with constraints with application to syllabification of Turkish words

G. Goenenc
Middle East Technical University
Ankara

Presentation of an algorithm for automatic.syllabification of printed Turkish words, based on the Even's test for unique decipherability and unique decipherability of finite delay but extended to codes with word-sequence constraints. Example printouts of automatically syllabified word strings included.

Establishing a German root system by computer

Howard H. Keller

Computers and the Humanities, 7, 4, 199 - 207, March 1973

Ways in which the computer can be used to generate a
German root system. The choice of corpus, data, storage system,
and presentation format all play an important role in obtaining
a finished product that is both useful as a reference work and
efficient as a starting point for further statistical research.
The uses of a computerized corpus in establishing a German-
English Root Dictionary (GERD) and in formulating a system of
entropy for measuring the efficiency of the German root system
as a subset of the Germanic semantic system.

Canadian bilingual legal vocabulary

(JURIVOC, Vocabulaire juridique bilingue Canadien)

D. Burke
Law Faculty
University of Ottawa

Mimeographed, 1974

An operational description of the JURIVOC automated English
and French bilingual legal lexicon production system under devel-
opment at the Law Faculty of the University of Ottawa. The lexi-
cographic analysis techniques and the analysts' operations are
briefly described.

MEMEM a new approach to lexicography

Richard W. Bailey and Jay L. Robinson
University of Michigan
Ann Arbor

Source, 2 - 6, April 1974

Charles C. Fries organized Early Modern English lexicog-
raphy at the University of Michigan in the 1920s. The first in-
stallment of Michigan Early Modern English Materials is a collec-
tion of 20,000 citations drawn from slips collected to illustrate
nine modal verbs: can, dare, gin, may, must, need, ought, shall,
and will. About forty words of context are included for each
slip. MEMEM is a collection of computer-generated microfiche and
a magnetic tape. A handbook, to be published in 1974, will des-
cribe the materials and include full bibliographical entries ex-
panding the abbreviated forms that appear on the fiche and tape.

FREQUENCY DICTIONARY OF MATHEMATICAL VOCABULARY
(Castotnyj slovar matematiceskoj leksiki)

E. T. Ter-Misakyants
Computing Center
Erevan State University

From a concordance of 140,000 entries taken from 120 books
published between 1930 and 1965. A list of 1277 words with fre-
quencies, from v 4480, i 3409, funktsiya 2210, čto 2207, etot
2058, dlya 1607, na 1379, tocka 1374, iz 1362, byt' 1328, s 1316,
pri 1293, esli 1249, my 1226, and to 1048, through 83 words with
frequency 10. Nearly 3500 other words occurréd. A list of 417
names in alphabetical order; the highest frequency is Koši 76.
Lists of foreign names, place names, abbreviations (t.e. 316).
Bibliography of sources. Analytic tables.

COMPUTER APPLICATIONS IN LEXICOGRAPHY

Richard L. Venezky
Department of Computer Sciences
University of Wisconsin
Madison

For modern English lexicography, the lack of programming
systems for automatic semantic classification and subsequent re-
duction in the volume of data is a serious problem. Further
qualitative advantages from automation depend on a revolution in
semantic analysis techniques and in the interaction between lexi-
cographer and computer. Computers are currently used for data
transfer and sorting; the process is limited by the editor's rate
of scanning. Even rough parsing could provide a rough filter.

What do standard transformational grammars produce?

Michael H. O'Malley
Phonetics Laboratory
University of Michigan
Ann Arbor

The base rules and transformations of a recently published elementary grammar of English are transcribed into a computer format and tested. Problems encountered in transcribing the grammar are explored and it is suggested that the emphasis on formal argumentation in linguistics is not justified by the current level of syntactic theory.

Surface-syntactic relations in English
(Poverxnostno-sintaksičeskie otnošenija anglijsko jazyke)

I. A. Mel'čuk and N. V. Percov
Russian Language Institute
Academy of Sciences of the USSR

Preliminary sketch of description of English surface syntax. A list of 32 surface-structure relations has been proposed for description of English surface-structure constructions. The goal of the description is automatic English text processing, especially automatic English-to-Russian translation. The description is based on analysis of English scientific and technical texts.

Some problems of the automatic analysis of English Prepositional constructions

Unsigned

Uses of in ,according to (a) the goverened expression, (b) the governor of the preposition, (c) words connected through the governor, (d) word order. Spatial, temporal, and modal meanings are common. Spatial: position, direction, path. Temporal: interval, duration, simultaneity. Modal: respect, purpose, cause, circumstance, means. Discussion of problems. Lists of noun classes; semantic features of nouns, verbs, and adjectives; rules.

COMPUTER-ASSISTED LINGUISTIC ANALYSIS

H. Cristoph Wolfart and Francis Pardo
University of Manitoba
Winnipeg

Description and users' manual for programs created for use in a study of Cree. Text processing, indexing, and concordance making. Application of rules of segmentation, including morpho-phonological alternations. The objectives include identification of new stems and affixes and testing a model of morphological analysis.

ON THE STRUCTURE OF CONTEXT-SENSITIVE GRAMMARS

Ronald V. Book
Center for Research in Computing Technology
Harvard University

Context-sensitive grammars generate languages which are not context-free, and it is undecidable whether a CSG generates a CF language. However, the mechanism by which the use of context allows a non-CF language to be generated is not well understood. A survey of some results which speak to two questions: (i) What constraints can be placed on the form of CSG rules without restricting weak generative capacity? (ii) What (nontrivial) constraints can be placed on the form of CSG rules such that only CF languages will be generated?

THE AUTOMATIC CONJUGATION OF DUTCH VERBS IN THE SIMPLE PRESENT

P. Van de Craen
Free University of Brussels

An algorithm for automatic derivation of 1st, 2nd, and 3rd person singular simple present forms based on recognition of a string of characters at the outset of a verb stem and a set of tables of verbs showing every kind of irregularity.

AUTOMATIC SYNTACTIC ANALYSIS
(Zur machinellen Syntaxanalyse)

I. MORPHOSYNTACTIC FOUNDATIONS
(Morphosyntaktische Voraussetzungen)

II. A LEXICON
(Ein Lexikon)

FOR AUTOMATIC LINGUISTIC ANALYSIS OF GERMAN
(für eine maschinelle Sprachanalyse des Deutschen)

Arbeitsgruppe MasA

Institut für deutsche Sprache

Tübinger Beiträge zur Linguistik, 1974

D-74 Tübingen   Postfach 256.7

I. Procedures for segmentation of sentences and attachment of morphosyntactic information to the segments. Flowcharts, program listings, examples.

    Problem description
    Grammatical meaning of a graphemic sequence
        Linguistic representation
        Formal representation
        Programming
    Grammatical meaning of a string of graphemic sequences
        Substantive and pronominal groups
        Governing verbal groups

II. Lists.

    Pronouns   Conjunctions   Verb prefixes
    Determiners   Adverbs   Interjections
    Prepositions   Pronominal adverbs   Helping verbs
    Partial list of verb stems

A COMPUTER ASSISTED LANGUAGE ANALYSIS SYSTEM

J. E. Rush, H. B. Pepinsky, B. C. Landry, N. M. Meara,
S. M. Strong, J. A. Valley, and C. E. Yong
Computer and Information Science Research Center
Ohio State University
Columbus

In the first pass, function words are looked up in a dictionary; rules using only function-word classes determine other class assignments. Phrase templates and a conjunction procedure give a grouping structure. Clauses are separated by occurrences of conjunctions, relative pronouns, etc. Case roles are assigned within each clause according to verb type, etc. The builders of the system were motivated by the problem of conversational analysis (sociological); they discuss the terms of this problem as a preface to their technical description of the parser.

AUTOMATIC SUFFIX ANALYSIS

(Automatische Suffixanalyse)

M. Pfeifer
Institute for Informatics
University of Stuttgart

Standard methods; mathematical formulation; lists of English suffixes from several sources; criteria for evaluation of such lists; experimental application of three lists to a collection of words. Algorithm for construction of a suffix list from a reverse dictionary. User's manual for a suffix-stripping program.

AN INTERACTIVE SYNTACTIC ANALYZER FOR MAN-MACHINE COMMUNICATION

(Un analyseur syntaxique interactif pour la communication homme-machine)

J. Courtin
Laboratoire d'Informatique
Université de Grenoble

Brief description of algorithms for morphological and syntactic analysis of natural language (French) texts. The morphological algorithm is based on finite-state grammar, the syntactic one on dependency grammar, with a context-free grammar as a device for structural verification of output substrings.

COMPUTER DETERMINATION OF THE STRUCTURE OF THE SIMPLE SENTENCE

(Issledovanie struktury prostogo predloženiya s pomoščyu EVM)

B. V. Suxotin
Russian Language Institute
Academy of Sciences of the USSR

The rules of grammar are followed more or less precisely in different occasions; a parser can allow for this variability. Exact tests are applied where reliable; probability functions are used in regions of uncertainty. Syntactic relations between words; probability interpretation; the role of syntactic classes; optimization of the procedure; an experiment on Pushkin; the search for the best tree; three algorithms in Algol 60; tables of conditional probability of occurrence of syntactic classes; analyses of 200 sentences.

SEGMENTATION OF FRENCH SENTENCES

B. Maegard and E. Spang-Hanssen

An informative description of a program for automatic segmentation of French compound sentences into constituent clauses based on recognition of finite-verb forms, conjunctions, prepositions, and punctuation marks.

CAUSALITY AND REASONING

Roger C. Schank
Instituto per gli Studi Semantici e Cognitivi
Castagnola, Switzerland

An attempt at a system for computer analysis of causal re-
lations in a natural language text. The system is based on the
language-independent Conceptual Dependency Theory. The conceptual
causal syntax and semantic interpretation of causal relations are
presented. This study is part of a general attack on the problem
of making sense of utterances.

CONSTRUCTIBLE REPRESENTATIONS FOR TWO SEMANTIC RELATIONS

F. G. Pagan
Computer Centre
University of Aston
Birmingham, England

Deals with the problem of representing semantic relations
among words for purposes of natural language processing. The cri-
terion of constructibility of such representations is introduced,
and the constructible representations for hyponymy and semantic
compatibility are elaborated.

SEMANTIC REPRESENTATION OF THE SENTENCE AND ITS RELATION TO LOGICAL REPRESENTATION

(Semantičeskaja zapi's predloženija i ee otnošenija k logičeskoj zapisi)

Petr Sgall
Charles University
Prague

Automatic Text Processing, 7-10

      The semantic representation of a sentence should correspond to the structure of predicate calculus formulas as well as to the deep syntactic structure of a natural language sentence. It has to reflect the syntactic dependency structure of a sentence and the theme-rheme (topic-focus) relation within it. The possibility of automatic translation of such semantic formulas into predicate calculus formulas has been proved at the Charles University in Prague.

THE SEMANTIC REPRESENTATION OF NEGATION IN THE FUNCTIONAL GENERATIVE DESCRIPTION

E. Hajicová
Charles University
Prague

Automatic Text Processing, 11-13

      There is a direct relationship between the topic-focus division of a sentence and the scope of the negation operator. In connection with negation, the problem of the meaning of a sentence, its presuppositions and its allegations is discussed. "A is an allegation of S if S entails A and non-S entails neither A nor non-A", where S is a sentence and A is one of the components of its content.

Set-theoretical semantics for elementary mathematical language

D. Rawson
California Institute for Mathematical Studies in Social Science
Stanford University

The development of computer language and analogs capable of interpreting and processing natural language found in elementary mathematics is discussed. Working with linguistic theories in combination with the special characteristics of elementary mathematics, the author has developed algorithms for the computer to accomplish the above task.

GRAMMAR, MEANING AND THE MACHINE ANALYSIS OF LANGUAGE
Yorick A. Wilks

Routledge & Kegan Paul, London, 1972

Reviewed by Stuart C. Shapiro
Computer Science Department
Indiana University, Bloomington

The author states immediately that "the main purpose of this book is not to survey research efforts, nor is it to give a conceptual analysis of the words in the title, but to describe a system of semantic analysis." [p1] The reader does well to remember this, because, although there are 314 entries in the bibliography, and the first chapter on the system itself begins on page 92 of 172 pages of text, no comprehensive survey is given of the relevant literature. Indeed, such a survey would be a massive book in itself since it would include much work from computer science, linguistics, psychology, philosophy and logic. The literature that is discussed is presented only in support of the author's argument that the aim of language analysis is, and always has been, not the judgement of the degree of grammaticality but the explication of meaning. The bibliography and literature discussion also remind the reader of the slow pace of publishing. The book was published in 1972 and reports on work substantially completed by 1968 [5]. Except for one paper by Wilks himself, no bibliography item is more recent than 1969. The meaning in the text of "recent" must be understood in light of this fact.

Wilks makes one mistake in his background discussion·that deserves comment since it leads him to reject a possible approach for a false reason. He, correctly, characterizes the Artificial Intelligence (AI) approach to the "problem of 'meaning and the machine'" [p. 3] as an attempt to get machines to "understand" languages, but he rejects this approach because he misunderstands one of the basic paradigms of AI-- the Turing test [3]. Wilks misstates the Turing test [p. 5,6] in such a way that Weizenbaum's ELIZA program [4] passed the misstated test [2]. Wilks states that ELIZA "hás passed Turing's test, and has done it by such simple minded procedures as to devalue the notion of 'understanding by machines' as an approach to the problem of 'meaning and the machine'...Weizenbaum's work has, from within AI itself, produced a disappointing answer for those who hoped that 'machine understanding' would provide a solution to the MT problem." [p. 6,7] Of course ELIZA did not pass Turing's test, and is accepted in the AI community as a demonstration of how much can be done by a program without understanding

The semantic system Wilks describes, Computable Semantic Derivations (CSD), is designed solely to disambiguate word-senses. This is, for Wilks, the major task in language analysis since he sees the main aim of language analysis as explication of meaning and, "to be meaningful is to have one and only one of a number of possible interpretations...with respect to some dictionary." [pp. 23,30] For Wilks, a sentence like "Colorless green ideas sleep furiously" is meaningless not because it has no reading, but because it has several. It is possible to embed the sentence in a context that will give it a unique reading--therefore, a meaning, and it is possible to do this in several ways.

CSD does not produce a parse tree of the sentences it analyzes, although it does a small amount of what might be considered syntactic analysis. It also does not produce a representation of the "meaning" of the sentences, for example, of the sort adequate for a question-answering system. What it does is to assign word senses to the content words of the text and, by use of various combination rules, attempt to assign a single word sense to each word occurrence in the text. A word sense is represented by a "semantic formula" and a "sense description." A semantic formula is a list structure whose atomic elements are selected from a set of 53 primitive semantic classifiers. The sense description is a list of English words providing an informal description of the word sense. The semantic classifiers correspond to Katz and Fodor's semantic markers [1] and the sense descriptions to KF's distinguishers. The semantic formulae are, however, more complex than KF's simple lists and Wilk's combination rules are more complex than KF's projection rules, and apply across sentence boundaries. Wilks's system recommends itself by being embodied in a computer program that runs and produces readings of paragraphs of text. It is only when a theory is embodied in a running program that one can be sure that it is well defined and can discover what it predicts for any case.

Another strong feature of CSD is its Expand facility designed to model "our undoubted ability to recognize and understand words being used in a new, or possibly metaphorical, sense, as in Dylan Thomas's 'A grief ago'" [p. 54]. When a reading cannot be produced for a paragraph, Expand constructs a new sense for some word and attempts to resolve the paragraph using this new word sense. However, the new sense is always a sense of some other word in the paragraph [p. 167].

The book is not self-contained. Details of CSD which are necessary for a full understanding are contained not in the book but in the appendices of the earlier technical report [5] to which the reader is frequently referred. The book was also not proofread well. For example, page 83 ends at the end of a paragraph and page 84 begins in the middle of a sentence.

The CSD approach is presently being applied to machine translation [6;7;8]. It is radically different from other current approaches rooted in linguistics and logic. Its adequacy as a model awaits psychological testing. Its adequacy as a program has been demonstrated but so far only on a few small examples.

References

1. Katz, Jerrold, J., and Fodor, Jerry A. The structure of a semantic
      theory. Language 39 (April-June, 1963), pp. 170-210. Also in
      Fodor, Jerry A., and Katz, Jerrold J., eds. The Structure of
      Language, Prentice-Hall, Englewood Cliffs, New Jersey, 1964, pp.
      479-518.

2. Special Interest Group on Artificial Intelligence of the ACM. ELIZA
      passes Turing test. SIGART Newsletter, no. 10 (June, 1968), pp.
      15-16.

3. Turing, A.M. Computing machinery and intelligence. Mind 59 (Octo-
      ber, 1950), pp. 433-460. Also in Feigenbaum, E.A., and Feldman,
      J., eds. Computers and Thought, McGraw-Hill, New York, 1963,
      pp. 11-35.

4. Weizenbaum, J. ELIZA--a computer program for the study of natural
      language communications between man and machine. Comm. ACM 9, 1
      (January, 1966), pp. 36-45.

5. Wilks, Yorick. Computable semantic derivations. SP-3017, System
      Development Corporation, Santa Monica, California, January, 1968.

6. _____. The stanford machine translation project. Natural Language
      Processing, Rustin Randall, ed. Algorithmic Press, New York,
      1973, pp. 243-290.

7. _____. Understanding without proofs. Proc. Third Int. Joint Con-
      ference on Artificial Intelligence, Stanford, California, August,
      1973, pp. 270-277.

8. _____. An artificial intelligence approach to machine translation
      Computer Models of Thought and Language, Schank, Roger C., and
      Colby, Kenneth Mark, eds. W.H. Freeman and Co., San Francisco,
      1973, pp. 114-151.

COMPUTER MODELS OF THOUGHT AND LANGUAGE

Roger C. Schank and Kenneth Mark Colby, Editors

W. H. Freeman and Company
San Francisco
1973

TABLE OF CONTENTS

SEMANTIC ANALYSIS OF ENGLISH TEXT BY COMPUTER

Anne-Louise Guichard Radimsky

Doctoral dissertation, 1973
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

SPEC is a SNOBOL IV program designed after W. L. Chafe's theory of syntax and semantics. A preprocessor detects misspellings, idioms, and abbreviations and creates a lexical table. A parser has a top-to-bottom CF component using a deep-structure grammar; it also has a transformational component to eliminate semantic elements superimposed on a kernel sentence. A semantic component of the parser checks whether given structures are meaningful and directs the selection of possibly applicable syntactic rules. In a third phase, anaphoric and linkage components recognize coreference and insert the new sentence in an already constructed structure.

AN ALGORITHM FOR GENERATING STRUCTURAL SURROGATES OF ENGLISH TEXT

S. M. Strong
Department of Computer and Information Sciences
Ohio State University
Columbus

An algorithm which generates non-linear representations of English text using the results of a syntactic analysis system and a set of rules which prescribe linkages to generate a graph of a sentence. The shape of these graphs corresponds to the syntax of the sentence; the labels correspond to the vocabulary of the sentence and the edge types correspond to case grammar roles. The sentence graphs can then be interconnected at common nodes and analyzed according to common edges. Preliminary experimentation has yielded promising results. It appears that the algorithm produces a representation of English text which could be quite useful in automatic language processing.

An information-structure based method for extracting meaning from input sentences

S. Ohsuga
Institute of Space & Aeronautical Science
University of Tokyo

Report of research on a natural language based information structure for storing semantic information and recognizing the logical relations between two statements. It comprises a system for transforming a simple sentence into a standard form and for deducinc from the latter an answer to a question.

Expectation as a basic mechanism of language comprehension

C. K. Riesbeck
Computer Science Department
Stanford University

A model of human language comprehension and a description of its implementation. The basic mechanism assumed is expectation. All processing is done in terms of expected situation and of actions to perform if these situations arise. The program uses this mechanism to build structures that represent the meanings of English sentences.

Topic analysis

Brian Phillips
State University of New York
Buffalo

Description of a computer system for analysis of the topic organization of discourse based on a network model of cognitive knowledge. The system works in four steps: parsing, normalization, topic analysis, output. The normalization processes, which are the main concern of the article, establish potential referents for proforms, link the discourse to the cognitive network (encyclopedia), and introduce general information omitted from discourse.

JUNCTION GRAMMAR AND MACHINE ASSISTED TRANSLATION

A. K. Melby
Brigham Young University

Junction Grammar is a generative grammar which recognizes 4 levels: the real world (object) level, conceptual level consisting of sememes interrelated by junctions, lexical level and surface level. The project at Brigham Young University performs analysis of surface level expressions into conceptual level, adjustments at this level for incompatabilities between languages, and synthesis to surface level.

AUTOMATIC INFERENCE OF SEMANTIC DEEP STRUCTURE RULES IN GENERATIVE SEMANTIC GRAMMAR

S. Klein
Computer Science Department
University of Wisconsin
Madison

A brief explication of a system for the automatic inference of semantic deep structure rules in generative semantic grammars. A representation of deep structure is given in the notation of a 4-dimensional network with properties of at least 2nd-order predicate calculus, and also in the notation of a compiler-driven behavioral simulation language that describes and modifies the linguistic conceptual universe of the speaker. The system is able to make grammatical-semantic inference within the frameworks of all current generative semantic linguistic models.

A MODEL FOR SYNTHESIS OF CZECH
(Model' sinteza ce̊skogo jazyka)

J. Panevová
Charles University
Prague

This model for Czech texts is based on the multilevel functional approach developed at the Charles University in Prague. The model consists of a set of puch-down automata which in several successive steps translate (transduce) a semantic representation of a sentence into its graphemic representation. An example of the generation of a sentence is included.

THE Q* ALGORITHM - A SEARCH STRATEGY FOR A DEDUCTIVE QUESTION-ANSWERING SYSTEM

Jack Minker, James R. McSkimin and Daniel H. Fishman
Department of Computer Science
and Computer Science Center
University of Maryland
College Park

Department of Computer Science
University of Massachusetts
Amherst

An approach for bringing semantic, as well as syntactic, information to bear on the problem of theorem-proving search for Question-Answering (QA) Systems is described. The approach is embodied in a search algorithm, the Q* search algorithm, developed to control deductive searches in an experimental system. The Q* algorithm is part of a system which incorporates both the Q* algorithm, which performs the search required to answer a query, and an inferential component, which performs the logical manipulations necessary to deduce a clause from one or two other clauses. The inferential component includes many refinements of resolution. The Q* algorithm generates nodes in the search space, applying semantic and syntatic information to direct the search. Use of semantics permits paths to be terminated and fruitful paths to be explored.

BREADTH-FIRST SEARCH: SOME SURPRISING RESULTS

L. Siklóssy, V. Marinov   and   A. Rich
Computer Sciences Department   Mathematics Department
The University of Texas   The University of Texas
Austin   Austin

Although breadth-first search procedures cannot explore truly large search spaces, actual implementations of such procedures can result in surprisingly powerful problem-solvers that outperform more sophisticated heuristic search procedures. We describe two breadth-first search procedures. The first one, S&R, proves theorems from Principia of Whitehead and Russell, and is compared to two versions of the Logic Theorist. Previous estimates of the size of the search space are significantly reduced. When theorems are proved in an optimal order, this order differs markedly from that found in Principia, while more general theorems than those of Principia are often found. The second system, S&M, adapts breadth-first search to locally infinite search spaces in systems of rewriting rules. S&M is compared extensively to the heuristic theorem-prover of Quinlan and Hunt, and to some other theorem provers.

TOWARDS MORE INTELLIGENT TEACHING SYSTEMS

J. R. Hartley and D. H. Sleeman
Research Councils' Computer Based Learning Project
The University of Leeds, U.K.

This paper suggests criteria against which the "intelligence" of a teaching machine can be judged. With the electronic computer in mind, distinctions are made between pre-structured, generative, adaptive, and self-improving teaching systems. From the work which has been carried out at Leeds, examples are taken which illustrate the characteristics and intelligence of these systems and the requirements for their implementation.

DESIGN AND TESTS OF COMPUTER-FURTHERED INSTRUCTION

Lawrence A. Sherr, Francis D. Tuggle and H. Gordon Fitch
University of Kansas
Lawrence

The philosophy of computer-furthered instruction (CFI), the design of a CFI system used in an undergraduate course in probability theory and statistics in a school of business, and performance and process tests of the effects of introducing the CFI system. CFI systems do not replace traditional methods used in teaching (e.g., lectures), but take advantage of real time computing to allow the students to see and perform assignments otherwise impossible. Contrasts between CFI and CAI systems, the course in its standard format, traditional problems in teaching this course, the programs used. Sample CFI conversations. Statistical tests of performance and process changes induced by the CFI system; their significance and implications.

A COMPUTER-AIDED LEARNING SYSTEM

Jean Lavoie
Département de Génie Electrique
Ecole Polytechnique
Montréal

        In the framework of an educational experiment in optimiz-
ing teaching by the audio-tutorial method, which implies frequent
and non-periodic evaluation of each student's progress, it is
necessary to have a computer system to realize three main objec-
tives: (a) easy preparation of multiple questionnaires in the
form of examinations or quizzes; (b) automatic grading of student
papers; (c) updating a file of each student's progress. The aims
of this paper are threefold: (1) to describe the techniques used
to fulfil the objectives of (a) above; (2) to point out the ad-
vantages and the difficulties observed while using this technique
of evaluation for a pilot group of students (about 30 tests given
to 300 students during one semester in a statics course); (3) to
discuss possible extensions to the present system, mainly concer-
ning the implementation of objectives (b) and (c) above.

A THEORY OF CONVERSATIONS AND INDIVIDUALS (EXEMPLIFIED BY THE LEARNING PROCESS ON CASTE)

G. Pask, B. C. E. Scott and D. Kallikourdis
System Research Ltd.
Richmond, Surrey, U.K.

        The minimal experimental situation for psychological ob-
servation is a conversation. The logical and structural require-
ments for making such observations are presented in icons which
dynamically represent formalisms in the abstract theory of self-
reproducing automata. Two sorts of stable, self-reproducing sys-
tems are distinguished: mechanically and psychologically charac-
terized individuals. (M- and P- Individuals). A conversation is
a P-Individual that is executed in one or more of a restricted
class of M-Individuals (processors). The theory is exemplified
by work on learning and teaching. Other exemplifications are
given us as interpretations of the paradigms extant in convention-
al experimental psychology.

TRANSFER FROM ALTERNATIVE PRESENTATIONS OF
SPELLING PATTERNS IN INITIAL READING

J. D. Fletcher
California Institute for Mathematical Studies in Social Science
Stanford University

Mimeographed, 1973

      The investigation was concerned with the class of grapheme-phoneme correspondence called spelling patterns and was intended to broaden empirical understanding of grapheme-phoneme correspondences used in teaching initial reading. 25 boys and 47 girls from three moderate-ability first grade classes were chosen as subjects. These students had accumulated six minutes of session time on the computer-assisted reading program (CAI) phonic strand before beginning the experiment. [ERIC ED083634]

GENERATION OF PUNCTUATION AND USAGE EXERCISES
IN FRESHMAN ENGLISH USING A SENTENCE POOL

M. M. Freed
Computer-Assisted Instruction Laboratory
University of Texas

Mimeographed, 1970

      The purpose of PUNCT2-CW is to test the feasibility and practicality of a data base system of computer-assisted instruction in English. To save time and to provide comparison, this course was based on the objectives, format and logic of an earlier course, PUNCT. The present data base is a sentence pool composed of certain sentence patterns called prototypes, each with a unique call number. A program author can call a type of sentence by using the prototype name. [EDRS ED084880; $.65-MF, 3.29-HC]

Structuring and Retrieving Information in Computer-Based Learning

Klaus Brunnstein and Joachim W. Schmidt
Institute of Informatics
Hamburg University

An attempt to use special techniques to introduce information systems into learner-controlled instruction on a broader basis than encountered today. To do this, "information networks" are defined to add to the information stored in the data base some kind of semantic understanding of the logical structures involved. This approach differs from others, in that the student (or user) works himself with the information network, while other approaches try to guide the learning process via conversational programs which are generated according to a given network.

Teaching Speech to the Deaf: Can a Computer Help?

Raymond S. Nickerson and Kenneth N. Stevens
Bolt Beranek and Newman Inc.
Cambridge, Massachusetts

An attempt to develop a computer-based system of speech-training aids for the deaf. Some of the problems associated with the speech of the deaf are briefly reviewed. Reasons for attempting to apply a digital computer to the problem of speech training are given. The system and its display capabilities are described. The importance of evolving such a system through a close interaction of developers and users is stressed.

REL -- English bulk data input

R. H. Bigelow
California Institute of Technology

Mimeographed, 1973

A bulk data input processor which is available for the rapidly extensible language (REL) English versions is described. In REL English versions, statements that declare names of data items and their interrelationships normally are lines from a terminal or cards in a batch input stream. These statements provide a convenient means of declaring some names and stating some facts, and they are especially useful in the interactive mode. [ERIC ED083015]

REL -- An information system for a dynamic environment

Bozena H. Dostert
California Institute of Technology

Mimeographed, 1973

Three outstanding features of the rapidly extensible language (REL) system are discussed. REL itself is an integrated software system designed to facilitate conversational interaction with the computer, especially by those working with dynamic, highly interrelated data. [ERIC ED083810]

The REL command language

S. Gomberg
California Institute of Technology

Mimeographed, 1973

Three basic notions -- version, session and command language -- are discussed which are essential for a user of the rapidly extensible language (REL) system. The first is defined as a language - data base package consisting of a language, data and definitions. A session is stipulated to be the period during which the user operates at a terminal, and the command language is the means by which he communicates with and manipulates other versions. Details are provided about the commands used. [ERIC ED083814]

STRUCTURES FOR AN INTERACTIVE ON-LINE THESAURUS

Jon Rickman   W. E. Walden
Department of Computer Science   Department of Computer Science
Pennsylvania State University   Washington State University
University Park   Pullman

A data structure for plant pathology thesaurus is imple-
mented as a continuous string of characters. A hashing function
is used to access terms and their related terms. A prefix hash-
ing schema is described which permits thesaurus file access with
partial spellings and defines file blocking to improve sequential
(alphabetic) scans. Sequential searches within blocks permit
full word matching when desired. A general system and file struc-
ture is proposed with display formats and a user command language.

THE COMPARATIVE EFFICIENCY OF TWO DICTIONARY STRUCTURES FOR DOCUMENT RETRIEVAL

P. Ein-Dor
School of Business Administration
Tel Aviv University

Inverted lists and tree structures are compared in terms
of their efficiency as indexes for document retrieval. The rel-
ative efficiency of search is shown to depend on the structure of
the information file, the form of queries posed to it and the in-
teraction between them. Formulae are developed for computing the
amount of search in each case and examples are provided.

On the Problems of Time, Retrieval of Temporal Relations, Causality, and Coexistence

Nicholas V. Findler and David Chen
Department of Computer Science
State University of New York
Buffalo

Intelligent question-answering programs do more than re-
trieve "raw" data; they make deductive inferences in order to
return all valid responses. They report logical inconsistencies,
possibly at the data input phase. Similarly, more information is
requested from the user if a question asked proves to be ambigu-
ous. A question-answering system of the above type has been de-
signed and implemented. Besides retrieving explicit and implicit
temporal relations, the system discovers potentially causal rela-
tionships which also satisfy different time restrictions. Ques-
tions concerning a generalized concept of coexistence can also be
answered. It is hoped that programs of a similar nature will be-
come of much pragmatic use to researchers in physics, chemistry,
biology, etc., in evaluating complex, interrelated experimental
data. Several additional applications for this type of program
are mentioned, ranging from problems in criminology to air traf-
fic control. The Associative Memory, Parallel Processing Lan-
guage, AMPPL-II, was found rather satisfactory for the project.
It is finally suggested that the system being described could
serve as a component in a complex cognitive mechanism.

A Method of Constructing Information Languages Having Grammar

G. E. Vleduts and N. A. Stokolova

(Translated from Russian by J. Lineweaver, Iowa State University)

Presents the main features of a method, used for construc-
ting information languages with 'grammar' for some fields of
natural sciences. As tools for recording syntagmatic relations...
'standard phrases' are used; these are multiplace predicates of a
special kind describing typical situations corresponding to cate-
gories of facts which are significant for the given subject area.
Statements obtained from some types of 'standard phrases' are
used as 'composite terms' in some other 'standard phrases', this
leading to a many-leveled hierarchy of 'standard phrases'. Exam-
ples from chemistry, biology and geography are given.

THE USE OF SELECTED PORTIONS OF TECHNICAL DOCUMENTS AS SOURCES OF INDEX TERMS AND EFFECT ON INPUT COSTS AND RETRIEVAL EFFECTIVENESS

H. H. Schumacher, J. F. March and F. L. Scheffler
Dayton University
Ohio

Mimeographed 1973

Looks at the effect of using various portions of technical reports as sources of indexing terms on input costs and retrieval effectiveness. Results of an experiment showed the body of the document could be excluded as a source of index terms. This reduced unit cost from $10 to $8.25. [NTIS Order XC-761-808;$4.85]

LINGUISTIC ASPECTS OF AUTOMATIC ANALYSIS OF ABSTRACTS
(Lingvisti̮ceskie aspekty avtomati̮ceskoj obrabotki annotacij)

E. Bene̮sova
Laborato̮r algebraické lingvistiky
Charles University
Prague

Automatic Text Processing, 79-85

Brief information on the system under development at Charles University, Prague, for automatic analysis of abstracts. The system, based on analysis of syntactic and semantic relations in a text, distinguishes two classes of text elements: (1) relational words and their semantic representations, (2) terms which can be used as arguments of the relational words. The need for standardization of the language of abstracts is stressed.

Automatic Indexing and Generation of Classification Systems by Algorithm

W. G. Hoyle
National Research Council of Canada
Ottawa

Brief description of system, based on Baye's theorem, for automatic indexing and document classification. For 124 documents assigned to 9 different categories, the system indexed 97 documents in agreement with human indexers. Classification into the 9 categories also agreed half of the time.

Computer System Support for Data Analysis

N. R. Greenfeld
California Institute of Technology

Data analysis and supporting computer software systems are studied. The idea of data analysis as an experimental science involving data-gathering and theory building is developed. Language's role as a determinant of meaningful theory and the informativeness of a language and data base pair are studied, along with the static and dynamic aspects of data analysis. Computer systems useful for data analysis are surveyed, with special attention paid to language restrictions. [EDRS ED083811 $0.65-MF, 6.58 HC]

Automatic Error Correction in Natural Text, Supplement

A. J. Szanser
Division of Computer Science
National Physical Laboratory
Teddington, England

This supplement to "Final Report on Automatic Error Correction" is a summary of the procedure for automatic short-word and abbreviation error-correction, correction of disjoined words and of fused words. The experimental results demonstrate the efficiency of the procedure chosen.

Procedural guide for the translation of foreign-language thesauri into German

Gerd Beling
Forschungsinstitut
für Funk und
Mathematik
Wachtberg-Werthoven

Hans Jochen Schuck
Bundessprachenamt
Hürth

Gernot Wersig
Institut für medizin-
ische Statistik und
Dokumentation der
Freien Universität
Berlin

The establishment of a multilingual thesaurus is one way of obtaining a documentation language suitable for use in several natural languages. Guidelines are needed most when the thesaurus is part of a documentation system. This Guide discusses terminological control, semantic factoring, hierarchies, transliteration, homography, synonymy, and indexing. It also deals with administration and step-by-step control of operations.

Structural approach to the linguistics and design of a hybrid information-retrieval language

(Strukturnyj podxod k Iingvistike i sozdanie gibridnogo inform-acionnogo jazyka)

E. Seljan

A project in Bulgaria to design an information retrieval language for Bulgarian texts. The language consists of the set of indexes characterizing the field of a document, and the vocabulary of extended roots of Bulgarian words. Some questions of synonymy and multi-word terms are discussed.

GROUPE DE RECHERCHES POUR LA TRADUCTION AUTOMATIQUE

Université de Montréal

PUBLICATIONS

Progress reports. January, April, July, October 1966; January, April, July, October, 1967; January, April, October 1968; January, April 1969; January 1971; May, August 1973; March 1974.

Colmerauer, A., Kittredge, R., and Stewart, G. Résultats préliminaires. January 1970.

De Chastellier, Guy. Description d'un programme de maintenance des fichiers de texte. October 1967.

De Chastellier, Guy. Edition des textes: Description du programme. April 1968.

Dugas, André. Composition partielle d'un dictionnaire: les prépositions. August 1967.

Dung, N. H., and Masson, P. Description d'un analyseur automatique basé sur l'algorithme de Kuno-Woods. September 1967.

Faubert, A. Analyse morphologique de l'anglais. December 1966.

Friant, J. Grammaires ordonnées, grammaires matricielles. October 1968.

Friant, Jean. Langages ultralinéaires et supralinéaires: nouvelles caractérisations. November 1968.

Quérido, A. A.-M. Deux modèles de description syntaxique: analyse fonctionnelle et grammaire générative-transformationnelle. July 1967.

Van Caneghem, Michel. Morphographie générative du Francais. January 1971.

Dansereau, Jules. Modèle génératif de la syntaxe francaise, 1972.

Harris, Brian, Hoffman, T. P., Courval, J., Weydt,H., and Weydt, S. Report in transcoding experiments in 1971. January 1971.

INTRODUCTION TO AUTOMATIC TRANSLATION
(Introduction à la traduction automatique)

A. R. Gouin
Department of National Defense
Ottawa

Elementary information on machine translation, its scope, limitations, goals, the state of the art. Short description of the SYSTRAN Russian-to-English system.

MECHANICAL TRANSLATION BETWEEN ENGLISH AND JAPANESE

Shigeharu Sugita
Department of Information Science
Kyoto University

Translation programs from English into Japanese and from Japanese into English were implemented in order to investigate the adequacy of context-free grammar as a model of English and Japanese. The computer system is NEAC 2200/200 with core memory of 32K characters of $2\mu s$ cycle time and disk memory of 9.2 Mch for dictionary storage. A minicomputer MACC-7F manages input and output of sentences and correction of the grammar table. The word dictionary contains 8000 head words and the grammar table contains about 900 rewriting rules. About 70% of sentences from scientific or technological papers are analyzed correctly from the syntactical point of view. Results of English into Japanese translation can be spoken through speaker by a voice synthesizer.

ENGLISH AS A VERY HIGH LEVEL LANGUAGE FOR SIMULATION PROGRAMMING

G. E. Heidorn
IBM Thomas J. Watson Research Center
Yorktown Heights, N.Y.

An automatic programming system which produces simulation programs from information obtained through natural language dialogue has been implemented under CP-CMS on the IBM 360-67. The information obtained from an English conversation about a simple queuing problem is used to build a language-independent entity-attribute-value data structure. From this structure both an English description of the problem and a CPSS simulation program for it can be produced. This processing is done by a FORTRAN program which interprets sets of decoding and encoding rules written in a specially developed grammar-rule language.

COMPUTER PROGRAMMING USING NATURAL LANGUAGE

Edmund C. Berkeley, Andy Langer, and Casper Otten
Associates, Frontiers Group
Newtonville, Massachusetts

The task is to "produce a definite, efficient, and rather small program which will take in ordinary natural language specifying any one of many kinds of calculations, and will put out automatically a computer program which will correctly handle any sample of any kind of those calculations."

The principles are "Appropriate Neglect, the Preferred Synonym, Relevant Context, a Series of Separate Statements, Statements of Reasonable Length, a Finite Vocabulary, Gradual Development, Questioning by the Clerk," etc.

"We have taken ordinary natural language and produced precise computer programs from it."

Modeling Simultaneous Actions and Continuous Processes

Gary G. Hendrix
Department of Computer Sciences and CAI Laboratory
The University of Texas
Austin

A new methodology for the construction of world models is presented. The central feature of this methodology is a mechanism which makes possible the modeling of (1) simultaneous, interactive processes, (2) processes characterized by a continuum of gradual change, (3) involuntarily activated processes (such as the growing of grass) and (4) time as a continuous phenomenon. Considerable attention is given to the application of this methodology in robotics.

METHODS OF DECIPHERING OF A MESSAGE FROM EXTRATERRESTRIAL INTELLIGENCES

(Metody deširovki)

B. V. Sukhotin
Institute of the Russian Language
USSR Academy of Sciences
Moscow

Deciphering problems may be combined into definite stages, each having a small number of crucial tasks: (1) Finding out the set of alphabetic symbols, (2) Disclosure of grammar, (3) Semantic analysis, (4) Mapping the language of the message into the language of the investigator, (5) Translation of the message. Easily operable algorithms exist for most of these stages. The problem of recognizing the artificiality of the message may be replaced by an easier task recognizing its 'intelligibility' The ratio of the quality of the best and the worst interpretations of an elementary deciphering problem may be characteristic of the intelligibility of the text.

A COMPUTER ASSISTED STUDY OF THE VOCABULARY OF YOUNG NAVAJO CHILDREN

Bernard Spolsky, Wayne Holm, Babette Holliday, and Jonathan Embry

In recent years there have appeared a large number of studies of child language suggesting to us a number of different approaches to our task. We might have chosen to study the language of one child in depth, or to survey a number of children. We could choose to work with a corpus of free speech, to elicit specific linguistic forms, or to test the child's understanding of selected items. We could focus our attention on phonology, syntax, lexicon, semantics, or style, or try to cover all of them. We met first with a group of Navajo language experts to explore each of these approaches. Taking into account their suggestions, our specific goals, and the practical limitations we faced, we finally decided to base our study on a recorded collection of speech of six-year-old children; with this corpus we could study letter and word frequency, spelling patterns, and morphological complexity and start looking at syntax.

A COMPREHENSIVE AUTOMATED PSYCHOLOGICAL EXAMINATION AND REPORT (CAPER)

Daniel Paitich

A program is described that generates psychological reports automatically for a battery of standard measures of intelligence and personality in adults. Problems are encountered in the programming of a battery of tests that are not present in the single test automated systems that are available. The standard measures include the Raven MMPI, and 16PF as measures of non-verbal intelligence, psychopathology and personality traits, respectively. Newly developed measures also are included; a multiple choice version of the WAIS vocabulary subtest, a questionnaire dealing with family interaction, and a sexual history questionnaire. All tests are interpreted by simple procedures that involve the printing of sentences or short paragraphs when scores fall in a given range. In the case of MMPI, the configural systems of Gilberstadt and Duker, Marks and Seeman, and the Mayo Clinic have been included. The system provides a comprehensive clinical report that seems adequate and requires little time for a psychology staff member.

KINSHIP STRUCTURES REVISITED

Nicholas V. Findler
Department of Computer Science
State University of New York
Buffalo

The present work adds new dimensions of power and flexibility to an earlier formulation of a programming system that could generate and query kinship structures. The use of the associative memory, parallel processing language AMPPL-II, and particularly its definitional capability, have proved to be of a great help in programming.

AN AUTOMATON FRAMEWORK FOR NEURAL NETS THAT LEARN

William L. Kilmer and Michael A. Arbib
Computer and Information Science
University of Massachusetts
Amherst

Brindley (1967, 1969, 1972) has discussed nets of several types of formal neurons, many of whose functions are modifiable by their own input stimuli. Because Brindley's results are widely referred to, for example Marr (1970, 1971) and include some of the scarce non-trivial theorems on learning nets, it is important that serious side-conditions be made explicit. The language of finite automata is used to mathematicize the problem of adaptation sufficiently to remove some ambiguities of Brindley's approach. We close the paper by relating our framework to other formal studies of adaptation.

RESPONSE ROUTING IN SELCUK NETWORKS AND LASHLEY'S DILEMMA

Kenan E. Sahin
University of Massachusetts
Amherst

Given a large net in which connectivity is primarily to near neighbors, how can responses to a generally broadcast request be routed to the site of the request? A routing procedure which does not require site information is based on the Selcuk Principle, each module remembering only on which input channels the request had first arrived. The procedure is successfully applied to a neural network model proposed by Eccles and to one advanced by Burns. The strongest requirement is switching among output channels and this appears consistent with findings of selective invasion of axonal arborization. The procedure is applicable to content-based addressing, to computer communication nets and to pattern recognition: the image on the "retina" is transformed into a time spectrum of responses, the analysis of which yields information on angle, size, curvature, and position of the edges.

A THEORY FOR THE NEURAL BASIS OF LANGUAGE.
PART 1: A NEURAL NETWORK MODEL

Robert J. Baron
Department of Computer Science
University of Iowa
Iowa City

A detailed functional description of elementary visual-linguistic processes: (1) selection and neural encoding of visual patterns; (2) representation of visual experience in memory; (3) mechanisms of association between visual and verbal information including (a) naming images, (b) naming positional relationships, (c) naming size and shape attributes, and (d) imaging pictorial information previously stored; (4) neural representation of phrases and simple sentences; (5) recognition of simple sentences and the concept of meaning; and (6) verbally-directed recall of visual experience. Control strategies for networks. Computer-simulation studies. Processes demonstrated by simulation: (1) verbally directed visual recall; (2) verbal understanding; (3) aspects of learning and forgetting; (4) dependence on contextual information; and (5) elementary concepts of sentence generation.

END

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A