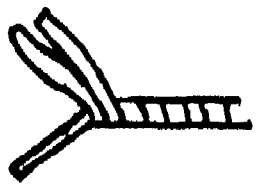
Journal of Computational Linguistics Microfiche 14 : 1

THE FINITE STRING



NEWSLETTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS

VOLUME 11 - NUMBER 4 DECEMBER 1974

This issue was released for production on March 28, 1975. It contains abstracts of preprints and publications available to the editors on March 20.

NOTICE TO AUTHORS

In science, those who read also publish. The help of the reader-become-author makes announcement journals more helpful,

> *On preparing drafts for semipublic circulation: send a copy to the editorial office to be abstracted.*
>
> *On receiving notice of acceptance for publication in a book or journal: inform the editor, giving full citation (as much as known)*

With the collaboration of the readers, The Finite String will provide much more timely information.

BIBLIOGRAPHY

CURRENT TRENDS IN LINGUISTICS

Thomas A. Sebeok, Editor
Research Center for the Language Sciences
Indiana University

VOLUME 12: LINGUISTICS AND ADJACENT ARTS AND SCIENCES

Associate Editors:
Arthur S. Abramson, Dell Hymes, Herbert Rubenstein
Edward Stankiewicz

Assistant Editor:
Bernard Spolsky

Assistants to the Editor:
Alexandra Di Luglio
Lucia Hadd Zoercher

MOUTON
The Hague · Paris
1974

CONTENTS

VOLUME 12: LINGUISTICS AND ADJACENT ARTS AND SCIENCES (CONT'D.)

VOLUME 12: LINGUISTICS AND ADJACENT ARTS AND SCIENCES (CONT'D.)

PART SIX: LINGUISTICS AND PSYCHOLOGY

PART SEVEN: ANTHROPOLOGY AND SOCIOLOGY

PART EIGHT: LINGUISTICS AND ECONOMICS

PART NINE: LINGUISTICS AND EDUCATION

VOLUME 12: LINGUISTICS AND ADJACENT ARTS AND SCIENCES (CONT D)

PART TEN PHONETICS

PART ELEVEN: BIO-MEDICAL APPLICATIONS

PART TWELVE: COMPUTER APPLICATIONS

VOLUME 12: LINGUISTICS AND ADJACENT ARTS AND SCIENCES (CONT D)

PART THIRTEEN: LINGUISTICS AS A PILOT SCIENCE

INFORMATION PROCESSING 74

PROCEEDINGS OF IFIP CONGRESS 74

Stockholm, August 5-10, 1974

Jack L. Rosenfeld, Editor
Computer Sciences Department
IBM Thomas J. Watson Research Center

North-Holland Publishing Company
Amsterdam & New York
1974

PARTIAL CONTENTS

INFORMATION PROCESSING 74.

PARTIAL CONTENTS

General

THE PRAGUE BULLETIN OF MATHEMATICAL LINGUISTICS

21

TABLE OF CONTENTS

General

LANGUAGES AMONG COMPUTERS, MACHINES, ANIMALS AND MEN

Lawrence M. Clark

Computers and People, 24, 1, January 1975, 7-13

Reports on a number of aspects of language, discusses some significant problems of designation of meaning, and indicates some probable future developments in language. Computers make extensive use of language to fulfill their functions. To place computer languages in the perspective of languages in general is helpful.

DECODING METHODOLOGY AS A TOOL FOR LINGUISTIC RESEARCH
Methode de dechiffrage, outil de recherche en linguistique

B. V. Suhotin
Russian Language Institute
Moscow

T. A. Informations, 1974, 2, 3-43

These methods of analysis are among those that will work for any language without prior knowledge other than what is gained from linguistically prior analyses. The examples presented are algorithms for (1) classification of letters into vowels and consonants, (2) morpheme classification, (3) determining the structure of simple phrases, (4) phonetically transcribing syllabic letters, (5) determining the pronunciation of letters.

Phonetics

COMPARISON OF THE FORMANT SPACES OF RETROFLEXED AND NON-RETROFLEXED VOWELS.

Iris Kameny

IEEE Transactions on Acoustics, Speech and Signal Processing ASSP-23, 1975, 38-49

The formant 1 (F1) and formant 2 (F2) frequency movements of vowels next to /r/ are compared with the same vowels next to other consonants. With the exception of /i/ the effect of initial /r/ on the following syllable nuclei is minimal. The effect of final /r/ on the syllable nuclei preceding it is appreciable. Algorithms are postulated to define a retroflexed vowel space for vowels preceding /r/ in terms of the non-retroflexed vowel space.

THE HUMAN VOCAL CORDS: A MATHEMATICAL MODEL

Ingo R. Titze
Dept. of Physics & Astronomy
Brigham Young University
Provo, Utah

Phonetica 28, 129-170, 1973

A mathematical model for digital computer simulation of human-like utterances is developed. The overall system consists of a period structure of 16 coupled masses for each of the vocal cords, an 18-section cylindrical tube approximation of the pharynx and mouth, and a similar 12-section nasal tract. Special care has been taken to model separately the functions of the vocal ligament the vacalic muscle, and the mucosa. Simulated phonation in modal, mixed and falsetto registers is possible. The parameters which control the nature of the phonation are lung pressure, external tension applied to ligament, vocalic, and mucosa, and the internal muscular action of the vocalis. Applications are cited in areas which include physiology, pathology and pedagogy.

Phonetics

NARROW PHONETIC TRANSCRIPTION ON THE COMPUTER: TAKING THE PHONE OFF THE HOOK

Gerald C. Keil
Leeds University
England

Computers and the Humanities 8, 4, 217-229

The IMPAC project of the Survey of English Dialects has developed a system for the machine representation of narrowly transcribed phonetic data. The external machine representation seeks to maintain some proximity to the IPA system and to have the minimum number of symbols used to represent each phone. The internal code can be considered as a matrix in which members of each row or column share a common property. An internal property table expresses the characteristics of each phone.

COMPUTER CONTROLLED RADIOGRAPHY FOR OBSERVATION OF MOVEMENTS OF ARTICULATORY AND OTHER HUMAN ORGANS

O. Fujimura, S. Kiritani and H. Ishida
Research Institute of Logopedics and Phoniatrics,
Faculty of Medicine
University of Tokyo

Computers in Biology and Medicine, 3, 371-384, 1973

On-line computer control of a flying spot X-ray microbeam generator is proposed for substantial reduction of radiation dose and automatic processing of radiographic data. A small X-ray microbeam generator was used in a pilot study. Preliminary experiments have demonstrated its applicability in studies of articulatory gestures and cerebral blood flow measurements. Monitoring of the position of a fiberscope in the pharynx during speech utterances has been tested in real time successfully with use of an integrated dose of approximately 16 mR cm^2/min at an

Phonology

COMPUTER EXPLORATION OF FAST-SPEECH RULES

Joyce Friedman
Department of Computer and Communication Sciences
University of Michigan
Ann Arbor

IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23, February 1975, 100-103

A set of fast-speech rules has been tested on the computer using the phonological grammar tester (PGT) program of Friedman and Morin. We examine the types of difficulties encountered in the rules and discuss ways in which the program can be made more useful for studying fast-speech rules.

Speech recognition

SYSTEM FOR ACOUSTIC-PHONETIC ANALYSIS OF CONTINUOUS SPEECH

Clifford J. Weinstein
Stephanie S. McCandles
Lee F. Mondshein
MIT Lincoln Laboratory
Lexington, Mass

and

Victor W. Zue
Department of Electrical Engineering
M.I.T
Cambridge, Mass

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, Februaru 1975, 54-67

Spectrum analysis via linear prediction, computation of parameters of the spectrum and fundamental frequency extraction. Preliminary segmentation and classification yields categories of vowel; volume dip within vowel; fricative; stop The decision tree is based on energy measurements in selected frequency bands, derivatives and ratios of these measurements, a voicing detector, and a few editing rules. More detailed classification of diphthongs, semivowels, and nasals; detected vowel segment to stored formant positions in a speaker-normalized vowel table; a fricative identifier, which employs measurement of relative spectral energies in several bands to group the fricative segments into phoneme-like categories; stop consonant classification based on the properties of the plosive burst.

MULTIDIMENSIONAL ANALYSIS OF THE PERCEPTUAL UNIQUENESS OF 31 ENGLISH CONSONANTAL CLUSTERS

John.W. Black, Sadanand Singh and Elizabeth Janocosek
Department of Speech Communication
Ohio State University, Columbus

Report No. TR-16, March 1974. AD-776 649/6GA

Acoustic recordings were made of 31 'doublet' consonant-clusters with five vowels by 12 speakers. The pairs of syllables were heard by twelve listeners who assigned values to the aural differences between each pair in the manner of magnitude estimation'. The clusters were treated as two groups, 18 non-sibilant clusters and 13 sibilant clusters. The responses of each listener became the data for multidimensional analysis. Four-dimensional space provided the most efficient analysis for both sets of data. The interpretations of these dimensions were in terms of the features either of the first or of the second member of the consonant clusters. In the instance of the non-sibilant clusters on dimension one the perceptual feature was determined by the groupings of the second members, on dimensions two and three by the first members.

Speech Recognition

PERCEPTUAL CONTINUOUS SPEECH RECOGNITION

H. Yilmaz, L Ferber, W. Park, H. Kellett, and E. Koprucu
Perception Technology Corporation
Winchester Mass

Report No. RADC-TR-74-180, July 1974

The objective is to study and investigate the recognition of connected speech composed of a context-free limited vocabulary A new method of segmentation is based on the recognition of vowels and vowel-like phonetic segments. This is coupled with a speaker transformation that maps the vowels of each speaker into a standard space thus reducing inter-speaker variations. A method of extending these principles to the recognition of consonants is presented. [AD-783 899/8GA; PC $3.75, MF $2.25]

THE OPTIMUM COMB METHOD OF PITCH PERIOD ANALYSIS OF CONTINUOUS DIGITIZED SPEECH

James A. Moorer
Department of Computer Science
Stanford University
Stanford, California

IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-22, 5, October 1974, 330-338

The method is shown to be of similar accuracy as the Cepstrum technique. Since the method involves only additions, no multiplication, it is shown to be faster than the SIFT algorithm. The basis of the method is searching for a minimum in the magnitude of the difference between a speech segment and a delayed speech segment. This is shown to be equivalent to selecting the comb filter which best annihilates the input signal. The computational complexity of the Cepstrum technique thus is proportional to N *log N where N is the number of points in the window in question.

Speech recognition

WHERE THE PHONEMES ARE: DEALING WITH AMBIGUITY IN ACOUSTIC-PHONETIC RECOGNITION

Richard Schwartz and John Makhoul
Bolt Beranek and Newman Inc.
Cambridge, Mass

IEEE Transactions of Acoustics, Speech, and Signal Processing ASSP-23, 1975, 50-53

Errors in acoustic phonetic recognition occur not only because of the limited scope of the recognition algorithm, but also because certain ambiguities are inherent in analyzing the speech signal. Examples of such ambiguities in segmentation and feature extraction are given. A lattice representation of the segmentation allows for multiple choices that can be sorted out by higher level processes.

INVERSE FILTER FOR SPEAKER IDENTIFICATION

Larry L. Pfeifer
Speech Communications Research Lab Inc.
Santa Barbara, California

Report No RADC-TR-74-214, August 1974

The objective is to determine if a sample can be associated with reference talkers. The method of comparing feature vectors from individual sound elements was chosen for experiments using the inverse filter analysis technique (the autocorrelation method of linear prediction). Ten male talkers supplied speech samples bandlimited to 3250 Hz. Thirteen sound units, 10 vowels and 3 nasals, were studied. Many additional identification tests were performed for the purpose of evaluating different distance functions and different feature vectors. [AD-787 860/6GA; PC $5.25, MF $225]

Speech recognition

THE ROLE OF PHONOLOGICAL RULES IN SPEECH UNDERSTANDING RESEARCH

Beatrice T. Oshika, Joseph Aurbach, Helene Neu
Speech Communications Research Laboratory
Santa Barbara, Cal.

Victor W. Zue
M.I.T. Lincoln Laboratory
Lexington, Mass

Rollin V. Weeks
System Development Corporation
Santa Monica, Cal.

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, February 1975, 104-112

This paper presents some phonological rules which describe systematic pronunciation variation occurring in natural continuous speech. It is argued that a speech understanding system must account for such variation by incorporating phonological rules, either implicitly or explicitly, into the system. Spectrographic evidence for the phonological phenomena described by the rules is included.

RESEARCH ON SPEECH COMMUNICATION AND AUTOMATIC SPEECH RECOGNITION

David J. Broad
Speech Communications Research Laboratory
Santa Barbara, California

Report No. AFOSR-TR-74-0582, February 1974

Theory of phonology: a theory of symbolization, a large computer-based quasi-phonemic/orthographic dictionary of American English, dialect description, and the formalization of a functional phonemic theory. Logical procedures for the interpretation of acoustic phonetic data: a massive investigation of formant frequency transitions in CVC syllables as well as an analysis of the segment durations in the same syllables: the use of formant frequency information in automatic speech recognition; segmentation using formant dynamics and the sources of formant frequency variability; the relations between applied basic problems in speech and language.

Speech recognition

SYLLABLE AS A UNIT OF SPEECH RECOGNITION

Osamu Fujimura
Bell Laboratories
Murray Hill, N.J.

IEEE Transactions on Acoustics, Speech, and Signal Processing
ASSP-23, February 1975, 82 87

Basic problems involved in automatic recognition of continuous speech are discussed with reference to the recently developed template matching technique using dynamic programming. Irregularities in phonetic manifestations of phonemes are discussed and it is argued that the syllable, phonologically redefined, will serve as the effective minimal unit in the time domain. English syllable structures are discussed from this point of view using the notions of "syllable features" and "vowel affinity".

A DESCRIPTION OF A PARAMETRICALLY CONTROLLED MODULAR STRUCTURE FOR SPEECH PROCESSING

N. Rex Dixon and Harvey F. Silverman
Computer, Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, N.Y.

IEEE Transactions on Acoustics, Speech, and Signal Processing
ASSP-23, February 1975, 87-91

The modular acoustic processor (MAP) has been designed for speech recognition. The parametrically controlled (spectral) analyzer (PCA), serves as input to an hierarchically operated string transcriber (HOST). PCA allows parametric selection of several analysis methods, including discrete Fourier transform, linear predictive coding, and chirp z-transform (CZT), and of smoothing, normalization, interpolation, and F_0 estimation methods. PCA develops spectrographic representations and performs spectral-similarity matching and training. HOST does segmentation, classification, and prosody analysis. PCA is a packaged, debugged, running system. A first version of HOST is operational.

Speech recognition

REAL-TIME LINEAR-PREDICTIVE CODING OF SPEECH ON THE SPS-41 TRIPLE-MICROPROCESSOR MACHINE

Michael J. Knudsen
Computer Science Department
Carnegie-Mellon University
Pittsburgh, PA

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, February 1975, 140-145

SPS-41, a commercially available system, is composed of three dissimilar micro-processors working in parallel. Using user-written microcode, one processor performs I/O and master control, the second handles loop indexing and counting, and the third does the actual arithmetic on data. Such parallelism allows 2×10^6 I/O operations and 4×10^6 multiplications/s, but actually realizing this potential requires fresh approaches to some old algorithms. Most important is a new autocorrelation scheme. The present program converts frames of 256 16-bit samples into 14 coefficients and then into 128 points of logarithmic power spectrum at 100 frames/s.

EXPERIMENTS WITH A TREE-SEARCH METHOD FOR CONVERTING NOISY PHONETIC REPRESENTATION INTO STANDARD ORTHOGRAPHY

C. C. Tappert
Speech Processing Group
IBM Thomas J. Watson Research Center
Yorktown Heights, N.Y.

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, February 1975, 129-135

A 250-word lexicon and a finite-state grammar specify the tree. The search is performed in a best-first manner. Phonetic variants for each word are generated automatically by a set of phonological rules. Substantial improvement over earlier performance on the same data was realized.

Speech recognition

MINIMUM PREDICTION RESIDUAL PRINCIPLE APPLIED TO SPEECH RECOGNITION

Fumitada Itakura
Acoustics Research Department
Bell Laboratories
Murray Hill, N.J.

&

Electrical Communications Lab
Nippon Telephone & Telegraph
Public Corporation
Musashino, Tokyo

IEEE Transactions on Acoustics, Speech, and Signal Processing
ASSP-23, February 1975, 67-72

Isolated words, spoken by a designated talker, are recognized through calculation of a minimum prediction residual. A reference pattern for each word is stored as a time pattern of linear prediction coefficients (LPC). The total log prediction residual of an input signal is minimized by optimally registering the reference LPC onto the input autocorrelation coefficients. The input signal is recognized as the reference word which produces the minimum prediction residual. In a 200-word recognition experiment, the recognition rate for a designated male talker is 97.3 percent for telephone input, and the recognition time is about 22 times real time.

PITCH DETECTION BY DATA REDUCTION

Neil J. Miller
Artificial Intelligence Laboratory
Department of Computer Science
Stanford University
Stanford, California

IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING
ASSP-23, February 1975, 72-79

An algorithm determines fundamental frequency by segmenting the signal into pitch periods. Segmentation is achieved by identifying the beginning of each pitch period. Segmentation has three phases. First, using zero crossing and energy measurements, a data structure is constructed. Next, the number of candidate pitch period markers is reduced utilizing syllabic segmentation, coarse pitch frequency estimations, and discrimination functions. Finally, the remaining markers are corrected. This algorithm processes both male and female speech, provides a voiced-unvoiced decision, and operates in real time on a medium speed, general purpose computer.

Speech Recognition

A PHONETIC-CONTEXT CONTROLLED STRATEGY FOR SEGMENTATION AND PHONETIC LABELING OF SPEECH

Paul Mermelstein
Haskins Laboratories
New Haven, Conn.

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, February 1975, 79-82

In a sequential strategy processes are applied to a labeled speech segment and result in a possible subsegmentation; the subsegments are labeled by the process. No more segments are considered than those actually differentiated by the analysis steps. The extraction of acoustic cues pertinent to a phonetic feature can be tuned to classes of sounds separated on the basis of other cues, increasing the reliability of segment labeling. The analysis sequence yields a structure for the syllabic units of the speech signal that can be used to retrieve similar syllabic units for detailed comparison.

AN OBJECTIVE PARALLEL EVALUATOR OF SEGMENTATION/ CLASSIFICATION PERFORMANCE FOR MULTIPLE SYSTEMS

Harvey F. Silverman and N. Rex Dixon
Speech Processing Group
IBM Thomas J. Watson Research Center
Yorktown Heights, N.Y.

IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23, February 1975, 92-99

The system provides for concurrent objective evaluation of up to five methods against a single referent. For segmentation, the evaluator provides first-order statistics, at the phonetic, class and summary levels, for four types of errors: Missed, Adventitious, Misplaced, and Adventitious and misplaced events. For classification, the evaluator gives confusion matrices at the phonetic, class and summary levels. The system, still in the developmental process, is operational and currently used.

PERFORMANCE OF A SPEECH SYNTHESIS SYSTEM

W. A. Ainsworth
Communication Department
University of Keele
Keele, Staffordshire, U.K.

International Journal of Man-Machine Studies 6, 493-511, 1974

A string of phonetic symbols representing the sentence to be uttered is transformed into the control signals required by a parametric speech synthesizer using a small digital computer. The performance of the system was investigated by listening tests. In the first set of experiments consonant-vowel syllables were synthesized, and presented to listeners for identification. The vowels were readily identified, but the fricatives less so. In the second set of experiments the intelligibility of synthesized sentences was examined. It was found that after about an hour of transcribing the sentences, listeners identified about 90% of the words correctly.

A PROGRAMMING SYSTEM FOR STUDIES IN SPEECH SYNTHESIS

P. V. S. Rao, R. B. Thosar
Tata Institute for Fundamental Research
Bombay,

IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-22, 3, 217-225, 1974

This paper describes a speech synthesis system which is particularly suitable for experimental investigations. The synthesis is accomplished in two stages. The concatenation stage generates a schematized spectrographic representation corresponding to the symbolic input. The second stage consists in generating the corresponding acoustic signal. The steady state characterization of each phoneme is supplied as data. Independent concatenation procedures incorporate context dependent effects such as format transitions, changes in the normal duration of vowels, etc. The parameter values for these procedures are obtained by a set of rules. Applicability of a rule is determined by attributes assigned to the phonemes.

Orthography

A THEORETICAL APPROACH FOR CHARACTER RECOGNITION BASED ON PHENOMENOLOGICAL ATTRIBUTES

B. Blesser, R. Shillman, T. Kuklinski, C. Cox, M. Eden and J. Ventura
Research Laboratory of Flectronics
Massachusetts Institute of Technology
Cambridge, Mass.

International Journal of Man-Machine Studies, 6, 701-714

A theory based on ambiguities, rather than on the classical archetypal shape of letters, leads to algorithms which will perform more accurately. Letters are described in terms of an abstract set of functional attributes, each of which can be related to a type of ambiguity between two letters. The relations between the functional attributes, which specify the letter's identity, and the physical attributes, which are derived from the physical image, are called graphical context rules. These rules can be determined from psychological experimentation.

PROBLEME DE LA CLASSIFICATION DES CARACTERES CHINOIS (THE CHINESE CHARACTER CLASSIFICATION PROBLEM)

M. R. Finley, Jr.
Department of Mathematics
University of Laval
Quebec

Proceedings of the Second Open Conference on Information Science in Canada, edited by A. Gamache & R. Penner. Ottawa: Canadian Association for Information Science, 1974, 163-180

The classification problem is presented for the set of 40,000 two-dimensional patterns known as the Chinese characters. The traditional classification according to certain meaning patterns termed radicals is sketched together with some variants derived from it. Using the notion of two-dimensional formal grammar, a classification is outlined yielding a quasi-algebraic formula for each character. Also mentioned are spin-offs on the development of Chinese text composition devices using graphic-display techniques and mini-computers.

Concordance

PHRASE DICTIONARY DISTRIBUTION ANALYSIS AND GROWTH PREDICTION REPORT

J. H. Waite, R. Boehm, J. G. Fisher, S. D. Epstein, D. J. Stewart
Cryptanalytic Computer Sciences Inc.
Cherry Hill, N.J

This study of the DDC Phrase Glossary includes a computer program to tabulate work frequencies for blocks of phrases of optional sizes. On the basis of these distributions, empirical and statistical analyses are made including two prediction models. Two-word distributions are also included. Based upon the available distributions, a two-word Phrase Glossary size of 320,000 two-word phrases was determined. Also included are analyses of various techniques, such as suffix truncation, imbedded phrases, and query effectiveness. Comparisons are made of the DDC system to other plain language machine retrieval systems. [AD-780 957/7GA PC $3.75, MF $1.45 April 1974]

THE LANGUAGE OF THE 'PETERBOROUGH CHRONICAL'

J. L. Mitchell

Computers in the Humanities, J.L. Mitchell, Editor, 1974, 132-145

As a necessary prerequisite to a syntactic investigation of the chronicle the following analyses are produced: alphabetized list of every word of the corpus, cumulative frequency, alphabetized frequency list, rank of every word, cumulative absolute frequency of every group of words, percentage and cumulative frequency of the text represented by each word and group of words, a concordance, a frequency list for grammatical categories, and the text with each word tagged for syntactic category.

JEUDEMO: A TEXT HANDLING SYSTEM

P. Bratley, S. Lusignon, and F. Oue lette

Computers in the Humanities, J. L. Mitchell, Editor, 234-249, 1974

In a typical text-processing task, the structure of the text must be described, the operations, with any restrictions to be performed defined, and the format of results given. The present system allows, for the first part, the definition of the alphabet of a corpus, the word separators, tags (for homograph separation, category markers, etc.), and the language of the text. Allowable operations are: the listing of all words present; KWIC concordances; the building of indexes; the searching for words, or word patterns; and searches as in the previous operation, but with restrictions to words of specified frequency, or to sections of the corpus.

FROM A WORD-FORM CONCORDANCE TO A DICTIONARY-FORM CONCORDANCE

D. J. Koubourlis

Computers in the Humanities, J.L. Mitchell, Editor, 225-233, 1974

A word-form concordance does not conjoin inflected forms of the same lexeme, nor does it separate homographs. By manually editing the output of a word-form concordance for these two phenomena and resorting, a dictionary-form concordance is produced.

DEVELOPMENT OF A TEXT CONCORDANCE AND STATISTICS PROGRAM

Lance S. Smith
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1973

The system has seven basic phases: (1) Entry of source text; (2) Main dictionary update; (3) Creation of upgraded text for processing; (4) Upgraded text pre-edit for processing; (5) Production of KWIC concordance; (6) Production of keyword in phrase concordance; (7) Production of word frequency, parts list, reverse alphabetical word list, and various statistics, e.g. average number of words per sentence. In (2) an exhaustive list of words in the text but not in the dictionary is produced, and an interactive program requests information for each new word. This information is merged into the main dictionary. In (3) information from the dictionary is appended to each word of the text and ambiguities resolved by interactively interrogating the user.

Lexicography

AUTOMATIC IDENTIFICATION OF PHRASAL VERBS.

Godelieve L. M. Berry-Rogghe

Computers in the Humanities, J. L. Mitchell, editor, 16-26, 1974

A phrasal verb is an idiomatic phrase of verb plus particle, e.g. look after'. The author seeks to automatically construct a lexicon of phrasal verbs given an adequately large quantity of data, and statistical procedures.

The statistical procedure used is 'collocation'--the probability of syntagmatic association of two items occurring separated by n items. An analysis of the particle 'in' shows that verbs from phrasal verbs are more closely collocated with this particle than non-idiomatic constituent verbs.

SHAD: A SHAKESPEARE DICTIONARY

M. Spevack, H. J. Neuhaus, and T. Finkenstaedt

Computers in the Humanities, J. L. Mitchell, editor, 111-123, 1974

SHAD merges information from a concordance of Shakespeare, a computer dictionary (drawn from the Shorter Oxford English Dictionary), and data from semi-automatic lemmatization.

Lexicography

A COMPUTERIZED LEXICON OF ENGLISH

E. R. Maxwell, and R. N. Smith

Computers in the Humanities, J. L. Mitchell, editor, 124-131, 1974

An autonomous theory for lexical structure, called Semantic Field Theory, is presented. There is a set of primitives (about 20): sameness, difference, events, st, motion, space, emotion, ets. Neutral concepts are defined by primitive relations to the primitives, e.g., cause, change. A semantic field (SF) is taxonomically related to a neutral concept. For example, the semantic field 'punch' is a member of the neutral concept 'hit', with 'fist' as 'instrument', and 'hit' is an 'event' involving a motion against an object (object being defined as experiencer's body or immediate environment) by another object, the two objects being different.

A COMMON STRUCTURE FOR LEXICOGRAPHIC DATA

D. Sherman

Computers in the Humanities, J. L. Mitchell, Editor, 215-224, 1974

Libraries in the US and the UK have developed a standard structure for data exchange, termed MARC--MAchine Readable Catalog. Each record has a fixed length leader of format defined by the MARC standard, which gives the general record status and control information. Following the leader is a table of contents and inventory of all data fields in the record. The data part of a record is divided into fields, each of which is tagged by a 3-digit number; sub-fields are headed by a symbol of the form $n, where n is an identifier code. The system is used to encode Webster's Seventh Collegiate Dictionary, giving the WEBMARC file.

Lexicography

THE 'THESEE' THESAURUS FOR ELECTRICITY AND ELECTRONICS: DESCRIPTION AND METHOD OF USE

'Thesee' thesaurus pour l'electricite et l'electronique: description et methode d'utilization

A. Dewèze
Grenoble

T. A. Informations, 1974, 1, 2-51

This trilingual thesaurus, with both orthographic and magnetic tape implementations, contains about 11,000 terms. Its five indexes are (1) dictionary, (2) relations of substitution, descriptors with terms, (3) related terms, (4) polyhierarchic classification trees, (5) semantic fields.

LEXICOGRAPHY FOR A STRING GRAMMAR OF FRENCH

La lexicographie pour une grammaire en chaînes du français

Morris Salkoff and Anne Zribi
Laboratoire d Automatique Documentaire et Linguistique
Universite de Paris

Rapport de Recherches No. 3

The structure of a dictionary to accompany Salkoff's grammar (cf. AJCL Microfiche 6:64). Subclasses, selectional constraints on verbs, treatment of multiple classification, idioms, homonyms, etc. An extensive analysis of the French lexicon.

Grammar

WORD ORDER AND WORD ORDER CHANGE

Charles N. Li, Editor
Linguistics Program
University of California
Santa Barbara

University of Texas Press
Austin and London
1975

TABLE OF CONTENTS

ISBN 0-292-79002-3
LC Card 74-17620 $12.50

Grammar

A System for Automatic Inflectional Analysis Implemented for Russian

Anna-Lena Sagvall

Almqvist & Wiksell
Stockholm
1973

ISBN 91-20-04159-4

Grammar

CONCERNING THE STRUCTURE OF A COHERENT TEXT ON THE MORPHOLOGICAL LEVEL

Felix Dreizin
Mathematics Department
Bar-Ilan University
Ramat-Gan, ISRAEL

Hebrew Computational Linguistics No. 8, 1974

The proposed set of rules for characterizing agreement (= concord) phenomena in Modern Hebrew is sub-divided as follows: a) Gender and Number concord within the NP; b) Definiteness assignment; c) Subject -- Main Verb concord -- in turn sub-divided into cases where the Main Verb is a "full" verb and those where it is a copula. The formation of these rules is preceded by a brief descriptive statement of types of redundancy rules which inter-act with agreement phenomena (e.g. certain types or nominals are inherently [+Definite]). The phenomenon of agreement assignment is viewed as a transformational process of "feature-assignment" whereby a specific feature or set of features (e.g. Gender, Number) is copied from one term to other terms within a given syntactic configuration. The process involved is essentially one of "daughter-adjunction". Note is made of the ordering agreement assignment with respect to other processes such as passivization or nominalization.

JUNCTION GRAMMAR AS A BASE FOR AUTOMATIC LANGUAGE PPOCESSING

E. Lytle, and D. Packard
Brigham Young University
Provo, Utah

Preprint, Annual Meeting, Association for Computational Linguistics 1974

Junction-rule schemata interrelate constituent operands via adjunction, conjunction, or subjunction. These fundamental relations are further subdivided in terms of: (1) operand categories; (2) operand junction attributes; and (3) operand scope. Constraints imposed upon the form and content of junctions exclude those not evidenced by natural language data. At the highest level of specificity, junction markers represent an interaction of individual indices, class indices, and a variety of functional entities.

Grammar

MACHINE TRANSLATION AND GERMAN VERB CLASS SYSTEM

Mark Strong
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1972

German verbs have 25 forms per verb, and can use either 'have' or 'be' as an auxiliary. A verb class system is proposed in which information about the complete conjugation can be stored in two hexadecimal digits (i.e. in less than 32 classes). This compares with a detailed traditional verb class system that makes 47 distinctions and could not be contained in the space available.

PROCEDURES FOR ORDERING WELL-FORMED SYNTACTIC STATEMENTS

Floyd H. Billings, Jr.
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1973

In junction grammar representations of sentence structure, the relations between lexical items is logical but not linear. Lexical ordering rules are thus independent from the syntactic representation. Some rules can be assigned by know ng which junction rule is involved e.g. in English the rule joining a preposition to its object is always left-to-right, but in German prepositions have to be classified as either preceding or following. A second type of ordering rule involves the operands of a junction rule being more closely related to each other than to any outside element. An example is the processing of articles; an unmodified noun immediately follows the article, but is realized elsewhere when the noun is modified. Discontinuous orderings are omitted at their actual locations and processed at a designated insertion point.

Grammar

CORPUS OF MODERN FRENCH VERB-AFFIX NOMINALIZATIONS

Corpus des nominalisations verbalo-affixales du Français moderne

Laurent Bourbeau
Groupe de recherches pour la traduction automatique
Université de Montreal

Working Papers in the Linguistics of Machine Translation, 1974

A study of productive affixes serving as lexical indicators. One program edits a deck and writes a matrix of verbs. A second lists the verbs by categories of various codes. A third lists all nominalizations indicated by codes. The results are tools for research. Tables of codes, card formats, etc.

LA NOMINALISATION

Laurent Bourbeau
Groupe de recherches pour la traduction automatique
Universite de Montreal
Working Papers in the Linguistics of Machine Translation, 1974

Theories of Grevisse, Tesniere, Dubois, Tutescu. Coding of verbs: morphology, complements, genders of nouns derived.

Parsers

A BEST-FIRST PARSER

William H. Paxton
Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California

SRI Publication No. Z108 April 1974

A parser for a speech understanding system is described. The parser uses a best-first strategy in which alternative paths are assigned priorities and paths are suspended as long as there is a higher priority alternative to explore. Discussions are included on the types of steps in a parse, the assignment of priorities, cooperation among competing parses, and experimental results.

THE LINGUISTIC STRING PARSER

R. Grishman, N. Sager, C. Raze, and B. Bookchin
New York University

Proceedings of the National Computer Conference, 1973, 427-434

The string grammar has three components: (1) A set of some 300 BNF definitions for combining elementary strings into sentences (2) A set of restrictions on the strings. (3) A word Dictionary. The parser is top-down serial with automatic backup. It produces all parse trees of a sentence. Restrictions are written in a subset of English and translated by the parser into lists of basic operations recognized by the restriction interpreter. For conjunctions, satisfaction of an element of their definition causes an interrupt to insert in the tree a process node that later causes zeroed slots to be filled in by links to other subtrees.

Semantics

LEXICAL SEMANTICS

Synonymic Means of Language

Leksiceskaja Semantika
Sinonimiceskie Sredstva Jazyka

Ju. D. Apresjan

Science Press
Moscow
1974

Table of Contents

LEXICAL SEMANTICS (CONTINUED)

LEXICAL SEMANTICS (CONTINUED)

LEXICAL SEMANTICS (CONTINUED)

THE STRUCTURE OF TASK ORIENTED DIALOGS

Barbara G. Deutsch
Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California

SRI Publication No. Z106 April 1974

The discourse and task information in task oriented dialogs and their use in a speech understanding system are discussed. The results of analyzing some task oriented dialogs are given. A preliminary model of the structure of these dialogs and heuristics for building and using it in a speech understanding system are presented.

Comprehension

THE SRI SPEECH UNDERSTANDING SYSTEM

Donald E. Walker
Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California

SRI Publication No. Z107 April 1974

This paper describes the structure of the SRI speech understanding system and presents the available data on its performance. The system is distinctive in the way that knowledge of various sources is coordinated by a "best-first" parser to predict the sequence of words in an utterance, and in the use of word functions --programs that represent the acoustic characteristics of a word-- to test the predictions.

THE BELIEVER SYSTEM

G. Brown
Department of Computer Science
Rutgers University
New Brunswick, N.J.

Report No. CBM-TR-34, 1974

Attempts to understand and interpret natural language descriptions of human action. Plans and episodes must replace sentences as units of human action to facilitate interpretation of motivation. A plan is an action or a sequence of actions caused by a person in order to achieve a foreseeable goal. An episode is a sequence of acts containing a single plan or several inter-related plans. The Believer System uses Schmidt's theory of intention and personal causation to isolate a plan from a series of acts which may be mostly unrelated to the plan.

Comprehension

SEMANTICS AND SPEECH UNDERSTANDING

Bonnie L. Nash-Webber
Bolt Beranek and Newman Inc.
Cambridge, Mass

Report No. BBN-2896, AI-19, October, 1974

Syntactic constraints and expectations are based on the patterns formed by a given set of linguistic objects, e.g. nouns verbs, adjectives, etc. Pragmatic ones arise from notions of conversational structure and the types of linguistic behavior appropriate to a given situation. The bases for semantic constraints and expectations are an a priori sense of what can be meaningful and the ways in which meaningful concepts can be realized in actual language. The paper describes how semantics is being used in several recent speech understanding systems and discusses in detail some actual problems that have arisen. [Ad-787 616/2GA; PC $4.25; MF $2.25]

THE CASPERS LINGUISTIC ANALYSIS SYSTEM

John W. Klovstad and Lee F. Mondshein
M.I.T. Lincoln Laboratory
Lexington, Mass.

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-23, February 1975, 118-123

CASPERS (Computer-Automated Speech Perception System) is a user-modifiable facility for translating strings of acoustic symbols into sentences. Three distinctive aspects of the system's design are the dynamic application of acoustic-phonological rules across word boundaries, an acoustic-unit splitting-and-merging strategy for treating the dictionary matching problem, and an extensive capability for handling semantic routines within an augmented context-free grammar.

Comprehension

MOTIVATION AND OVERVIEW OF SPEECHLIS: AN EXPERIMENTAL PROTOTYPE FOR SPEECH UNDERSTANDING RESEARCH

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

IEEE Transactions on Acoustics, Speech, and Signal Processing
ASSP-23, 1975, 2-10

Syntactic, semantic, pragmatic and lexical knowledge interact with acoustical and phonological information in the process of speech understanding. A feature-extraction component produces a segment lattice of phonetic descriptions of the acoustic signal. Words from a lexicon are matched against the input signal. A syntactic component judges the grammaticality of hypothesized interpretations, and a semantic component judges meaningfulness. A pragmatic component judges the likelihood of a sentence being uttered in the situation by the speaker. The control component uses the above components predictively to construct 'theories' that can be evaluated by other components or used to set monitors, that, when triggered, initiate procedures to assimilate the event.

A PROSODICALLY GUIDED SPEECH UNDERSTANDING STRATEGY

Wayne A. Lea, Mark F. Medress, and Toby E. Skinner
Sperry Univac
St. Paul, Minnesota

IEEE Transactions on Acoustics Speech, and Signal Processing
ASSP-23, 1975, 30-38

Prosodic features break up continuous speech into sentences and phrases, and locate stressed syllables in those phrases. The most reliable phonetic data are obtained by performing a distinguishing feature analysis within the stressed syllables and by locating sibilants and other robust feature information in unstressed syllables. The numbers and locations of syntactic boundaries and stressed syllables are used to select likely syntactic and semantic structures, with which words are hypothesized to correspond to the partial distinguishing features matrices obtained from the segmental analyses.

Comprehension

CONTROL CONCEPTS IN A SPEECH UNDERSTANDING SYSTEM

Paul Rovner, Bonnie Nash-Webber, and William A. Woods
Bolt Beranek and Newman Inc.
Cambridge, Mass

IEEE Transactions on Acoustics, Speech; and Signal Processing
ASSP-23, 1975, 136-140

An entirely accurate and precise acoustic transcription of speech is unattainable. Applying knowledge about the phonology, syntax, and semantics of a language, and the pragmatic constraints imposed by a task domain can resolve much of the acoustic ambiguity. Lexical retrieval and word matching programs map segments of phonetic transcriptions of the acoustic signal. Syntactic, semantic and pragmatic components of the system form hypotheses about the original utterance which change as evidence for or against them is found. Theories set traps to catch evidence; when a trap is triggered, an evaluation is made to decide if or when to reprocess. A component can also make a 'proposal' which is a request to try immediately to match words against part of the utterance.

SEMANTIC SUPPORT FOR A SPEECH UNDERSTANDING SYSTEM

Bonnie Nash-Webber
Bolt Beranek and Newman Inc.
Cambridge, Mass

IEEE Transactions on Acoustics, Speech, and Signal Processing
ASSP-23, 1975, 124-129

The principal data structures of the BBN SPEECHLIS semantic system are a semantic network and case frame tokens. The network represents associations among words and concepts. The case frames describe how the semantic relationships may be expressed in an utterance. The semantic system proposes words that might have occurred in the original utterance but have not yet been recognized, constructs meaningful sets of word matches from possible ones, and evaluates the consistency of syntactic structures and semantic hypothesis. Two further tasks for it are under investigation: (1) to guide syntax, and (2) to turn the best theory about an utterance into a formal procedure for operating on a data base.

Comprehension

A PRACTICAL APPLICATION OF A REAL-TIME ISOLATED-WORD RECOGNITION SYSTEM USING SYNTACTIC CONSTRAINTS

Jean-Paul Haton
Laboratoire d'Electricite et d' Automatique
Universite de Nancy,
Nancy, France

IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-22, December 1974, 416-419

The recognition of sentences of a language used in numerical command of machine tools is described. The acoustic level operates with dynamic matching procedure and knowledge about syntactics and semantics of the language is used to predict the incoming words. With such a syntax-directed system, real-time recognition of sentences pronounced word-by-word is very accurately achieved, even for several speakers.

A COMPUTER PROGRAM THAT LEARNS TO UNDERSTAND NATURAL LANGUAGE.

Sara R. Jordan

Computers in the Humanities, J. L. Mitchell, editor, 205-214, 1974

Underlying the system is a memory in graph structure form. Concept nodes are related by the following relations: transforms to, combines to form fact, member/subset of this class, class membership in, equivalence, description. Each concept node 'transforms to' words in natural languages, in this case French, German, and English. Mechanical translation is performed by 'transforming to' one language from another for the concepts of the input. Question answering is carried out by inferences on set membership.

Comprehension

PROGRESS IN NATURAL LANGUAGE UNDERSTANDING--AN APPLICATION

W. A. Woods
Bolt Beranek and Newman
Cambridge, Massachusetts

Proceedings of the National Computer Conference, 1973, 441-450

The BBN LUNAR system is described. It enables users to interrogate two data files, using a large subset of English: (1) Analyses of Apollo 11 lunar rock samples; (2) A key phrase index to reports of the first annual Lunar Science Conference. An augmented transition network grammar translates natural language requests into a formal query language, a generalization of the predicate calculus, which operates on the data base.

SOME PRINCIPLES OF COMPUTERIZED LANGUAGE ANALYSIS

R. Brent Thompson
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1972

The four stages in automatic language analysis are (1) Determining the kinds of linguistic elements in the input sentence; (2) Determining the structural relationships among elements; (3) Finding the referential information that may be present; (4) Finding the contextual information that may be present. The BYU group uses junction grammar for (2) and is beginning to work on (3-4)

TOWARD COMPUTER RECOGNITION OF SPOKEN LANGUAGE

R. Byron Purves
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1973

A syntactic recognizer utilizing binary rules is applied to the testing of random strings drawn from a vocabulary of 800 words. Use of the recognizer in a set of 80 utterances gave 65% error-free results; this compared with 35% when only a lexical recognizer was used.

Expression

SENTENCE PARAPHRASING FROM A CONCEPTUAL BASE

Neil M. Goldman
Stanford University
Stanford, California

Communications of the A.C.M. 18, 1975, 96-106

A model of natural language based on Schank's representation of meaning. A program produces sentence paraphrases which demonstrate understanding with respect to a given context. This generator operates in conjunction with a combined memory and inference model. The model encompasses several diverse classes of linguistic knowledge, which include (1) executable tests of conceptual properties stored in discrimination nets, (2) information relating conceptual and syntactic rules, stored in a word sense dictionary, and (3) surface grammatical knowledge stored in a formal grammar.

MODELLING PROPP AND LEVI-STRAUSS IN A META-SYMBOLIC SIMULATION SYSTEM

Sheldon Klein, John F. Aeschlimann, Matthew A. Appelbaum, David F. Balsiger, Elizabeth J. Curtis, Mark Foster, S. David Kalish, Scott J. Kamin, Ying-Da Lee, Lynne A. Price, David F. Salsieder
Computer Sciences Department
University of Wisconsin
Madison

Technical Report #226, October 1974 WIS-CS-226-74

The plot of several myths is given in relational form; the structure Propp suggested is given similarly. Compatibility of selections of characters, objects, and functions is controlled by subscripts, etc. Programs, grammars, traces, and output are exhibited and commented.

INFORMATION SYSTEMS: RECORDS, RELATIONS, SETS, ENTITIES, AND THINGS

Michael E. Senko
Mathematical Sciences Department
IBM Research Laboratory
Yorktown Heights, N.Y.

Information Systems, 1, 1975, 3-13

This article reviews progress in the creation of a scientific discipline for information systems. It discusses contributions from four sources of fundamental knowledge: (1) Information Systems Technology; (2) Scientific Computation Technology; (3) Linguistics; and (4) Mathematics. It then selectively reviews progress on an information systems science in the most active areas of study: name-based representations, stored representations, access languages, and information systems performance. This discussion relies on the definition of a series of abstract, structured levels for the description, design, and implementation of generalized data base management systems.

Inference

A HEURISTIC APPROACH TO INDUCTIVE INFERENCE IN FACT RETRIEVAL SYSTEMS

C. William Skinner
North Carolina State University
Raleigh

Communications of the ACM, 17, December 1974, 707-712

The procedures make use of a similarity structure which is imposed on the data base using nonnumerical clustering algorithms. They are implemented in a model fact retrieval system which uses a formal query language and a property-list data structure. The procedures are used in a program of experiments with test data bases which are altered by deleting part of the data and by purposely introducing false data. The system can infer the correct response under a variety of conditions involving incomplete and inconsistent data.

REPRESENTATIONS OF THE LANGUAGE RECOGNITION PROBLEM FOR A THEOREM PROVER

Jack Minker and Gordon J. VanderBrug
Department of Computer Science
University of Maryland
College Park

International Journal of Computer and Information Sciences, 3, 3, 1974 217-250

Two representations of the language recognition problem for a theorem prover in first-order logic are presented and contrasted. One of the representations is based on the familiar method of generating sentential forms of the language, and the other is based on the Cocke parsing algorithm. An augmented theorem prover is described which permits recognition of recursive languages. The state-transformation method developed by Cordell Green to construct problem solutions in resolution-based systems can be used to obtain the parse tree. In particular the end-order traversal of the parse tree is derived in one of the representations. The paper defines an inference system, termed the cycle inference system, which makes it possible for the theorem prover to model the method on which the representation is based. The general applicability of the cycle inference system to state-space problems is discussed. Given an unsatisfiable set S, where each clause has at most one positive literal, it is shown that there exists an input proof The clauses for the two representations satisfy these conditions as do many state-space problems.

Dialectology

A COMPUTER MODEL FOR THE ONTOGENY OF PIDGIN & CREOLE LANGUAGES

Sheldon Klein
Linguistics Department
University of Wisconsin
Madison

and

V. Rozencvejg
I MGPIIYa
Laboratoriya Machinnogo Perevoda
Moskva

Technical Report #238, December 1974, Computer Sciences Department University of Wisconsin

A system for simulation of language contact as a function of sociocultural, demographic and historical factors; computer model for the generation and growth of Pidgin and Creole languages purely in terms of structural principles and mechanisms. A generative semantic grammar is required for each language. The system contains representations of speakers interacting conversationally. They negotiate and bargain, trying to communicate, selecting constructions that minimize the problems of semantic parsing.

A METHOD FOR ASSESSING VARIABLE RULE AND IMPLICATION SCALE ANALYSES OF LINGUISTIC VARIATION

D. Sankoff, and P. Rousseau

In: Computers in the Humanities, J. L Mitchell, editor, 3-15, 197

There are currently two theories of linguistic variation. One suggests that there is an underlying probabilistic component in the competence of each speaker; the other considers the variation to be an artefact of grouping speakers with discretely different grammars--there being an implicational scaling relation among the possible grammars.

A rigorous comparison is made by making a variable rule analysis and an implicational scaling model for data on the deletion of the complementizer QUE in Montreal French. Using the probabilities predicted by the variable rule, the errors expected in fitting the data to the proposed scale are calculated by Monte Carlo simulation techniques. If the actual number of scaling errors is significantly less than predicted, then the variable rule analysis should be rejected. But the number is found to be exactly as expected from a variable rule analysis.

Acquisition

NATURAL LANGUAGE ACQUISITION

Larry R. Harris
Dartmouth College
Hanover, N.H.

Report No. TR-74-1, October, 1974

Adaptive techniques for lexical correlation and grammatical inference suitable for natural language processing are described. These techniques form the basis for a natural language understanding system that improves its performance with time. [AD-787 805/1GA; PC $3.75, MF $2.25]

Instruction

INFORMATION PROCESSING MODELS AND COMPUTER AIDS FOR HUMAN PERFORMANCE: SECOND-LANGUAGE LEARNING

Daniel N. Kalikow
Bolt Beranek and Newman Inc.
Cambridge, Mass.

Report No. BBN-2841; AFOSR-TR-74-1730, June 1974

Description of the second field evaluation experiment on the Mark II model of the Automated Pronunciation Instructor (API) system. Two matched groups of students were studied. All were native speakers of Spanish, and all were enrolled in the Intensive English Program at the University of Miami. One group was tested and trained with the API system; the other was simply tested within the same time frame. [AD-787 876/2GA; PC $6.25, MF $2.25]

A SEMANTICALLY CENTERED PARSING SYSTEM FOR MIXED INITIATIVE CAI SYSTEMS

R. R. Burton
Bolt Beranek and Newman
Cambridge, Mass.

Preprint, Annual Meeting, Association for Computational Linguistics, 1974

SOPHIE is a CAI system for teaching electronics. It uses AI techniques to perform question-answering, hypothesis verification and theory formation. Most of its capabilities are derived from using simulation models. It includes a highly tuned structural parser for allowing the student to communicate in a subset of English.

Instruction

A MODEL DRIVEN QUESTION-ANSWERING SYSTEM FOR A CAI ENVIRONMENT

J. S. Brown
Air Force Human Resources Laboratory
Lowry AFB
Colorado

Mimeographed 1973

A question answering system which permits a computer-assisted instruction (CAI) student greater initiative in the variety of questions he can ask is described. A method is presented to represent the dynamic processes of a subject matter area by augmented finite state automata, which permits efficient inferencing about dynamic processes and provides a satisfactory deep structure for paragraph generation. A CAI system dealing with meteorology is described which uses this automation model.
[EDRS EDO77195; $0.65-MF, $3.29-HC]

TOM SWIFT AND HIS ELECTRIC BILINGUAL GRANDMOTHER

T. Manwell

ACM SIGCUE Bulletin, 7, January 1973, 5-17

A description of a prototype computer assisted instruction system for teaching Russian. Although the course stresses the Russian morphological system, the construction of the computer program is applicable to other languages and language teaching specialties.

Documentation

ON AUTOMATIC QUESTION MODIFICATION IN FREE TEXT SEARCH

Christine Schaab

ZMD-A-26 Zentralstelle fur maschinelle Dokumentation Beuth-Vertrieb GmbH. Berlin 30. May 31, 1974

To increase the hope of a hit in free-text search, questions can be expanded by the addition of broader terms, narrower terms, related terms, and synonyms. To reduce the chance of false drops, expansions can carry fractional weights; a match is accepted only above a threshold. Examples are analyzed.
[ISBN 3-410-44026-7]

SUPPLEMENTARY PROGRAMS FOR INPUT TO STAIRS IN THE I & D - AREA

ZMD-A-27 Zentralstelle fur maschinelle Dokumentation Beuth Verlag GMBH. Berlin 30. July 15, 1974

STAIRS is an IBM program. The additions improve I-O operations in information and documentation use. Content: Data bank construction; dictionary improvements; data input; perameter input; dialogue retrieval; batch retrieval; user accounting.
[ISBN 3-410-44027-5]

STATISTICAL RELATIONS BETWEEN TEXT WORDS AND DESCRIPTORS

Rainer Kragenings

ZMD-A-25 Zentralstelle für maschinelle Dokumentation Beuth-Vertrieb GmbH. Berlin 30. May 2, 1974

From six issues of Food Science and Technology Abstracts, containing 8405 abstracts, the conditional probability **Z** of a descriptor given a text word was computed. In a seventh issue, containing 12,705 descriptor applications, the automatic assignment of descriptors was tested. With Z = 0.3 as cutoff, 28,251 applications include 9,139 correct; with Z = 0.7, 4,936 applications include 3,782 correct. [ISBN 3-410-44025-9]

Documentation

AUTOMATIC EXTRACTION OF CONTENT-SIGNIFICANT SENTENCES

J. M. Carroll and J. Cakarnis
Computer Science Department
University of Western Ontario

In Proceedings of the Second Open Conference on Information Science in Canada; edited by A. Gamache & R. Penner. Ottawa: Canadian Association for Information Science, 1974, 73-78

Content-significant sentences can be extracted automatically in decreasing order of importance from scientific papers available in machine-sensible format. Ordering is accomplished according to multiple regression of a non-linear combination of variables. The dependent variable is a subjective weighting of sentence importance. The independent variables include each word's relative frequency, the type-to-token ratio, mean-word length, and the predominant parts of speech encountered.

Translation

SOME SEMANTIC CONSIDERATIONS IN RUSSIAN-ENGLISH MACHINE TRANSLATION

Larissa Toma, Paul Garrett, Ludek Kozlik, Donald Perwin,
and Chuck Starr
Latsec Inc.
La Jolla, California

Report No. RADC-TR-74-189, August 1974

The final RADC supported optimization phase of the SYSTRAN Russian-English translation system. The primary thrust of this effort was directed at implementing the use of semantic analysis in both source language analysis and target language synthesis. This project has shown semo-syntactic analysis to be a highly feasible means of sophisticating machine translation and decreasing the need for post-editing. [AD-787 671/7GA; PC $5.75,MF $2.25]

MACHINE TRANSLATION. A BIBLIOGRAPHY WITH ABSTRACTS

E. J. Lehmann
National Technical Information Service

Report No. COM-73-1171/8, October 1973 Price $20.00

Bibliography containing 100 selected abstracts of research reports retrieved using the National Technical Information Service on-line search system. Research on machine translation of various languages is covered. Topics concerning syntax, computer programming, computer hardware and semantics are included.

Translation

USERS EVALUATION OF MACHINE TRANSLATION, GEORGETOWN GEORGETOWN MT SYSTEM, 1963 - 1973

Bozena Henisz Dostert
Texas A&M University

Report No. AD-768 451

The utility of unedited Russian-English machine translation in operational environment. The study is based on the performance of the Georgetown MT system at the AEC Oak Ridge National Laboratory and EURATOM Common Research Center. Production volume, suppliers/ users of machine translation services, production cost, computer environment, and improvements since 1963. Methodology of collection and analysis of individual assessments. A wide range of uniformly favorable responses elicited by questionnaire and personal interview from users in the USA, Italy, Belgium, Germany and Holland.

MACHINE-AIDED EDITING

Peter P. Toma, Jerry A. Carlson, David R. Stoughton, Joann P. Ryan
Latsec Inc.
La Jolla, California

Report No. RADC-TR-73-368, December 1973

SYSTRAN system displays simultaneously on a CRT a Russian input sentence and the English output, plus some context. The editor can delete, insert, rearrange, etc, via key-board operations. Recommendations for optimization of this system include how editing functions might best be carried out on video editing terminals in conjunction with a computerized typesetting system. Various criteria for choosing a typesetter are considered, and a Photon Pacesetter Mark II is recommended to best serve the future needs of the Technical Translation Division at Wright-Patterson AFB. Explores the feasibility of automatically flagging SYSTRAN output through application of English well-formedness criteria. Includes a survey of the distribution in English of adnominal genitive constructions, compound nouns, and possessive-noun plus noun constructions and an illustration of how the findings of this survey might be applied to improve the English output of Russian adnominal genitive structures.
[AD-775 160/5GA, PC $500, MF $1.45]

Translation

DEVELOPMENT OF CHINESE-ENGLISH MACHINE TRANSLATION SYSTEM

William S-Y Wang, and Stephen W. Chan
University of California
Berkeley

Report No. RADC-TR-74-22, February 1974

Describes a 2-1/3 year effort to further develop the prototype Chinese-English Machine Translation System. Additional rules were incorporated into the existing grammar for Chinese analysis and interlingual transfer, with emphasis on the latter. CHIDIC was updated and revised. Approximately 16,000 new entries were added to CHIDIC, bringing the total available entries to over 73,000. Linguistic work on a random access dictionary incorporating feature notation was carried out. A new design for the translation system was initiated and partially programmed for conversion of the current system from a CDC 6400 version into an IBM version. Better control of the parsing process was achieved by improving the segmentation procedures during input, and by addition of more revealing diagnostic printouts as steps toward reduction of spurious ambiguities. The Model 600D Chinese Teleprinter System was used for the first time to prepare large batches of texts for input. A total of 307 pages of machine readable texts, comprising 300,000 characters were prepared during this report [AD-776 813/8GA; PC $4.75, MF $1.45]

THE EVOLUTION OF A COMPUTER MODEL OF AUTOMATIC TRANSLATION BASED ON JUNCTION GRAMMAR

Lance S. Smith
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1972

Initially the system was applied to Russian analysis and English synthesis. Later versions have Russian and English synthesis, and English, French, Spanish, Portuguese, German, and Japanese synthesis. Analysis is based on the junction grammar model of syntax, and has a vertical cycle, i.e., all the tests are applied to a single node before moving on to the next node. Reverse Polish notation is used to represent the analysis. A skeleton supervisor calls in language specific routines for language specifrc problems, for example, in the interlingual transfer process.

Translation

INTERACTIVE SENTENCE PARSING AND TRANSFER

Alan K. Melby
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1973

The system for manipulating syntactic analyses generated using the junction grammar formalization. The system is used to construct inputs for synthesis routines and to study comparative grammars of languages. In addition to operating on syntactic structures, semantic ambiguities are resolved by requests to the user.

ON THE FRENCH EQUIVALENTS OF ENGLISH PASSIVE CONSTRUCTIONS

Irena Bellert
Groupe de Recherches pour la Traduction Automatique
Universite de Montreal

Etudes de linguistique appliquee a la traduction automatique, 1974

Selection of passive, impersonal (il, on), reflexive, active. Keys are anaphora, modal, adverbs, quantifiers, by-phrase, heaviness.

COMPUTERS, PROGRAMMING, AND NATURAL LANGUAGES

Ordinateurs, Programmation, et Langues Naturelles

Jacques Andre and Catherine Fuchs

Maison 4M Mame
49 Boulevard Preuilly
37017 Tours Cedex
France

TABLE OF CONTENTS

Described as a book for students in linguistics, sociology, psychology, and history. Analysis of linguistic deep structure using a language based on Algol 68.

35.00 Francs

Programming

STUDY AND COMPILATION OF COMPUTER LANGUAGES

Y. Wallach
Department of Electrical Engineering
Technion-Israel Institute of Technology
Haifa, Israel

Gordon and Breach Science Publishers
New York, London, and Paris
1974

TABLE OF CONTENTS

ISBN 0 677 04010 5
LC Card 73-89202

$45.00

Programming

STRING AND LIST PROCESSING IN SNOBOL4: TECHNIQUES AND APPLICATIONS

Ralph E. Griswold
Department of Computer Science
The University of Arizona

Prentice-Hall, Inc.
Englewood Cliffs, New Jersey
1975

TABLE OF CONTENTS

"SNOBOL4 is usually described as a string-processing language...[but] is...a general-purpose language that stresses 'nonnumerical' facilities. The list-processing facilities in SNOBOL4 are not as well known as the string-processing facilities." p. xi

ISBN 0-13-853010-6

Programming

FORTPAN TECHNIQUES

WITH SPECIAL REFERENCE TO NON-NUMERICAL APPLICATIONS

A. Colin Day

Cambridge
at the University Press
1972

TABLE OF CONTENTS

ISBN 0 521 08549 7 hard cover
0 521 09719 3 paperback
LC Card 72-78891

$3.95 paperback

Programming

A MULTI-PROCESSING APPROACH TO NATURAL LANGUAGE

Ronald M. Kaplan
Harvard University
Cambridge, Massachusetts

Proceedings of the 1973 National Computer Conference, 435-440.

The General Syntactic Processor (GSP) incorporates basic facilities of Woods's augmented transition network grammar and Kay's 'powerful parser'. The former is a top-down system that can needlessly repeat computation on backtracking. The latter is a bottom-up parser that exhaustively forms every wellformed string in a sentence. GSP overcomes the disadvantages of these systems through use of a transition netowrk as the formalism for the grammar, and charts as parse structures. In a chart each constituent is represented only once; thus a chart represents in one structure all possible parses of a sentence. The subnetworks of GSP are conceived as a collection of asynchronous processes which operate on overlapping chart sections and use the chart to communicate with each other.

ENGLISH ANALYSIS: WHAT YOU SEE IS WHAT YOU GET

Daryl Gibb
Brigham Young University

Proceedings of the [BYU] Linguistics Symposium, 1973

In processing an English sentence, humans bring into play a vast amount of experience and logic. In the BYU system, the automatic analysis is assisted by human interaction to resolve problems of antecedents of proforms, the noun phrase modified by a prepositional phrase, etc. The system is being used to study the problems that arise in language analysis and to form a base for extending the automated system.

Picture analysis

ISIS: AN INTERACTIVE FACILITY FOR SCENE ANALYSIS RESEARCH

J. M. Tenenbaum, T. D. Garvey, S. A. Weyl, and H. C. Wolf
Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California

SRI Publication No. Z111 June 1974

Summarizing initial progress in developing a computer system that can be rapidly programmed to analyze any class of pictorial scenes. Scene analysis programs have been awkward to develop using conventional programming systems because of the difficulty of formulating pictorial descriptions in symbolic terms. Picture processing techniques are inherently ad hoc and must be deduced empirically for each application.

An interactive system specifically designed for expressing and experimenting with perceptual strategies, it allows an experimenter to describe basic perceptual concepts to a computer in terms of pictorial examples. Examples are designated graphically by encircling areas of a displayed scene with a cursor. A concept is represented internally by values of primitive feature-extraction operators that distinguish it from examples of previously defined concepts. Concepts so defined constitute a common vocabulary, shared by man and machine, that can be used symbolically in describing objects and specifying scene analysis procedures.

The system has been used to formulate interactively descriptions that distinguish objects in indoor room scenes and programs that locate these objects in images.

Picture analysis

A RELATIONAL DATA BASE SCHEMA FOR DESCRIBING COMPLEX PICTURES WITH COLOR AND TEXTURE

Tosiyasu L. Kunii,
Information Science
Laboratories
University of Tokyo

Stephen Weyl and Jay M. Tenenbaum
Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California

SRI Publication No. Z109 June 1974

The potential for applying computers to large masses of pictorial information, such as remotely sensed earth resource data or medical photographs, leads to the necessity for carefully designing underlying data structures. This paper presents a relational schema for describing complex pictures having color and texture. The schema is in abstract form free from any specific implemented storage structure. It provides a means for extending the lifetime of pictorial data by divorcing it from particular hardware and applications programs. Furthermore, it allows integration with shared data bases that include symbolic and numerical information.

The schema presented uses Codd's relational formalism to achieve modularity and associativity of data. Since this formalism was originally developed for commercial applications, this paper provides an outline of the relevant concepts. To accommodate picture-processing applications, procedures for reducing relations to canonical form, based on a world model, are considered from an inductive viewpoint.

Literature

COMPUTERS IN THE HUMANITIES

J. L. Mitchell, Editor
University of Minnesota

Edinburgh University Press
University of Minnesota Press
Minneapolis
1974

TABLE OF CONTENTS

A summary of this paper appears on the indicated frame of this fiche.

COMPUTERS IN THE HUMANITIES

TABLE OF CONTENTS (Continued)

COMPUTERS IN THE HUMANITIES

TABLE OF CONTENTS (Continued)

Literature

ASSOCIATION FOR LITERARY AND LINGUISTIC COMPUTING

BULLETIN

Volume 2 Number 3
Michaelmas Term 1974

CONTENTS

Literature

On the authenticity of the Baligant episode in the 'Chanson de Rolande

J. R. Allen

Computers in the Humanities, J. L. Mitchell, editor, 65-72, 1974

There is a dispute over the authenticity of one episode of the poem. The author uses a vocabulary distribution test to show that there is a significant indication of stylistic differences between the Baligant episode and the rest of the poem.

Authorship attribution in Jacobean dramatic texts.

W M. Baillie

Computers in the Humanities J. L. Mitchell, editor, 73-81, 1974

Using the EYEBALL program for stylistic analysis, statistics are found for function-word modifiers, complements, noun modifiers and coordinators/subordinators that distinguish the writings of Fletcher from those of Shakespeare. This data is to be applied to Henry VIII, the authorship of which is disputed between the two writers.

An EYEBALL view of Blake's Songs of Innocence and of Experience

D. Ross

Computers in the Humanities, J. L. Mitchell editor, 94-108, 1974

Starting from a text in natural language the EYEBALL system provides statistics of vocabulary distribution, the number of syllables per word, and an augmented text with annotations for each word which indicate syllable length, grammatical category and function, locations in clause, sentence and text. Analyses of word and clause length, distribution of word classes, and properties of combined word classes are also produced. The system is illustrated by application to Blake's early poems.

Literature

THE SEMANTIC SIGNIFICANCE OF SPATIAL MOVEMENT IN NARRATIVE VERSE: PATTERNS OF REGRESSIVE IMAGERY IN THE DIVINE COMEDY

C. Martindale

Computers in the Humanities, J. L. Mitchell, editor, 57-64, 1974

Using COUNT, a content analysis program of the General Inquirer ilk, with a content analysis dictionary, the Regressive Imagery Dictionary, the hypothesis that a quantitative translation from the imagery of downward spatial movement into psychoanalytic regression, and ascending movement into movement away from regression, is substantiated.

ANNUAL BIBLIOGRAPHY FOR 1973: LANGUAGE AND LITERATURE

Computers and the Humanities, 8, 1974, 100-108

An unannotated listing by author of some 310 articles.

Mathematics

MATHEMATICAL DESCRIPTION OF RELATIONS IN AUTOMATIC INDEXING AND RETRIEVAL

Hubert Hüther

ZMD-A-24 Zentralstelle ur maschinelle Dokumentation Beuth-Vertrieb GmbH. Berlin 30. April 30, 1974

The algebraic theory of relations on a set: transitivity, symmetry, equivalence, order. Statistical theory: distribution theory for relations, equivalence classes, growth. Application of matrix theory. No examples from documentation.
[ISBN 3-410-44024-X]

NAME INDEX
A J C L 1974

Names accompanied only by the title of a contribution or equally limited information are omitted. C is an opaque card.

A

B

C is an opaque card
M is a microfiche

13:13 is microfiche 13, frame 13

C

D

E

F

G

I

J

K

L

M

N

O

Þ

R

S

T

U

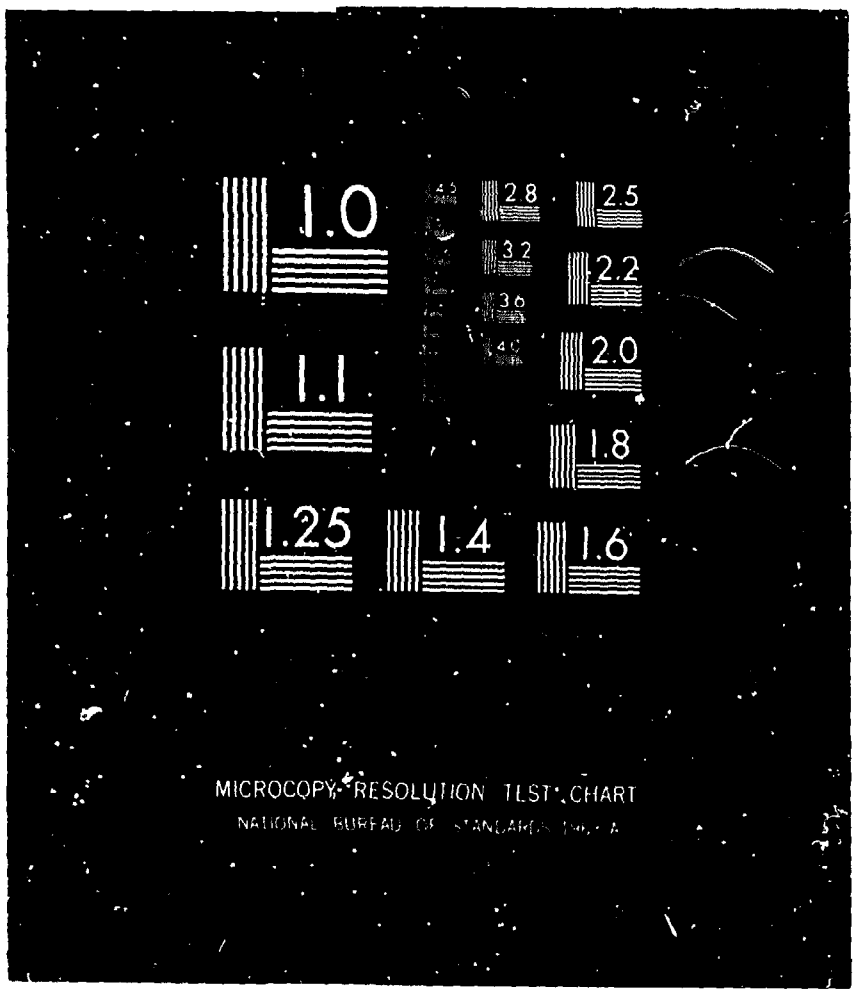
V

W

Y

Z

END


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