American Journal of Computational Linguistics Microfiche 53

NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

VOLUME 13 - NUMBER 8 DECEMBER 1976

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

SECRETARY-TREASURER: Donald E. Walker, Stanford Research Institute, Menlo Park, California 94025.

EDITOR: David G. Hays, 5048 Lake Shore Road, Hamburg, New York, 14075.

EDITORIAL ASSISTANT: William Benson,

American Journal of Computational Linguistics *Microfiche 58 : 2*

CURRENT BIBLIOGRAPHY

AJCL thanks Martin and Iris Kay and the Xerox Palo Alto Research Center for their participation in the making of this bibliography.

Practical Natural Language Processing: The REL System as Prototype

Frederick B. Thompson, and Bozena Henisz Thompson
California Institute of Technology, Pasadena, California

Morris Rubinoff and Marshall C. Yovits, Eds., Advances in Computers 13, Academic Press: 109-168, 1975
ISBN 0-12-012113-1

REL (Rapidly Extensible Language) is a system for natural language communication with a computer for specialists. Each user of the system will have an idiosyncratic package reflecting the particular vocabulary and semantic processes of his specialty. The vocabulary of a package can be easily expanded by the user through use of the definition facilities of the language. Three base languages, each of which can be used in the creation of a number of specialized user packages, have been developed: REL English (with nomenclature and processing routines for statistical analysis). Animated Film language (for interactive graphics). REL Simulation Language (for designing, testing, and applying discrete simulation models). Other topics covered: Semantics and Data Structures (primitives, nets interpretive routines), Deduction, English for Computer (case grammar, verb semantics, quantification. Practical NL processing.

GENERAL

Artificial Intelligence--The Past Decade

B. Chandrasekaran
Department of Computer and Information Science, The Ohio State University, Columbus, Ohio

Morris Rubinoff and Marsahall C. Yovits, Eds., Advances in Computers 13, Academic Press: 169-232, 1975
ISBN 0-12-012113-1

Topics considered: Language Processing, Some Aspects of Representation, Inference, and Planning, Automatic Programming, Game-Playing Programs, Some Learning Programs, Heuristic Search, Pattern Recognition and Scene Analysis, Cognitive Psychology and AI. Work of the following has been given particular attention: Winograd, Quillian, Simmons, Schank et al., Fikes et al. (STRIPS), Hewitt (PLANNER), Biermann, Winston, Evans (ANALOGY). 161 refs.

On Automatic Speech-Understanding Systems

Georgette Silva
System Development Corporation, Santa Monica, California

Computers and the Humanities 9: 237-244, September 1975

A Speech-understanding system must capture the meaning of the entire utterance, even if some parts of it cannot be clearly recognized; it must be a *meaning-extraction* system. Topics discussed: What knowledge must a speech-understanding system incorporate? How is the knowledge coordinated? What applications do we anticipate? The SDC-SRI system.

Artificial Intelligence

Earl B. Hunt

Academic Press, New York, 1975
ISBN 0-12-362340-5
478 pp. $29.00
Contents: *Introduction.* The Scope of Artificial Intelligence. Programming, Program Structure, and Computability. *Pattern Recognition.* General Considerations in Pattern Recognition. Pattern Classification and Recognition Methods Based on Euclidean Description Spaces. Non-Euclidean Parallel Procedures: The Perceptron. Sequential Pattern Recognition. Grammatical Pattern Classification. Feature Extraction. *Theorem Proving and Problem Solving.* Computer Manipulable Representations in Problem Solving. Heuristic Problem-Solving Programs. Theorem Proving. *Comprehension.* Computer Perception. Question Answering. Comprehension of Natural Language. Review and Prospectus.

Abstracts from the 1974-1975 Rand Information Sciences Conference

G. F. Groner, Editor
The Rand Corporation, Santa Monica, California

Rand Paper P-5233/1, July 1975

The Rand Information Sciences Conference was initiated in Fall, 1973 to promote interaction among Information Sciences Department and Rand Computation Center staff members about their work in mathematics, computer science research, programming and analysis. Papers of interest include: Dan Relles, Statlib - A Statistical Computing Library; Peter Weiner, The Rand Text Editor; R. H. Anderson, J. J. Gillogly, Intelligent Terminal Research at Rand; Ivan E. Sutherland, Sorting and the Hidden Surface Algorithm; Peter H. Farquhar, Multi-Attribute Decision Analysis and Applications; R. Stockton Gaines, the Sad State of Debugging; David Drew, Evaluating the Impact of Federal Funds upon Universities; Ray Pyles, MODIA and DOSS - Software Engineering in a Research Environment.

The Credibility of Machine Intelligence

J. R. Ullmann
Division of Computer Sciences, National Physics Laboratory, Teddington, England

Nature 257: 547-549, October 16, 1976

Not all problem solving algorithms should be called intelligent. "Intelligence" denotes a very distinctive class of problem-solving processes: probably, in the brain, using highly parallel computation and distributed logic; in some cases, a sequential process that does not use stored programs. A programmed computer should not be deemed intelligent unless it is functionally similar to the brain in its use of previous experience.

Speech Input/Output System Employing a Minicompuer

H Fujisawa, and K. Shirai
Department of Electrical Engineering, Waseda University, Japan

Electrical Engineer ng in Japan 94:103-110, July-August 1974

The authors describe a speech input/output system which is designed for giving commands to robot WABOT-1. A problem concerning sentence recognition is also discussed together with the proposal of a concrete means for sentence recognition.

GENERAL

Speech Recognition by Machine: A Review

D. Raj Reddy
Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

Proceedings of the IEEE 64: 501-531, April 1976

Topics covered: *I. Introduction:* the nature of the speech recognition problem, the uses of speech recognition; *II. Systems:* word recogntion systems connected speech recognition, speech understanding systems; *III. Signal Processing for Speech Recognition:* parametric analysis of speech, end-point detection, noise normalization; *IV. Task-Independent Knowledge:* phonemic labeling phonological rules, prosodics, word hypothesis, word verification; *V. Task-Dependent Knowledge.* vocabulary syntax, semantics, pragmatics and discourse analysis; *VI. System Organization.* control strategies in the presence of ambiguity and error, real-time input, knowledge acquisition; *VII. Conclusions.*

Color Their Prose Gray

M. E. Orth
Department of English, University of Wyoming, Laramie

IEEE Transactions on Professional Communication 18: 65-6, June 1975

Though surrounded by fascinating and challenging subjects, too many engineers and scientists write dull reports and papers. More attention to verbs can help alleviate dullness and can promote such qualities as vividness, directness, force, and interest. Passive verb forms can be changed to active; combinations of passive verbs with other lifeless verbs can be. reduced through subordination and elemination; and nominalized verbs can be used as true verbs. An improved approach to technical writing is suggested.

Proceedings of the ASIS 37th Annual Meeting, vol. 11

Washington, D. C.: American Society for Information Science, 1974
x + 278 ISBN 0-87715-411-2

Conference held at Atlanta, Ga., 13-17 October 1974. The following topics were dealt with: information utility; data base management; processing for information retrieval; facilities and services; information use; user needs; bibliometric analysis of trends; semiotic implications; human information processing; education and research. Abstracts of some papers appear elsewhere on this fiche.

Some General Remarks About Pattern Recognition, Its Definition; Its Relation with Other Disciplines: A Literature Survey

C. J. D. M. Verhagen
Department of Applied Physics, Delft University of Technology, Netherlands

Pattern Recognition 7: 109-16, September 1975

A survey is given of definitions and descriptions, taken from the literature, concerning the terms: pattern,. recognition, pattern recognition, classification, cognition, etc. Both very general and more specific definitions are quoted; no deep agreement on the meaning of these terms seems to exist. Relations between pattern recognition and other disciplines are discussed and graphicaaly indicated. Some tentative conclusions are given about possible future activities concerning the definition of general terms in the field.

Computational Linguistics or What's in a Name?

W. Martin
Institute of Applied Linguistics, University of Leuven, Belgium

Association for Literary and Linguistic Computing:
Bulletin 3: 124-132, 1975

A linguist can use a computer as a classifying, calculating, control, and simulation machine. *Senso largo*, all linguistic investigations which make use of a computer are called computational linguistics. It is only in the last two senses, however, and more especially in the last, that one can speak of computational or algorithmic linguistics as a possible autonomous discipline within the field of linguistics.

Explizite Beschreibung der Sprache und automatische Textenbearbeitung, I: Terminologisches Worterbuch. Explicit Description of Language and Automatic Text-Processing, I: Terminological Dictionary.

Matematicko-fyzikalni fakulta University Karrlovy Praha, 1975

The Terminological Dictionary contains terminology from algebraic linguistics and some terms from related disciplines, such as structural linguistics, logic, algebra, statistics, information theory, etc. A certain number of practical catch-words are also included, but any terms which are contested idiosyncratic, or otherwise problematical are generally omitted. The dictionary is primarily intended to meet the needs of research work, but it may also be of use tc translators, students, etc. As a relatively new and quickly-developing one, the field faces the problems of establishing terminology and of determining the extent to which terms in various languages correspond to one another. On practical grounds the dictionary works from English terminology, by which entries are alphabetized; the equivalents in French, German, Russian, and Czech then follow.

Explizite Beschreibung der Sprache und automatische Textenbearbeitung, II: Bibliographie der linguistischen Gruppe des ZNM MPF KU Explicit Description of Language and the Automatic Processing of Texts, II: Bibliography of the Linguistics Group of the Center for Numerical Mathematics of the Mathematical-Physics Faculty of Charles University

Matematicko-fyzikalni fakulta University Karlovy Praha, 1975

The bibliography proper is subdivided into five general topics: formal description of language; empirical linguistic problems; surveys and reports; reviews; bibliographies, translations. It is preceded by a lengthy introduction to the so-called functional generative description of language as developed by the Prague group for nearly twenty years. This description is but one type of stratification description, i.e. it works with ordered planes which extend from meaning on the one hand to expression on the other. Functional generative grammar's specific nature makes it suited to automatization-related applications of linguistics, such as machine translation and information science (automatic abstracting, indexing etc. Although this work needs much further development, it is already of real use for purely intellectualized texts which are stylistically straightforward and contain relatively clearly-defined terminology. The system of levels in functional generative description is informally outlined, with special emphasis being laid on the semantic plane. (In German)

Cahiers de linguistique theorique et appliquee
Papers in Theoretical and Applied Linguistics

G. C. Moisil, and A. Rosetti, Eds.
Centrul de Cercetari Fonetice si Dialectale, 194, Calea Victorie, Bucarest (22), Roumanie

Vol. 10,.No. 1. 1973

Contents:

Cahiers de linguistique theorique et appliquee
Studies in Theoretical and Applied Linguistics

G.C. Moisil, and A. Rosetti, Eds.
Centrul de Cercetari Fonetice si Dialectale, 194; Calea Victorie: Bucarest (22), Roumanie

Vol 10, No. 2. 1973

Contents:

Cahiers de linguistique theorique et appliquee
Papers in Theoretical and Applied Linguistics

A. Rosetti, Ed.
Centrul de Cercetari Fonetice si dialectale, 194, Calea Victoriei, Bucarest (22), Roumanie

Vol. 11, No. 1, 1974

Cahiers de linguistique theorique et appliquee
Papers in theoretical and Applied Linguistics

A. Rosetti, Ed.
Institutul de Cercetari Ethnologie si dialectologice, 194, Calea Victoriei, Bucarest (22), Roumania

Vol. 11, No. 2, 1974

Recent Computer Science Research in Language Processing

Allen Klinger
Department of Computer Science, University of California at Los Angeles

Report UCLA-ENG-7541 AFOSR-TR-75-1153 June 1975
NTIS: AD-A013 856/0GA
PC $3.75/MF $2.25

The machine translation problem has recently been replaced by much narrower goals and computer processing of language has become part of artificial intelligence (AI), speech recognition, and structural pattern recognition The narrower goals involve making it possible for a computer user to employ a near natural-language mode for problem-solving, information retrieval, and other applications. Natural computer responses have also been created and a special term understand has been used to describe the resulting computer/human dialogues. The purpose of this paper is to survey these recent developments to make the AI literature accessible to researchers mainly interested in computation on written text or spoken language.

Directions in Artificial Intelligence: Natural Language Processing

Ralph Grishman, Ed.
Courant Institute of Mathematical Sciences, New York University

Report NSO-7, August 1975
NTIS: AD-A014 605/0GA
PC $5.25/MF $2.25

The report contains proceedings of a symposium on natural language processing held at the Courant Institute of Mathematical Sciencies New York University on December 6, 1974. The talks were concerned with the analysis of the structure among definitions in a dictionary, the automatic generation of semantic word classes by text analysis the design of semantic hierarchies, and transformational language analysis procedures and underlying structures for information retrieval.

Computers that Talk and Listen: Man-Machine Communication by Voice

James L Fla agan
Acoustics Research Department, Bell Laboratories, Murray Hill, N. J. 07974

Proceedings of the IEEE 64: 405-415, April 1976.

Three modes of man-machine voice interaction: machine voice read-out of stored information, speaker verification, automatic recognition of spoken commands. Voice response. adaptive differential pulse-code modulation (ADPCM) formant synthesis, text synthesis. Speaker verification: verification (cooperative individual having computer verify his claimed identity), identification of unknown individual via voice pattern. *Speech Recognition*: the Holy Grail of the field: A three-mode experimental system for airlines information and ticketing is described. Digital interfaces for man-machine communication: block addressed packet-switched communication. Current capabilities: for systems with designated speakers, vocabularies of several hundred individual words are tractable; for speaker independent systems, small vocabularies, such as spoken digits, are tractable.

Lip and Jaw Motor Control During Speech: Responses to Resistive Loading of the Jaw

J. Folkins, and J. H. Abbs
University of Washington, Seattle

Journal of Speech and Hearing Research 18, 1: 207-220, March 1975

Resistive loads were applied to the jaw during speech production. Loads were initiated during the jaw closing movement associated with the production of bilabial stops, creating a situation in which bilabial closure would be disrupted if motor control were independent of peripheral feedback. Three subjects were observed during control and experimental conditions. In all utterances in which a load was appropriatedly introduced, closure of the lips was achieved and the bilabial stop was adequately produced. To assess the nature of this control, displacement of the upper lip, lower lip, and jaw, in the inferior-superior dimension, were recorded along with EMG from medial pterygoid, anterior temporalis, masseter, and orbicularis oris superior muscles. Based on observation of these variables, it appears that the muscles of the lips and jaws are capable of on-line compensatory motor reorganisation.

A New Time-Domain Analysis of Human Speech and Other Complex Waveforms

Janet MacIver Baker
Department of Computer Science, Carnegie-Mellon University, Pittsburgh PA

National Technical Information Service: AD-A013 583/0GA
PC $6.25/MF$2.25
May 1975

Three separate investigations are presented: 1) Cycle-based time-domain parameters were extracted from the speech waveforms of many hundreds of utterances, and were then analyzed by hand and by machine. 2) Based solely on time-domain phenomena found in 1) a segmentation program was written for continuous speech. 3) Examination of time-domain characteristics of 228 allophones of fricatives and stop consonants for 2 males and 1 female. Comparison of frequency-domain techniques with time-domain techniques. Applications to other complex waveforms.

Development of Analytical Methods for a Semi-Automatic Speaker Identification System

J. E. Paul, Jr., A S. Rabinowitz, J. P. Riganati, and J. M. Richardson

In 1975 Carnahan Conference on Crime Countermeasures, Lexington, Ky.: University of Kentucky Press, 1975: 52-64

Segments of speech are extracted from two speech utterances and are computer-analyzed to yield a statistical measure indicating whether the utterances were said by the same or different speakers. The analytical studies involved the collection and processing of speech data from over 250 speakers from which was extracted over 35,000 phonetic-event tokens. A study of the discriminating power of individual phonetic events, along with a study of coarticulation effects on these events, was conducted. A unique set of 30 features was defined for each of the resultant 13 phonetic categories used by the SA IS in carrying out speaker identification. A weighted Euclidean distance measure was selected for measuring speaker similarity within a phonetic category and a desensitized Fisher discriminant was incorporated to combine these individual distance measures into an overall measure of similarity between the speakers of two spoken utterances. A set of likelihood tables, indexed by similarity measure and the phonetic events used, was derived to determine the probability that the utterances were uttered by the same individual.

Semi-automatic Speaker Identification System

P. K. Broderick
Aerospace Corporation, El Segundo, California

In 1975 Carnahan Conference of Crime Countermeasures,
Lexington, Ky.: University of Kentucky Press, 1975: 29-37

A system is described which analyzes speech samples to identify and extract speaker-dependent features and to perform a statistical comparison of the features from different samples. The purpose is to enable law enforcement personnel to compare the recorded voice of a criminal (e.g., from a bomb threat recording) with recorded voice samples from suspects to identify the perpetrators of crimes. A minicomputer and associated peripherals accept analog speech signals for processing and statistical comparison. Specific phonetic events found to have a high degree of discriminating power are identified and labeled by the operator using an interactive graphic display terminal. In the comparison phase each selected event from the criminal sample is compared with a like event from a suspect sample. The points of comparison are well defined and yield quantitative results on a repeatable basis.

Automatic Speaker Recongition by Computer

E. Bunge
Philips Forschungslab, Hamburg GmbH, Germany

In 1975 Carnahan Conference on Crime Countermeasures,
Lexington, Ky.: University of Kentucky Press, 1975: 23-28

For security systems in banking and in law enforcement, the human voice is an essential aid for verifying a person s identity. The paper describes the principles of an automatically working speaker recognition system and outlines the basic difficulties and pitfalls. A promising approach is the combination of fast signal processing techniques and trainable pattern recognition procedures. Results of a comparative study are discussed.

A Programme for Synthesizing Hebrew Speech

Asher Laufer
University College, London, and the Hebrew University, Jerusalem

Phonetica 32:292-299, 1975

The program has two sets of inputs: 1) a linguistic description (essentially a phonemic transcription) of the utterance to be synthesized, 2) a phoneme table containing information about the category of phoneme, transition duration, average duration, average values of formant frequencies and their amplitudes which are the target values for the phonemes. After inputs are fed into the computer the program changes the basic data of the phonemes according to the allophone rules, which modify the target values of phonemes according to context. Parameter values are then calculated according to transition rules and intonation rules. The output of this program is a punched paper tape which is acceptable to the synthesizer (a computer-simulated parallel-formant synthesizer). Various intelligibility tests were carried out, the final test including sentences and a story. This last test scored an average of 97% words correctly recognized by 33 participants.

Digital Techniques for Computer Voice Response

Lawrence R. Rabiner
Bell Laboratories, Murray Hill, N. J. 07974

Ronald W. Schafer
Georgia Institute of Technology, Atlanta, 30332

Proceedings of the IEEE 64: 416-433, April 1976

Digital waveform coding: Pulse-code modulation (PCM), Differential PCM, Adaptive DPCM, Analysis/Synthesis Techniques. Design Considerations: memory and speech coding, vocabulary preparation, message composition. A digital voice response system capable of handling interaction on 10 channels simultaneously is described and applications of the system are discussed: 1) computer-aided wiring of circuits, 2) directory assistance, 3) stock price quotation, 4) data set testing information system, 5) flight information, 6) speaker verification.

Synthesis of Speech from Unrestricted Text

Jonothan Allen
Research Laboratory of Electronics, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology 02139

Proceedings of the IEEE 64: 433-442, April 1976

When we remove all limitations on the input, then what we require of a text-to-speech system is no less than a cognitive model for reading aloud. Text input is analyzed into an abstract linguistic specification from which a surface phonetic realization is generated. Word level synthesis is the first step. High frequency words can be recognized as a unit, less frequent words can be recognized as a sequence of familiar morphemes; a 12,000 item morpheme lexicon and a morphological analyzer does this. If this fails the system reverts to direct letter-to-sound synthesis rules. The next level of synthesis is the assignment of morphophonemic and lexical stress and the highest level concerns clause-level phenomena (surprasegmentals, prosodic features). At this highest level three functions, suggested by Halliday, are carried by the stress contour: 1) ideational, 2) interpersonal, 3) discourse.

A Model of Articulatory Dynamics and Control

Cecil H. Coker
Acoustics Research Department, Bell Laboratories, Murray Hill, NJ 07974

Proceedings of the IEEE 64: 452-460, April 1976

The system includes: 1) a physical model of the vocal system, with spatial constraints very close to those of natural articulation; 2) a representation of the motional constraints of the articulators which, when moving from one stated shape to another interpolates realistic intermediate shapes; 3) a similar model of the movements of the excitation system, including subglottal pressure, vocal cord angle and tension; and 4) a controller for the mechanism which produces sequences of articulatory commands which cause this dynamic system to execute properly timed articulatory motions from input phonetic strings.

Linguistic Rules for Text-to-Speech Synthesis

Noriko Umeda
Bell Laboratories, Murray Hill, N. J. 07974

Proceedings of the IEEE 64: 443-51, April 1976

Prosody plays the role in speech of producing a spoken message meaningful to a particular circumstance, from otherwise unweighted abstract units called phonemes. In the process of message formation phonemes are given proper acoustic characterization according to the syntactic and semantic structure of the message, the location of stress and boundaries, and surrounding phonemes. This paper concerns work at Bell Laboratories on automatic text synthesis and summarizes the development of related prosody rules from natural speech data.

Application of Fuzzy Set Theory to Syntactic Pattern Recognition of Handwritten Capitals

Walter J. M. Kikcert
Department of Electrical and Electronic Engineering, Queen Mary College, London University

Henk Koppelar
Department of Psychology, State University of Utrecht, The Netherlands

IEEE Transactions on Systems, Man, and Cybernetics 6: 148-151

The recognizer works with a few idealized segments where the letter composition process is governed by a context free grammar. The recognition is performed sequentially: the pattern is scanned from segment to segment, and each segment is separately recognized as a member of the previously defined set of ideal letter segments. The recognition rules are divided according to the different places in the total pattern where a segment can occur (11 arbitrarily ordered rules). The fuzzy membership function represents the degree to which the actual pattern is a member of the fuzzy set "line" or "arc". The fuzzy set evaluation procedure uses only min and max operators.

Real-Time Recognition of Handwritten Numerals

S. Impedove, B. Marangelli, and V. L. Plantamura
Institute of Physics, University of Bari, Italy

IEEE Transactions on Systems, Man, and Cybernetics 6: 145-148, February 1976

Experiments were set up to find a simple and inexpensive way to test many types of algorithms in real time, in particular, those which analyze nonstylized two-dimensional images by contour description. In this case the processing of the contour may help, while it is being drawn, to avoid ambiguity arising in *a posteriori* analysis from the crossing points. Our equipment allowed a simple implementation of an algorithm that starting from the tangents to the contour points, discretizes their slopes and then identifies successions of slopes which are characteristics for each class of slopes.

A Bibliography in Character Recognition: Techniques for Describing Characters

R. Shillman, C. Cox, T. Kiklinski, J. Ventura, M. Eden, and B. Blesser
Cognitive Information Processing Group, Massachusetts Institute of Technology

Visible Language 8: 151-166, Spring 1974

I. General References and Review Papers, 32 references
II. Engineering Descriptions of Characters, 74
III. Psychological Descriptions of Characters, 89
IV. Insight Through the Study of Character Formation, 10
V. Relevant Journals, 15

Algebraic Structurre for the Recognition of Korean Characters

Joo K. Lee, and Hoon Choo
Department of Electronic Engineering, Inha University, Korea

Journal of the Korean Institute of Electronics Engineers 12, No. 2: 44-50, April 1975

The algebraic structure of the basic Korean characters is analyzed in terms of the concave structure, line structure, and node relationships of character graphs. Characters are classified by degree of complexity. Equivalence relationships defined by rotational transformation group by Affine transformation of one element into another exist between the 10 vowels. Hence geometric properties, as well as topological properties, are important for recognition.

Automatic Signature Verification

N. M. Herbst, and C. N. Liu
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

IBM Research Report RC 5810, 19 January 1976

The fine structure of the muscle forces exerted during a signature is very precise for most people and not subject to conscious control. The overall time for a signature as measured from start to finish is remarkably consistent as are the durations of the individual muscle forces. As the signature is a constant time phenomenon the magnitude of the forces is then related only to the size of the resultant trace. Based on this model, an automatic signature verification system is developed which uses the acceleration-time function as the principal measurement. The system is operational in the laboratory and an experiment involving 0 test subjects tested over a four week period indicated a false rejection rate of 2.87% and a forger acceptance rate of 2.06% with an average of 1.18 trials necessary for verification. Examples given.

An Experimental device for the Recognition of Handwritten Numerals

E. Backer, B. M. van der Boom, I J. Boxma, and J. C. Venniker
Laboratory for Information Theory, Department of Electrical Engineering, Delft University of Technology, Netherlands

Delft Progress Report Series B (Netherlands) 1, No. 2: 25-30, July 1975

Recognition of handwritten numerals, written without any constraints. The recognition procedure is based on the logical n-tuple method. The device first calculates set (triplet) of the three most likely categories of numerals with the aid of a general mask, based on all possible categories. This triplet is then processed with the aid of a second mask (the 'triplet mask'), which contains the tuples giving the best discrimination for the triplet in question. The hardware required for this application is comparatively inexpensive and gives 90% reliability.

Optimisation Algorithms of Linguistic Deciphering

B. V. Sukhotin

Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, No. 5: 36-42, 1975

It is natural to formulate the problems of language deciphering in terms of an optimisation algorithm. A definition of a language phenomenon that can be used to recognize this phenomenon in text comprises three components: a description of the set of admissible soltutions; a statement of the evaluation function; and an optimisation algorithm to find the extremum of this function. The author gives the deciphering definitions of vowels and consonants, of the subordinative relationship of words in a simple sentence, and of the morphological analysis of a text written without spaces between words. Algorithms for the recoding of a syllabic writing into a phonemic one and for transcribing a text in an unknown language into the alphabet of a known one are given.

Syntagmatic concordances and surface analysis (Concordance syntagmatiques et analyse de surface)

P. Laurette

Computers and the Humanities 8:147-151, May 1974

The paper describes two procedures for the establishment of syntagmatic concordances through pre-coding and automatic analysis for the nominal syntagm.

Statistical Analysis of Lexical Data Using Chi-Squared and Related Distributions

Barron Brainerd
University of Toronto

Computers and Humanities 9: 161-178, July 1975

Quantitative data from literary sources often appear in the form of word-counts or counts of the occurrence of literary and/or grammatical structures. In their raw form such counts cannot always be assumed to have the properties (e.g. approximate normality, homoscedasticity) that make them directly analyzable in terms of statistical procedures involving the classical sampling distributions. There are, however, so called distribution-free procedures available for the rigorous treatment of such data. This paper discusses procedures involving the chi-square distribution. Topics include: significance tests, degrees of freedom, long-range variation in contiguous samples, independence tests, randomness of occurrence of events in a text, goodness-of-fit, dispersion. Examples are given.

Linguistic Subgrouping and Lexicostatistics

Isodore Dyen
Yale University

Janua Linguarum, Series Minor, 175. Mouton, 1975
ISBN 90-279-3054-6 pp. 251

Towards an Algorithmic Methodology of Lemmatization

M. L. Hann
Department of European Studies and Modern Languages, University of Manchester Institute of Science and Technology, England

Association for Literary and Linguistic Computing, Bulletin 3: 140-150, 1975

The article introduces some algorithmic techniques which enable a computer to determine those inflected word forms in a textual corpus which are lexically equivalent, and to generate automatically character-string transformations for reducing the raw source text to a series of lemmata. The methods proposed make no specific reference to the language of the source text and are applicable for the reduction of both grammatical and syntactic inflexions. The algorithms are exemplified in use on German text.

Use of an Interactive Program in Analyzing Data for a Dialect Dictionary

Michael M. T. Henderson
Assistant Editor, Dictionary of American Regional English, University of Wisconsin, Madison.

Computers and the Humanities 9: 105-113, May 1975

DARE (Dictionary of American Regional English) uses written data and oral data obtained from 2752 informants in 1002 communities with a 1847 item questionnaire. The *words* are stored on paper where editors can get at them easily, while *numbers* are stored in the computer file. The data is available in tabular form and can be examined regionally by maps displayed on a CRT screen. The blank map has as many blank spaces in each state as there were questionnarires taken in that state. The distribution of a response can then be visually examined by having characters appear on the map which represent responses to the item being studied. If desired the computer can give a map of responses given by people fitting only certain classifications--male, old and female, black high-school graduate, etc.

Automated Morphosyntactic Analysis of Medical English

M. G. Pacak
Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20014

W. C. White
Centre Hospitalier Universitaire, Sherbrooke, Canada

Information Processing & Management 12: 71-76, 1976

The procedure for morphosyntactic analysis of medical diagnoses is based on the identification of terminal morphemes which convey information about part-of-speech class membership and, in some instances, also information about a semantic category to which the given word form may belong. The rules for the recognition of productive terminal suffixes and assignment of syntactic and or semantic values are represented by a tree-like right-to-left branching graph whose roots are the terminal characters of input words as they occur in the running text. The examination of characters proceeds from right-to-left to locate a set of defined preterminal characters which constitute a potential suffix. Advantage is taken of the fact that semantic values can be assigned to a set of highly productive Greek and Latin suffixes which occur frequently in medical languages. The program achieved an accuracy of 98.6% for 300 MEDLINE titles and 99.4% for medical diagnoses.

Augmented Phrase Structure Grammars

George E. Heidorn
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

IBM Research Report RC 5787, 31 December 1975

An augmented phrase structure grammar (APSG) consists of a collection of phrase structure rules which are augmented by arbitrary conditions and structure building actions. The data structure used by APSG is a form of semantic network, consisting of "records" which are collections of attribute-value pairs. *Records* represent entities, either physical or abstract, and such diverse things as vehicles, actions words and verb phrases. There are three kinds of *attributes*: *relations*, which have as their values pointers to other records; *properties*, which have as their values either numbers or character strings; and *indicators* which have bit string values and usually serve in a role similar to features in linguistic terminology. The decoding algorithm is basically a bottom-up, left-to-right, parallel-processing, syntax-directed compiler.

Case Systems for Natural Language

Bertram Bruce
Bolt, Beranek and Newman Inc., Cambridge, Massachusetts 02138

Artificial Intelligence 6:327-360, Winter 1975

Deep cases bear a close relationship to the modifiers of a concept. In fact one could consider a deep case to be a special, or distinguishing, modifier. Several criteria for recognizing deep cases are considered here in the context of the problem of describing an event. Unfortunately none of the criteria serves as a completely adequate decision procedure. A notion based on the context-dependent "importance" of a relation appears as useful as any rule for selecting deep cases. Theoretical case systems by Fillmore, Celce, Grimes, Schank and Rumelhart, Lindsay and Norman, are discussed. Also: Simmons; Hendrix, Thompson and Slocum; the glaucoma model of Kulikowski and Weiss and its implementation in the CHRONOS system by Chokhani; Martin's work on automatic programming; Cohen, using procedural semantics; Brown, Bruce and Trigoboff (CHRONOS); Baranofsky (the SRI system); Nash-Webber (SPEECHLIS). 54 references.

From Remote Structures to Surface Structures Without the Cycle: A Computational Study

Donald L. Smith
Curriculum in Linguistics, University of Georgia, Athens 30601

International Journal of Man-Machine Studies 7:751-800, November 1975

A set of rules which account for roughly the same sentence types that Burt attempts to account for in her introduction to transformational syntax (1971) has been successfully developed and computer tested. In contrast to Burt's approach, this grammar does not utilize the transformational cycle; all conditions on the applicability of and the operation of the rules are made completely explicit. The rules are simply arranged on the basis of their functions: ranking, grafting (simplification), agreement, suffixing, and surface-structure transformations. The only transformations which reorder constituents are the surface-structural ones. The other transformations serve only to rank constituents or to alter trees. In formulating the transformations, the primary concern is what logical, stylistic, or psychological factors (such as perceptual strategy, short-term memory, etc.) might underlie them.

Data Base for Natural Language Information Processing. Normalization of Natural Language

S. Yoshida, T. Fujita, and H. Tsurumaru
Kyushu Institute of Technology, Fukuoka, Japan

Bulletin of the Kyushu Institute of Technology (Science and Technology)
30: 101-8, March 1975

Ordinary Japanese sentences can be decomposed into a set of simple sentential forms, which are called normal form sentences, and a set of connective relations between the normal form sentences. In general, a Japanese sentence is composed of one or more bunsetsu's (sentence units), each of which is composed of an independent word and zero or more annex words. Decomposition into the normal form is performed through the syntactic analysis related to the dependency relations among bunsetsu's in a sentence and the segmentation of words within each bunsetsu. (In Japanese)

SEMANTICS-DISCOURSE

A new journal: *Semantikos*

Ryszard Zuber, Managing Editor

Business and editorial correspondence should be addressed to Semantikos, 8 rue des Boulangers, 75005 Paris, France

The first issue, which appeared in June 1975, contains articles by Anscombe, Atlas, Ducrot, and Dahl. Articles will appear in English or in French.

EDITORIAL BOARD: Richard Carter, Edward L. Keenan, Noam Chomsky, Sige-Yuko Kuroda, Osten S. Dahl, George Lakoff, Oswald Ducrot, Barbara Partee, Gilles Fauconnier, John R. Searle, Charles J. Fillmore, Robert Stalnaker, Lauri Karttunen, Richard Thomason, Jerrold J. Katz, Arnim von Stechow.

Subscriptions: Students 51 F or $ 15 per volume
Non-students 59 F or $ 18 per volume
Institutions 89 or $ 26 per volume

For non-European countries add 3 F for postage.
Subscription checks should be made payable to *Semantikos*.
Interbanks transfers should be addressed to

Societe Generale, Agence H V, 23 Bd St-Germain, 75005 Paris, account no. 3726045 3

Indexical Symbolism: A Primitive Semiotic System

J. Gough, Jr., and M. Valach
Georgia Institute of Technology, Atlanta

Proceedings of the ASIS 37th Annual Meeting, vol.11, Washington, D. C.
American Society for Informations Sciences, 1974, ISBN 0-87715-411-2:187-90

The indexical symbolism in natural language is man's primitive information processing language. It is such by virtue of the egocentric processor oriented semiotic system in it and appears to be a linguistic parallel to man's directed sensorial system.

Semantics, Preference and Inference

Yorick Wilks, and M. King
Institute for Semantic and Cognitive Studies, Castagnola, Switzerland

Working Paper No. 16

This paper gives full details of the program and the linguistic protocols for the system described in *Communications of the ACM* 18: 264-274, May 1975 (abstracted on AJCL microfiche 30, frame 52) and *Artificial Intelligence* 6:53-74, 1975 (AJCL M 30, F 49).

An Algorithm for Natural-to-Predicate Calculus Language Translation

T. D. Korel'skaya

Nauchno-Tekhnischeskaya Informatsiya, Seriya 2, 6:24-34, 1975

For a subset of Russian, the language of geometry, an algorithm produces a predicate calculus formula expressing the same meaning as a given sentence. The algorithm is based on a system of formalised synonymic transformations which translate Russian syntactic structures into those of the predicate calculus. The dependencies connecting transformations in this system are disoussed, and, taking account of them, the process of applying transformations to a syntactic representation is arranged so that the end result--a predicate calculus formula--is obtained after a minimum of trials. (9 refs., in Russian)

A Framework for Representing Knowledge

Marvin Minsky
Massachusetts Institute of Technology

In Patrick H. Winston, ed., The Psychology of Computer Vision, McGraw-Hill, 1975
pp. 211-280, ISBN 0-07-071048-1

A frame can be thought of as a network of nodes and relations. The "top levels" of a frame are fixed, and represent things that are always true about the supposed situation. The lower levels have many *terminals*--"slots" that must be filled by specific instances or data. Each terminal can specify conditions its assignments must meet. Frames are linked together into *frame systems.* Topics discussed include: local and global theories of vision, parallelism, problem solving, is vision symbolic?, seeing a room, scene analysis, perspective and viewpoint transformations, imagery, frames and Piaget (the need for structures corresponding to "representations of representations"), words and meanings, discourse, active vs. passive, scenarios, memory, matching, advice and similarity networks, higher level organization of "similarity paths" in the memory, analogies and alternative descriptions, heuristic search, frames as paradigms (Kuhn), control structures, spatial imagery, global space frame system (GSF), embedding in GSF, evolution.

Spatial Networks as a Site for the Study of Language and Thought

Charlotte Linde
Hunter College, City University of New York

William Labov
University of Pennsylvania

Language 51:924-939, December 1975

When asked to describe their apartments, people (in NYC) generally give descriptions organized as an imaginary TOUR of the apartment--though a few people organize their description as a MAP. A TOUR is a speech act which provides a minimal set of paths by which each room could be entered The path is presented as a series of units of the form *direction* (or *vector*) and *room*. Analysis of 72 layouts organized in this way reveals that : a) the imaginary tour always begins at the front room of the apartment, b) if the (imaginary) visitor comes to a one-room branch, he does not enter it, c) if he comes to a branch with rooms beyond the first room he always enters, d) when he reaches the end of a branch and there are other branches to be traversed he does not turn around and go back, instead he is moved back instantaneously to the fork point where the other branches originate. Phrase-structure rules are used for displaying these networks and for formally characterizing their structure.

Montague Grammar

Barbara H. Partee, Ed.
Department of Linguistics and Philosophy, University of Massachusetts, Amherst

Academic Press, New York, 1976 ISBN 0-12-545850-9
370 pp. $19.50

Contents:

Lattice Grammars

Hyo Heng Kim, Masaharu Mizumoto, Junichi Toyoda, and Kokichi Tanaka
Faculty of Engineering Sciences, Osaka University, Toyonaka, Japan 560

Systems - Computers - Controls 5, 3: 1-8, 1974

Formal grammars such as stochastic grammars, fuzzy grammars, weighted grammars, etc., are considered grammars in which a certain algebraic system is assigned to rewriting rules of formal grammars. In lattice grammars elements constructed by various parameters can be given as dimensions of lattices in membership space and the grade of words obtainable by these elements can be regarded as estimation specified by these parameters. The concept of L-fuzzy sets (Goguen 1967) is introduced into phrase-structural grammars to formulate lattice grammars wherein lattices are an algebraic system assigned to rewriting rules of formal grammars. Type-1 languages can be generated by type-2 lattice grammars. Chomsky and Greibach normal forms for type-2 lattice grammars are constructed.

On the Ambiguity Function of Context-Free Languages

Tadao Takaoka
Faculty of Engineering Ibaraki University, Hitachi 316, Japan

Makoto Amamiya
Musashino Electrical Communication Laboratory, N.T.T. Musashino 180, Japan

Systems - Computers - Controls 6, No. 1: 31-35

The number of distinct derivation trees of a word x under a CFG (context-free grammar) is denoted by aG(x). The lower ambiguity function aG(n) and the upper ambiguity function AG(n) of the grammar G are respectively defined by the minimum value and the maximum value of aG(x) such that lg(x) = n. Choosing an appropriate CFG G, the functions aG(n) and AG(n) can have the functional order of n to the k for any nonnegative integer k or the exponential order. The functional order of AG(n) is in general undecidable. However, it can be bounded by some range by utilizing the generating function of the grammar when the number of words in L(G) whose length is equal to n, N(L(G), n), is computable. The ambiguity function of a CFL (context-free language) L is said to have a functional order of f(n) if the functional order of AG(n) or any CFG G such that L(G) = L is larger than or equal to f(n) and, for at least one CFG G the functional order of AG(n) is equal to f(n). There exist CFLs whose ambiguity function has the functional order of n to the k for k = 0, 1, 2, . . . or the exponential order.

Do Quantifiers Branch?

Gilles Fauconnier
Laboratoire d'Automatique, Documentaire et Linguistique, 2 Pl. Jussieu, Paris 5e, France

Linguistic Inquiry 6: 555-578, Fall 1975

J. Hintikka (LIINBL 6: 153) claimed that some sentence types exhibit the logical properties of finite partially ordered (FPO) quantification. It is shown that the fragments of English considered by Hintikka offer no evidence that the formulas of FPO quantification theory are repoducible as the semantic structures of some English sentences. However, to say that standard logic is powerful enough for these fragments is not to say that standard logic itself is adequate. The question of how the logical properties of NL sentences can best be accounted for within linguistic theory therefore remains open.

Automata Theory (A Bibliography with Abstracts)

David W. Grooms
National Technical Information Service, Springfield VA

NTIS/PS-75/707/0GA,
PC $25.00/MF $25.00,
164 p. September 1975

Research is cited on pushdown automata, tessellation automata, web automata and finite state automata. Studies on finite state machines, Turing machines, and sequential machines are included. Research on boolean functions, recursive functions, the moore model, and the mealey model as applied to automata theory is also covered. (Contains 159 abstracts)

On Non-numeric Architecture

C. Jack Lipovski, and Stanley Y. W. Su
Department of Electrical Engineering, University of Florida, Gainsville

SIGIR Forum 10, 1:5-20, Summer 1975

The main primitive operation in non-numeric processing is the search. But we must use standard von Neumann computers whose main primitive operation is addition. The main hardware problem in non-numeric processing: the entire data base cannot be put into primary memory so that the CPU can search it there. If the secondary memory had some searching capability only highly relevant information would have to be sent to primary memory. An alternative to the conventional approach is to search data with many processors acting in parallel. In designing a non-numeric machine one must consider the design of both software and hardware because: 1) The operations of the machine should be as close as possible to the operations of the end user. 2) The way the data is stored in the machine should be as close as possible to the way the user sees it, 3) but the user shouldn't have to be aware of how the machine stores the data in its storage structure. Other design considerations are discussed and project CASSM, a non-numeric processor we are building, is discussed.

Computer Communication Networks

R. L. Grimsdale, Ed.
Department of Electrical Engineering, University of Sussex, U.K.

F. F. Kuo, Ed.
Department of Electrical Engineering, University of Hawaii

Noordhoff International, Leyden, 1975
ISBN 90-286-0593-2

Proceedings of the NATO Advanced Study Institute on Computer Communication Networks, Sussex, UK, September 9-15, 1973. Papers include: "The ARPA Network," F. E. Heart; "Performance Models and Measurements of the ARPA Computer Network," Leonard Kleinrock; "Topological Design Considerations in Computer Communication Networks," V. G. Cerf, D. D. Cowan, R. C. Mullin; "A new Mini-computer/Multiprocessor for the ARPA Network," F. E. Heart, S. M. Ornstein, W. R. Crowther, W. B. Barker, "A parallel Processing Approach to Computer Communication," Dan Cohen, Edward Taft; "Colloquies in Computer Networks," G. Le Moli, "Presentation and Major Design Aspects of the Cyclades Computer Network," Louis Pouzin.

Summary of MYCROFT: A System for Understanding Simple Picture Programs

Ira P. Goldstein
Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge 02139

Artificial Intelligence 6: 249-288, Fall 1975

MYCROFT operates in the restricted world of simple programs for LOGO turtles. Input to MYCROFT consists of the user's programs and a *model* of the intended outcome. Model: conjunction of geometric predicates describing important properties of the intended picture. MYCROFT then analyzes the program, building both a Cartesian *annotation* (using an imperative semantics assscіated with each turtle primitive) of the picture that is actually drawn and a *plan* explaining the relationship between the program and the model. The system interprets the program's performance in terms of the model and produces a description of the discrepancies expressed as a list of violated model statements. The debugger repairs each violation, making the procedural description produce a graphical result that satisfies the set of predicates describing intent. The final output is an edited turtle program which satisfies the model

A Logic for Semantic Networds

Robert Bechtel, and Stuart C. Shapiro
Computer Science Department, Indiana University, Bloomington 47401

Technical Report No. 47 March 1976

NL understandding systems require the use of nonstandard logic. We start with Belnap's four-valued logic: True, False, Both, None. Two non-standard connectives, AND-OR and THRESH, which are generalizations of the familiar symmetric binary connectives, are introduced. The classical quantifiers ALL and EXISTS, are retained and three new quantifiers are introduced: NONE states that there is no substitution which can validate the expression in its scope; ONE states that there is one and only one substitution which will make the expression within its scope valid; ALMOST-ALL (ALA) states that there maybe some substitution which invalidates the expression within its scope. A form of implication is defined and representations of constructions of the logic are developed for a semantic network.

The Relevance of Relevance

Stuart C. Shapiro, and Mitchell Wand
Computer Science Department, Indiana University, Bloomington 47401

Technical Report No. 46, March 1976

Classical material implication is unsuitable for many AI and NL purposes becauses it leads to the deduction of theorems which are irrèlevant in that they are semantically unrelated to the statements from which they are deduced (how useful is the knowledge that a false statement implies any statement?). A "fail-safe" heuristic is: "do not assert 'A implies B' unless the hypothesis A was actually used in the proof of B." A formal system embodying this heuristic is the system R of Relevant Implication of Anderson & Belnap (1975). Two examples are discussed. One involves the introduction of universes of discourse, and the other involves the deduction of new rules of inference.

International Symposium on Theoretical Programming

Andrei Ershov, and Valery A. Nepomniaschy
Computer Center, Informatics Division, Novosibirsk 630090, USSR

Springer-Verlag, 1974
ISBN 3-540-06720, Berlin - Heidelberg - New York, DM 30,-
ISBN 0-387-06720-5, New York - Heidelberg - Berlin,

Proceedings of the Symposium on Theoretical Programming held in Novosibirsk, August 7 - 11, 1972. Papers include: "An axiomatic definition of the programming language PASCAL," C.A.R. Hoare; "Central technical issues in programming language design," J. Schwartz; "On universal classes of program schemas," B.A. Trachtenbrot; "Criteria for the algorithmic completeness of the systems of operations," V.A. Nepomniaschy; "Looking for an approach to a theory of models for parallel computation" A.S. Narinyani; "Configurable computers: a new class of general purpose machines," R.E. Miller, J. Cocke; "Towards automatical construction of parallel programs," V.E. Kotov; "A calculus for the mathematical theory of computation," R. Milner; "Some features of a language for a proof-checking programming system," G S. Tseytin.

Breaking the Complexity Barrier Again

Terry Winograd
Artificial Intelligence Laboratory, Computer Science Department, Stanford

SIGPLAN Notices 10, No. 1, January 1975, combined with SIGIR Forum 9, No. 3, Winter 1974, 13-30.

Large programs such as the SHRDLU system, are so complex that they are almost impossible to handle. In order to deal with programs an order of magnitude larger than current programs we need an interactive programming system which can function as a moderately stupid, but very patient, assistant which combines the powers of a compiler, and interpreter, and editor, a debugging system, a documentation system, and a problem solver. The user should be able to intermix writing a program, editing it, running pieces of it, asking questions about it, stating new information increasing the store of abstract concepts for describing it, and debugging, all without switching between systems, writing and loading files, etc. The system should be based on an explicit modeling of the programming world, as SHRDLU is based on a model of the BLOCKS world. This project is so complex that it presents us with the paradox of fighting complexity with complexity. The article is followed by transcripts of discussion.

The Utility of Natural Language Systems

Ashok Malhotra, and Irving Wladawsky
Computer Science Department, IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

IBM Research Paper RG 5739, November 25, 1975

Natural languages and computer languages are suited to quite different kinds of applications. Formal languages are best where the user knows exactly what the system can do, how to use it, and instructs it to his own ends. Natural language systems are more appropriate where the user is uncertain about the contents and use of the system and/or uncertain about the problem domain. A NL interface would allow the user to start work on a problem in spite of these uncertainties and it would provide facilities for learning about the system and the problem domain. Such system should contain knowledge about the problem domain, the requests appropriate in it, and the linguistic strategies used to specify them. It should be able to use this knowledge and information about the context and the user's intentions to understand the sentence, including being able to complete and understand partially specified requests, fuzzy words,. and perhaps ambiguous requests.

On the Formal Documentation of Programs

Martin Mikelsons, and Irving Wladawsky
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

IBM Research Report RC 5788, 5 January 1976

Programs can be formally documented through two interlinked models: 1) the program model, which contains the semantic properties of the different tokens in the programming language, and 2) the application model, which contains the properties of a set of application concepts meaningful to the users. Application and programming knowledge are represented in a semantic network of nodes and directed arcs. The program model consists of the basic concepts of the programming language organized as a tree, roughly corresponding to the abstract of the language. The application model follows the organization of the program and program model. Without the application model, the only link between the nodes in the program and program model and the application would be a meaningful label, i.e. "Invoice" for a document in the program corresponding to the equivalent business document. The application model contains whatever additional information we wish to associate with the concept "Invoice" as used in this program. The system is being implemented for accounting programs written in BDL; examples of networks are given.

The Realization of the Anglo-Russian Multiaspect Automatic Dictionary with Grammatical Servicing on a Third-Generation Computer

S. A. Yakhontov

Machine Translation and Applied Linguistics 17: 106-116. 1974

Some ideas about the practical realization of a MT system on a type 'R'AD' computer are described. Specifically, the set of programs and the structure of the ARMAD dictionary needed for morphological analysis of NL texts. The set of programs is written in assembly language. It makes extensive use of the facilities of Disk Operating System for computers of the third generation. (In Russian)

Programming Languages, Natural Languages, and Mathematics

Peter Naur
Datalogisk Institut, Copenhagen University

Communications of the ACM 18:676-683, December 1975

Several of the social aspects of mathematics and natural language show a meaningful analogy with similar aspects of programming languages. It therefore makes sense to extrapolate the analogy to other aspects. On this basis the following conclusions may be drawn: the split between the more academic, pure computer science oriented study of programming languages and the world of practical programming will persist indefinitely; the era of influential programming language construction is past, Fortran and Cobol will retain their dominance; the existing programming languages will develop slowly, with only weak interaction among them; the areas of widest influence from scholarly studies on programming at large will be program literature and style, the most important medium of influence being textbooks at the fairly elementary level; greater interest in empirical studies of programming may be expected to develop in the future.

The List Processing Language SLPL - An Informal Outline

L. V. Atkinson
Department of Applied Mathematics and Computing Science, University of Sheffield, S10 2TN, UK

The Computer Journal 19: 32-39, 1976

SLPL was designed as an extension of ALGOL 60 concepts to provide a more versatile language to be used by undergraduates, primarily in conjunction with courses in list processing, AI, and compiling techniques. SLPL combines the security of compile-time mode compatibility with the flexibility of run-time mode checking. The degree of the flexibility/security trade-off is under the control of the user. SLPL is designed to encourage structured programming; list structures and a number of predefined operators and system primitives to process these structures are provided. The main concepts of the language are outlined and a sample program is given in the appendix.

Some Steps Towards a Better Pascal

Bruce Knobe
Intermetrics, Inc., 701 Concord Ave., Cambridge, Massachusetts 02138

Gideon Yuval
Computing Science Department, The Hebrew University of Jerusalem, Israel

Journal of Computer Languages 1: 277-286, 1975

A change is *compatible* if it is close enough to the original PASCAL that the transformation to the new syntax (or semantics) is a trivial, automatable project. Compatible changes are suggested concerning: program readability: constant declarations, automatic indentation to facilitate readability of the listing; variable dimensions: array allocation, function types and infix operators, FOR statements; in-line code; 3 more functions of the WRITELN type; compiler directives; cross reference listing. These features can easily be added to the ETH implementation for the CDC 6000 series.

Cognitive Psychology and Programming Language Design

B. Shneiderman
Department of Computer Science, Indiana University, Bloomington

SIGPLAN 10,7: 46-7, July 1975

Programming language designers must become familiar with the ideas and techniques of cognitive psychology. This will help in the development of the next generation of programming languages and will facilitate more widespread computer literacy.

Use and Maintenance of a Data Dictionary

William C. Dunn, and Beatrice Yormark
The Rand Corporation, Santa Monica, California

Rand Paper P-5324, November 1974

The Health Insurance Study (HIS) is a complex multi-yeared longitudinal study of the role of health insurance in the utilization of health care services Because of the nature and size of the data base it is important that a systematized archive be maintained which presents an everview of the data contained in the data files The Data Element Dictionary is central to this archive, providing a formal reference point for the proper usage, meaning and manipulation of the data elements The DED consists of two physical files: a data description file, a statistics file By including a statistics file as part of the logical DED, the concept of a "dynamic" dictionary is introduced; one which is defined partly at the time new data elements are defined and partly upon inspection of the data. The contents of the DED can be classified under two categories: one providing the researcher/data element interface and the other the program/data interface. To meet these needs the dictionary entry for each data element contains: context information, statistical information, attribute information.

System Organizations for Speech Understanding: Implications of Network and Multiprocessor Computer Architectures

Lee D. Erman, Richard D. Fennell, Victor R. Lesser, and D. Raj Reddy
Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

IEEE Transactions on Computers 25: 414-421, April 1976

The structure of the Hearsay system is based on a set of cooperating, independent processes using the hypothesize-and-test paradigm The system is implemented as a small number of parallel coroutines (modules), each realized as a separate job in the PDP-10 time-sharing system, thus the time-sharing monitor is the primary scheduler for the modules. The acoustic, syntax, and semantic modules are each linked to ROVER (*R*ecognition *OVER*lord), which handles the inferface between them. Futher exploitation of parallelism in speech understanding is being investigated by implementing a parallel decomposition of Hearsay on C mmp, a closely coupled network of PDP-11's which communicate through a shared memory. Finally, the problem of resource sharing in a large loosely coupled computer network (the ARPA-net) is discussed.

Learning Structural Descriptions from Examples

Patrick Henry Winston
Massachusetts Institute of Technology, Cambridge

In Patrick H. Winston, Ed., The Psychology of Computer Vision, McGraw-Hill, 1975
pp. 157-209, ISBN 0-07-071048-1

A semantic network is used to build representations of simple three-dimensional objects and scenes of such objects. Scenes can be matched by examining the networks describing the two scenes; the *skeleton* describes those parts of the compared networks that correspond while *comparison notes* (C-NOTE) are attached to the skeleton and describe the nature of the correspondence. The program can learn to recognize simple configurations by exploiting the *near miss*. Examples are given in a carefully constructed training sequence so that the machine can learn about an object by exploiting the similarity between a valid example and one which is *almost* a valid example. Some of the concepts built up so far: house, tent, arch, table. The program also has the capacity to recognize symmetry along three axes by generating a copy of the scene description and converting LEFT-OF pointers to RIGHT-OF pointers, and vice versa, (or ABOVE BELOW IN FRONT OF/BEHIND) and then comparing the original description against the modified copy. If the match is exact, then the scene is symmetric along the tested axis. Evan's analogy program is also discussed.

The Psychology of Computer Vision

Patrick Henry Winston, Ed.
Massachusetts Institute of Technology

McGraw-Hill Book Company, 1975
ISBN 0-07-071048-1 282 pp. $19.50

* These chapters have been abstracted elsewhere on the fiche.

A Multidimensional Approach to Syntactic Pattern Recognition

K. L. Williams
Department of Mathematics, Western Michigan University, Kalamazoo

Pattern Recognition 7: 125-37, September 1975

A syntactic method for representing the primitive parts of a pattern as nodes of a type of directed graph is described. A linear representation of the digraph can then be presented to a regular unordered tree automaton for classification. Regular unordered tree automata can be simulated by deterministic pushdown automata, so this procedure can be implemented easily. Regular u-tree automata and the corresponding generative systems, regular u-grammars, are formally defined. Several results are shown which are applicable to all syntactic pattern recognition schemes involving the use of primitives.

Recent Progress to Formal Approach of Pattern Recognition and Scene Analysis

J. C. Simon
Inst. de Programmation, Univ. Paris VI, France

Pattern Recognition 7: 117-24, September 1975

PR techniques and programs may currently be considered as a bag of tricks, justified by their experimental results. Formalization and theory have been lagging. Propositions are made for a formal PR language, similar to the language of logic, which could be used to describe procisely the PR algorithms. Interpretations of the terms of this language have to be made in other languages: natural, machine or programming languages. A PR function should be constructive, i.e. find its interpretation in an algorithm or program. A training set is far from enough to define such a function. Properties of the experimental domain should guide the PR specialist. They are examined and discussed, in an effort to find some unity in the PR techniques.

Library Automation and Information Science in Ethiopia

C. Darch
University of Dar es Salaam, Tanzania

Network 2, No. 3-4: 19, 30, March-April 1975

The role of information science and its techniques, in particular, the position of computerized libraries and information systems, in Ethiopia is surveyed. The library system of the National University of Ethiopia, which is the only library and information network outside the government which is large enough to either need or consider computer applications, is outlined.

The Use of Title and Cited Titles as Document Representation For Automatic Classification

Division of Medical Systems, Memorial Sloan-Kettering Cancer Center, U.S.A.

Information Processing and Management 11: 201-206, 1975

The use of title and cited title words as document representation offers a method intermediate between the use of title and abstract of a document and that of citation identities, retaining some advantages of both. Compared with title and abstract it leads to more compact and uniform document representation with a high concentration of indicative words, gives more consistent coupling strengths to profiles with results agreeing with that employing citations, and offers a more consistent ability for inter-group diferentiation when the groups are close to each other. Compared with the use of citations, it gives results with less specificity and operationally requires an extra step to input and analyse the full citation titles. However, the group profiles derived from title and cited titles are words and can be used to classify documents that have descriptive abstracts but no or few citations

A Vector Space Model for Automatic Indexing

G. Salton, A. Wong, and C. S. Yang

Communications of the ACM 18:613-620, November 1975

The best indexing (property) space is one where each entity lies as far away from the others as possible; in these circumstances the value of an indexing system may be expressible as a function of the density of the object space; in particular, retrieval performance may correlate inversely with space density. Good discriminator terms must be those with uneven occurrence frequency distributions which cause the space to spread out when assigned by decreasing the similarity between the individual documents. Terms of medium frequency should be used for content identification directly, without further transformation. Terms with very high document frequency should be transformed into entities of lower frequency--perhaps by using them as components of indexing phrases. Terms with very low frequency should be transformed into entities of higher frequency--perhaps by grouping them into common term (thesaurus) classes.

A Theory of Indexing

Gerard Salton
Cornell University

Society for Industrial and Applied Mathematics, Regional Conference Series in Applied Mathematics 18:1975

Five types of significance measure are discussed and evaluated: discrimination values, inverse document frequencies, signal-noise values, variance-based measures, and information values. Good and bad index terms are characterized by objective measures, leading to the conclusion that the best index terms are those with medium document frequency and skewed frequency distributions. A discrimination value model is introduced which makes it possible to construct effective indexing vocabularies by using phrase and thesaurus transformations to modify poor discriminators--those whose document frequency is too high or too low--into better discriminators, and hence more useful index terms. Test results are included which illustrate the effectiveness of the theory. 56 pp., 32 refs.

The Computation of Discrimination Values

Robert G. Crawford
Department of Computing and Information Science, Queens University, Kingston, Ontario, Canada

Information Processing and Management 11: 249-253, 1975

Two algorithms for computing the discr mination values are given. The more efficient one is:

0. Given: A file of document vectors *d*1, *d*2,....., *d*N. An inverted file listing for each term those documents in which it occurs. *c*, the centroid of the document space. SUMSQC [a property of the centroid]

1. For each document j = 1, ..., N DO

    1.1 Compute SUMSQDj

    1.2 Compute DOTPRODj

    1.3 Update QSUM (QSUM = QSUM + COS (*c*, *d*j))

2. For each term i = 1,..., M DO

    2.1 Qi = QSUM

    2.2 For each document j *in which term i occurs* DO

    2.21 Qi = Qi - COS (*c*, *d*j) + COS (*c* to the i, *d*j to the i)

    2.3 compute discrimination value for term i. Di = 100* (Qi - QSUM)QSUM

Automatic Indexing Using Term Discrimination and Term Precision Measurements

G. Salton, and A. Wong
Department of Computer Science, Cornell University, Ithaca, New York 14853

G. T. Yu
Department of Computing Science, University of Alberta, Edmonton, Canada

Information Processing & Management 12: 43-51, 1976

A term discrimination model and a term precision system are described and experimental evidence is cited showing that a combination of both theories produces better retrieval performance than either one alone. In the discrimination model the full indexing vocabulary is retained while the terms whose frequency characteristics are nonoptimal are transformed into indexing units with better assignment frequencies. This model does not account for the linguistic or semantic aspects of the texts being processed. The term precision method remedies this situation somewhat by utilizing customer opinions concerning the relevance or nonrelevance of certain documents to the queries submitted to the system. A precision weight attached to each query term is used as a partial indication of the lingistic characterization of the terms.

Automatic Indexing System for German-Language Documents

I B. Arzumanova, L A. Khoklova, and B. R. Pevzner

Nauchno-Teknischeskaya Informatsiya, Seriya 2, 6:21-23, 1975

The core of the system is a German to descriptor dictionary in electrical engineering and an algorithm for morphological analysis of compound words, which splits these words into their semantic components. (6 refs., in Russian)

Natural Language Access to a Large Data Base

David L. Waltz
University of Illinois at Urbana-Champaign, Coordinated Science Laboratory

National Technical Information Service: AD-A013 578/0GA, April 1975

The report describes the first year's accomplishments toward a natural language system which answers questions about a data base of naval aircraft maintenance and flight data. The system is designed to: Allow a user to ask questions in natural English; provide answers to questions requiring averaging, statistical analysis, comparison of sets of data, and other complex functions as well as answers to simpler questions about specific data base records; provide aid in evaluating and predicting the causes of failures and of the need for excessive amount of maintenance work.

The Importance of User Education and Training in a multi-data Online Information Network

J. Egeland
State University of New York, Albany

Proceedings of the ASIS 37th Annual Meeting, vol. 11,
Washington, D. C., American Society for Information Science, 1974
ISBN 0-87715-411-2 : 137-140

The SUNY Biomedical Communication Network provides online access to three different data bases to 32 member institutions across a 10-state area. Through a single terminal located in their library, users may search the MEDLARS, ERIC, and Psychological Abstracts data bases. As part of this service to member institutions, SUNY also provides training sessions in the use of the system. This paper focuses on the importance of user awareness of the differences in the indexing policies and vocabulary structures for each of the files. More attention should be directed to the development of this type of training for users of multi-data base online systems. The need for cooperation between data base vendors and producers in this regard is noted.

Automatic Abstracting by Applying Graphical Techniques to Semantic Networks

S. L. Taylor
Northwestern University, Evanston, Illinois

Thesis, University Microfilms, Order No. 76-8002, 1974

A method is presented for processing semantic networks using graphical techniques in order to produce a reduced semantic network while retaining the important concepts or topics of the original semantic network. The processing method is considered as part of an abstracting system composed of three parts: a parser that constructs a semantic network from discourse, a reduction that executes the processing method, and a generator that generates discourse from the reduced network. The processing method consists of two phases. The first phase processes the semantic network as a directed graph using graph clustering techniques. The second phase then analyses the largest cluster found using signal flow graph analysis.

Methods of Automatic Abstracting (USA, 1958-1974)

V. P. Leonov

Nauchno-Tekhncheskaya Informatsiya, Seriya 2, 6: 16-20, 1975

The current methods of abstracting in the USA are discussed: statistical, logico-mathematical, and linguistic. The linguistic is especially promising. General views are given on the quality and outlook for machine abstracting. (24 refs, in Russian)

Chinese-English Translation System

William S-Y Wang, Stephen W. Chan, and Philip Robyn
Department of Linguistics, University of California, Berkeley 94720

Rome Air Development Center, Technical Report RADC-TR-76-21, February 1976

Research was focussed on integrating the QUINCE system modules into a completely sequenced system of programs during execution of the translation process. Interface modules for the front end of the system were written to enable all machine coded texts to be normalized before input to the actual analysis process. All texts entering the system are designated by a decimal reference so that each sentence or subpart of a sentence could be cross-referenced for retrieval and additional analysis. Programs for interfacing the sentence dictionary with the text normalization programs were completely designed but not implemented (funding problems prevented this). The complete text of a physics monograph was fully coded, revised for errors and machine processed into a normal form into data suitable for direct input into the QUINCE system. Appendices include: Grammar Code Description; Grammar Rules; Programs, Texts and Data on File. 94 pp.

Search. Material for the Anglo-Russian Multiaspect Automatic Dictionary

L. L. El'nitskii

Machine Translation and Applied Linguistics 17: 100-5, 1975

Presents three dictionary entries describing two of the meanings of the English verb/search/ and a noun derived from one of them. Among other things, the entries contain an exhaustive formal definition of the primary lexical meaning of the verb as well as some rules for idiomatic translation of the words described. (In Russian)

Signal 1 and Signal 2. Materials for the Anglo-Russian Multiaspect Automatic Dictionary

T. V. Pivovarova

Machine Translation and Applied Linguistics 17: 94-9, 1974

Dictionary entries for two meanings of /signal/ as used in radio and electronics. The first meaning is characterized by a wide range of lexical functions while the second one is an example of a common value of the lexical function Magn. (In Russian).

Require. Materials for the Anglo-Russian Multiaspect Automatic Dictionary

T. G. Potemkina

Machine Translation and Applied Linguistics 17: 86-93, 1974

A dictionary entry is given for one of the meanings of the English word /require/. Each section of the entry is supplied with a commentary. The following sections are of particular interest: the rules relating syntactic structure to semantic representation; the syntactic zone, where two modifications of the syntactic pattern of the word in question are described; sections dealing with the transformations of Russian and English structures incorporating this lexical unit. (In Russian).

Pronoun IT in the Anglo-Russian Multiaspect Automatic Dictionary

A S. Chekhov

Machine Translation and Applied Linguistics 17: 68-85, 1974

The paper contains lexical entries of the English-Russian Multiaspect Automatic Dictionary which describe the pronoun IT and some commentaries thereto. (In Russian)

The Anglo-Russian Multiaspect Automatic Dictionary (ARMAD)

Z. M. Shalyapina

Machine Translation and Applied Linguistics 17: 7-67, 1974

The Anglo-Russian Multiaspect Automatic Dictionary (ARMAD) is described with respect to the MT system in which it is used. ARMAD treats NL texts at three levels of representation: morphological, syntactical, and semantic. Each of these is discussed at some length. (In Russian)

Braille Programs

P. Coleman
Warwick University, Coventry, England

Computer Weekly 19, 459: 6, 21 August 1975

Programs for Braille translation are needed. Some of the possible programming and language techniques are discussed.

Documentation in Social Science Experiments

Michel M. Rogson
The Rand Corporation, Santa Monica, California

Rand Paper P-5494-1, January 1976

Systematic and well documented data accrual processes are essential. In a large experiment a maximal set of documentation causes the volume of this documentation to approximate that of the data collected. This documentation should be organized by a glossary of key terms, a dictionary that defines all data elements, and codebooks that describe the data elements and transactions affecting them. All access to the data would be effected through the central dictionary. If a suitable dictionary scanning tool is prepared, access to the data base documentation could also effected through the dictionary. Given this structure of documentation organization, the researcher's view of the data base organization and its physical representation are independent of each other. This independence permits the implementation of standardized access and update methods that need not depend on the data base content or organization at any point in time.

A Representation of Systems of Concepts by Relational Structures

Klaus Witz
University of Illinois, Urbana

John Earls
Department of Anthropology, University of Michigan

Paul A. Ballonoff, Ed., Mathematical Models of Social and Cognitive Structures, Illinois Studies in Anthropology, No. 9, University of Illinois Press: 104-120, 1974
ISBN 0-252-00415-9

Conceptual-semantic systems can be represented by *relational structures* which permit the use of mathematical structures such as homomorphisms, products, and partial orderings in the analysis of conceptual structures. A relational formalism is sketched out briefly and applied to: 1) Thai ideas about the relationship between a person and the territory he controls, 2) Homomorphism in the conceptual structure underlying metaphysics of a fragment of the Tewa origin myth, 3) Analysis of the concept of authority based on transcripts of discussions with a single informant in a small Illinois town.

Social Structure, Social Classifications, and the Logic of Analogy

Francois Lorrain
Society of Fellows, University of Michigan

Paul A. Ballonoff, Ed., Mathematical Models of Social and Cognitive Structures, Illinois Studies in Anthropology, No. 9, University of Illinois Press: 89-103 1974
ISBN 0-252-00415-9

When considering kinship of myth, some anthropologists--especially Levi-Strauss--have stated analogy relations of the type "A is to B as C is to D." Phonologists of the structuralist school often stated such analogy relations among phonemes. If certain natural axioms, stated below, govern these analogy relations, then the system described by these relations exhibits strict structural properties; it consists of various criss-crossing series of parallel lines or cycles. These structures are investigated in: a) componential analysis, classificatory kinship, elementary structures of kinship (section systems and the relationship nomenclatures associated with them), b) structural time-invariance, c) egocentric and sociocentric representations of a network of social relationships.

A Spreading-Activation Theory of Semantic Processing

Allan M. Collins
Bolt Beranek and Newman Inc., Cambridge, Mass.

Elizabeth F. Loftus
University of Washington

Psychological Review 82:407-428, 1975

Quillian's theory of semantic memory search and semantic preparation or priming is the basis for a spreading activation theory of human semantic processing which can be applied to a wide range of recent experimental results. A number of additional assumptions are proposed for Quillian's theory in order to apply it to recent experiments: production experiments by Loftus, Juola and Atkinson's multiple category experiments, Conrad's sentence-verification experiments, and several categorization experiments on the effect of semantic relatedness and typicality by Hollyoak and Class, Rips, Shoben, and Smith, and Rosch. A critique of the Smith, Shoben model for categorization judgements, in which concepts are represented as bundles of semantic features, is offered. A revised version of the Smith *et al.* model can be seen as a special case of a more general procedure specifiable in our theory.

Set-Theoretic and Network Models Reconsidered: A Comment on Hollan's "Features and Semantic Memory"

Lance J. Rips
Department of Behavioral Sciences, University of Chicago

Edward E. Smith
Stanford University

Edward J. Shoben
University of Illinois

Psychological Review 82:156-157, 1975

The choice of a network or set-theoretic representation correlates with two substantive differences among models of semantic memory. First, in explaining certain reaction time effects network models emphasize retrieval processes, while set-theoretic models often emphasize comparison processes. Second, set-theoretic models posit more semantic analysis during sentence verification than do network models. Hollan's recent demonstration of an isomorphism between set-theoretic and network representations cannot resolve these important issues.

Features and Semantic Memory: Set Theoretic or Network Models?

James D Hollan
Department of Social Sciences, Clarkson College of Technology, Potsdam, New York 13676

Psychological Review, 82:154-155, 1975

Smith, Shoben, and Rips (1974) propose a set-theoretic model (sets of semantic features) of semantic memory. Their set-theoretic model can, without loss of explicatory power, be formulated as a network model and therefore the distinction between set-theoretic and network models is vacuous. The network model is preferable because: 1) it can call on the extensive literature of graph theory, 2) graph theoretical models can easily be implemented as computer programs.

The Measurement of Meaning in Psychoanalysis by Computer Analysis of Verbal Contexts

Hartvig Dahl
Downstate Medical Center, 450 Clarkson Avenue, Brooklyn, New York 11203

Journal of the American Psychoanalytic Association 22:37-57, 1974

Twenty-five hours of abbreviated transcripts out of 363 on tape of the analysis of one patient were analyzed: 10 hours showing significant "analytic work" by the patient, 10 hours of "resistance" to analytic work, and 5 "middle" hours. A computer searched the texts of the patient's portion of the 25 hours for words from the Harvard III Psychosociological Dictionary (Stone et al., 1966). For 20 of the 83 categories from the dictionary there were highly significant linear relationships between the measure of analytic work and the percentages of the categories in each hour. Further analysis used 4 methods for defining contexts of word usage: 1) Are particular words more highly correlated with either the denotative or the connotative category to which the Dictionary assigns them? 2) A listing of words with which a particular word is correlated--done only with words selected on the basis of their high correlation with one of the dictionary categories. 3) Factor analysis of a correlation matrix of 47 words using a Principal Components solution with a varimax rotation of 10 factors. 4) In which one takes words from a given factor, finds an hour in which they all occur, and then lists the words with the line number of the text where they were found. The evaluation of the results of this procedure depends on clinical judgement.

A Quantitative Study of a Psychoanalysis

Hartvig Dahl
Research Center for Mental Health, New York University, New York City

R. Holt and E. Peterfreund, Eds., Psychoanalysis and Contemporary Science, Vol. 1; The Macmillan Company, New York: 237-257, 1972

Data from 363 tape-recorded psychoanalytic sessions with one patient were analyzed. P-technique Factor Analysis of a correlation matrix of 53 analyst-coded variables yielded 6 factors (Resistance I, Family-Genetic, Sexuality, Dreams, Anxiety, Interaction-Resistance II) and produced estimates of the amount of each factor in each hour. These scores were plotted against time. A measure of "analytic work" vs. "resistance to analytic work" was derived by combining the resistance factors into one factor *R* and the other four into another factor *C*. *C* - *R* measures the amount of analytic work done by the patient. Twenty-five hours were selected for analysis using the Harvard III Psychosociological Dictionary. "High work" hours were significantly different from "resistance" hours in terms of the 83 categories of the dictionary. Finally, analysts were asked to rate 8 of the 25 hours for predominance of analytic work or resistance. This measure confirmed the results of the factor analysis and the content analysis.

Cognitive Structures in Human Story Comprehension and Memory

Perry W. Thorndyke
The Rand Corporation, Santa Monica, California 90406

Rand Paper P-5513, September 1975
174 pp., $7.00

Simple narrative stories contain setting, main character, problem to be solved, plot sequence, and resolution. A process model for the comprehension of these stories assumes that stories are encoded in a hierarchical organizational framework which represents the abstract structural relationships of the plot. Four experiments on text comprehension were run. In Experiment I subject's recall of a story was found to be a function of the amount of inherent plot structure in the story. Experiment II extended the results of I and found that repeating story structure across two passages produced proactive intereference. Experiment IV tested a model of the use of inferences in comprehension. False recognition rates for implicit inferences from a story depended on the plausibility of the inferences and their role as organizing and integrating devices for other information in the story.

Computer Simulation of Two-Person Interactions

Thomas W. Malone
Rice University, Houston, Texas

Behavioral Science 20: 260-267, July 1975

Using roles as determinants of action makes possible the elimination of complicated goal-directed decision making and consideration of belief structures. Four data elements: 1) Description of the roles of the two people, 2) Description of responses they make to each other, 3) Description of the situation, 4) Description of any necessary universal constants. These descriptions are made as ratings on various axes. Two functions: 1) Response generator using 1-4 (above) to produce the next description of a response, 2) A learning function modifies roles as a function of 1-4. An optional third function would interpret the response descriptions into English dialogue. The program DYAD is based on Leary's personality theory (1957) which has 16 categories of interpersonal response arranged on 2 axes, dominance-submission, attack-affection. Each person's role is represented as a probability distribution over the 16 categories. The response generator is based on a principle of reciprocal interactions which is simulated as a transition matrix and the learning function is based on a model of positive and negative reinforcement. Experiments with the model are described.

Representing Logical and Semantic Structure of Knowledge Acquired from Discourse

Carl H. Frederiksen
Department of Psychology, University of California, Berkeley, 94720

Cognitive Psychology 7:371-458, July 1975

The model: 1) Semantic structures: propositions are represented as networks of concepts. 2) Logical structures: networks of propositions are connected by various labeled logical, causal, and algebraic relations. 3) Detailed consideration is given to the algebraic (and metric) properties of noncase (classification, attributive, degree, temporal, locative, quantification) relations and certain relations having specified algebraic properties (e.g. transitivity, symmetry, reflexivity) which may be used to connect propositions containing these noncase relations are defined. The result is a mechanism for representing comparative relations, and relations of relative time and location, tense and aspect. 4) A stochastic element is included to deal with imprecision and linguistic "hedges." 5) Distinction is made between "symbolic" and "non-symbolic" objects and "cognitive and "noncognitive" actions, both of which involve symbolic content. To assess acquired knowledge, a procedure is presented for coding a subject's verbal reconstruction of knowledge acquired from a presented text (or other input) against the logical and semantic structure from which the text was derived. Experimental data are presented.

In Search of Buber's Ghosts: A Calculus for Interpersonal Phenomenology

Burton L. Alperson
California State University at Los Angeles

Behavioral Science 20: 179-190, May 1975

A Boolean analysis of the following three independently developed methods for the study of interpersonal phenomenology reveals that they are isomorphic to one another: Interpersonal Perception Method (Laing, Phillipson, and Lee, 1966), Interpersonal Perception Technique (Drewery, 1969,), Family Relationship Test (Scott, Ashworth, Casson 1970). Their shared structure makes it possible to develop a single lucid and rigorous language for the full interpretation of each. This language removes the need for intuitive derivations of terms, reduces semantic confusion, clarifies relations among terms, provides a basis for new applications, and reduces scoring effort by over 95 percent. Possible applications: the study of process and outcome in marital therapy, the phenomenology of different ethnic groups, the study of communication and attribution in the families of schizophrenics.

The Psychological Unreality of Semantic Representations

J. D. Fodor
Department of Linguistics, University of Connecticut, Storrs 06268

J. A. Fodor, and M F. Garrett
Department of Psychology, MIT, Cambridge, Mass. 02139

Linguistic Inquiry 6: 515-531, Fall 1975

Both generative and interpretive semantics assert the necessity for rules of eliminative definition. However, there is no convincing evidence for the psychological reality of such rules. Intuitive arguments are given against the reality of eliminative definition (consider the eliminative definitions of *the* and *a* for a start) and experimental evidence concerning reaction time to attain a correct evaluation of sentences containing various types of negatives suggests that such a level is unreal. If our arguments are sound, then it appears practically mandatory to assume that *meaning postulates* (similar in thrust, but not in detail, to Kintsch, 1974) mediate whatever entailment relations between sentences turn upon their lexical content.

System Developments in the ACT Language: Towards Machine Independence

J. R. Millenson
McGill University, Montreal, Quebec, Canada

Behavior Research Methods and Instrumentation 7, No. 2: 164-73, March 1975

ACT (the Automated Contingency Translator) is a list processing state oriented sequence control language for on line control and data acquistion of psychological experiments. Since 1967 ACT software has been progressively expanded and the machine base extended from the PDP 8 family to the PDP9, NOVA, and the PDP11. The most recent variant of this language, ACTN, removes previous arbitrary restrictions on state network complexity, expands the conversational repertoire, and adds a subset of compatible BASIC to ACT, thereby giving the package greatly increased computational powers and data storage facilities.

Hierarchical Man: A Comparison of Three Cybernetic Systems

W. Stallings
Honeywell Information Systems Inc., Waltham, Massachusetts

Kybernetes 3: 195-201, October 1974

The cybernetic view of man holds that man's behavior and experience can be accounted for by feedback-control processes which are hierarchically organized. The ideas of Koestler, Laszlo, and Powers are examined. Despite differences in detail they articulate remarkably consistent theories of the nature of man.

An Information Processing Constraints Approach to the Conjunction of Macroeconomic and Macropolitical Theory

W. E. McAlpine
Georgia Institute of Technology, Atlanta, 30332

Information Processing and Management 12: 1-17, 1976

Human information processing is a fundamental and critically important resource for political life. It is moreover a constrained resource. Using this fact, connection can be established between ideas employed in political science and concepts in macroeconomic theory. A concept of "interest rates" can be derived from information processing constraints. These "interest rates" are analogous to the "arousal level" construct of motivational psychology. A generalized Keynesian "general equilibrium" model can be derived largely from macropolitical considerations. In the space defined by activity in the system (a generalization of GNP) and interest rate one has a governance of production" curve G (a generalization of the Keynesian LMc curve) and an "assimilating production" curve A. The positions of G and A in the space are determined by: 1) the "well directedness" of intellectual structures with respect to either governance or assimilation, 2) the "risk-readiness" of society with respect to failures of governance or assimilation, and 3) the character of the environment in terms of the type of problems generated for governance and for assimilation.

Situation and Prospects of Computer-Aided Literary Research in Spanish

Leopoldo Saez-Godoz
Institute for Communications Research and Phonetics, University of Bonn

Computers and Humanities 9: 245-6, September 1975

There has been relatively little computer-aided work done on Spanish. And there is little in the way of analysis or explanation of the theoretical basis of the work (mostly concordances) which has been done. In order to facilitate more work in the field, an information center on the use of computers in Spanish has been founded at the Institute for Communication Research and Phonetics at the University of Bonn.

Trends in Computer Applications to Literature

R. L. Widmann
Department of English, University of Colorado

Computers and the Humanities 9: 231-235, September 1975

Too many humanities students remain ignorant of computer methodologies. Perhaps graduate students should be allowed to offer competence in a computer language instead of the more traditional French, German, or Russian. Other topics discussed: concordances, historical lexicology, stylistics, content analysis, MT.

Directory of Scholars Active

Computers and the Humanities 9: 187-196, July 1975

Areas of work: Education, General, Language and Literature, Music, Philosophy, Social Science, Visual Arts

Annual Bibliography for 1974 and Supplement to Proceding Years

Stephen V. F. Waite
Kiewit Computation Center, Darmouth College

Computers and the Humanities 9: 127-144, May 1975

Topics covered and number of references per topic: General-42, Archaeology-84 History-245. Language and Literature-265, Music-37, Visual Arts-69.

The Humanist in the Computer Lab

Joseph Rabin
Department of English, Queens College of the City University of New York, Flushing 11367

Visible Language 8:167-177, Spring 1974

The most substantial accomplishments to date have been the rationalized lists of words (dictionaries, indexes and concordances) for which the computer's capacity to sort rapidly without fatigue or error makes it an ideal servant. A new brand of humanistic scholar now evolving--highly trained in the humanities and at the same time in those aspects of computer science genuinely relevant to his studies--will contribute to the creation of new programming languages specially designed for this work, assist in the training of others who follow and help to guide computer-assisted instruction beyond the mechanistic mode in which it currently operates.

Association for Literary and Linguistic Computing: Bulletin, Vol. 4, No. 1, 1976

Contents

HUMANITIES: ANALYSIS

Poetry Generation and Analysis

James Joyce
Computer Sciences Division, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

Morris Rubinoff and Marshall C. Yovits, Eds., Advances in Computers 13, Academic Press: 43-72, 1975.
ISBN 0-12-012113-1

The following topics and investigators are considered: *Poetry Generation*: Marc Adrian's work on concrete poetry, R. Caskins' generation of haiku, P. Kilgannon's use of computer generated poems as a basis for his own poems L. T. Milic. *Concordances*: the WYLBUR editor, the Cornell Concordance series, Shinagel's Swift concordance, Ingram and Swaim's Milton concordance, Spevack's Shakespeare concordance, Misek's concordance to *Paradise Lost*. *Stylistic Analysis*: the EYEBALL programs of Ross and Rasche, Green on formulas and syntax in Old English poetry. *Prosody*: Dilligan on quantative verse in Hopkins and Bridge's translation Book VI of the *Aeneid*. *Literary Influence*: Raben and Smith on Milton and Shelley. *Statistical Analysis*: Sainte Marie et al. on Moliere. *Mathematical and Statistical Modeling*: Edward Kahn's modelling of narrative structure in *The Faerie Queene*, Joyce's work on periodicity of pauses in poetry and prose. *Textual Bibliography*.

The Use of Function Word Frequencies as Indicators of Style

Fred J. Damerau
IBM Thomas J. Watson Research Center, Yorktown Heights, New York

Computers and the Humanities 9: 271-280, November 1975

Style is presumed to be independent of content. The evaluation is based on the observation that words not context dependent would occur in a Poisson distribution; context independence indicates that the usage of a word is not dependent on content. 60,000 word samples were taken from *Slaughterhouse Five, For Whom the Bell Tolls, Tropic of Cancer,* and 2 separate samples from *Vanity Fair.* Words which may have an ordinary looking Poisson distribution for one author are widely divergent from such distribution for another author. For some authors many words are Poisson distributed, for others few are. Analysis of the two *VF* samples indicates significant differences between them. It thus appears that the usage of function words is not context independent and therefore doesn't constitute a useful measure of style. An appendix gives a list of all words which occurred at least 5 times per 10,000 words in any sample. They are coded for the significance measure, modified mean, for each word in each sample.

A Shakespeare Dictionary (SHAD): Some Preliminaries for a Semantic Description

H. Joachim Neuhaus, and Marvin Spevack
Department of English, Westfalische Wilhelms-Universitat, Muenster, Germany

Computers and the Humanities 9: 263-270, November 1975

From lemmatization to final semantic description: A list of lemmata maybe taken as a skeleton dictionary. Lemmatization produces some internal structuring of entries: paradigmatic (differently inflected word forms of a lemma); variational; syntagmatic (contexts of token occurrence). Semantics: An adequate dictionary semantics is possible only when there is a shift away from the lemma as the focus of semantic analysis and description towards an analysis of the function of a token in particular contexts. Topics discussed: defective distributions (ex: *un-* words); variational phenomena (ex: *-eth* and *-es* in 3rd person singular); underlying systems (ex: *-ish* suffix in forming adjectives; *child, childish*).

The Algorithm and Some Results of a Statistical Investigation of Rhythm on the 'Minsk-32' Computer

V. S. Baevskii, and L. Ya. Osipova

Machine Translation and Applied Linguistics 17: 174-95, 1974

The representation of alternating verse rhythm in based upon a gradual numerical evaluation of degrees of intensity of syllables. The procedure includes the computation of a number of statistical characteristics. The algorithm of verse analysis was realized on the 'Minsk-32' computer. The results of a computer experiment on the study of trochaic tetrameter of Russian poets of the middle of the 20th century are given.

Literary Statistics. VI. On the Future of Lieterary Statistics

N. D. Thomson
IBM United Kingdom Laboratories, Ltd., Winchester, England

Association for Literary and Linguistic Computing, Bulletin 3: 166-171, 1975

Discusses the future of literary statistics as a fully-fledged branch of applied statistics with its own special methodologies and a considered rationale for their use. The controversy centered around the philosophy of science issues concerned with the nature of inference and the application of significance tests is discussed.

Carlyle and the Machine: A Quantitative Analysis of Syntax in Prose Style

R. L. Oakman
University of South Carolina, Columbia

Association for Literary and Linguistic Computing.
Bulletin 3: 100-114, 1975

A large selection of Carlyle's prose is analyzed with linguistic and quantitative methods of syntactic analysis and a computerized parsing procedure. Two objectives. what are the stylistically significant elements of Carlyle's syntax and is large scale automatic syntactic analysis a profitable technique for use in describing prose style?

Using Computers in a Natural Language Mode for Elementary Education .

Alan L. Tharp, and Woodrow E. Robbins
Computer Science Department, North Carolina State University, Raleigh, 27607

International Journal of Man-Machine Studies 7:703-725, November 1975

The fact retrieval system (about history and geography) was designed for NL interactive use with fourth grade students. The parser uses an augmented transition network. The storage and retrieval system inspects output from the parser and enters information into the system if the input was a declarative sentence and retrieves the desired information if the input was a question. The original system has been modified to include: 1) An interactive program facilitating the addition of new words to the dictionary so that the teacher and more advanced students would be able to do this. 2) The parser was modified to convert input strings into a deep case structure rather than a deep structure. 3) An assembler language module was constructed to act as an interface permitting data and dictionary items to be individually accessed directly from disk during execution. 4) Information about synonym, antonym and implication relations was added to. the dictionary. 5) The capacity to decay the reliability of a fact with time (some things are true forever, other things are not).

Interactive Computer Simulations for Teacher Education

J. L. Flake
Mathematics Education Department, Florida State University, Tallahassee

Educational Technology 15, No. 3: 54-7, March 1975

A part of laboratory experiences for teacher education students includes interactive computer simulations of various classroom situations. Such simulations can be used to help teacher education students reach a high awareness level and can also be used to study individual teacher behavior. Simulations discussed here have focused upon teaching strategies and questioning behaviors.

MODELR--Model Building and Model Modification for Instruction

G. H. Shure, and K. Brainerd
University of California, Los Angeles

Behavior Research Methods and Instrumentation 7, No. 2: 221-5, March 1975

An integrated set of three computer programs that enable the implementation and student investigation of computer models of behavioral phenomena is described. These models are intended as tools for aiding in the instruction of undergraduate social science courses in research methods. Each program provides for natural language interaction with one of three classes of users: A 'Modeler', who implements a model of some behavioral phenomenon; an 'Instructor', who tailors models for his course of instruction, and a 'Student Experimenter', who applies an experimental design to the model and receives synthetic data in return. None of these users need have any prior computer expertise nor do they depend on external documentation on how to use the programs.

Interpretative Systems in Instruction

W. Glatthaar
University of Stuttgart, Germany

Angewandte Informatik 17, No. 3: 113-15, March 1975

Computers in education should not only be used for organizing instruction, but they should be available to support user's problem solving. This is achieved by programs explaining themselves. In case of operations of an interpretative system such as APL this can be performed by the proposed teaching strategy. (3 refs., in German)

Model-Building and Computer Simulation for Non-Programming Users

W.L. Bewley
Minnesota Educational Computing Consortium, Lauderdale

AEDS 13th Annual Convention, Washington, D. C ; Association for Educational Data Systems, 1975, 108-114

Learning involves building internal (mental) models of reality, testing these models against reality, and correcting (debugging) the models when errors are found. The computer implementation of this notion assumes no user knowledge of computer programming. The computer is programmed to perform several information processing operations appropriate for a certain class of models, e.g. queueing models. The user builds a model by specifying the sequence in which the operations are to be performed. The program then runs a simulation of the model. Because the user has not written a computer program, any errors revealed by the simulation can be attributed to the logic of the model. The user debugs the model by changing the sequence of operations. Two sample model building programs, Q and EAT, are described.

Exploring New Worlds

R. L. Ballard
Physics Computer Development Project, University of California, Irvine

AEDS 13th Annual Convention, Washington, C. C ; Association for Educational Data Systems, 1975, 95-98

"New Worlds" is the title of one Physics Computer Development Project proposal. It aims at developing breakthrough subjects, graphic communications, and intelligent NL interaction. Collectively, these elements become literal play worlds wherein subject knowledge can be represented, manipulated, and made into new conceptual experiences. Teaching and concept testing in higher education can look to Piaget-like approaches. The strategies for doing this are based on firm successes with several existing educational programs. The paper looks at the microstructure and macrostructure of one particular New World approach.

Teaching Computational Linguistics: A Continuation of the Discussion

M. King
Institute per gli Studi Semantici e Cognitivi, Castagnold, Switzerland

Association for Literary and Linguistic Computing, Bulletin 3: 161-165, 1975

In designing courses in computational linguistics for arts students, in particular, an optional postgraduate course in CL within the overall framework of an M.A. course in General Linguistics, it is both feasible and valuable to base such courses on a strong foundation of practical computing, providing that sufficient attention is paid to teaching students to program well and that students are encouraged to discriminate between problems suitable for computer solution and problems more aptly tackled by other means.

The Use of a Computer in Devising A Beginners' Latin Course

C. W. E. Peckett

Association for Literary and Linguistic Computing Bulletin 3: 158-60, 1975

There was a need to devise a one-year course of forty lessons for beginners at university level. It was clear that emphasis should be placed on the ability to translate, furthermore, the Direct Method could not be used. Target texts were chosen and analyzed by computer to ascertain the vocabulary and grammar. Texts were pre-edited to enable word forms to be identified. The computer produced several lists, for example of all the word forms of all the words and frequency of use, verbs used in various forms of the subjunctive. Analysis revealed facts about Virgil's style which may be of interest to both linguists and classicists. Details regarding the method of translation are given, all words being considered in the order in which they come in the sentence--the order in which the Romans heard and understood them.

Computers and Mathematics Instruction

J. Nievergelt
Department of Computer Science, University of Illinois, Urbana

Computers and Mathematics with Applications.1: 121-32, January 1975

1) It is important that every educated person understand some of the principles on which computers operate. 2) There is an important relationship between mathematics education and computers, both in the sense that the mathematics curriculum can contribute significantly towards teaching students about computers, as well as in the sense that computers have a role to play in mathematics education. The latter point is shown by several examples which are discussed in detail. It is also argued that high school mathematics teachers can obtain the required knowledge to use computers effectively in their teaching by taking two one-semester courses as part of an in service training program.

Schemas: The Brain's Representation of Domains of Interaction

Michael Arbib
Center for Systems Neuroscience and Department of Computer and Information Science, University of Massachusetts, Amherst, 01002

Brain Theory Newsletter 1: 37-42, March 1976

The brains of humans and animals contain a large number of different subsystems - schemas corresponding to different domains of interaction in which the animal might find itself. At any given time, these schemas are at different levels of activation - and it is the most active schemas which between themselves constitute the current representation of the animal's environment. These schemas must comprise three types of routines: (1) *inputmatching* routines that sample environmental stimuli as well as the activity of other schemas to determine whether in fact activation of that schema appears appropriate; (ii) *action* routines which can control movement or activate other schemas, in a way that is appropriate for interaction with the domain of interaction which the schema represents; and (iii) *competition and cooperation* routines which serve to increase the activation of other schemas consistent with the given schema, while depressing the activity of schemas which are mutually exclusive.

BRAIN THEORY

Brain Theory Newsletter

Fred K. Lenherr, Editor
Center for Systems Neuroscience, Graduate Research Center, University of Massachusetts, Amherst 01002

Volume 1, Number 3, March 1976

The Fundamental Theses of Neuro- and Psychocybernetics (Neurocybernetics)

C. Balaceanu, and G. Dona

Kybernetes 3: 203-6, October 1974

Neurological sciences, cybernetics and general system theory achieved enough progress in the last decade to permit the establishment of a satisfactory theory of the functioning of the nervous system. An axiomatic-like theoretical skeleton for neurocybernetics and psychocybernetics is proposed in the form of eight theses which can be considered as the essential aspects of the formal framework needed for the actual interpretation and the future development of our knowledge in the field of brain research.

American Journal of Computational Linguistics *Microfiche 58 : 78*

MATHEMATICAL MODELS OF LANGUAGE

SOVIET PAPERS IN FORMAL LINGUISTICS

EDITED BY FERENC KIEFER

Sprakforlaget Skriptor AB
Box
S 104 65 Stockholm
1973

284 pages SKr 60

REVIEWED BY DAVID B. BENSON

Computer Science
Washington State University
Pullmar 99163

(Currently visiting at the University of Colorado)

From the Preface:

"The volume is more or less a random sample of the great number of works done in the field of mathematical linguistics by Soviet scholars. The random character of the selection is due to the difficulties which an editor inevitably encounters in compiling an anthology like the present volume. If I were to start working on this volume now I would certainly choose more recent papers, perhaps ones in one or another aspect more representative than those included in this volume. Nonetheless, these articles are at least in one respect representative. They clearly testify to the breadth of interest and variety of approaches in Soviet mathematical linguistics. This anthology is intended to convince

the reader who has not mastered Russian and is perhaps not familiar with works by Soviet "mathematical linguists" that they deserve much more attention than they have received up to now."

The papers in this volume are indeed beginning to show their age. From internal evidence, primarily the bibliography or notes at the end of each paper, these papers were written in 1967-1972. As a "random sample", the only way to review these papers is to take each in turn.

1. M. V. Arapov - E. N. Efimova

On the Complexity of Government Trees.

pages 3-36.

A government tree is a derivation tree deprived of its labels. Thus the complexity relates solely to the structure of the tree without regard to phrase-names (nonterminals), lexical considerations and the like.

"On the one hand, the government tree contains information about the structure of the text which must be taken account of in any model. On the other hand, it is a comparatively simple object for which it is easier to develop a suitable mathematical apparatus."

The complexity depends on the internal arrangement of vertices, thus for a sentence of length n, there are government trees with n leaves which have minimal complexity. "Here we shall proceed from the assumption that those structures which have minimal or close to minimal complexity are realized in natural language." Very reasonable.

The complexity measure used is developed as follows: For each vertex let k_i be the out-degree of i, that is, the number of descendents of i, and let i* be the father of i. Let the root of the tree be node 0. The complexity of each vertex is defined as:

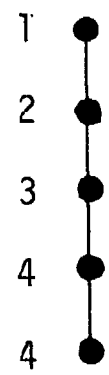
$$F(0) = k_0$$

$$F(i) = k_i + F(i^*) \text{ for } i>0.$$

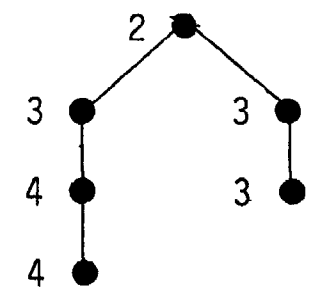
For tree Λ_N with N vertices, the complexity is:

$$F(\Delta_N) = \sum_{i=0}^{N-1} F(i)$$

For example:



$F(\Delta_5) = 14$



$F(\Delta_6) = 19$

Consider the set of all trees with N vertices, M_N and define $\Delta \varepsilon M_N$ to be the minimal if and only if

$$F(\Delta) \leq F(\Delta')$$

for all $\Delta' \varepsilon M_N$. Let $M_N \subset M_N$ be the set of minimal trees with N vertices and let $M = \bigcup_N M_N$ be the set of all minimal trees. Then for each minimal tree $\Delta_N \varepsilon M$ $\phi(N) = (\Delta_N)$ ϕ is the complexity measure studied. This does not directly find the complexity of the minimum tree with n leaves, which is a more interesting question given the paper's stated orientation toward linguistics. Nonetheless, the authors find several suggestive results about the structure of minimal trees.

Theorem I. If $\Delta \varepsilon M$, then $k_0 \leq 3$

Assuming that this notion of minimality is indeed a principle of economy, then no sentence has more than three main constituents.

Theorem II. If $\Delta \varepsilon M$, for each vertex $i>0$, $k_i^* \geq k_i$.

The deeper one goes in a minimal tree, the lower the out-degree.

"..the monotonous decrease in the number of arrows issuing from a vertex proportionate to its... distance from the root of the tree essentially agrees with the empirical facts. In fact, the number of completed valences for the verb-predicate (which are usually placed in the root of the tree) is on the average larger than for a noun which is subordinate to it, larger for the noun than for an adjective subordinate to the noun, and this number is more often than not equal to zero for an adverb governed by such an adjective. Of course, such a monotony is in reality only approximate."

Theorem III. For any $\Delta_N \varepsilon M$ with $N>81$, $k_0 = 3$.

Theorem IV. $\phi(N)$ is of the order $N \ell n N$.

The authors point out that a detailed comparison of minimal government trees with 'concrete' syntactic structures is without much meaning. Nonetheless, this is the first paper that I know of which broaches the notion of an economy of syntactic effort. Whether the theorems are indeed suggestive of linguistic reality is a matter for future research.

2. V. B. Borscev - M. V. Xomjakov

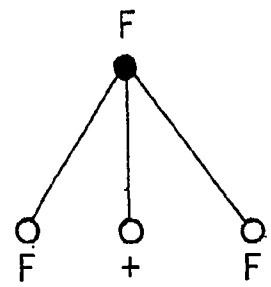
Axiomatic approach to a Description of Formalized Languages and Translation Neighborhood Languages.

pages 37-114.

This lengthy contribution consists of four chapters of detailed development. The basic plan is an interpretation of P. M. Cohn's Universal Algebra (Harper and Row, New York, 1965) as relational systems to treat "texts" and grammars. While I enjoyed reading Cohn's excellent treatment of universal algebra, I did not enjoy this paper. It tends to wander, whereas I prefer papers which build to a definite climax. Further, most of the authors' ideas have been presented in the Western literature, so I found at most two new nuggets of wisdom. Nonetheless, here is the substance of the paper.

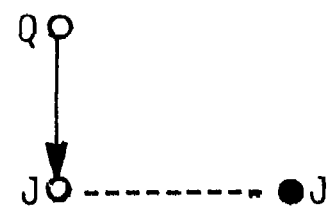
Chapters I and II build a notion of "text" and grammar via systems of relations. One has relations of "to the left of" and "below" in trees as well as other relations, such as "isomorphic subtree". Even the notions of terminal and non-terminal alphabets are treated as relations. This uniformity might offer some advantages for the abstract development about classes of sign systems, texts and grammars, but makes the concrete cases and examples hard to follow. In fact there are other uniform treatments, mentioned below, which are undoubtedly better for the particular cases in question.

The authors treat neighborhood grammars in these two chapters. A neighborhood of a vertex in a tree consists of some of the connecting arcs and nearby nodes. For example, a neighborhood of F is (page 53):



where the distinguished node whose neighborhood is in question is marked by Given a collection of neighborhoods, a tree is in the neighborhood language if all the neighborhood constraints specified in formula which constitute the

grammar are satisfied all nodes of the tree. The major virtue of this approach is in enabling one to specify other connections between the nodes of trees other than the usual descendant relation. Thus



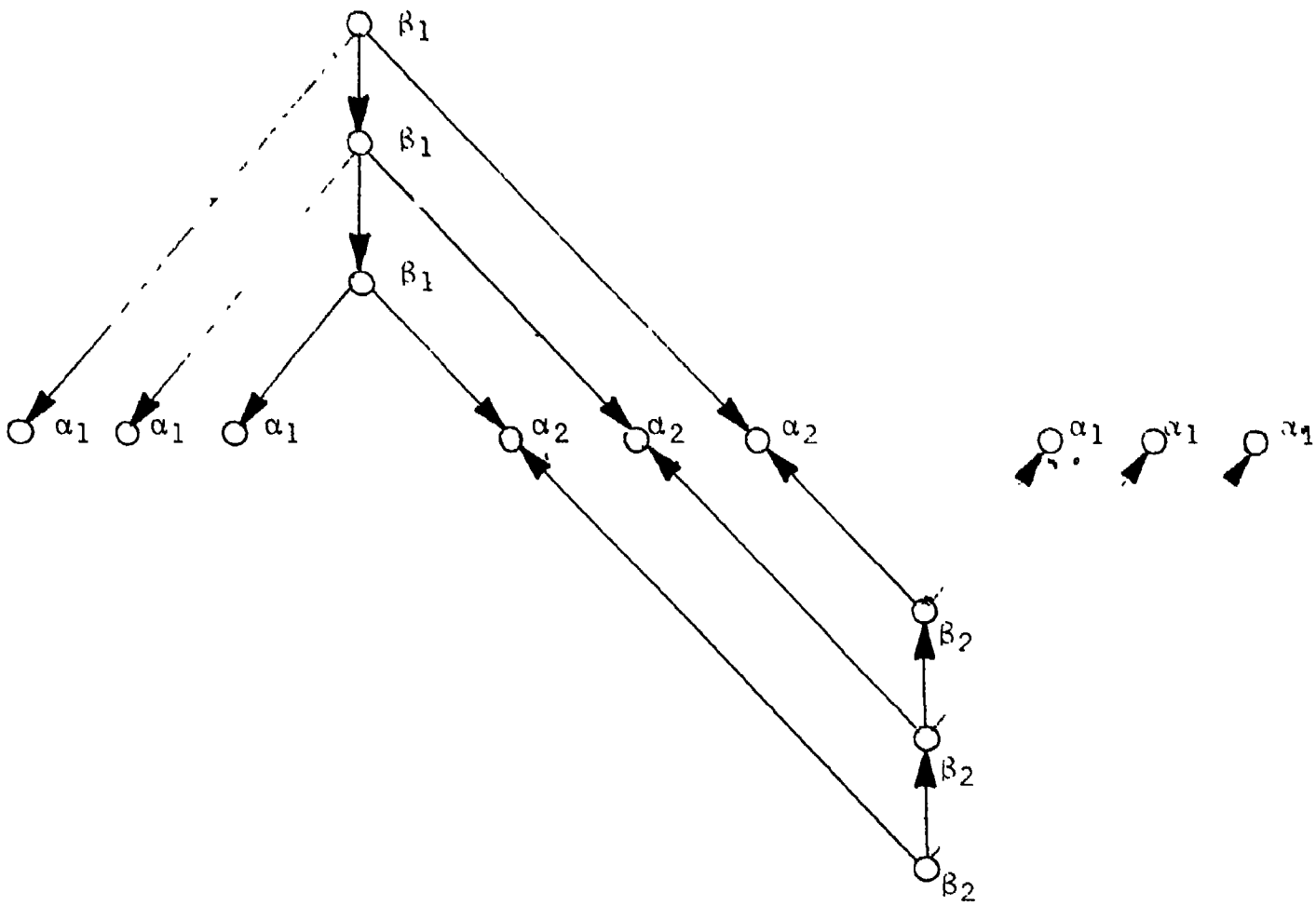
is a neighborhood specifying -- via other information too complex to describe here -- that both copies of J dominate isomorphic subtrees. This enables one to specify "syntactically" that every variable in a programming language must be declared before it is used.

However, there are better methods to handle these non-tree restrictions. For example, property grammars (Aho and Ullman, The Theory of Parsing, Translation, and Compiling: Vol. 2, Compiling, Prentice-Hall, Englewood Cliffs, N. J., 1973), macro-grammars (M. J. Fischer, Grammars with macro-like productions. Ph.D. Thesis, Harvard University, 1968), and mathematical semantics (R. D. Tennent, The Denotational Semantics of Programming Languages, Comm. ACM 19:8 (Aug.1976), 437-453.)

Chapter III, "concrete sign systems" develops phrase structure grammars and nominal neighborhood grammars. The type 0 phrase structure grammars produce "phrase structures", as generalizations of trees. These phrase structures have appeared in the Western literature in at least the following papers: J. Loeckx, The Parsing for General Phrase-Structure Grammars, Inform. & Control 16:5 (Jul 1970), 443-464, H. W. Buttelmann, On the Syntactic-Structures of Unrestricted Grammars, I. Generative Grammars and Phrase Structure Grammars, Inform. & Control 29:1 (Sept 1975), 29-80; D. B. Benson, Syntax and Semantics: A categorical view, Inform. & Control 17 (1970), 145-160, all three of which

were originally written in 1969-1970. This development was clearly ripe at that time in Russia, Europe and the U. S.

The nominal neighborhood grammars are an extension of neighborhood grammars which allow fairly complex structures. For example, the following is taken from page 90:



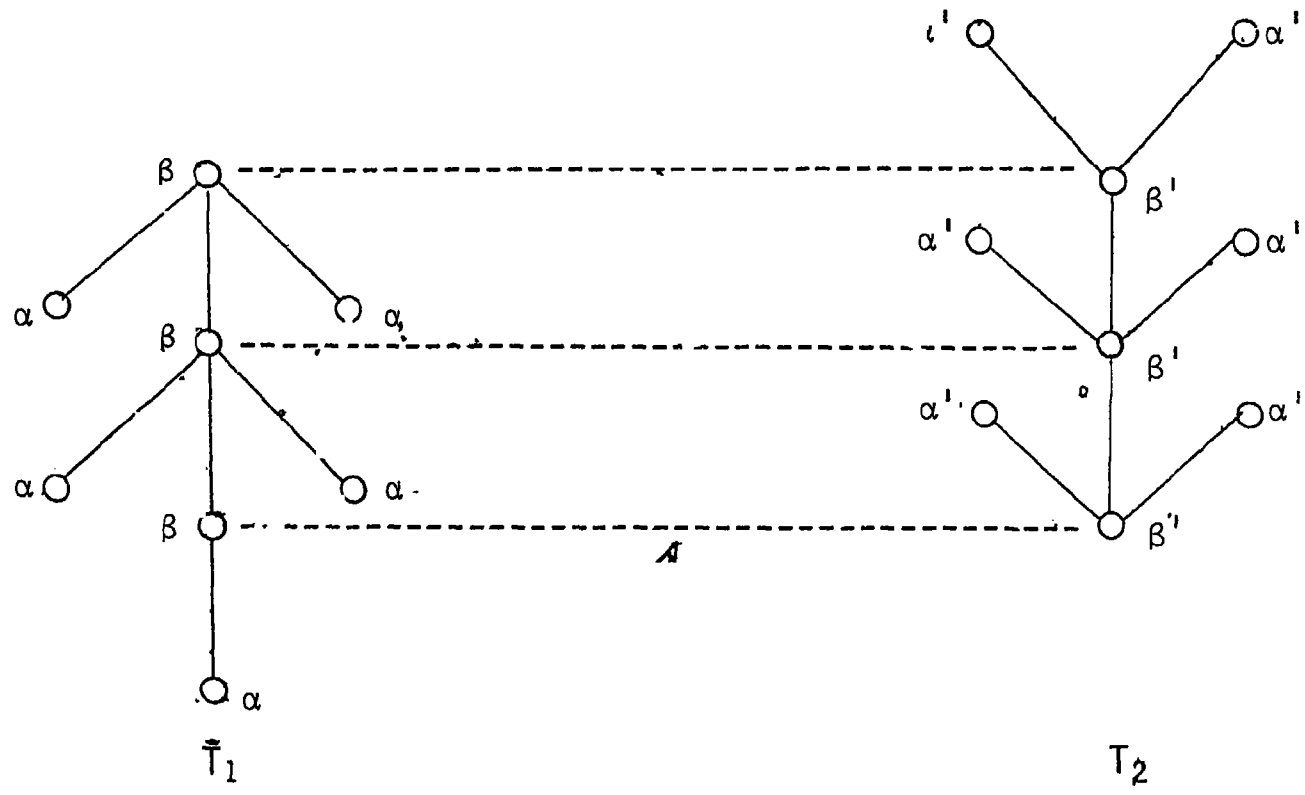
It appears that the language

$$\{\alpha_1^n \alpha_2^n \alpha_1^n \mid n \geq 1\}$$

can be generated by nominal neighborhood grammars in an essentially context-free manner. The nominal neighborhood grammars are new to me and appear to offer considerable generating power at the usual expense of a complex definition.

Chapter IV treats syntax-directed translations and certain extensions thereof using the idea of neighborhoods Most of their development is now standard (Aho and Ullman, The Theory of Parsing, Translation, and Compiling:

Vol. 1, Parsing, Prentice-Hall, Englewood Cliffs, N. J., 1972) and has been advanced to truely elegant abstractions by Alagic (Natural State Transformations, J. Comp. Sys. Sci. 10: 2 (Apr 1975), 266-307.) However, the use of neighborhoods allows for the extension of syntax-directed translations in new directions, best indicated here by the authors' diagram of a translation from tree T_1 to tree T_2.



I can't think of any use for this order-reversal in carrying out the translation, but it is an interesting idea nonetheless.

3. S.-Ja. Fitialov

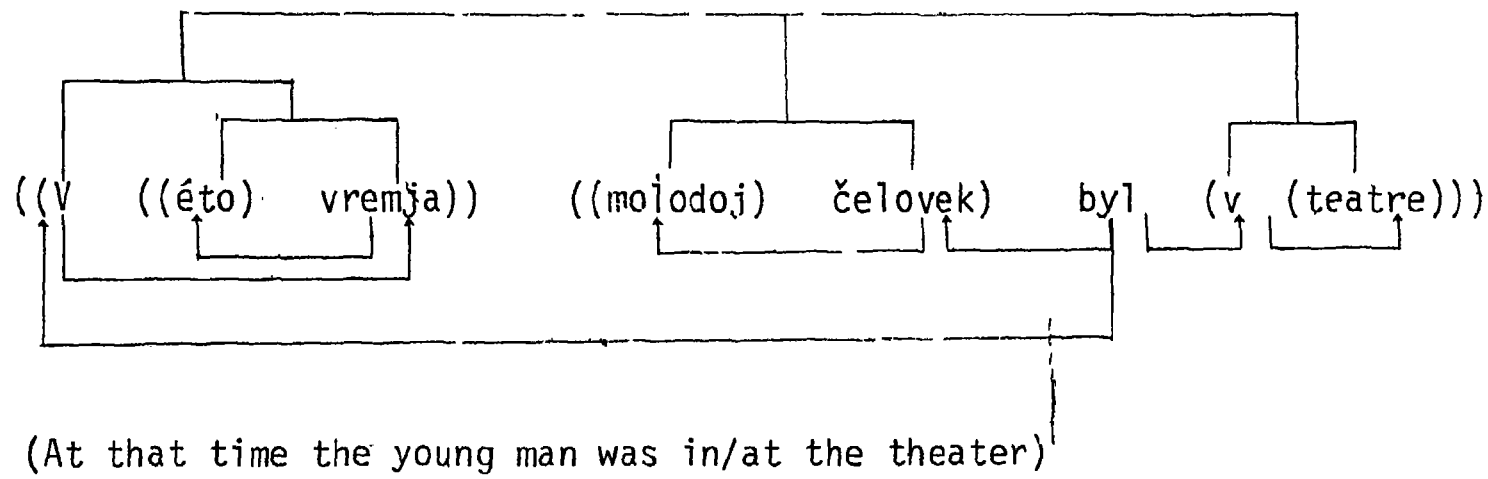
   On the Equivalence of IC Grammars and Dependency Grammars

   pages 115-158.

   According to the author, both the direction and nesting of syntactic

relationships should be accounted for in a sufficiently adequate and complete linguistic description. As dependency grammars handle direction and Immediate Constituent grammars handle nesting, the question of the relationship between the two descriptive mechanisms arises.

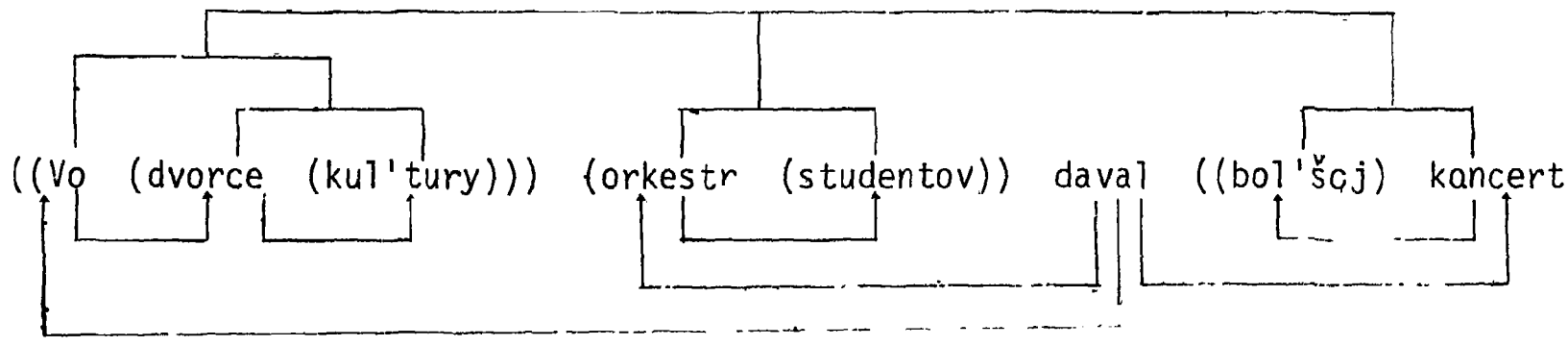
As the phrase names (non-terminal symbols) can not be determined from the dependencies, the IC structure considered consists solely of the tree. This is best illustrated by the following example. The element groups in the dependency structure are enclosed in parentheses, the dependent directions is shown below the sentence and the IC tree is shown above. From page 128:



(At that time the young man was in/at the theater)

Fig. 1.

Two dependency structures can give rise to a single IC structure. Compare with Fig. 1.



(The student orchestra gave a big concert in the Palace of Culture)

Fig. 2.

Furthermore, Fitialov gives examples in which the same sentence can have two different IC trees. Thus the "equivalence" is many-to-many. The author then sets up an algorithm to construct a dependency grammar from certain IC grammars. The IC grammar must have "finite degree", a technical concept that need not detain us. The final topic is carrying the idea of "degree of nesting" from IC structures with non-terminals over to dependency structures.

Much of this paper is apparently devoted to clarifying the ideas presented by Gaifman (Dependency Systems and Phrase-Structure Systems, Inform. & Control 8(1965), 304-337). I found little of interest in this selection.

4 A. Gladkij

An Attempt at the Formal Definition of Case and Gender of the Noun. pages 159-204.

With the recent interest in case in computational linguistics, this paper by the foremost Russian formal languages expert should appeal to those who wish to build logically coherent case structures. The mathematics is minimal but suggestive. The focus of the work is on a classification of (Russian) nouns. I am in no position to comment on the quality of the classification system proposed. Nonetheless, here is a sketch of the method.

Let V be the set of words. These are called segments by Gladkij to stress the graphical sense of word he is using. Each subset of V having "identical lexical meaning" is called a neighborhood. Thus:

{DOM, DOMA, DOMU, DOMOM, DOME, DOMA, DOMOV, DOMAM, DOMAMI, DOMAX} (house)

Let , a subset of V, be the set of nouns. "The set S should be a union of some neighborhoods." The next notion is subordination or dependency. Say

that x (potentially) subordinates y if there is a sentence in which some occurrence of segment x "syntactically directly subordinates" some occurrence of segment y. Now let O be any neighborhood. Say that O subordinates y if y is subordinate to at least one segment in O. Let N_0 be the set of all S-segments (noun words) which are subordinate to O. A set N_0 is said to be minimal if N_0 is not empty and there are no non-empty $N_{0'}$ which are proper subsets of N_0. The minimal sets N_0 are said to be cases. "If two different neighborhoods O and O' of the sets N_0 and $N_{0'}$ coincide, we will not consider N_0 and $N_{0'}$ to be different cases, but one and the same case."

Gladkij gives examples of all these concepts, including the distinction between minimal and non-minimal neighborhoods. He goes on to show that the cases are not necessarily mutually disjoint, and then uses the development to explicate the "special position" occupied by the second prepositional and second gentive cases in Russian grammar. Gladkij then shows, to no one's surprise, that there are instances in which meaning, even the meaning of the prior several sentences, must be taken into consideration to determine the case of certain words. If one's purpose is to understand the text, then in these instances the case structure won't help. In most sentences however, it will clarify the relationships of the segments in the sentence and thus aid understanding. Whether Gladkij's formulation is more useful than unaided intuition and knowledge of the language is for others to judge.

The last sixteen pages of the paper develop a similar formalism for the concept of coordinated class, apparently as an aid to arriving, at the very end of the paper, in a definition of gender. The mathematics is very easy, but the Russian examples are not--for this reviewer,

5. Ju. K. Lekomcev

On Models for a Syntax with Explicitly Differentiated Elements (D-Syntax). pages 205-239.

This paper is, by Western standards, fussy and pedantic. One must suppose that the editor's selection was rather more random than less. Despite the following quotation from the introduction--

> "Concerning the characteristic of a D syntax model, it should be noted that our model is a continuation of the glossematic variant of the Saussurian trend, partly complemented by Russian and American concepts. The notions of syntagmatic-paradigmatic relations and of distinctive features lie at the heart of the concept."

--I was disappointed. The mathematical model, stated in the complete formality of first-order predicate logic, actually says very little. The foundation of the paper's development is a notion of differentiation system (DS). A DS is basically a system of lists of the values of attributes. Thus two element (i.e., lists) differ if some value of some common attribute differs. Actually the paper develops somewhat more complex differentiation systems, but the additional complexities are obvious, not requiring such an overly formal development.

This notion is then applied to the question of generating (resp., analyzing) words from phonemes, in a fashion that would have produced more insight if it had been treated in automata-theoretic terms. The concluding remarks--on applying DS to semantics--seem to this reviewer to be irrelevant, or else more clearly presented elsewhere.

6. Ju. A. Srejder

On the Contrast between the Concepts 'Language Model' and 'Mathematical Model'

pages 241-267.

> "The concept 'language model" is widely used in structural and mathematical linguistics. In a certain sense, this concept is the cornerstone of these branches or linguistics, where so called formalized or precise methods have taken root. It is of some use, therefore, to gain an understanding of just what is meant with the words "language model"."

The author, evidently a mathematician, contrasts the notion of a mathematical theory and a language model. In the terms of mathematical logic, a theory consists of names for relations or functions, names for variables, a method of contructing well-formed formulae (wff), the system of formal deduction to be used, and the axioms of theory. A model of a theory is a system of sets and relations such that "if relation R_i is compared to every name of relation $\mathbb{R}_i$ in such a way that if variables x,y,z,... are explained as elements of set $\mathbb{M}$ all formulae of the given Theory are true."

After several examples of mathematical theories and models--the theory of partial orders and a model of it in the natural numbers is one--the author gives a fairly strong argument that what many linguistics "call a model is in mathematics known as a thoery." He then gives some examples of 'language models' to give substanence to this thesis. The most interesting is an all-too-brief discussion of the poem "Eugene Onegin" in which the theory is

> "An accented syllable can be located only on an even-numbered place from the beginning of the line."

for which presumably a standard edition of "Eugene Onegin" stands as a model. He continues by giving a short neighborhood grammar as the axioms of the theory. In an appendix he shows that Chomsky's "generative model of context-free grammars is in fact a particular mathematical theory.

However, linguistics is not mathematics and the models (i.e., the actual utterances or texts) fail to satisfy all details of the theory. Thus the author suggests

"1) The quasi-model of a Theory, i.e., the set in which the theory is almost fulfilled (for this it is necessary to introduce a measure onto the Theory) and:

2) The measure on a class of quasi-models of a given Theory, which allows us to say that the Theory can be fulfilled for almost all quasi-models."

Unfortunately, these fine ideas are not developed. Nonetheless, this paper does help explain the terminological differences between mathematicians and linguists.

.7. E. D. Stockij

Generalized Grammars and Their Properties.

pages 269-284.

"Let us assume that in grammar Γ not all derivations are permissible, but only those which can themselves be generated by another grammar Γ', which is working as the device for programming of the derivation. We shall investigate the question of how the selection of a strategy of phrase generation in grammar

Γ (in other words, selecting grammar Γ') affects the generative capacities of grammar Γ."

Let $X_0 \Longrightarrow X_1 \Longrightarrow \ldots X_n$ be a derivation in grammar Γ. Let V' be a set of names for the rules of Γ. Thus the derivation corresponds to a word $\rho_1\rho_2\ldots\rho_n \varepsilon (V')^*$ where ρ_i is the name of the rule doing the rewriting $X_{i-1} \Longrightarrow X_i$. Each word over V' that corresponds to a derviation is called a control word. In general there is no one-to-one correspondence between the derivations and their control words.

A generalized grammar is a pair of grammars (Γ,Γ') such that the second grammar is used to control the first. Specifically, $X_0 \Longrightarrow X_1 \Longrightarrow \ldots \Longrightarrow X_n$ is an allowable derivation of Γ if and only if there is at least one control word corresponding to it in $L(\Gamma')$. The language of the generalized grammar is that subset of $L(\Gamma)$ for which each word has at least one allowable derivation. Note that there is no requirement that the derivations be canonical (left-most).

Grammars in this paper are classified by the usual Chomsky Hierarchy into types 0, 1, 2, 3. Then generalized grammars have type (i,j) where i is the type of the language-producing grammar and j is the type of the controlling grammar. Let D_{ij} be the class of languages generated by all generalized grammars of type (i,j), and D_i be the class of languages generated by all (ordinary) grammars of type i.

The main portion of the paper presents, without proof, the relationships known among the D_{ij} as of April, 1969, excluding some American work such as Ginsburg and Spanier's (Derivation bounded languages, J. Comp. Sys. Sci. 2:3(1968), 228-250). Most of the references cited--which contain the proofs--are to Stockij's own work on these questions. Example results are:

$D_{3j} = D_j$, j = 1,2,3.

$D_{13} = D_1$

$D_{23} \subset D_1$

$D_2 \subset D_{23}$

$D_{00} = D_{01} = D_{02} = D_{03} = D_0$

$D_{10} = D_{20} = D_{30} = D_0$ (without null words)

This last is a consequence of Stockij s disallowing rewritings to the null word in grammars of type 1, 2 and 3.

Now consider the set of control words, P(Γ), of an ordinary uncontrolled grammar, Γ. Let tP(i) denote the type of the language P(Γ) for grammar Γ of type i. The following are representative results.

tP(3) = 3

tP(2) = 1

tP(0) = 1

The study of controlled grammars arises from the psycholinguistic idea that derivations are controlled by a "generation program" which determines the semantics of the phrases and their grammatical structure. Thereby, the results presented here presumably explicate the potential grammatical structures which such a generation program could possibly produce. Whether or not one accepts the "generating program" hypothesis, these are nice results in formal language theory.

American Journal of Computational Linguistics *Microfiche 58 : 94*

DESCRIPTION GRAMMATICALE DU PARLER DE L'ILE-AUX-COUDRES, QUEBEC

EMILE SEUTIN

Les Presses de l'Universite de Montréal
C.P. 6128 Montréal H3C 3J7
1975

459 pages ISBN 0-8405-0299-0

REVIEWED BY DAVID A. DINNEEN

Department of Linguistics
University of Kansas
Lawrence 66045

Seutin has succeeded in condensing his 600-page dissertation into a useful grammatical description of the French spoken on the Ile-aux-Coudres (IAC) in the St Lawrence. Structuralist and rigidly descriptive, it lists forms, frequencies of occurrence, and variants, with few and isolated explanatory comments. Despite a discussion in the introduction to the second part of the need for a different approach to syntax, the second like the first part--consists of a straight presentation of data in a structuralist framework The only harm in this is that Seutin seems to be trying to do something different. There is no need. the data gathered are significant in themselves and are presented in such a way as to be useful to

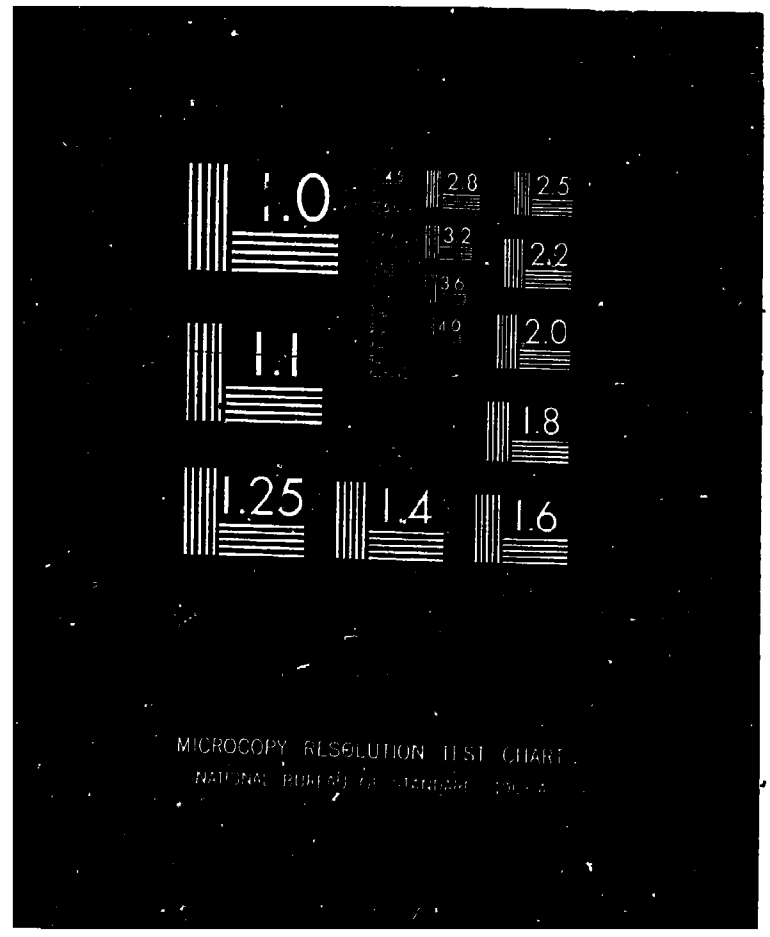
researchers in dialect study, historical and comparative linguistics, sociolinguistics and other subdisciplines.

Seutin not only describes the morphology and syntax of IAC but he compares the usage in the island with that of standard French as described in *Français Fondamental* (FF) He recognizes and accounts for the differences between each corpus and avoids making generalizations when the two cannot be reasonably compared. He does, however, seem to forget from time to time that FF is not as "colloquial" as his corpus, even though he mentions more than once the need for a current description of "familiar" French.

The work is a good example of the use of the computer for recording and storing grammatical and lexical data and, more interesting, for searching and analyzing the corpus and comparing data from other sources. Little description is given of the program but the results indicate that Seutin and his colleagues were able to handle a very large corpus and extract from it the data in which they were interested. It would appear that most of the syntactic analysis was done by hand, but once forms were encoded, the program could find all examples of each structure being studied and print them out in a usable format. Since, as in most concordances, forms are (can be) printed out in context, including an indicator of the speaker, the same data could and should be used in the future for many different kinds of studies.

One finds discussion of a number of particular but unrelated phenomena, with some excellent insights, but no general summary of projects that might be undertaken in consideration of these phenomena On pages 372-3, S. speaks of the relatively high frequency of occurrence of *propositions nominales* in IAC, but does not pursue the matter further, possibly because the traditional framework does not encourage any generalizations. In the lengthy list of words in the beginning of the second part, many specific comments would be of great interest to semanticians and sociolinguists. For those interested in de-acquisition, the list of losses or significant decreases ir use of certain verb forms on page 305, as well as many other statements on disappearing forms, would be of much use. This is a model description: more studies, of various French dialects, following the same basic pattern, should be undertaken.

END


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A