

LISTES OF CONTRIBUTIONS
TO THE
INTERNATIONAL CONFERENCE
ON THE APPLICATION OF MATHEMATICAL MODELS
TO COMPUTERS IN LINGUISTICS
Human Alphabet Contributions

MAY 3 - 9, 1975

A. Ljudskanov
Mathematics and Mechanics
of the Organizing Committee

Copyright 1975

For Computational Linguistics

EDITORIAL NOTE

The conference was large and vigorous; 189 persons attended, of whom 69 were foreigners to Bulgaria--from 16 different countries. The substance of the conference consisted of 101 papers, 24 of them invited, and three roundtables. In addition, the organizers provided the most agreeable amenities in generous measure.

This microfiche carries the Roman-alphabet summaries of 69 contributions, as submitted by their authors to the organizers of the Conference. Translations of Cyrillic-alphabet contributions are in preparation; they will be published on another fiche when available.

Photographs taken at the Conference are included at the end of this fiche.

AJCL thanks the organizers for permission to reprint the summaries and for the hospitality extended to the editor at Varna.

--D.G.H.

TABLE OF CONTENTS

GENERAL

SUR LES MODELES HEURISTIQUES EN LINGUISTIQUE	<i>A. Ljudskanov</i>	8
SOME METHODOLOGICAL ASPECTS OF MEASUREMENT IN LINGUISTICS AND PSYCHOLINGUISTICS	<i>Encho Gerganov</i>	10
MEASUREMENT OF LINGUISTIC COMPETENCE	<i>L. Mavlov, V. Rainov, and E. Gerganov</i>	11

PHONETICS - PHONOLOGY

AN ALGORITHM USING LINGUISTIC INFORMATION AND ITS APPLICATION TO THE ANALYSIS OF SPEECH IN THE SPECTRAL DOMAIN	<i>P. Christov</i>	12
STATISTICAL DIFFERENTIATION OF TEXTS ON THE PHONEMIC LEVEL	<i>M. Ludvíkova</i>	13
AUTOMATISCHE TRANSKRIPTION THAILANDISCHER UND LAOTISCHER ORTOGRAPHISCHER TEXTE	<i>Fischer Gero</i>	14
ON PERCEPTION OF SPEECH SOUNDS	<i>Encho Gerganov, Tzvetanka Pashova, and Vaklin Nikolov</i>	15
THE IMPORTANCE OF THE SECONDARY ACOUSTICAL PARAMETERS TO THE SPEECH SYNTHESIS	<i>A. Misheva</i>	16
SYNTHESIS OF BULGARIAN VOWELS AND THEIR PERCEPTION BY BULGARIAN INDIVIDUALS	<i>Radka Kirlova and Maria Popova</i>	17
SPECTRAL ANALYSIS OF SOME ACOUSTICAL CHARACTERISTICS OF VOWELS AND THEIR IMPORTANCE FOR SPEECH QUALITY	<i>Maria Popova</i>	18
THE ROLE OF THE VOWEL FORMANT TRANSITIONS AND NOISE CHARACTE- RISTICS OF THE BULGARIAN CONSONANTS IN THE PROCESS OF RECOGNITION	<i>D. Tilkov and A. Misheva</i>	19

WRITING

HYPHENATION PROGRAMMES AND THEIR APPLICATION IN AUTOMATED TYPESETTING SYSTEMS	<i>Konstantin Bossilkov, Kiril Naidenov, and Marin Kotarov</i>	20
ZUR AUTOMATISCHEN SILBENTRENNUNG	<i>J. Fischer</i>	21

LEXICOGRAPHY - LEXICOLOGY

DAS ELEKTRONISCHE WORTERBUCH DER FACHSPRACHEN - STRUKTUR UND NUTZUNGSMÖGLICHKEITEN <i>G. Neubert</i>	22
LE DICTIONNAIRE AUTOMATIQUE TRILINGUE DE "THÉSÉE", THESAURUS COMPATIBLE AVEC LES SYSTÈMES D'INDEXATION FRANÇAIS-ANGLAIS ALLEMAND <i>A. Deweze</i>	23
LE PROBLÈME DU FICHER MULTILINGUE DANS L'IMPLANTATION DE CANTERM <i>G. Rondeau</i>	24
LA CATÉGORIE DE DIMINUTIVITÉ DU POINT DE VUE DE LA LINGUISTIQUE QUANTITATIVE <i>Borimir Kristev</i>	26
STATUS OF NUMBER NAMES IN THE WORD COUNTS <i>Boyan Nikolayev</i>	27
GRAMMAR	
CLUB-SYSTEMS <i>V. C. Borscev and M. V. Chomjakov</i>	28
THE PROOF THAT LANGUAGE IS NOT CONTEXT-FREE <i>H. Brandt Corstius</i>	30
ON COMPARATIVE STUDY OF VARIOUS TYPES OF SYNTACTIC CONFIGURA TIONS IN ALGEBRAIC LINGUISTICS <i>M. Semeniuk-Polkowska</i>	31
LES SYSTÈMES A.T.E.F. ET C.E.T.A. <i>J. Chauché</i>	32
DEFINITORISCHE BEGRÜNDUNG DER ABHÄNGIGKEITSGRAMMATIK <i>Jürgen Kurtze</i>	33
UNTERORDNUNGSKRITERIEN FÜR DIE ABHÄNGIGKEITSGRAMMATIK UND EINIGE BEISPIELE AUS DEM BULGARISCHEN <i>Jürgen Kunze and Alexander Ljudskanov</i>	35
SYNTAKTISCHE MEHRWEGANALYSE AUF DER GRUNDLAGE VON BÜSCHELN <i>Dorothee Arold</i>	36
LA DÉCOMPOSITION AUTOMATIQUE DE LA PHRASE COMPOSÉE BULGARÉ EN PROPOSITIONS SIMPLIES <i>Edouard Selian</i>	37
AUTOMATIC ANALYSIS OF BULGARIAN SCIENTIFIC TEXTS <i>Hristo C. Georgiev</i>	38
A TECHNIQUE FOR AUTOMATIC SEGMENTATION OF POLISH VERBS <i>Wiesław Lubaszewski</i>	39
SOME PROBLEMS OF MORPHOLOGICAL ANALYSIS OF POLISH LANGUAGE <i>W. Łukaszewicz</i>	40
FLEXIVISCHE UND WORTBILDUNGSANALYSE DES DEUTSCHEN <i>Barbara Rüdiger</i>	41

SEMANTICS - DISCOURSE

COGNITIVE NETWORKS <i>David G. Hays</i>	42
LINGUISTICS AND ARTIFICIAL INTELLIGENCE <i>Petr Sgall</i>	44
SOME QUESTIONS OF FORMAL MODELS IN LINGUISTIC SEMANTICS <i>E. Hajičová</i>	46
ON THE QUESTION OF THE DESCRIPTION OF POSSESSIVITY IN THE AUTOMATIC PROCESSING OF TEXTS <i>Petr Pitha</i>	47
AN INTEGRATIVE SEMANTIC APPROACH TO INTERSENTENCE PHENOMENA <i>Th. R. Hofmann</i>	48
RECENT RESEARCH RESULTS ON THE COMPUTATIONAL SIMULATION OF LANGUAGE UNDERSTANDING SYSTEMS <i>Luc Steels</i>	49
USER-ORIENTED AUTOMATIC SEMANTIC ANALYSIS OF TEXTS <i>Jacques Noël</i> .	50
FORMALE SEMANTIK UND LINGUISTISCHE MODELLIERUNG <i>Alfred Nozicska</i>	52
SIGNIFICANCE OF ISOLATED MORPHOLOGIES TO FORMALIZE NATURAL LANGUAGES <i>Kimmel Peter</i>	53
ZUM PROBLEM EINER FORMALISIERTEN BESCHREIBUNG DER SEMANTIK DES FINITEN VERBS (AM BEISPIEL DES BULGARISCHEN) <i>H. Walter</i> . . .	54
ON SEMANTICAL ORGANIZATION OF TEXT <i>W. Marciszewski</i>	56
ZUR PRONOMINALEN REFERENZ IN DEUTSCHEN WISSENSCHAFTLICHEN UND LITERARISCHEN TEXTEN <i>Sw. Gizowa</i>	57
SUR LES RÉFÉRENCES NON-PRONOMINALES <i>A. Ljudskanov</i>	58
MEASUREMENT OF READABILITY OF WRITTEN BULGARIAN <i>Encho Gerganov, Andriana Mateeva, and Borimir Kristev</i>	59
LINGUISTICS	
MACHINE DE CONTRÔLE POUR L'ENSEIGNEMENT DE LA DIALECTOLOGIE BULGARE (M.C.E.D.B.) <i>Tsanco Mladenoff and Danaile Dancoff</i>	60
UN MODELE LOGIQUE DE RECONSTRUCTION CHRONOLOGIQUE EN PHONOLOGIE NATURELLE <i>T. Charpentier, J. Y. Morin and M. Picard</i>	61
COMPUTATION	
COMPUTER EXECUTION OF FUZZY ALGORITHMS <i>V. Dimitrov, W. Wechler, D. Drjankov, and A. Petrov</i>	62

DOCUMENTATION

INFORMATION RETRIEVAL SYSTEMS <i>D. M. Dobrev</i>	63
FACETS, FORMS, AND RINGS <i>B. Harris</i>	64
PROGRAMS OF EXPERIMENTAL RADIOELECTRONICS AUTOMATED INFORMATION RETRIEVAL SYSTEM AND ITS FUNCTIONS <i>K. Ilieva, A. Kuzmanov, I. Mitsev, T. Djakov, E. Jerassy, G. Bakalsky, S. Anastassova, T. Anghelova I. Stamenova, and B. Gheorghieva</i>	66
ELIMINATION OF POLYSEMIA AND HOMONYMITY BY A DIALOGUE IN THE RADIOELECTRONICS AUTOMATED INFORMATION-RETRIEVAL SYSTEM <i>K. Ilieva and A. Ljudskanov</i>	67
CONCEPTUAL MODEL OF THE AUTOMATED LEGAL INFORMATION RETRIEVAL SYSTEM <i>Jan Bobrowski</i>	68

TRANSLATION

LA STRATEGIE SELECTIVE DANS LA LUMIERE DES NOUVEAUX DEVELOPPE- MENTS DE LA TRADUCTION AUTOMATIQUE <i>A. Ljudskanov</i>	69
ON FULLY AUTOMATIC HIGH QUALITY TRANSLATION <i>Irina Uzunova</i>	71
OPTIMUM LINKING OF INDEPENDENT ANALYSIS AND SYNTHESIS <i>Svatava Machova</i>	72
ZU EINIGEN VORAUSSETZUNGEN FUR EIN LINGUISTISCHES MODELL DER TRANSLATORISCHEN KOMPETENZ <i>Gert Jäger</i>	73

SOCIAL - BEHAVIORAL SCIENCE

SOME PROBLEMS OF MODELLING OF APHATIC DISORDERS OF THE SPEECH IN ITS PHYSIOLOGICAL, PSYCHOLOGICAL AND LINGUISTIC ASPECTS <i>R. Raychev and A. Ljudskanov</i>	75
DISORDERS OF LINGUISTIC COMPETENCE AND PERFORMANCE IN LANGUAGE PATHOLOGY <i>Ludmil Mavlov</i>	77
AN ATTEMPT FOR ANALYSIS OF THE SYNTACTIC LEVEL OF THE SPEECH IN PATIENTS WITH APHASIA <i>P. Ovcharova, A. Ljudskanov, R. Raychev, S. Karagyozova</i>	78
INFORMATIONAL AND STATISTICAL INDICES OF THE REPORTED SPEECH OF APHASICS <i>P. Ovcharova, L. Tzoneva-Pencheva, and M. Vulcheva</i>	79

HUMANITIES

GRAMMAIRE DES CONTES DU TYPE 300 DU-CATALOGUE AARNE-THOMPSON (APPLICATION AUX CONTES POPULAIRES ROUMAINS) <i>Mihaela Dumitru</i> . . .	80
A SYSTEM APPROACH TO JULIO CORTÁZAR NOVELS <i>Aurora Niculescu-Mizil</i>	81
MATHEMATICAL HIERARCHIZATION OF ARCHITECTURAL STYLES (APPLIED TO COLUMNS) <i>G. Chioca</i>	82
CHRONOLOGICAL SERIATION OF THE WORKS OF AN AUTHOR BY MEANS OF COMPUTER <i>L. Boneva</i>	83

INSTRUCTION

INFORMATION-COMPACTIZING OF STUDENTS' COMPOSITIONS <i>Rosalina Novachkova</i>	84
PRACTICAL JUSTIFICATION OF NOTATIONAL CONVENTIONS <i>Theodora Mikhailova</i>	85
AUTHOR INDEX	86
PHOTOGRAPHS FROM THE CONFERENCE	88

SUR LES MODELES HEURISTIQUES EN LINGUISTIQUE

A.Ljudskanov /RP de Bulgarie/

I. Quelques précisions épistémologiques

1. La description et la science; critères de vérité; fondements idéologiques; force prognostive et gnocéologique; sciences descriptives /taxonomiques/ et explicatives /théoriques/; méthodes inductives, déductives et hypotheticodéductives; théorie. Quelle science est la linguistique?

2. Voies "directe" et "indirecte" d'acquisition des nouvelles connaissances scientifiques.

II. Bref aperçu historique et logique

1. Les modèles dans les domaines "traditionnels" et en logique mathématique.

2. Pénétration des modèles dans des domaines nouveaux - les domaines "non-traditionnels". Influence de la cybernétique.

3. Comparaison des modèles, qui s'utilisent dans ces deux domaines. Constatations; contradictions et problèmes non-résolus /p.e. définition ; possibilité d'introduire une notion générique de modèles valable pour les deux domaines; types des originaux; niveau d'abstraction; types des relations; caractère ontologique et deontologique du point de vue de la théorie marxiste de la réverbération; but - interprétation des théories ou élaboration des théories, etc./.

III. Les modèles en linguistique contemporaine

Pénétration et quelques points de vue sur la méthode des modèles en linguistique: les modèles sont inacceptables en linguistique; la linguistique est impossible sans l'élaboration des modèles; l'introduction des modèles n'est que le résultat de la mode; les modèles sont admissibles seulement dans la linguistique appliquée, etc. Quelques définitions et analyse critique.

IV. Classification provisoire des modèles - modèles physiques et idéaux: modèles-images et modèles symboliques; modèles pragmatiques et heuristiques. La mathématique est science des

modèles.

V. Le processus de l'élaboration des modèles heuristiques et la notion de modèle-objet.

1. Original, modèle-objet (relations), modèle (relations), modèles mathématiques (types); traitement, réalisation et vérification par la pratique; acceptation et incrément dans 'une théorie plus vaste; correction ou refus.

2. Force heuristique. Exemple.

VI. Nécessité et possibilités des modèles heuristiques en linguistique théorique.

1. Le type, l'objet, la base logique et le but d'une science particulière conditionnent la méthodologie comme ensemble d'approches de recherche.

2. Le point de vue sur l'objet et le but de la linguistique contemporaine conditionne la nécessité d'élaborer des modèles heuristiques. Donc la méthode des modèles est conditionnée par le développement interne de la linguistique et n'est pas le résultat de la mode.

3. La force théorique des modèles et la constructivité.

4. L'historisme et la méthode des modèles (l'élaboration des modèles-objets ne nie pas, mais au contraire suppose les connaissances acquises par voie traditionnelle).

5. Problèmes "extralinguistiques" et restrictions des possibilités de la méthode des modèles (au moins de nos jours).

6. La connaissance scientifique de tous les aspects de la langue comme phénomène social exige la convergence de différentes méthodes.

**SOME METHODOLOGICAL ASPECTS OF MEASUREMENT
IN LINGUISTICS AND PSYCHOLINGUISTICS**

Encho Gerganov (Bulgaria)

In contemporary linguistics and psycholinguistics there are a great number of studies, in which the investigators use numbers and operate with them by means of the tool of mathematics. However, the linguists and psycholinguists have not always paid attention to the following very important questions. In what way have the numbers been obtained and what do they mean? These questions relate to the problems of measurement. In this paper measurement in linguistics and psycholinguistics is discussed from the point of view of the Suppes and Zinnes' basic measurement theory, Coombs' theory of data and Torgerson's theory of scaling. Various definitions of measurement are considered. Types of scales and kinds of measurement are discussed by using typical examples from linguistic and psycholinguistic works. Coombs four kinds of data (four quadrants) are given: preferential choice data, single stimulus data, stimulus comparison data and similarities data. Various linguistic and psycholinguistic data are mapped into these quadrants. Measurement of attributes of language objects: familiarity of words, abstractness of nouns, pronunciability of syllables etc. is mapped either into single stimulus data or into stimulus comparison data according to the scaling method used: the law of categorical judgment or the law of comparative judgment. Stylistic preferences of one means of expression over another are mapped into preferential choice data. Similarities data are used widely in studying of speech perception.

MEASUREMENT OF LINGUISTIC COMPETENCE

L. Navlov, V. Rainov and E. Gerganov

(Bulgaria)

Language communication requires knowing of the rules of a certain language system (competence), as well as their utilization (performance). The aim of the study is to measure the degree of competence development in individual normal subjects and degradation of competence in patients with brain lesions.

A test battery for measurement of linguistics competence has been devised. The battery includes 7 tests for measuring the following aspects of linguistic competence : semantic correctness of verbs and adjectives in sentences; grammatical correctness of verbs, adjectives, prepositions, pronouns, syntactic structures, articles in sentences. Each test consists of 10 items and each item includes 5 quasisynonymous alternatives differing with respect to their degree of correctness.

The correctness has been measured by the method of paired comparisons (Torgerson, 1958). It was found that the normal subjects differ in their competence

AN ALGORITHM USING LINGUISTIC INFORMATION AND ITS APPLICATION
TO THE ANALYSIS OF SPEECH IN THE SPECTRAL DOMAIN

Philip Christov, Bulgaria

An algorithm is described in this paper which performs spectral analysis of speech and automatic extraction of the formant frequencies from the speech spectra. The algorithm is programmed in FORTRAN-IV IBM 360/370 version. It has the shape of an application program package. The most important parts of this package are:

Pitch extraction program

Waveform analysis program

Quasi-stationary region and section determining program

Pitch synchronous spectral analysis program

Formant frequencies extraction program

The programs of the application package above mentioned are tested by the processing of a rather large experimental material of 614 Bulgarian vowel sounds uttered by 60 speakers. The vowels are imbedded in words in comparable allophones and verified by 20 listeners. It has been established during the data processing that the performance of the algorithm can be significantly improved by the use of a priori linguistic information in the program of formant extraction. Thus in the latter subroutines have been inserted for preliminary classification of the vowels in 'front', 'central' and 'back' and for processing each of these subclasses according to different rules. In result the total percent of uncorrect decisions about the values of the formant frequencies dropped to 18,06%. A thorough analysis of the errors has shown that further improvement of the algorithm can be achieved by the use of more detailed linguistic information.

STATISTICAL DIFFERENTIATION OF TEXTS ON THE PHONEMIC LEVEL

M. LUDVÍKOVÁ (ČSSR)

An investigation was carried out dealing with the problem which statistically evaluated linguistic properties on the phonemic level might be considered as differentiating spoken texts of different stylistic areas. The unit of investigation was the syllable. To this purpose we used the material of a tape recorded scientific lecture and that of a conversation supposed to represent two style extremes. The total length was 10.000 syllables (5000 syllables for each sample). - First, we examined the texts for homogeneity; a statistical test confirmed that the frequency distributions of syllables for the two samples differ significantly from each other - the texts belong to two different corpuses. Further statistical searches discovered substantial divergences in other points, too, e.g. in the extent of syllabic inventory, syllabic word length, number of monosyllables, index of repetition of syllables, number of syllables with a long nucleus: a diverging value was also ascertained with the entropy of both texts. Coinciding values were obtained e.g. for the covering of texts by sequential frequency groups, repertory and distribution of syllable types etc. - The research confirms that there exist quantitatively expressed characteristics on the phonemic-syllabic level which are suspect to be stylistically distinctive.

AUTOMATISCHE TRANSKRIPTION THAILANDISCHER
UND LAOTISCHER ORTHOGRAPHISCHER TEXTE

Fischer Gero (Osterreich)

Die thailändische und die mit ihr verwandte laotische Schrift leiten sich von der Devanagari-Schrift ab. Beide Schriftsystemen haben u.a. eine Reihe von gemeinsame Eigenschaften, die im Bericht beschrieben werden.

Die Algorithmierung des Transkriptionsprozesses vollzieht sich ausschließlich auf der graphematisch-phonematischen Ebene, dies bedeutet nun die Lösung der folgenden Aufgaben: 1) Strukturelle und positionelle Untersuchungen zur Graphematik für die mechanische Identifizierung der Grapheme und Rekognisierung der Wort-/Silbengrenzen; 2) Entwicklung eines Algorithmus mit Input: orthographischer Text und Output: phonologisch transkribierter in Silben segmentierter Text, der im wesentlichen aus zwei Teilen besteht: (a) Analyse: Ermittlung der aktuellen Oberflächenstruktur des orthographischen Textes; Identifizierung der Grapheme und Segmentierung des Textes in kleinere Einheiten; (b) Transformationssynthese: Überführung der graphematischen Oberflächenstrukturen in die phonologischen Silbenstrukturen bzw. mit der phonologischen Notation als Output.

Ergebnisse: Die thailändische Orthographie realisiert in geringerem Ausmass als die laotische syllabophonetische Prinzipien. Zufriedenstellende Ergebnisse sind im Thailändischen dort zu erwarten, wo vorwiegend thailändisches Wortgut die Texte bestimmt, fehlerquellen treten vor allem bei Fremd- und Lehnwörtern auf und das speziell bei der Segmentierung. Deshalb sind für das Thai bei bestimmten Texttypen zusätzliche Operationen notwendig, die die auftretenden Homographien, Polystrukturalitäten etc. eliminieren. Für das Laotische hingegen bestehen - nach dem derzeitigen Stand der Tests - diese Probleme nicht.

PERCEPTION OF SPEECH SOUNDS

Zvetanka Pashova, Vaklin Nikolov (Bulgaria)

of speech sounds is described in terms of a system of psychophysical operators (SPO) and a feature system (FWS). The input to SPO consists of phonological features - axes of an acoustical space, and speech sounds are presented as points. SPO maps this space into psychophysical space with the same axes. The axes of the latter in turn constitute a system which weighs each axis and at the output a space with a new dimensionality is formed. This is the space (PsSp) of perception of speech sounds. It is hence the same for all subjects (Ss) with one native language, while FWS has developed during the acquisition of a native language in childhood. It is supposed, that one native language may differ from FWS of another native language, especially when the two are very different phonological systems. As a result, the structures of the PsSp of the two groups of Ss may differ. This difference could manifest either in the dimensionality of the two spaces, or the same dimensionality but different weights of axes, or both. A psychological scaling technique is used to obtain similarity data for Bulgarian vowels and Arabic vowels. The complete method of triads (Torgerson, 1958) is used with 100 Bulgarian Ss and 100 Arabian Ss. These data are analyzed. It is found, that Bulgarian Ss have two-dimensional perception of Bulgarian vowels, while Arabian Ss have one-dimensional PsSp. Consequently, our results are reported.

THE IMPORTANCE OF THE SECONDARY ACOUSTICAL PARAMETERS
TO THE SPEECH SYNTHESIS

A.Misheva /Bulgaria/

In speech synthesis the main efforts of research workers are devoted to the realization of schemes and circuits by means of which a signal with a spectrum equivalent to the natural speech spectrum is created. However, each speech sound is characterized not only by its spectrum but also by its intensity, fundamental frequency and duration. The absolute values of these parameters are conditioned mainly by the individual qualities of the utterance - the voice loudness, its pitch, intonation and tempo. That is why these characteristics are called secondary, i.e. not of primary importance to the phoneme recognition. On the other hand, the investigation of the secondary acoustical parameters of the different phonemes presents grounds to consider their relative values as specific, inherent to each phoneme or groups of phonemes.

The experimental quantitative analysis of statistical data with a sufficient large volume, which is carried out in the Phonetic Laboratory of the Bulgarian Language Institute, shows that there exist significant differences between the relative values of the secondary acoustical features of the Bulgarian phonemes. This is valid for the separate vowels as well as for the different groups of consonants - nasals, fricatives, stops and so on.

The results of this study make it possible to assign in advance the secondary acoustical parameters to each Bulgarian phoneme in the speech synthesizing machine, i.e. to involve them in the programme of the synthesis by rules. This will result in an increased intelligibility and improved naturalness of the synthetic speech. Moreover, the relative values of the secondary features form the acoustical base on which the theory of the physical correlates of the stress in the Bulgarian language will be developed thus providing a most economical programming in the speech synthesis systems.

SYNTHESIS OF BULGARIAN VOWELS AND THEIR
PERCEPTION BY BULGARIAN INDIVIDUALS

Radka Kirlova and Maria Popova (Bulgaria)

Synthetic vowels give wide scope of opportunities for investigating various aspects of perception. The major problems of synthesis of Bulgarian vowels are: a) defining the average formants of vowels; b) defining of the band-width of formants' filters and c) construction of link by means of which it becomes possible to change and to regulate physical parameters of vowels (frequency and intensity of formants, fundamental vocal frequency). The average formants F_1 , F_2 , F_3 were determined on the basis of the pronunciations of one hundred and fifty Bulgarian speakers of different age and sex. The frequencies of the first three major vocal resonances effectively conditioned the resonances of the transfer function for the electrical network. The band-width of formants filters was computed using the method described by Flanagan (JASA 592-596, 1956). The possibility of using active filters as formants' filters was put to investigation. The active filters provide a simple means of the changing physical formants' parameters of synthetic vowels.

The study of synthetic vowels' perception was undertaken in two directions: estimating the changes of physical parameters of synthetic vowels and determining the effect of these changes on perception of synthetic Bulgarian vowels.

The results of the perception experiment are presented.

SPECTRAL ANALYSIS OF SOME ACOUSTICAL CHARACTERISTICS OF VOWELS AND
THEIR IMPORTANCE FOR SPEECH QUALITY

Maria Popova (Bulgaria)

The study is devoted to the problem of determining the relationship between personal speech features and some acoustical characteristics, as well as the variation of these characteristics during speech transmission through a communicative channel. The study is made on the basis of spectral analysis. The Key Electric Co. SONAGRAPH was used. The speakers are males with different voice pitches. The tested linguistic material consists of monosyllables. The speech analysis is carried out in a frequency band from 85 to 8000 Hz. The results have been based on two kinds of experiments: 1) Spectral analysis of voices with modified timbres. Two cases have been considered: a) modification of the timbre during speech transmission through a communicative channel, and b) modification of the timbre in acoustical pathology. The comparative analysis of the acoustical characteristics of a voice with an original and modified timbre presents data about their relationship with the timbre features. 2) Spectral analysis of voices belonging to speakers of different ages: 11 young males of an average age of 18 and 11 elderly males (average age 65).

The results show: 1) The modification in the timbre results in a change of the ratio between the high formant level and low formant level. 2) The speakers age exerts influence on the formant picture mainly on the second formant and on the interval between the first and the second formant.

The obtained results show that the personal voice features reflected in the formant acoustical picture are changing regularly and unidirectionally under the influence of the age and the sonority of the timbre.

THE ROLE OF THE VOWEL FORMANT TRANSITIONS AND
NOISE CHARACTERISTICS OF THE BULGARIAN CONSONANTS
IN THE PROCESS OF RECOGNITION

D.Tilkov, A.Misheva /Bulgaria/

Basic to automatical speech recognition is the determination of the acoustical cues of the speech sounds utilized in the process of perception. The purpose of this study is to establish the role of the proper characteristics of the consonants and the formant transitions of the neighbouring vowel in the auditory identification of the consonants. With that end in view several series of auditory tests have been made. The test material has been prepared on the basis of "consonant-vowel" syllables recorded on a tape. Each recorded syllable is divided into three parts, corresponding to the proper noise characteristics of the consonant, vowel formant transition and the stationary part of the vowel, with an accuracy of 2 msec. To determine the role of the first two segments separately as well as which one of them is of primary importance in each of the cases, the following auditory tests were made: a/identification of the consonants according to their articulatory place with removed noise segment; b/identification of the consonants with removed vowel formant transition; c/identification of the consonants with transplantation of the noise segment on the transition, corresponding to another consonant; d/identification of the consonants with transplantation of the noise segment preceding a low and a high vowel.

Analysis of the results indicate in the first place that both characteristics taken separately are capable of ensuring consonant identification with a probability higher than their random recognition. The removal of the formant transition makes the perception of the palatal consonant difficult. Under the simultaneous action of conflicting acoustical characteristics the perception of the consonant is determined by the kind of the given consonant and its neighbouring vowel. Not only a confusion of the two consonants, whose acoustical characteristics are presented, is observed, but also an auditive impression is created of a third consonant different from them.

HYPHENATION PROGRAMMES AND THEIR APPLICATION
IN AUTOMATED TYPESETTING SYSTEMS

Konstantin BOSSILKOV, Kiril NAIDENOV, Marin KOTAROV (BULGARIA)

The introduction of computers in the printing industry and the digital automation of typesetting processes call for the development of hyphenation programmes. The solution of this applied problem is brought down to the searching for an optimum between three criteria- linguistic, graphic and economic. With a view to achieving a solid theoretical basis necessary for the development of a specific programme, these criteria are analyzed in detail with respect to their complicated interconnection. Different possible ways are considered of building up an algorithm for hyphenation based on the three principles of Bulgarian spelling- formal, phonetic and morphologic.

The results from statistical investigations on Bulgarian language are given which illustrate the relative importance of the last two principles in word hyphenation.

The possibility is studied of building-up a series of hierarchically arranged programme modules. An attempt is made at a statistical evaluation of the contribution of each of them in achieving a hyphenation accuracy, maximum graphic syllables, minimum storage and suitable processing speed.

An attempt is made at quantitative and qualitative evaluation of the optimality of hyphenation programmes in view of typesetting practice.

specifically worked out variant of a programme already introduced in the practice is announced and practical results are analyzed.

ZUR AUTOMATISCHEN SILBENTRENNUNG

J. Fischer (DDR)

Es werden Grundvoraussetzungen für die automatische Silbentrennung angegeben. Diese zielen im wesentlichen darauf ab, welche Zeichenfolge in Texten überhaupt für die Silbentrennung zugelassen werden, und in welchem Bereich eines Wortes getrennt werden darf.

Hinzu kommen Forderungen der polygrafischen Industrie an Silbentrennprogramme. Das betrifft einmal rein technische Dinge wie den Speicherplatz und die Rechenzeit des vorhandenen Computers. Es betrifft jedoch auch die Zahl der gefundenen Trennstellen und die der fehlerhaften Trennungen, damit verbunden die Korrekturmöglichkeiten aufgrund der vorhandenen Technik.

Weiter werden Möglichkeiten für Verfahren zur automatischen Silbentrennung aufgezeigt. Es wird dabei von der Wörterbuchmethode ausgegangen und auf verschiedene Formen von logischen Verfahren wie Konsonanten-, Präfix- und Suffixmethode verwiesen. Anhand dieser Verfahren wird die Zweckmäßigkeit der einzelnen Varianten für verschiedene Sprachen untersucht.

Am Beispiel eines Silbentrennprogramms für die deutsche Sprache werden weitere Probleme aufgezeigt, die vor allem bedingt sind durch die vielen Zusammensetzungen im Deutschen. Bei großer Satzbreite haben die Trennstellen an Kompositionsgrenzen Präferenz, während das bei geringerer Satzbreite nur bedingt der Fall ist.

DAS ELEKTRONISCHE WÖRTERBUCH DER FACHSPRACHEN - STRUKTUR UND NUTZUNGSMÖGLICHKEITEN

G. Neubert (DDR)

Die fachsprachliche Lexikographie hat zwei neue Aufgaben: Erhöhung des Aktualitätsgrades und vielseitige Nutzbarkeit. Die elektronische Datenverarbeitung als eines der Mittel, die gewachsenen Aufgaben der rationellen Speicherung, Sortierung und Verarbeitung von Sprachdaten zu erfüllen, bedingt eine enge Verknüpfung der inhaltlichen und formalen Darstellung im Wortschatzspeicher mit der Methodik der Wortschatzauswertung. Aus den Nutzungsforderungen an das Elektronische Wörterbuch der Fachsprachen Russisch/Englisch/Deutsch (EWF), das am Institut für Angewandte Sprachwissenschaft der Technischen Universität Dresden in Kooperation mit dem Universitätsrechenzentrum entwickelt worden ist, leiten sich Bedingungen für seine lexikographische Struktur ab.

Um auf jeder Entwicklungsstufe praktische Anwendbarkeit zu gewährleisten, wurde das EWF als ein "Baukasten" von selbständig nutzbaren und kompatiblen Teilspeichern und -programmen zusammengestellt. In der ersten Ausbaustufe werden zwei Grundtypen von Abfrageprogrammen verwendet: Programme zur Ausgabe textgebundener Wortzuordnungslisten und Programme für Listen von Termini, die nach Merkmalen des Begriffs oder der Benennung klassifiziert sind. Weiterhin existieren Einspeicherungs- und Laufendhaltungsprogramme, die eine Rechnerunterstützung der lexikographischen Wortschatzerarbeitung ermöglichen.

Es werden Aufgaben der fachsprachlichen Lexikologie, der fachsprachlichen Aus- und Weiterbildung, der Informations- und Dokumentationswissenschaft und -technik, der Terminologiestandardisierung, der Sprachmittlung usw. bearbeitet.

LE DICTIONNAIRE AUTOMATIQUE TRILINGUE
DE "THESEE" THESAURUS COMPATIBLE AVEC
LES SYSTEMES D'INDEXATION
FRANÇAIS-ANGLAIS-ALLEMAND

A. Deweze (France)

Le thésaurus dont on examine ici les possibilités d'application à la traduction automatique et, en particulier, son dictionnaire trilingue, est le thésaurus **THÉSÉE** (Thésaurus pour l'Electricité et l'Electronique).

La réalisation de Thésée a été financée par la DGRST (Délégation Générale à la Recherche Scientifique et Technique) et le BNIST (Bureau National d'Information Scientifique et Technique) pour le réseau de documentation sectoriel ELDOC. Il permet l'indexation et la diffusion sélective de documents en français, en anglais ou en allemand, donc également d'interroger des bandes françaises, anglo-saxonnes (NTIS, INSPEC) et, éventuellement, allemandes (ZDE). On attire ici l'attention sur les problèmes sémantiques rencontrés dans des systèmes d'indexation multilingues. On donne la description organique du dictionnaire trilingue, qui peut être utilisé pour enregistrer les variantes morphologiques des langues à flexion. La première réalisation de traduction automatique en vue est celle de résumés normalisés.

LE PROBLEME DU FICHER MULTILINGUE DANS
L'IMPLANTATION DE CANTERM

G. Rondeau (Canada)

L'auteur est conseiller auprès du Secrétariat d'Etat canadien en matière d'automatisation des services terminologiques. La mise en place d'un organisme central de dépistage, collecte, classement et diffusion des données terminologiques et documentaires (CANTERM) a été amorcée en février 1975; cet organisme devra desservir, en plus des 800 traducteurs du gouvernement, tous les fonctionnaires qui ont pour tâche l'application de la loi sur les langues officielles du Canada (1969).

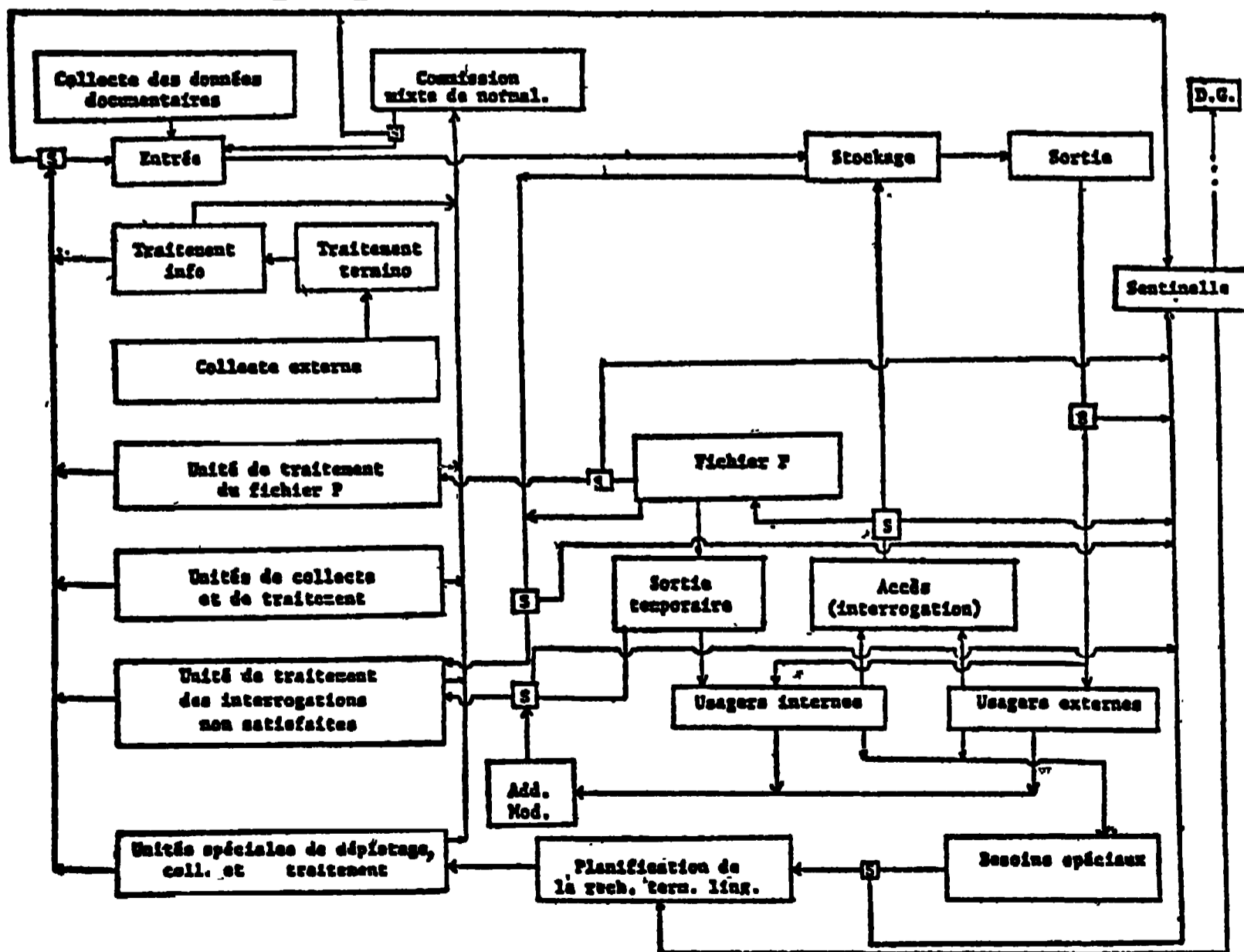
Le cadre. On situera d'abord le problème dans le cadre de CANTERM à l'aide du tableau ci-joint.

Le problème. Il y a deux langues officielles au Canada, le français et l'anglais. Mais les besoins en terminologie dépassent les limites de ces deux langues en ce sens que de nombreux documents rédigés dans d'autres langues doivent être traduits dans l'une ou l'autre (ou les deux) des langues officielles. Il s'ensuit des problèmes qui ont trait à la collecte, au classement et à la diffusion des données terminologiques. Plus précisément, la fiche de CANTERM doit pouvoir s'accomoder des situations suivantes quant à la collecte, au classement et à la diffusion: a) des données terminologiques sont disponibles à l'origine seulement dans les deux langues officielles; mais, en cours de route, des équivalents en une ou plusieurs des langues non-officielles sont fournis à CANTERM par les usagers; b) des données terminologiques sont disponibles à l'origine dans une seule langue non-officielle ou dans un couple de langues non-officieles; c) des données terminologiques sont disponibles à l'origine en couples de langues dont l'une seulement est une langue officielle du Canada; mais, en cours de route, des équivalents dans l'autre langue officielle sont fournis à CANTERM par les usagers. Dans tous les cas, le problème consiste à créer un système dynamique qui permette d'absorber les données terminologiques d'origine ainsi que leurs développements ultérieurs d'une part, et d'autre part, un système de dépistage de l'information qui permette de faire les raccords nécessaires entre les familles sémantiques de fiches sans encombrement de la mémoire. Notons que le problème

comporte un double aspect: a) terminologique, en ce sens qu'une unité néologique créée dans une langue A risque d'être étudiée dans une langue C mais par rapport à une langue naturelle intermédiaire B plutôt que par rapport à la langue A; ce n'est pas ce point de vue qui sera abordé ici; b) computationnel, en ce sens, qu'il faut mettre au point une technique de stockage et de repérage des unités terminologiques qui permette non seulement d'enmagasiner et de repérer des données disponibles immédiatement, mais encore de stocker des données nouvelles.

A priori, deux hypothèses de travail s'établissent: a) centraliser l'ensemble du fichier multilingue sur un fichier fondamental anglais-français; b) établir un fichier multilingue du type de celui de la CEE, avec des cases vides. On proposera, pour discussion, une solution à ce problème.

Organigramme des fonctions du CANTERM



LA CATEGORIE DE DIMINUTIVITE DU POINT DE VUE DE LA
LINGUISTIQUE QUANTITATIVE

Borimir Kristev (Bulgarie)

Sur la base des données de la fréquence des lexèmes dans le bulgare contemporain on a prouvé que la catégorie de diminutivité a la tendance de s'exprimer par des complexes de morphèmes et que la fréquence est régulièrement liée à la composition des morphèmes de la diminutivité.

Plus souvent se fait l'emploi d'un diminutif, plus simple est sa composition de morphèmes.

Tel est le cas, par exemple, de la diminutivité par le suffixe "-e".

Mais à cette tendance normale - moins de morphèmes, plus haute fréquence - s'oppose une autre tendance: l'émotivité du diminutif exige une morphématique plus compliquée.

Tel en est le cas avec le suffixe "-ичка" ou s'accumulent deux suffixes "-иц-" et "-к(а)".

STATUS OF NUMBER NAMES IN THE WORD COUNTS

Boyan Nikolayev (Bulgaria)

In the existing word counts number names are especially treated. One could get the impression that their quantity is insignificant. This is due mainly to the usual symbolic representation of the numerals.

In this paper a survey of Bulgarian data shows that the frequency of number names is higher than its usual estimation. All means of expressing definite quantity are taken into account (i.e. "substantive numerals", as defined by Bulgarian grammar, are also being considered).

CLUB SYSTEMS

V.C. Borscev et M.V. Chomjakov (USSR)

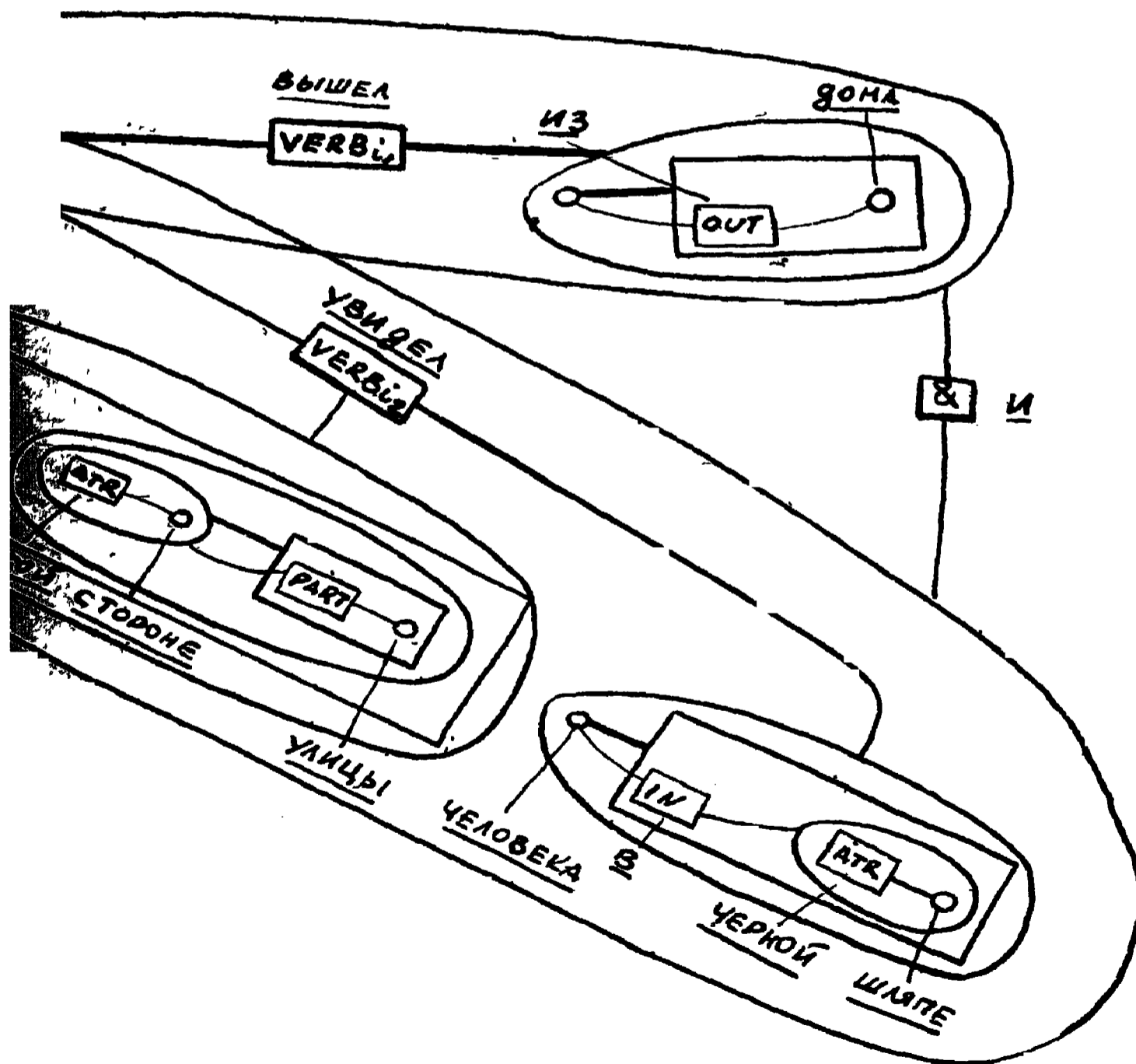
Several years ago the authors felt that the notion of a model over some signature was the best way to describe objects with a complicated structure (in particular, natural language texts). It seemed that many others shared this opinion (perhaps subconsciously), as witnessed by the widespread interest in semantic graphs, logical nets, semantic nets of various kinds, etc. Nevertheless, the application of this apparatus (despite its universality) encounters some difficulties, especially in cases when a certain part of the structure in question (which may be very complicated itself) plays an independent role in the whole structure.

For example, a dependency grammar sometimes looks clumsy just because it rejects the idea of constituents. It seems to us (and others as well) that an appropriate symbiosis of the two grammars involved (i.e. dependency relations between constituents containing other constituents linked by dependency relations) will fit syntactical structures of natural language better than each of these grammars separately.

Another example. Any man within a society is a member of many 'clubs': he has a family, works somewhere, is a member of a trade union, keeps money in a bank, plays tennis in a tennis club, etc. Every club is not only the set of its members. It is also the roles its members play and the relations between them that create the club. Clubs, in turn, establish relations of different kinds with each other and create clubs of a higher level (for example, corporations, banks, trade unions linked together form the economic structure of society).

We propose the notion of club system to describe such phenomena. A club system consists of clubs and arrows. An n -ary arrow ($n = 1, 2, \dots$) is an n -tuple of clubs linked by

relations. There are terminal and non-terminal
 terminal club is a set of arrows. A terminal
 inside it, i.e. in this club system we are
 its internal structure. Thus, the club
 generalisation of the notion of model (where all
 model's elements -- are terminal). In this
 formal properties of club systems and their
 linguistics. The figure below shows a tenta-
 the Russian sentence "Петя вышел из дома
 стороне улицы увидел человека в черной шляпе"



THE PROOF THAT LANGUAGE IS NOT
CONTEXT-FREE

H. Brandt Corstius (Holland)

Chomsky (1) /and Postal (2)/ have argued that since English (and Mohawk) contain certain sets of sentences that can not be described by regular (context-free) grammars, English (or Mohawk) cannot be regular (context-free). These proofs are incorrect, but a correct proof can be given using well-known intersection properties of regular and context-free languages (3).

(1) N. Chomsky, Syntactic Structures, The Hague, 1957.

(2) P. Postal, 'Limitations of phrases structure grammars' in J.A. Fodor and J.J. Katz, The structure of language, Englewood Cliffs, 1964.

(3). H. Brandt Corstius, Algebraische Taalkunde, Utrecht, 1974.

ON COMPARATIVE STUDY OF VARIOUS TYPES OF SYNTACTIC
CONFIGURATIONS IN ALGEBRAIC LINGUISTICS

M. Semeniuk-Polkowska (Poland)

The notion of syntactic configuration is one of the fundamental notions of algebraic linguistics. It was introduced by O.S.Kulagina. Somewhat different ones were suggested by A.V.Gladkiĭ and M.Novotny.

The aim of this note is to present comparative study of various types of syntactic configurations. Two new types of configurations connected with Gladkiĭ's and Novotny's configurations are introduced and the problem of their relation with the already existing types of configurations is solved. Starting from the remark that all existing types of configurations are related to the relation of domination in the sense of Dobrushin new types of configurations obtained by replacing Dobrushin's domination by Trybulec's domination and Franclova's domination respectively are introduced. The relationships between these new types of configurations and between them and the old types of configurations are established.

LES SYSTÈMES A.T.E.F. et C.E.T.A

J. Chauché (C.E.T.A., France)

Le traitement sur ordinateur des langues naturelles nécessite des modèles algorithmiques plus ou moins perfectionnés. Les deux systèmes présentés ici correspondent à un choix d'une classe importante des algorithmes envisagés ces dernières années pour résoudre ces problèmes. Le principal choix effectué par ces systèmes réside dans l'utilisation formelle d'arborences étiquetées. La liberté du choix des étiquettes ainsi que des structures possibles donne à ces systèmes de vastes champs d'application dans beaucoup de domaines et notamment dans celui du traitement automatique des langues naturelles. Le système A.T.E.F. a pour but de transformer une chaîne de mots en une arborescence qui elle est manipulable par le système C.E.T.A. La définition des arborescences étiquetées définit les objets manipulables par le système C.E.T.A. et les objectifs du système A.T.E.F.. Cet exposé commence donc par la définition des arborescences étiquetées. Pour obtenir une arborescence de ce type à partir d'une chaîne d'entrée, le système A.T.E.F. utilise un dictionnaire et une grammaire d'états finis. Le résultat de ce système peut être manipulé par le système C.E.T.A. afin d'obtenir le type de structure désiré. L'exemple d'analyse donné montre les possibilités du système C.E.T.A. avec deux stratégies différentes de manipulation (l'une recherche une structure de constituants, l'autre une structure de dépendances).

DEFINITORISCHE BEGRÜNDUNG DER ABHÄNGIGKEITSGRAMMATIK,
JÜRGEN KUNZE (DDR)

Der Vortrag behandelt eine Möglichkeit zur formalen und inhaltlichen Präzisierung der verschiedenen Komponenten der Abhängigkeitsgrammatik. Die gesamte Definitionsprozedur zerfällt in folgende Hauptschritte:

1. Kriterien zur Bestimmung reiner Abhängigkeitsstrukturen

Hierbei handelt es sich um das Problem, die Zuordnung zwischen Sätzen und unmarkierten Baumgraphen zu objektivieren. Die Einzelheiten werden im Vortrag von J. Kunze und A. Ljudskanov "Unterordnungskriterien für die Abhängigkeitsgrammatik und einige Beispiele aus dem Bulgarischen" dargestellt. Nach Abschluß dieser Definitionsstufe ist jedem Satz φ der Sprache L ein unmarkierter Baumgraph β zugeordnet (im Falle gewisser Mehrdeutigkeiten von φ auch mehrere β), wobei die Knoten von β den Wortformen von φ entsprechen.

2. Definition der paradigmatischen Komponente

Der Erfassung der Paradigmatik dienen die paradigmatischen Kategorien (für das Deutsche 10) sowie die (stets zweistelligen) paradigmatischen Beziehungen. Letztere beschreiben Erscheinungen, die als Kongruenzen und Rektionen betrachtet werden können. Jede paradigmatische Beziehung ist definiert durch die betroffenen Kategorien und eine Klasse von Wirkungswegen (in Abhängigkeitsbäumen β), die die Ausgangs- und Endpunkte der Beziehung miteinander verbinden.

3. Definition der selektiven Komponente

Analog zu den paradigmatischen werden selektive Beziehungen eingeführt, die mit Merkmalen anstelle der Kategorienwerte operieren. Sie dienen so zur Erfassung sowohl semantischer als auch grammatischer Beschränkungen. Von den selektiven Beziehungen läßt sich eine Brücke zu den semantischen Netzen schlagen.

4. Markierung der Abhängigkeitsbäume

Die Kanten der aus 1. stammenden Bäume / β / werden aufgrund einer Markierungsvorschrift mit Unterordnungsrelationen markiert. Diese Markierungsvorschrift muß einem (als erfüllbar nachweisbaren) Differenzierungsprinzip genügen, das mit dem Verlauf von Wirkungswegen und der Substituierbarkeit von Teilbäumen zusammenhängt.

5. Definition der Merkmalkombinationen

Die Eigenschaften von Wortformen werden durch fünfkomponentige Merkmalkombinationen erfaßt, wobei sich zwei von ihnen naturgemäß aus 2. und 3. ergeben, eine weitere für Anordnungs-eigenschaften zuständig ist und die beiden restlichen das eigentlich syntaktische Verhalten beschreiben.

6. Definition der Büschel

Aus 2. bis 5. ergibt sich die Möglichkeit, einen allgemeinen Regelbegriff für Abhängigkeitsgrammatiken einzuführen. Er basiert auf der expliziten Definition der Merkmalkombinationen und dem Verlauf der Wirkungswege.

UNTERORDNUNGSKRITERIEN FÜR DIE ABHÄNGIGKEITSGRAMMATIK UND
EINIGE BEISPIELE AUS DEM BULGARISCHEN.

JÜRGEN KUNZE (DDR)

ALEXANDER LJUDSKANOV (VR BULGARIEN)

Eine Teilaufgabe bei der Aufstellung einer Abhängigkeitsgrammatik für eine bestimmte Sprache ist die Festlegung der Unterordnungen, d.h. die Beantwortung der Frage, wie (und "warum") die Wortformen des Satzes einander untergeordnet werden. Die bisherigen intuitiven oder durch listenmäßige Aufzählung erfaßten Zuordnungen lassen sich durch vier Unterordnungskriterien ersetzen, mit denen auch ein besseres Verständnis der Unterordnungen erreicht wird. Diese Kriterien setzen nur wenige formale Begriffe voraus. Sie basieren auf der syntaktischen Substituierbarkeit und Weglaßbarkeit (im weiteren Sinne) und ergeben bei geeigneter kombinierter Anwendung für syntaktisch eindeutige Sätze φ genau einen Abhängigkeitsbaum β . Die benötigten nicht-formalen Begriffe sind als a-priori-Begriffe ohnehin heuristische Grundlage jeder Grammatiktheorie.

Die Unterordnungskriterien sind keine algorithmischen Regeln, sie haben nur einen definitorischen Status. Ihre Anwendung kann jedoch streng formal geschehen. Die Kriterien werden abschließend auf einige bulgarische Beispielsätze angewendet.

SYNTAKTISCHE MEHRWEGANALYSE AUF DER GRUNDLAGE VON BÜSCHELN,
DOROTHEE AROLD (DDR)

Aufgabe der syntaktischen Analyse ist es, ausgehend von einer funktional-grammatischen Beschreibung der Wortformen eines Satzes, die Beschreibung seiner grammatischen Struktur (im Rahmen eines vorgegebenen Modells) herzustellen.

Wir stellen die Struktur von Sätzen in Form markierter Abhängigkeitsbäume dar. Als Analyseregeln verwenden wir sogenannte Büschel, die je einen Spitzenknoten und alle ihm in einem Abhängigkeitsbaum direkt untergeordneten Knoten sowie die als Markierungen verwendeten Unterordnungsrelationen enthalten.

Da es im allgemeinen zu jedem zu analysierenden Satz mehrere markierte Abhängigkeitsbäume geben wird und außerdem im Laufe des Analyseprozesses Fehler auftreten können, die nur sehr schwer zu korrigieren sind, sehen wir eine Mehrweganalyse vor. Diese beruht auf dem folgenden Prinzip: In jedem Analyseschritt werden stets alle Büschel zur Anwendung in Betracht gezogen, deren Anwendung in diesem Schritt möglich ist (d.h. alle Bedingungen, die das Büschel stellt, sind erfüllt). Es ist nun wegen des großen Aufwandes eines solchen Vorgehens erforderlich, Kriterien zu finden, die eine Optimierung des Analyseprozesses bewirken. Grundlage dafür ist ein während des Analyseprozesses aufzubauender Analysegraph, der aus mit Baumfragmenten (Zwischenschritten im Analyseprozeß) interpretierten Knoten und Kanten, die für jeweils eine Büschelanwendung stehen, besteht. Am Analysegraphen sind drei Erscheinungen (Maschen, Parallelanalysen und Sackgassen) zu erkennen, die eine Optimierung zulassen. Maschen und Parallelanalysen sind auf Grund der Verwendung von Büschel als Analyseregeln leicht zu vermeiden, zur Vermeidung von Sackgassen können wir für einige häufige Fälle Kriterien angeben, u.a. dafür, wann "zu kleine" Büschel nicht angewendet werden brauchen und wann eine Büschelanwendung stets zu einem richtigen Ergebnis führt, sie also vor anderen möglichen präferiert werden kann.

LA DECOMPOSITION AUTOMATIQUE DE LA PHRASE
COMPOSEE BULGARE EN PROPOSITIONS SIMPLES

Edouard Sélian /Bulgarie/

On fonde la nécessité de décomposer la phrase composée en propositions simples lors du processus d'élaboration d'un système de recherche automatique d'informations. La décomposition automatique est envisagée du point de vue de la méthode non-statistique d'indexation et en tant qu'une étape du processus de la création d'un langage d'information avec grammaire. En connexion avec cela on décrit certaines possibilités de simulation, menant à la création d'une méthode d'indexation automatique à deux niveaux appliquant tant les critères non-statistiques que les critères statistiques.

A la base d'un matériel en langue bulgare on a fait une expérience dans le but d'illustrer la possibilité de simulation de la partie non-statistique de l'activité sur l'indexation à deux niveaux. On a défini la phrase à la base des limites supérieure, intérieure et extérieure. On donne certaines assertions constituant la base de l'algorithme en question.

L'expérience a une importance pratique lors de l'élaboration automatique d'information - texte, tandis que la méthodologie employée peut être utile du point de vue de la formalisation ultérieure et de l'algorithmisation de la syntaxe de la langue bulgare.

AUTOMATIC ANALYSIS OF BULGARIAN SCIENTIFIC TEXTS

Hristo C Georgiev (Bulgaria)

The present report is devoted to the possibility to formalize one small aspect of the over-all analysis of a written text, that is - the dissection of a text on verbal (V) and nominal (N) word-groups. That kind of analysis is aimed to aid machine translation and the automatic retrieval of information.

Undertaking such a formalization, I start from the assumption that Bulgarian parts of speech (verbs, nouns etc.) bear enough syntactical and morphological markers which allow to make them recognized in a purely formal way. The dissection of the text on V and N word-groups is based on the idea of the border marks between these two groups.

In this study, by a N group I understand the combination of a noun and adjective, pronoun, numeral or a syntactic word(s) - as the latter is in the role of a left or right border marker of the N group. To the N word-group are also attached the adverbs - relatively only - despite that they define the verb. The term V group is used to designate the verb alone, or the combination of the verb with the particles attached to it, with the conjunctions "ze" and "ga", or with the reflexive pronoun "ce", all in the role of left or right border markers of the V group. In the present study the verbal adjunct is not regarded as belonging to the V group.

The algorithm consists of 90 digital operations and 12 tables of about 200 markers. The formalization achieved is 98 per cent correct.

A TECHNIQUE FOR AUTOMATIC SEGMENTATION OF POLISH VERBS

Wiesław Lubaszewski (Poland)

A guiding principle of the program is to divide each verb into its constituents: (1) pre-root and post-root affixes; (2) root in its synchronic sense. The program is based on context-free "segmentation grammar" which consists of: "words" - Polish verbs in their orthographic form without any additional description;

"segments" - pre-root and post-root verbal affixes. In the program the segments are stored not in the "dictionary" (table) but each segment is represented by one rule; that is to say that each segment is transposed into the appropriate rule (the root segment is equal to "remainder" after splitting off affixes);

"context-free segmentation, restriction, and correction rules" - in the program each rule is represented by one procedure (i.e. a set of FORTRAN instructions), and because of that representation the grammar consulting routine is not a dictionary-lookup routine but the control program which initializes the appropriate procedures.

(1) The function of the segmentation rule is to identify (even if not in keeping with linguistic facts) the single segment and to split it, if the identification was successful.

(2) To eliminate erroneous splitting the program uses restriction rules; each restriction rule contains the erroneous root form and compares it with actual root form. If both of these are equal the segmentation was faulty.

(3) Whether the segmentation is faulty the program uses the appropriate correction rule which corrects the erroneous split.

SOME PROBLEMS OF MORPHOLOGICAL ANALYSIS OF POLISH LANGUAGE

W.Łukaszewicz (Poland)

One of the part of MARYSIA system, which is under design at University of Warsaw, namely the morphological analysis component will be presented here.

MARYSIA is general-purpose conversation system. A typical application of MARYSIA system is to use it as a part of a larger system (CAI System, Information Retrieval System, etc.), where it acts as a "buffer" between a user of that system and that system itself.

The morphological component of MARYSIA system will be described in detail..

This module receives natural language utterances as an input, segments them into smaller units and by means of dictionary entries provides appropriate information for further components of system.

The most important part of morphological analysis. i e. matching the units of input utterance with appropriate dictionary entries, is done by means of automata of Mealy.

The algorithm was designed mainly for Polish language but its flexible structure makes possible to use it for other natural languages.

FLEXIVISCHE UND WORTBILDUNGSANALYSE DES DEUTSCHEN,
BARBARA RÜDIGER (DDR)

Die flexivische Analyse bildet eine Vorstufe der syntaktischen Analyse. Alle für die syntaktische Analyse wesentlichen Eigenschaften von Wortformen werden durch Merkmalkombinationen erfaßt, die aus fünf Komponenten bestehen, deren eine die paradigmatische Komponente ist. Sie beschreibt die flexivischen Merkmale einer Wortform durch die zu ihr gehörenden Kategorien Genus, Numerus, Person, Kasus, Tempus, Modus und Deklinationsart.

Die Werte der für eine Wortform relevanten Kategorien werden durch die flexivische Analyse algorithmisch ermittelt.

Die algorithmische Darstellung der flexivischen Analyse ist mit einer Wortbildungsanalyse verbunden, die ihr vorangeht. Ihre Aufgabe besteht darin, zusammengesetzte und abgeleitete Wortformen in ihre Bestandteile zu zerlegen. Bei Ableitungen wird darüber hinaus angegeben, in welcher Weise die Wortbedeutung durch Affixe verändert wird. Dabei wird das Verhältnis zwischen Affix und Kern (das ist das, was nach Abtrennen eines Affixes übrig bleibt) durch Paraphrasierungen dargestellt, z.B. tragbar (trag(en) = Kern, bar = Affix) durch kann getragen werden. Für eine solche Darstellung sind nur bestimmte Affixe der deutschen Sprache geeignet.

Die ermittelten Ableitungsmuster (= Affixe mit ihren Paraphrasierungen) werden zu Ableitungstypen zusammengefaßt. Ein Typ ist durch die gleiche Wortklasse, von der die Ableitung gebildet wird, und die gleiche Art der Paraphrasierung gekennzeichnet.

Kann das Verhältnis zwischen Kern und Ableitung durch Paraphrasierungen beschrieben werden, die in Texten auftreten können, wird das zur Herstellung von Äquivalenzbeziehungen zwischen syntaktischen Fügungen ausgenutzt. Zur Feststellung dieser Beziehungen wird die Ableitung in die für sie typischen Kontexte eingebettet; dann wird untersucht, ob und wie diese Kontexte verändert werden müssen, wenn die Ableitung paraphrasiert wird.

COGNITIVE NETWORKS

David G. Hays (USA)

The classic foundations of western philosophy are goodness, truth, beauty, and consistency. The origins of two additional concepts are not so easy to specify: pragmatic effectiveness and social interaction-distribution are still less familiar and perhaps less completely understood. The study of language can be formulated as the pursuit of solutions for two kinds of problems that arise with respect to each fundamental concept:

How is it that language permits the formulation and application of the concept?

How is it that the concept applies to language?

Thus, we ask how language permits us to formulate the idea of goodness and consider whether different things are good; but we also ask whether a language is good as a language, and whether a speech act is good. Linguistics does not have to decide whether a language is good, consistent, etc.; but it must say what language is that such decisions can be made.

A theory of language that has to do only with the relationships among elements of language cannot express the answers to such questions. Neither can a theory that has to do further with the relationships between elements of language and elements of reality. Goodness, consistency, and the rest are not categories of reality, but of thought; reality is not consistent, it is merely real. A theory of language in which answers to the questions I pose can be expressed is a theory of relationships among elements of language and of relationships between language and a system of thought.

To create such a theory, one must decide how the elements of thought and language differ, and what operations belong to systems of thought and systems of language. The boundary can be fixed in one place or another, resulting in

ories of thought and of language.

Technical apparatus being created under the rubric
networks' supports the development of theories
require.

of robots leads to analyses of problems of prag-
matics in language and thought.

of poetics leads to analyses of problems of

of philosophy of science leads to analyses of
truth.

of philosophy of human action leads to analyses
of goodness.

Active networks are representations of knowledge
elements are nodes and links; with some of the
associated perceptual or effective processes
either on the external environment or on the
self. It appears that powerful systems can be
built from simple processes. The processes are familiar
transformational or stratificational grammar-
workers in artificial intelligence. Computer pro-
cesses required type have been demonstrated.

LINGUISTICS AND ARTIFICIAL INTELLIGENCE

Petr Sgall (Prague)

1. The future of linguistics is not connected with artificial intelligence to such an extent as to envisage a "new paradigm" of linguistics directly is studied concerned with man-machine communication, but it is of highest importance for the both domains to cooperate as closely as possible. The main differences between the two fields are connected with the theoretical character of linguistics (describing linguistic competence, its evolution and its relationships to human thought and all kinds of mental activity, as against the technical applications most of which must model, at least to some degree, also linguistic performance (combining linguistic analysis in proper sense with a handling of cognitive and ontological items); thus, a generative grammar aims at a theoretical description of every detail of the intricate structure of natural language (excluding non-linguistic issues), while linguistically oriented artificial intelligence (Winograd, S.Klein, Wilks etc.) wants to trace shortcuts given by non-linguistic environment, which are necessary for disambiguation. Machine translation (cf. the approaches to syntactic analysis by Vanquois, Mel'chuk, Kittredge, Woods, Petrick and others) stands in between, since it cannot use much ontological information, but has to simplify the theoretical description, if it wants to be effective.

2. Some of the approaches to theoretical linguistics seem to offer more convenient points of departure for technical applications than others: TG seems to be in a certain disadvantage, from this point of view, if compared with those types of description which work with a linear ordering of levels (stratificational, "meaning-text", functional, and also generative semantics), since there the relationship between semantic and phonetic representations is accounted for by a coherent sequence of operations. Another such distinction may concern the two ways of treating the relationship

between text and grammar: if TG works with ordered transformations, the repeated passing through the whole sentence (at every point of the list of rules) might make the analysis procedure unnecessarily complex.

3. Other open question that will be discussed concern the notion of synonymy (broader with Mel'chuk and other Soviet linguists, narrower e.g. in the functional description), the extent to which the typological character of a language is decisive for the choice of an appropriate strategy of syntactic analysis, and the main interlingual differences in semantic structure, which concern first of all the points of contact between semantics proper and (point of) reference: delimiting features of NP's, aspects of verb.

4. The main task may be seen in bringing together consciously two major trends: (a) the smaller or larger experience in main-machine communication oriented to specific areas and (b) the well-known but not yet mastered process of structural modification that makes programming languages more and more similar to (substructures of) natural languages, which are the only fully appropriate human means of communication (be it with other human beings, animals, or machines).

SOME QUESTIONS OF FORMAL MODELS IN LINGUISTIC SEMANTICS

E. Hajičová (Prague)

1. Most of the (recently so numerous) linguistic approaches to the relationship of grammar and semantics propose some kind of modified predicate calculus as the proper treatment of semantics of natural languages more or less neglecting that for decades one of the main aims of modern logic has been to elaborate this or that kind of such modifications. The difficulties underlying such approaches of linguists involve the difference between the way the semantic relationships are rendered by the structure of natural languages and the achievements and conventions of logic (e.g. the nature of participants of verbs, delimiting features, topic (comment articulation of sentence, connected with quantification, etc.)
2. A programme for semantically oriented linguistic research, formulated some twelve years ago by Sgall and carried out step by step by his Prague group of algebraic linguistics, which was then regarded by many linguists as unrealistic, may be now compared with other approaches to "logically oriented grammar", "logical structure of sentences", "natural logic", etc. most of which have emerged in the late sixties. It comprises a formulation of a metalanguage for semantic representations (SR's), which is intended, first of all, as a means of a theoretical description of semantics of natural language. These SR's do not constitute a logical language and the goal is understood as an algorithmic procedure translating SR's into a universal language of logic rather than a modification of some logical language; an elementary procedure was formulated that uses a generalized version of epsilon operator of Hilbert and Bernays.

ON THE QUESTION OF THE DESCRIPTION OF POSSESSIVITY
IN THE AUTOMATIC PROCESSING OF TEXTS

Petr Piřha (Prague)

0. This work is a part of theoretical research having two main aims - a) an explicit description of Czech, b) background for different types of mechanical processing of texts. Its linguistic and methodological background is the functional generative description of language as proposed and developed by P. Sgall's Prague group.

1. There are two ways, how possessive relation is expressed: a) by special grammatical units which, of course, can serve other functions as well; b) by other grammatical or lexical units in sentences whose meaning is interpreted as "possessive" only according to certain contexts and situations. The second type is connected with cognitive content and extralinguistic knowledge. However, the boundary between both types is not sharp and may vary in different languages, as e.g. *Dativus possessivus*.

2. Possessive relation can be described as a special unit of language meaning. The best way how to do it is to describe homonymies, synonymies and polysemies of grammatical units syntactical or morphological ; that is by comparing morphological and syntactical representations of given sentences on the semantic level. In this I use the strategy "from form to function" which enables us to classify the notion of possessivity on the semantic level as accurately as it is reflected in the language system.

3. A survey of grammatical tools expressing possessivity in Czech is presented and their basic semantic relations are described. (The main discussion is devoted to verbs *mít* - to have, *patřit* - belong, *patřit k* - belong to).

4. A preliminary sketch of formal description of possessive relation is proposed in the form simplified for certain applications (analysis and synthesis of Czech sentences for machine translation).

AN INTEGRATIVE SEMANTIC APPROACH
TO INTERSENTENCE PHENOMENA

Th. R. Hofmann (Canada)

We define a semantic theory as integrative if it makes provision for a single, integral meaning for a coherent text. This property is shown to be equivalent to representing the cognitive content of a discourse without representing its expression, i.e. without identifying which semantic elements derive from which sentences, words etc. In other terms, a semantic representation is integrative if and only if it is a "pure" semantic representation, i.e. if it does not represent syntactic and morphological features of expression. The most natural model for this type of semantic theory integrates the meaning of each successive sentence into the product of previous integrations, and ends up with the overall meaning of a text, after integrating its last sentence into the product of all integrations.

Of the several integrative theories which have been proposed, including Leech's deep semantics, and Mel'uk and Zolkovski's semantic graphs, it is C-net theory (developed primarily by Hofmann and Paillet) which is used in this paper to illustrate the detection and measurement of redundancy, identification of focus, and measurement of coherence of a text.

Intersentential anaphora with nominal, clausal and non-existent antecedents is accounted for, and "backwards pronominalization" is discussed. There appears to be no need for a syntactic transformation to account for pronominalization.

Lastly, a means is sketched for reconstructing those suppressed logical relationships between sentences which characterize literary as opposed to scientific writing.

RECENT RESEARCH RESULTS ON THE COMPUTATIONAL
SIMULATION OF LANGUAGE UNDERSTANDING SYSTEMS

Luc Steels (Belgique)

(a) Syntax and morphology. It will be argued that there is a theoretical relation between computational systems for the following linguistic tasks: 1. Morphographemic rewriting, where the alphanumeric input is encoded and orthographic problems are solved; 2. Lexical analysis where input words are splitted in possible subparts and located in the dictionary; 3. Morphological analysis, where the subset of the well-formed wordsplittings is defined and the internal word structure is reconstructed; 4. Syntactical analysis, where a constituent analysis for the input sentence is computed. It turns out that all these tasks can be performed using models from the theory of abstract automata. More specific, a deterministic finite automaton will be proposed to perform orthographic rewriting and encoding, the dictionary will be considered and implemented as a nondeterministic finite automaton and a recursive finite automaton (equivalent with a context free grammar) will be used to obtain the morphology and phrase structure syntax analyses.

To make the theoretical notions valid for practical use the conceptual base and a FORTRAN IV implementation of a nondeterministic parsing mechanism will be described. This parser takes as input the described automata and computes the required structural information. Also an extension of the system for the incorporation of features in the description will be developed. The main result is that the same system can be used for lexical as well as morphological and syntactical analysis, thus making interesting simplifications in the design of over all linguistic systems possible.

(b) Semantics. Also some results on higher levels will be discussed briefly. The semantic component is considered to have two parts: (a) A system for functional analysis and code generation such that the obtained structural information is used to perform a functional analysis and to generate a program written in a formal query language. (b) a system for code generation where the program is then executed such that information is stored in a database or questions are answered.

USER-ORIENTED AUTOMATIC SEMANTIC ANALYSIS OF TEXTS

Jacques Noel (Belgium)

This paper deals with the problem of setting up algorithms for converting texts into user-oriented representations of their content. It concentrates on the question of sense relations (such as implication, inference, projection, pre-supposition, and causation) because of their importance for the study of discourse and for the representation of knowledge. The data analysed include a poem by Prévert ("Déjeuner du matin") and English abstracts in the field of information science, which were processed in a mechanized indexing experiment presented by Noel 1974 (unpublished dissertation, to be distributed by ERIC in 1975). The present paper discusses the empirical and theoretical foundation of this experiment in the light of some recent work. Concerning the empirical foundations, arguments are presented which suggest:

(i) the validity of a goal- and user-oriented, and field-specific approach to content analysis based on the use of a metalanguage;

(ii) the need for a mediate, procedural (systematic) conception of the relations between expression (texts) and content (metalanguages) vs. the direct relation as advocated in generative semantics. (Lado and Higgs on the priority of thought in memory performance).

Recent theoretical investigations are adduced to discuss and assess:

(i) the crucial importance and the difficulties of causation in dealing with sense relations and their repercussions in discourse structure (e.g. the post hoc, ergo propter hoc problems; Grimes 1972 on rhetorical predicates);

(ii) the use in Noel 1974 of a version of Sager's restrictionless string grammar to model expression in texts (Paillet 1973); the subcategorization problem and a refutation

of Salkoff 1973's grammatical approach; the need for codes extending selectional restrictions to conceptual and expression units larger than the sentences);

(iii) the need to include a multilevel notion of effect in semantic analysis procedures of texts (e.g. the problem of 'reading between the lines'; cp. Freud on jokes) and the related need for a memory keeping track of discontinuities.

The conclusions concern the abstracts form of semantic analysis procedures of texts, some empirical prerequisites for setting them up, and the possibility of generating more texts than can be observed in a corpus.

FORMALE SEMANTIK UND LINGUISTISCHE MODELLIERUNG

Alfred Nozicska (Österreich)

Unter Semantik verstehen wir jeden Prozeß, der zwischen Sprachen und Dingen vermittelt. Ziel des Referates ist es, Aspekte der Semantik natürlicher Sprachen vom modelltheoretischen Standpunkt aus zu untersuchen. Dabei werden folgende Problem-bereiche berührt:

(A) DAS EPISTEMOLOGISCHE PROBLEM. Im Gegensatz zur Theorie der Syntax z.B. entsteht mit der Formalisierung der Semantik eine "Antimonie des Beobachtens": die Tätigkeit des "Beobachters" affiziert (und verändert damit) das beobachtete Objekt (bzw. dessen Elemente und Relationen), da letzteres in der Tätigkeit selbst involviert ist. Es wird versucht, diese Antimonie exakt zu formulieren.

(B) DAS MODELLTHEORETISCHE PROBLEM zentriert sich in zwei Punkten:

(1) in der Übersetzung der natürlichen Sprache (bzw. Fragmente davon) in eine künstliche Sprache \mathcal{L} oder eine Folge zunehmend komplexer künstlicher Sprachen $\mathcal{L}_1, \dots, \mathcal{L}_n$. Die Bildungsregeln für \mathcal{L} simulieren diejenigen von \mathcal{L} : sie sind ein Modell (Grammatik) für \mathcal{L} . Das Kernproblem besteht hier in der Auffindung genügend reicher \mathcal{L}_i , die die Bedeutungsrelationen von \mathcal{L} repräsentieren. (Das Problem der Universalität und Adäquatheit der Repräsentationsrelation $\mathcal{L} \mapsto \mathcal{L}$ taucht hier auf).

(2) Im Auffinden einer formalen Interpretation \mathcal{J} von \mathcal{L} und der Bildung von Modellen $\text{Mod}_{\mathcal{L}}$ für Sätze von \mathcal{L} als deren semantische Bezugssysteme. Hier fällt u.a. das Problem des "Teils" und des "Ganzen" in einer formal aufgebauten Semantik, insofern die Elemente von \mathcal{L} und deren Kombinatorik die möglichen Modelle $\text{Mod}_{\mathcal{L}}$ indizieren.

SIGNIFICANCE OF ISOLATED MORPHOLOGIES
TO FORMALIZE NATURAL LANGUAGES
Kümmel Peter (FRG)

Among natural language carrier systems like a) English, French, German, Indian and Russian, morphological phenomena exhibit structures, which due to syntactical entanglements within texts are quite tightly glued together. Different natural language carrier systems exist b), in which syntactical rules allow a more limited tie-up between morphological components. E.g. hieroglyphic and texts of Chinese, ancient "pictographic texts", sequences of traffic signs, ideoexpressive sequences of dearnute, gesture languages, morphosyntactically curtailed children expressions, etc. The more extensively glued together systems under a) were designated "systems of agglutinated morphologies" and the second array under b) "systems of isolated morphologies". Only a comparative evaluation of both discloses fundamental rules valid for content as well as morphological expression formalizations. In particular ideoexpressive and fully content related morphological structures gave hints to define the term: "sign" or "expression unit" legal and good for the whole scale of natural and artificial language expressions perceivable by all five senses. Further the definition of the phenomenon "expression quotient of one content unit" was made possible with the data $n/1$, 1 and $1/n$ concerning the usage ratio of expression units as carriers for one content unit. Finally one content unit, one meaning or Deuter has been formalized due to the natural phenomenon of "classifiers" in the treasures of Chinese characters. The more complex definition of one Deuter comprehends six sub-criteria in form of the so called "Deuter-Criteria": 1. IDENTITY, 2. TRUTH, 3. ASSOCIATION, 4. AGE, 5. FREQUENCY and 6. PRIORITY. These Deuter-Criteria have been enumerated and put into mutual metric relation within a sub-thesaurus as well as in the whole thesaurus of human knowledge.

ZUM PROBLEM EINER FORMALISIERTEN BESCHREIBUNG DER SEMANTIK
DES FINITEN VERBS (AM BEISPIEL DES BULGARISCHEN)

H. Walter (DDR)

1. Die finiten Formen des (bulgarischen) Verbs sind polysem. In ihrer Semantik verflechten sich Bedeutungselemente, die traditionell unter den Kategorien Tempus, Modus und Aspekt eingeordnet werden. Eine genaue Beschreibung dieser Semantik ist demnach nur möglich, wenn die temporale, modale und aspektliche Semantik gleichzeitig bei jeder zu beschreibenden Form berücksichtigt wird. Die Einbeziehung der aspektlichen Semantik erfordert eine getrennte Beschreibung der finiten Formen des perfektiven und des imperfektiven Aspekts.

2. Die Beschreibung erfolgt mit Hilfe der folgenden 20 semantischen Merkmale, die als Fragment einer Metasprache anzusehen sind: G - Gegenwart, Z - Zukunft, V - Vergangenheit, akt - aktuell, vorz - vorzeitig, gleichz - gleichzeitig, nachz - nachzeitlich, r - resultativ, du - durativ, ab - abstrakt-konstatierend, it - iterativ, st - stativ, kt - konstativ-tendentiv, Ev - Evidenz (svidetelstvenost), dir - direkte (persönliche) Aussage, pot - potentiell, hyp - hypothetisch, irr - unreal, opt - optativ, hort - hortativ (auffordernd). Jedes dieser Merkmale wird auf der Grundlage der bisherigen Forschungsergebnisse definiert.

3. Bei der Beschreibung werden neue Deutungen der Begriffe 'merkmalhaltig' und 'merkmallos' in Anwendung gebracht. Dabei werden die Markierungen der Merkmale ('u' - merkmallos, 'm' merkmalhaltig) von den Werten der Merkmale ('+', '-') getrennt. So kann eine Form hinsichtlich eines Merkmals mit seinem Wert 'merkmalhaltig' oder 'merkmallos' sein. Ist ein Merkmal mit einem der beiden Werte für die gegebene Form 'normal', so ist diese Form bezüglich dieses Merkmals mit seinem Wert merkmallos. Verhält sich aber die Form in bezug auf dieses Merkmal anders, als zu erwarten ist, so wird sie als merkmalhaltig eingeordnet. Dementsprechend wird für die Semantik jeder Form

mit zwei Matrizen gearbeitet. In die eine wird +/- und in die andere die dem Wert entsprechende Markierung (u/m) eingetragen. Es gibt Fälle, in denen nicht entscheidbar ist, ob ein Merkmal mit seinem Wert für eine Form 'normal' ist oder nicht. Dann wird die Eintragung aus der +/- -Matrix in die u/m -Matrix übernommen. Zum Beispiel: Es ist für die traditionellen "preizkazni formi" normal, /- dir/ auszudrücken, dem in der u/m -Matrix /u dir/ entspricht. Solche Formen können aber auch /+ dir/ ausdrücken, weshalb dann in der u/m -Matrix /m dir/ einzutragen ist (bei der sog. "admirativischen" Bedeutung). Das Merkmal 'it' kann bei allen finiten Formen des bg. Verbs den Wert '+' oder '-' erhalten. Deshalb wird die Eintragung aus der +/- -Matrix bei 'it' immer in die u/m -Matrix übernommen.

4.1. Mit dieser Beschreibungsmethode kann das Problem der Monosemierung der finiten Formen der Lösung näher gebracht werden, da die in der u/m -Matrix mit 'm' oder '+/-' versehenen Merkmale mit dem gleichen Wert auch im Kontext vorhanden sein müssen.

4.2. Das Merkmalinventar kann in Verbindung mit der vorgeschlagenen Beschreibungsmethode als 'tertium comparationis' beim Sprachvergleich benutzt werden, was den Gedanken nahelegt zu überprüfen, inwieweit die vorgeschlagene Beschreibungsmethode beim Aufbau einer 'Mittelsprache' im Sinne der Übersetzungstheorie berücksichtigt werden kann.

4.3. Das Modell hat insofern auch heuristischen Charakter, als es die Möglichkeit bietet, durch Abarbeiten aller möglichen Eintragungen bisher unberücksichtigt gebliebene semantische Fakten innerhalb des Systems des Verbum finitum zu ermitteln.

ON SEMANTICAL ORGANIZATION OF TEXT

W. Marciszewski (Poland)

The semantical organization is conceived as something opposite to the disconnectedness. Four factors of semantical coherence are distinguished and numerical indices for them are proposed. As a device in the investigating two of these factors the relation table is introduced that is constructed by means of two identical series of the text key concepts: one written down in the first left column and the second in the top verse. At the crossing of verses and column the predicates are placed which denote the corresponding arguments. In this way we arrive at a register of main propositions of the text. Those points of intersection which remains empty suggest new problems: one may ask for the relations between the concepts in question.

The relation table helps in the investigating the two factors of semantical organization which in the paper are called compactness and monotopicalness. The more a text is compact the greater is the number of relations between key concepts. The text is monotopical if there is one and only one series of predicates filling entirely either one verse or one column. Another factor of semantical organization is a proper division of text. The idea of proper division is explicated in terms of structural adequacy that depends on the isomorphism between physical and thematical structure of the text. The physical structure is formed by the ramified sequences of chapters, paragraphs, sections, etc. The thematical structure arises from branching off the main problem into more and more minute subsidiary problems.

The last factor analysed is the ordering of text by an ordering relation which holds between sentences, e.g. entailment in deductive theory, a sequence of events reported in a story. The degree of ordering $\text{Ord}(T)$ may be defined by the function

$$\text{Ord}(T) = 1 - \frac{i}{s}; \quad \text{where:}$$

i - is the number of irrelevant permutations;

s - is the number of sentences in the text.

TITEL. REFERENZ IN DEUTSCHEN WISSENSCHAFTLICHEN UND
LITERARISCHEN TEXTEN

Sw. Gizowa (VR Bulgarien)

Der Bericht legt Ergebnisse der Diplomarbeit

Referenz in deutschen wissenschaftlichen und li-
terarischen, dar, welche am Lehrstuhl "Deutsche Philologie"

"Westliche Philologien" an der Sofiaer Univer-

Leitung von Herrn Dr. A. Ljudskanow ausgearbei-

versuchung ist die Ermittlung von syntaktischen und

Antezedenten in Bezug auf die Relationen der Pronomen und

Antezedenten in deutschen wissenschaftlichen und literari-

schlichen Voraussetzungen für die automatische Identifi-
kation Antezedenten einiger pronominalen Klassen

geschaffen werden.

versuchung wurden je 50 Seiten wissenschaftliche und

literarische ausgewertet. Die Analyse beschränkt sich auf

possessiv- und Relativpronomen in der Einzahl.

enthält einen Versuch dar, durch eine Metasprache

die Funktion der pronominalen Referenz in deutschen

literarischen Texten zu beschreiben, wie dies für bulgarische

Texte bereits erfolgt ist.

Parameter und Parameter der Metasprache sind:

linksgerichtet und rechtsgerichtet, Antezedent, aktuel-

potenzieller Antezedent, Q, Q-max E, E-max, Segment.

besteht in der Identifizierung dieser Parameter in

welchen ihnen quantitative Werte zugeschrieben werden.

geworden Werte für einige wichtige Parameter gegeben

welchen entsprechenden Resultaten aus den bulgarischen

gestellt. Die Gegenüberstellung zeigt, daß die

welchen, wie auch die Werte für Q-max und E-max in deut-

lich bedeutend größer sind, als die Werte, die bei den bulga-

rischen festgestellt wurden. Es wird auch der Versuch unter-

suchung linguistisch zu interpretieren.

SUR LES REFERENCES NON-PRONOMINALES

A.Ljudskanov (RP de Bulgarie)

- I. 1. Genèse et logique interne de notre recherche sur les références dans les langues naturelles.
2. Le problème de l'identification automatique des antécédents actuels des pronoms dans le processus de la TA et de l'analyse automatique des textes.
3. Analyse de quelques algorithmes. Nécessité d'une information linguiste plus complète. Méthode d'analyse et metalanque de description.
- II. 1. Les références non-pronominales sont un phénomène des langues naturelles très peu étudié.
2. Les quasi-antécédents du mécanisme de la référence pronominale et les références non-pronominales.
3. Description par la metalanque proposée. Quelques résultats numériques de l'analyse des textes bulgares et russes et brève description grammaticale et sémantique.
- III. 1. Impossibilité de déduire une définition des références non-pronominales des définitions de l'anaphore.
2. **Analyse de trois phénomènes: la répétition, la référence pronominale et la référence non-pronominale. "Lignes" anaphoriques.**
- IV. 1. Précision de quelques concepts de base du point de vue de la gnoécologie matérialiste; signification linguistique; référant et nomination; nomination pluraliste; nouvelle nomination; nouvelle nomination par abstraction et par "prénomination" c.à.d. par changement du sous-ensemble des traits pertinents.
- V. Quelques problèmes de l'automatisation.

MEASUREMENT OF READABILITY OF WRITTEN BULGARIAN

Encho Gerganov, Andriana Mateeva, Borimir Kristev (Bulgaria)

Readability of 10 samples of written Bulgarian was measured by two methods - testing of reading comprehension and scaling of difficulty. In the first method 3 questions about the content of each passage were constructed. Each question involved 5 answers - one correct, one wrong and the others with various degree of correctness. The correctness of the answers was measured by the method of paired comparisons and scale value to each answer was estimated with respect to the correctness. Then the passages with the questions and the answers involved were presented to 300 readers, whose task was to read each passage and to answer the questions by choosing the most correct answers. The scale values of the answers the readers chose were assigned to the passage. It is postulated, that a given passage would be completely uncomprehended by a given group of readers if all the readers have answered all the questions by choosing answers randomly. The mean (M) of distribution of scale values of the randomly chosen answers is considered as a zero amount of comprehension. Index (D_i) of difficulty of a given passage i was defined by the formula $D_i = \frac{|M_r - M|}{\sqrt{\sigma_r^2 + \sigma^2}}$, where $M_r, M, \sigma_r^2, \sigma^2$ denote means and dispersions of the corresponding real-choice and random-choice distributions of answers' scale values for passage i . In the second method difficulty of the passage was measured by the method of paired comparisons and scale values of difficulty for 10 passages, were estimated. Product-moment correlation between D and scaled difficulty is very high ($r=0,81; p < 0,01$). A formula of readability of written Bulgarian has been derived.

MACHINE DE CONTROLE POUR L'ENSEIGNEMENT DE LA
DIALECTOLOGIE BULGARE (M.C.E.D.B.)
(PROBLÈMES LINGUISTIQUES ET TECHNIQUES)

Tsanko Mladenoff - Danaïle Dancoff (Bulgarie)

M.C.E.D.B. diffère essentiellement des appareils révardus, qui calculent simplement les résultats d'une enquête fondée sur des questions binaires ou sur un choix de réponses possibles. Le but de M.C.E.D.B., c'est de modéliser une situation réelle de l'examen universitaire, en reproduisant le rapport "professeur - étudiant". M.C.E.D.B. sert aux exercices et aux appréciations automatiques des connaissances à deux de rés:

(1) identifier un dialecte donné seulement par son nom (c'est-à-dire, examiner et apprécier les connaissances sur une matière définie: les isoglosses essentielles);

(2) apprécier les connaissances sur une quantité d'isoglosses d'importance identificative secondaire (au volume prévu du cours universitaire de dialectologie bulgare).

M.C.E.D.B. dispose d'une applicabilité relative est plus universelle: si nous remplaçons les cartes perforées par un nouveau complexe, en modifiant la signature sur le clavier, M.C.E.D.B. se transforme en examinatrice d'une autre discipline (par ex.: pour l'apprentissage des paradigmes).

L'étudiant lui-même choisit les questions (éléments phonétiques et (ou)morphologiques vieux - bulgares) et puis donne les réponses au moyen d'un alphabet phonétique (éléments correspondants dans le parler).

Pour les cas de réponses absurdes, les cartes perforées contiennent une information préliminaire qui fait rompre automatiquement l'examen.

L'appréciation des résultats est double:

- (1) selon la vérité
- (2) selon l'importance.

La probabilité des réponses exactes subies par hasard ne domine pas celle des réponses reçues par hasard en circonstances habituelles.

UN MODELE LOGIQUE DE RECONSTRUCTION CHRONOLOGIQUE
EN PHONOLOGIE NATURELLE

T. Charpentier, J.Y. Morin, M. Picard (Canada)

- Si l'on considère comme données les hypothèses suivantes:
- a) Non-adaptabilité des règles phonologiques: une règle phonologique ne s'adapte pas à l'output des changements qui l'affectent i.e. ne s'applique pas à l'output des changements qui la nourrissent, e.g. la règle (i) si elle précède historiquement (ii) (i) $V \rightarrow V / _ N \left\{ \begin{array}{l} \# \\ C \end{array} \right\}$, (ii) $\partial \rightarrow \emptyset / _ \#$, (iii) $/an\partial/$, (iv) $/\tilde{a}/$ ne peut être nourrie par elle dans la dérivation à partir de (iii), si on a (iv), c'est que (ii) précède (i) historiquement. Inversement, aucune règle ne peut être asséchée par un processus plus récent dans l'histoire de la langue, e.g. la règle (i) ci-haut ne peut rester telle quelle si une règle $N \rightarrow \emptyset / _ \#$ est introduite dans la grammaire à une époque ultérieure.
- b) Aucune règle phonologique n'est indépendante du contexte, e.g. un processus $h \rightarrow \emptyset$ est un changement ponctuel qui ne peut faire partie d'aucune grammaire synchronique (si ce n'est comme règle adaptive).
- c) Un processus affectant un même segment dans plusieurs directions différentes ne peut pas avoir été introduit en une seule étape, e.g. une règle $K \rightarrow \left\{ \begin{array}{l} \emptyset \\ \# \\ C \end{array} \right\}$ doit constituer une fusion de processus chronologiquement distincts.

De même que les contraintes sur la forme des règles et des changements phonologiques de la phonologie naturelle, il devient possible d'élaborer un modèle logique et heuristique pour la reconstruction interne de la chronologie des changements phonologiques (grammatisés ou non) dans une langue donnée. Un tel modèle, éventuellement automatisable, prendrait comme input les correspondances entre langue-source et langue-cible telles que définies par leurs grammaires synchroniques respectives (la langue-source peut aussi bien être reconstruite qu'attestée) et fournirait comme output une chronologie relative des changements phonologiques et des différentes grammaires synchroniques intermédiaires (pour les changements qui ont été grammatisés).

COMPUTER EXECUTION of FUZZY ALGORITHMS

V.Dimitrov, W.Wechler, D.Drjankov, A.Petrov
/Bulgaria, GDR/

A fuzzy instruction contains in its formulation linguistic uncertainties like "probably", "possibly", "likely", "maybe", "more or less", "almost", "nearly", etc., or words which represent labels of fuzzy classes i.e. classes of objects in which there is no sharp boundary between those objects that belong to the class and those that do not, for example: rich, old, clever, honest, beautiful, large, great, slow, successful, etc. The sequence of fuzzy instructions, after whose execution an approximate solution of some complex problem is obtained, is called fuzzy algorithm. The fuzzy algorithms are useful in modelling and studying humanistic systems, that is systems in which human behaviour plays a major role.

In the proposed paper various methods of computer realization of fuzzy algorithms that are composed by the following types of fuzzy instructions are analysed: assigning instructions /e.g. "assign to X a not very high value"/, conditional instructions /e.g. "if X becomes very large, decrease Y slightly"/; unconditional instructions /e.g. "multiply X several times by Y"/. A program written in FORTRAN for ES 1040 describing a purposeful movement of "robot" commanded by fuzzy instructions is enclosed.

Information Retrieval Systems

D.M.Dobrev

The increasing needs for creation and exploitation of Information retrieval systems hardly demands software facilities for generating such systems. Such software has to face a broad spectrum of demands. The system has to be maximum flexible and simple for generating and operating, it has to be able to provide effective offline and online search and data transformation functions.

There exist two main methods of creating software for the needs of Information retrieval systems creation of a special artificial language or creation of a program package.

This report considers the structure and possibilities of a program package for creation and exploitation of Information retrieval systems of a rather broad class. This program package allows not only search capabilities and flexible input and output options, but a large number of operations for the databank.

FACETS, FORMS AND RINGS

B. Harris (Canada)

This paper considers some of the theoretical and practical conclusions drawn from work done at the Linguistics Documentation Centre on indexing and abstracting (the FIRL and FABS projects).

The notion of 'faceted indexing' is closely linked with that of 'coordinate indexing' (Ranganathan) and immediately suggests Cartesian coordinates in multidimensional space as a model (May). The number of facets is not a problem in this model, and we may wish to take account of negative coordinates, multiple coordinates on the same axis, etc.

The prime motivation for faceting however, is not mathematical but semantic and cognitive: if we postulate a priori a definite number of facets, and if we choose to linearize them in a particular order (Ranganathan), or group them (FIRL), it is because they capture something useful in the classification of description. Faceting is therefore a matter of semiotics. Furthermore the surface presentation of the facets will certainly be linear, because users cannot conceptually manipulate multidimensional space above three or four dimensions; and the linearized expression may be 'translated' into natural language (FABS). Indeed we can speak of 'natural' vs. 'artificial' faceting, with a spectrum between.

In the middle of that spectrum is the 'form to be filled up', which we are all familiar with as an instrument of administration. It is more generally an instrument of search for knowledge. It directs search to a class of knowledge. It enables the computer, or any mechanical system, to interrogate in a meaningful but open-ended way (as opposed to multiple-choice prompting, in which the interrogee must be presented explicitly with all the options). This strategy has become the basis of the data entry system for FIRL and for our bibliographic indexing system (INFOKWOCK). FABS on the other hand combines facets and multiple choice.

Facets are not formally hierarchical. Systems like colon classification may achieve hierarchy, however, by underlying semantics or -- most typically -- by the arrangement of the indexing terms in a separate 'thesaurus'. Thesauri are usually tree-like, though they are never strictly confined to tree structures, and indeed are always multi-rooted. The notion of 'equi-relevance' (not the same as linguistic synonymy) leads to the need for a structure which is both vine-like and ring-like. This is the structure in the FURL thesaurus.

PROGRAMS OF EXPERIMENTAL RADIOELECTRONICS AUTOMATED
INFORMATION-RETRIEVAL SYSTEM AND ITS FUNCTIONS
K.Ilieva, A.Kuzmanov, I.Mitsev, T.Djakov,
E.Jerassy, G.Bakalsky, S.Anastassova, T.Anghe-
lova, I.Stamenova, B.Gheorghieva (Bulgaria)

The paper considers the principal programs of the above mentioned system and its functions as follows:

1. The Program, forming the machine dictionary: a) arranges automatically the word-list of descriptors in the computer memory by given parameters; b) divides the descriptors into two parts: "gists" (the first four symbols) and "tails" (the other part of the descriptor).

2. The Program for automated indexing: a) records the factologic parts of information cards and information forms in the memory of the computer, encoding some of its headings as country, language, firm, etc. and recording the others of them - as author, title etc. - without any change; b) realizes the automated indexing, i.e. using the descriptor dictionary, translates the summaries and inquiries from natural languages (Bulgarian and Russian) into metalanguage of the system, forming the matrix of descriptor codes for each summary and inquiry.

3. The Program for searching and giving out answers: a) compares the matrices of inquiries (MI) with the matrices of summaries (MS) including the document in the relevant submultitude when all descriptors from MI are found in the MD; b) prints from the full records of documents only the content of those headings marked off by the user as questions.

The programs of the system are written in Language of Symbolic Encoding for the computer Minsk-32.

Some experimental data are reported in our paper.

ELIMINATION OF POLYSEMIA AND HOMONYMITY BY A DIALOGUE IN
THE RADIOELECTRONICS AUTOMATED INFORMATION-RETRIEVAL SYS-
TEM K.Ilieva. A.Ljudskanov (BULGARIA)

1. The experimental variant of the completely automated system - RAIRS, with bilingual input (Bulgarian-Russian) is based on a dictionary of a particular type, assigning a hybrid descriptor language.
2. There are 13 types of homonymity and 16 types of Polysemia in it. The homonymity, caused also by appointed segmentation, consists in a possible insertion of many textual units, belonging to different syntactical classes, into one vocabulary unit. The polysemous descriptors have not more than two meanings in the dictionary.
3. The particular algorithms "Homonymity" and "Polysemia" have been made for automated elimination of these phenomena, and have been included in the indexing flow chart. But they are not able to eliminate all the complicated cases of homonymity and polysemia without syntactical and semantical analysis.
4. Two other algorithms have been made in this situation: "Dialogue-Homonymity" and "Dialogue-Polysemia". Their purpose is to duplicate the action of machine algorithms for a comparison of results from "man" and "machine" work, for a possible correction of machine algorithms. The subject, "man-machine" dialogue will be included in the industrial variant of the system in this case only when the experiments will demonstrate its real effectiveness.
5. The long-term hardware ensuring a dialogue is a terminal input-output device.
6. The software for elimination of homonymity and polysemia is based on the descriptor dictionary, machine or man-machine algorithms, list of four standard phrases, etc.
6. The indexing flow chart controls the action of both machine and man-machine algorithms which eliminate "Homonymity" and "Polysemia".

CONCEPTUAL MODEL OF THE AUTOMATED LEGAL INFORMATION RETRIEVAL SYSTEM

Jan Bobrowski /POLAND/

The Polish Parliament Library leads the pilot project of the automated legal information retrieval system. The following principles have been established:

1. Searching the data files can be run as batch process or real time conversation.
2. Data files consist of two types of entities:
 - /a/ legal acts in full text form treated as the facts,
 - /b/ catalogue entries furnished with descriptive indexes informing about the context of primary sources such as parliamentary and law literature and unpublished documents created in legal process.
3. Lexical set of information retrieval language consists of descriptors with related nonedescriptors compiled in thesaurus. The arrays of flexion forms of lexical set are applied to full text information retrieval.
4. Retrieval procedures. There are two separate retrieval procedures:
 - /a/ full text retrieval - the query terms together with arrays of flexion forms will serve as a basis for generating full set of query terms / also synonyms /. Search process will match all words relevant to the generated set of query terms.
 - /b/ catalogue entries retrieval - the set of the query terms will be enlarged by insertion of the related broader and narrower terms, which enables retrieval in conversational process.
5. Technical feasible design of described above system is expected to be finished till the end of 1978.

LA STRATEGIE SELECTIVE DANS LA LUMIERE DES NOUVEAUX
DEVELOPPEMENTS DE LA TRADUCTION AUTOMATIQUE

A.Ljudskanov (RP de Bulgarie)

I. Bref aperçu historique.

1. Malgré les résultats obtenus, l'état des choses dans le domaine de la TA était insatisfaisant vers les années 1965-1970.
2. Le rapport ALPAC et les points de vue pessimistes du prof. B.Hillel et de certains linguistes. L'analyse des causes qu'ils proposaient n'était pas complète.
3. Les stratégies "totales" - une des causes principales de l'état insatisfaisant. Genèse et logique interne de la stratégie totale. La stratégie totale, les grammaires génératives et l'idée de la traduction FQ totale. La stratégie totale n'est pas fondée sur l'analyse du mécanisme linguistique et des possibilités de l'opération traduisante entre les langues naturelles.

II. Analyse du mécanisme linguistique de l'opération traduisante entre les langues naturelles.

1. Communication unilingue et "bilingue"; schémas logiques et différence des buts.
2. La traduction et la compréhension; compréhension substantielle et compréhension linguistique comme choix des significations actuelles des éléments linguistiques du message d'entrée.
3. Notion de l'information nécessaire à la traduction. - I (Tn); son caractère objectif et prédictible; différence de la qualité et de la quantité de l'I (Tn) pour le passage entre des paires de langues naturelles différentes.

4. Accumulation de l'I (Tn); types d'objets et types d'analyses (extralinguistique et linguistique; non-contextuelle et contextuelle; analyse de surface et de profondeur; analyse de contenu et analyse formelle). L'automatisation de l'accumulation de l'I (Tn) - problème fondamental de tout système de TA.
5. Les différentes classes d'éléments linguistiques exigent des différents types d'analyses. Précision du caractère sélectif de l'opération traduisante. Conséquence - la stratégie de la TA devrait être sélective.

III. La stratégie sélective.

1. Description sommaire de la logique de la stratégie sélective, Possibilité de transformer les types d'analyses plus "profonds" en analyses plus "superficielles".
2. Est-ce-que l'accumulation de l'I (Tn) exige nécessairement une analyse syntaxique et sémantique totale, ou bien on peut se contenter d'une analyse "orientée"?

IV. Les développements nouveaux.

1. Les idées fondamentales du système de TA de Meltchouk et de Zolkovski; la TA "industrielle" du prof. Piotrovski; le nouveau point de vue du prof. B.Hillel; etc.
2. Ces développements approuvent au moins en principe l'idée de la stratégie sélective, fondée sur la notion de I (Tn) minimale et nécessaire comme une des voies possible et optimale de la réalisation de la TA.
3. Perspectives.

AUTOMATIC HIGH-QUALITY TRANSLATION
Irina Uzunova (BULGARIA)

generally known Professor Bar-Hillel, one of the pioneers of machine translation in the fifties has, in recent years, come out with a number of extremely pessimistic articles on the possibilities and future of MT. In his last article entitled "The Future of MT", published in 1969 in the Russian journal "Prasi-Yazikoznanya", he says quite explicitly that "Automatic high quality translation is not feasible. Present research in MT has come to an impasse".

The author examines the negative impact of these statements on the discouraging background, created by the ALPAC report of 1968. She goes on to comment that one of the aims of her report is also to summarize the main viewpoints given in the reports in England and the United States from 1971 to 1973. At the same time a critical analysis is made of investigations carried out in recent years in socialist countries and, in particular, of Dr. A. Lyudskanov's selective strategy. The second section of the report examines the concept of Fully Automatic High Quality Translation. It also includes a discussion of Professor Bar-Hillel's apostasy and the viewpoints of Reizei, Gattford, Filkmore, Lyons, Stachowitch, Zgusta and others, on the impact of Dr. A. Lyudskanov's works on the above mentioned Anglo-American study. The final section deals with the variability of standards, translatability, the human role in translation systems, man-assisted or machine-assisted translation. Views are assumed a priori and not as a result of an analysis of the actual and potential possibilities of the translation mechanism in natural languages. At the end of the report the author presents her final comments.

OPTIMUM LINKING OF INDEPENDENT ANALYSIS
AND SYNTHESIS

Svatava Machová (Prague)

Examples of working cooperation of two or more teams have been rather unique in linguistic science so far. (By working cooperation I mean a purposeful adoption of procedures and methods between the teams.) It is only after painstaking tests that a team of linguists can become certain that some procedures or methods adopted from some other team can be used in an adequate manner. In spite of this it is evident that this kind of cooperation is highly desirable at the present time.

In preparing the MT from English into Czech, the Prague Group of Mathematical Linguistics intends to make use of the analysis of the English language worked out at the Montreal University. (R. Kittredge) linked with the synthesis of the Czech language proposed by P. Sgall at Prague University. Thus procedures worked out and proposed independently of one another are to be utilized. An attempt will be made in the communication to present some ways of achieving an effective linkage of classifications of linguistic entities based on different systems.

ZU EINIGEN VORAUSSETZUNGEN FÜR EIN LINGUISTISCHES
MODELL DER TRANSLATORISCHEN KOMPETENZ

Gert Jäger (DDR)

1. Ein Modell der translatorischen Kompetenz soll die Fähigkeit eines Translators erklären, zwischen zwei gegebenen Sprachen L_1 und L_2 Translationen zu vollziehen, d.h. L_1 -Texten L_2 -Translate und L_2 -Texten L_1 -translate zuzuordnen. Als Translationen werden alle Prozesse (und nur diese) betrachtet, bei denen die quellsprachlichen Texte und die ihnen zugeordneten zielsprachlichen Texte hinsichtlich ihres kommunikativen Wertes übereinstimmen, d.h. kommunikativ äquivalent sind.

2. Für die Schaffung eines linguistischen Modells der translatorischen Kompetenz muß ein linguistisches Explikat für die Begriffe 'kommunikativer Wert' und 'kommunikative Äquivalenz' zugrunde gelegt werden. Als Explikat für 'kommunikativer Wert' kann 'funktioneller Wert', verstanden als Gesamtheit der einem Text zukommenden aktuellen signifikativen und intralingualen pragmatischen Bedeutungen sowie der mit der aktuellen Gliederung verbundenen Bedeutungen, angenommen werden. 'Funktionelle Äquivalenz' kann jedoch nicht als Explikat für 'kommunikative Äquivalenz' betrachtet werden: Sie trifft für ein idealisiertes Objekt, die ideale Translation, zu. Die realen Translationen nähern sich der idealen Translation (in unterschiedlichem Grade) an, da bei ihnen die Beziehung der funktionellen Äquivalenz durch das Wirken von Faktoren (Bedingungen) überlagert wird, die im gegenseitigen Verhältnis der Gegebenheiten von jeweils zwei Sprachen liegen.

3. Einem linguistischen Modell der translatorischen Kompetenz muß daher ein Translationsbegriff zugrunde gelegt werden, der auf dem Begriff der maximalen Äquivalenz, verstanden als durch das Wirken der genannten Faktoren modifizierte funktionelle Äquivalenz, beruht. Da diese Faktoren auf den verschiedenen Repräsentationsebenen der Sprache liegen, muß ein stratifikatives Modell angenommen werden, das es gestattet, die beiden gegebenen Sprachen auf den jeweiligen Repräsentationsebenen zueinander in Beziehung zu setzen. Dies erfordert auch die Erklärung der Nichttransferierbarkeit, die in einem linguistischen Modell der translatorischen Kompetenz gegeben werden muß.

4. Die Repräsentationsebenen sind hierarchisch geordnet, wobei jeweils zwei unmittelbar benachbarte Ebenen durch Transduktoren miteinander verbunden sind, die einen Übergang zwischen den Ebenen in beiden Richtungen ermöglichen. An der Spitze der Hierarchie steht die Bedeutungs-R-Ebene. Über die Transduktoren bestimmt die jeweils übergeordnete Ebene die Norm (verstanden als tatsächliche Ausnutzung der durch das Subsystem gegebenen Möglichkeiten) der folgenden Ebene.

5. Eine Erklärung der translatorischen Kompetenz schließt die Erklärung der Fähigkeit eines Translators, sogen. Lücken im gegenseitigen Verhältnis von zwei Sprachen "auszufüllen", ein. In einem stratifikativen Modell der translatorischen Kompetenz kann das durch die Annahme einer generativen und einer rekognoskativen Komponente (die das jeweilige Subsystem beschreiben) für jede Repräsentationsebene berücksichtigt werden.

SOME PROBLEMS OF MODELLING OF APHATIC DISORDERS OF THE SPEECH
IN ITS PHYSIOLOGICAL, PSYCHOLOGICAL AND LINGUISTIC ASPECTS

R. Raychev, A. Lyudskanov (Bulgaria)

1. Preliminary notes. The basic problems in the Marxist materialistic gnosiology show the unbreakable connection between the material (the brain) and the ideal (the thinking), between thinking and speech etc. In its nature, the aphasia represents a disorder of definite mechanisms of the speech at its different levels in the presence of lesions in certain brain structures of the speech carrier. The study of aphasias as pathological speech manifestations and the comparison of the clinical phenomena with the facts for the structural and functional brain organization as well as the psychological and physiological mechanisms of the higher integrative functions of the Central Nervous System in man, not only from the theoretical point of view, but also for the elaboration of rational rehabilitation programme - that is the essence of the subject of aphasiology. Hence, as an object of investigation, aphasias are distinguished by two basic characteristics: interdisciplinary and indirect assessment of the disturbed speech mechanisms.

2. The modern psychophysiological aspects of speech pathology - neuropathology - proceed from the assumption of the modern physiology that the speech is a complicated functional system with multilayer organization, high automation, working with great positiveness, excellent harmony in work between the individual links and levels and constant autoregulation (models of N.A. Bernshtain, P.K. Anohin; block-schemes after Rosenblatt; logical and probable and statistical models of the neuronal links after McCulloch and Pitts, Shimbel and Rapoport and many others). These directions permit the regularities of the speech activity (its pathology resp.) to be drawn out from the general principles of physiology and pathophysiology of the Central Nervous System in man resp. This suggests the big complexity and interrelations of the system of

blocks for the individual psychophysiological operations inside in each level as well as between the different levels.

3. The fundamental linguistic problem in aphasias is to deduce which structures and mechanisms of the speech are disturbed and how much they are disturbed in the different kinds of aphasias, in certain comparison, eventually, with the localisation of the brain lesion and the pathophysiology of the disturbed cerebral function. So far, the analytical (inductive) methods of the traditional linguistic and of some statistical methods which have given definite results - significance of the incidence factor, phoneme entropy in the aphatic people's vocabulary, elucidation of the characteristics of agrammatism in the different kinds of aphasias, etc. - have been used for the analysis of the pathological speech manifestations.

The experience of cybernetics, of generative and mathematical linguistics and others, shows that such objects are best investigated by using the method of elaboration of model

In this paper are discussed the questions about the consideration of the speech as a generating and identifying the speech mechanism (codifying and decodifying mechanism), as well as the stratification conceptions and the consideration of the speech as a system of convergent mechanisms; about the problem of the dichotomy "competence-performance", and the generation as a transition from profound to superficial structures; about the equivalent or dominating, but not interpreting significance of the semantic factor in the natural languages; and the description of the languages by means of the Nervous System models - grammars and valence grammars; about the problem of the equivalence of "generating" and "identifying" description in aphasias, and the semantic multilayer models, etc. The question is raised for the "artificial intellect" and the model of aphasias, as well as for the eventual possibility for objective analysis of the problem "speech-thinking".

DISORDERS OF LINGUISTIC COMPETENCE AND PERFORMANCE
IN LANGUAGE PATHOLOGY
Ludmil Mavlov (Bulgaria)

Linguistic competence is the knowledg^we of a set of semantic and grammatical rules (algorithms) valid for a given language system. Linguistic performance is the usage of these rules in the generation (i.e. coding) of proper sentences (expressive performance) and in the comprehension (i.e. decoding) of other sentences (impressive performance).

The different focal brain lesions result into language function disorders called aphasias. Thus, it is in the aphasias that linguistic competence and performance are deteriorated, so the aphasias present a valuable model for the investigation of the verbal behaviour and structural organization of linguistic functions.

In this study the theoretical consequences of isolated disorders of competence, of expressive performance and of impressive performance are first postulated. Some kinds of aphasia disorders have then been analysed. We are of the opinion that they can be regarded as a pure deterioration of competence (e.g. the so called "transcortical sensory aphasia" according to the classical scheme of Lichtheim-Wernicke, or "acoustico-mnesic aphasia" according to Luria), as well as a pure deterioration of the expressive performance (e.g. "agrammatism" or "telegraphic style" and "anomia" or "amnesic aphasia"). It has been pointed out that in language pathology there is no aphasia presenting models of an isolated disorder of impressive performance. Proper methodological approach is proposed. The degree of development (or degradation) of competence and performance can thus be tested and quantitatively measured. The differences between linguistic disorders in the generation and comprehension, due to competence and performance deterioration on the one hand, and the instrumental sensory and motor disorders in the production and perception of verbal signals due to verbal agnosias and apraxias on the other, have been brought out.

AN ATTEMPT FOR ANALYSIS OF THE SYNTACTIC LEVEL
OF THE SPEECH IN PATIENTS WITH APHASIA

P.Ovcharova, A.Lyudskanov, R.Raychev
S.Karagyozova (Bulgaria)

- I. Formulation of the problem. Our purpose was to elaborate a model for the description of the syntactic structures of the sentences generated by patients with motor and sensory aphasia, to compare them with the corresponding structures of test sentences from healthy people and to make some psychophysiological and linguistic interpretations on the basis of this comparison.
- II. Methods. 1. Elaboration of a syntactic model for the Bulgarian language - grammar of dependencies. 2. Selection of test material from 14 healthy informers (description of picture) and corresponding description of the same picture from 35 patients (18 with sensory aphasia and 17 with motor aphasia). 3. Representation of the syntactic structures of the corresponding sentences (graphical and sagittated). 4. Definition and introduction of "external" (at the level of realisation) and "internal" (at the level of the dependent structure) criteria for juxtaposing the structures in healthy and ill persons.
- III. Results. Differences between healthy and ill persons, as well as between the individual forms and degrees of aphasia have been established, both in the external criteria were: twice as big number of sentences in the aphatic patients depending, to a certain extent, on the degree of the syndrome's manifestation; marked differences between healthy and ill persons in the lexical and grammatical correctness. More significant differences in the internal criteria: manifestations of non-projectiveness; hanging; infringement of valence, especially in patients with sensory aphasia.
- IV. Preliminary inferences: The differences in the speech between patients with aphasia and healthy people on the level of the syntactic structures are considerable; there is a close relation between the syntactic and the semantic factors.

INFORMATIONAL AND STATISTICAL INDICES OF THE REPORTED SPEECH
OF APHASICS

P.Ovcharova, L.Tzoneva-Pencheva, M.Vulcheva (BULGARIA)

The informational and statistical indices of reported speech are determined experimentally by Shennon's Prediction Method and the data obtained are mathematically processed. The experiment was carried out with thirty normal subjects on the basis of sixty retold texts, thirty of which were elicited from aphasic subjects and thirty from normal ones. The results of the experiment enabled the authors to draw up the following distinctions between normal and aphasic speech:

The words in aphasic utterances are characterized by greater indefiniteness (entropy) when compared to those in normal utterances. This is symptomatic of the fact that the words spoken by the aphasics are much more context-sensitive.

The context relatedness of a word in aphasic speech (i.e. the mathematical difference between the optimal quantity of information borne by context-free word and the information charge of the context-related word) is considerably lower than that of normal speech. This typically shows that in aphasic speech the placement of words along the syntagmatic axis is ,to a much lesser extent , controlled by the constraints of statistical probability, semantics and grammar. The degree of impairment of the context relatedness of aphasic speech correlates with severity of the aphasic disturbance and much less with the particular form it chooses to accept. The informational and statistical speech indices could be used as objective criteria for an assessment of and a developmental approach to the aphasic disorder.

GRAMMAIRE DES CONTES DU TYPE 300 DU
CATALOGUE AARNE-THOMPSON
(Application aux contes populaires roumains)

Mihaela Dumitru (Roumanie)

Considérant comme prémisse l'idée que le conte compris comme langage infini, présente une série de répétition dont la logique est déterminée par le type de conte auquel il appartient (dans notre cas, le type 300 Ath), on a pu trouver un critère de classification générative des variantes d'un type.

L'application de la méthode générative détermine que la relation entre le type de conte et ses variantes devient en fait, la relation entre compétence et performance, c'est-à-dire, la relation entre la capacité de connaître et d'utiliser les règles et l'acte de la réalisation concrète de ceux-ci. La grammaire de synthèse, ou plutôt, la grammaire paradigmatique définit un informateur idéal, avec une compétence narrative plus complexe.

En analysant la grammaire paradigmatique obtenue, nous pouvons considérer un critère de classification typologique générative des contes fantastiques contemporains, même la compétence narrative de l'informateur.

La grammaire paradigmatique conditionne les réalisations concrètes du conte, soit qu'elles appartiennent à de diverses classes.

Les paramètres en fonction desquels on définit les types de grammaires ou les types de compétence narratives, sont preuve de cette affirmation.

A SYSTEM APPROACH TO JULIO CORTÁZAR' NOVELS

Aurora Niculescu-Mizil (Roumania)

Systems theory has a large variety of application in the contemporary sciences, its importance is constantly growing. The artistic works can also be thought as different systems which can include themselves; one also can speak about artistic meta-systems.

In the present paper, by using the mathematical systems theory we propose ourselves to find the "preferential constants" in a literature work; we use as an example 39 novels by the Argentinian writer Julio Cortázar.

Using concepts of systems theory we verified these constants' orderer interdependence. their dynamic and functional relationship.

Using a computer programme the frequency coefficients of system model were established.

It is important to point out that the model and its general framework of applicability gave satisfaction; the model coherence is holding and is proving that one can apply systems theory concepts with success even in the field of arts.

The mathematical modelling technique, has shown to be useful in finding new insight for analysis and understanding in the literature field, to find the preferential constants in the work of an author or a group of authors.

MATHEMATICAL HIERARCHIZATION OF ARCHITECTURAL
STYLES (applied to columns)

G. Chioca (Roumania)

The paper presents a mathematical hierarchy of some representative architectural columns of Ancient architecture. Semantic closeness between heterogeneous columns was looked for.

The research from papers (1) and (5) and from books (3) and (4) leads to the finding of a measure for the differences of architectural styles by applying the theory of binary codes and of the Hamming distance to some properties of the type considered in (1), and then treated as the models of (3-5). The chronological criterion appears only at the moment of selecting the ten types of columns to be considered. Further on, only architectural, historic and aesthetic criteria are used.

With the help of various methods of the mathematical theory of classification six hierarchies are obtained: four of them are obtained using Hamming's distance (three of them are established in relation to a single criterion - architectural, historical or aesthetical - and the other one by the simultaneous use of all criteria), a hierarchy applies the technique of "pilots" (introduced by Marie-Salomé Lagrange in (4) and a hierarchy is based on the transformation process of an initial binary matrix (one of its elements is equal to 1, when the column to which it refers has the property indicated by the line of the element, and to zero in the reverse case), by successive permutations of lines and columns in a diagonalized matrix. The obtained hierarchies (Table I) become criteria of comparative estimation of columns. Based on them, with the help of a procedure from (5), one obtains a hierarchy which synthesized all preceding ones.

Chronological Seriation of the Works of an Author
by Means of Computer

L. Boneva / Bulgaria

In the course of time an algorithm of putting a sample of works of one and the same genre of an author in its chronological sequence of composition turned out to be more and more useful. For instance an author may be well known but, for some reason or another, a dating of his works might be completely impossible.

The aim of this paper is to show that a sufficient chronological order can be obtained by comparison of some "varieties" of the author's style. These varieties could be established by the help of a suitable fonetico-syntactical criterion while the comparison, which could be carried out with a desirable accuracy only on a computer, is based on the well known MDSCAL algorithm, or precisely, on its modification known as the "Horse-shoe" method.

Some of the results obtained in this way are given in several tables and figures.

INFORMATIONAL IMPACTING OF STUDENTS' COMPOSITIONS
ina Novachkova (Bulgaria)

An attempt is made at characterizing students' compositions by quantitative indices with a view to observing the speech development of students.

Several indices are being considered: verb noun and adjective percentage, and variety of word-forms.

The excerpting and processing of the texts is most labour-consuming and cannot be carried out by the teachers alone. This requires that the students prepare the data themselves and assess their own text indices. But they cannot process each of their compositions either, since this would be rather time-consuming for them.

Possibilities are outlined for this processing to be performed by computer in the future so that the observations on the quantitative indices of compositions may be used in the management of verbal learning.

PRACTICAL JUSTIFICATION OF NOTATIONAL CONVENTIONS

Theodora Mikhailova, Bulgaria

grammatical device is being accounted for, which enables single representations in the Target Language /TL/ of utterances varying in possible arrangement of components in the Source Language /SL/ provides solution in cases where the two languages differ in ways of representing one and the same informational content.

double effect is achieved by a set of notational conventions used in a formulaic representation of the word order operative in the TL.

Formula has been worked out by means of selecting an appropriate set of symbols, arranging the symbols in an order to reflect syntactic system of the TL, and within this framework a specific meaning has been assigned to the individual symbols as well as to their positions.

essential prerequisite is the assumption that thinking processes are 'linguakized' along native tongue patterns, and therefore, a 'translator' can be devised which, on being fed with sequences of the SL, would yield as output equivalent sequences rearranged according to the requirements of the TL norms.

grammar, thus conceived, permits of partial algorithmization of the teaching/learning process.

This device is intended to facilitate and intensify foreign-lan-

AUTHOR INDEX

- Anastassova, S 66
 Anghelova, T. 66
 Arold, D. 36
 Bakalsky, G. 66
 Bobrowski, I. 68
 Boneva, L. 83
 Borscev, V. C. 28
 Bossilkov, K. 20
 Charpentier, T. 61
 Chauche, J. 32
 Chioca, G. 82
 Chomjakov, M. W. 28
 Christov, P. 12
 Corstius, B. 30
 Dancoff, D. 60
 Deweze, A. 23
 Dimitrov, V. 62
 Djakov, T. 66
 Dobrev, D. 63
 Drjankov, D. 62
 Dumitru, M. 80
 Fischer, I. 21
 Georgiev, H. 38
 Gerganov, E. 10, 11, 15, 59
 Gero, F. 14
 Gheorghieva, B. 66
 Gizowa, S. 57
 Hajicová, E. 46
 Harris, B. 64
 Hays D 42
 Ilieva, K. 66, 67
 Jäger, G. 73
 Jerassy, E. 66
 Karagyozova, S. 78
 Kirlova, R. 17
 Kotarov, M. 20
 Kristev, B. 26 59
 Klümmel, P. 53
 Kunze, I. 33, 35
 Kuzmanov, A. 66
 Ljudskanov, A. 8, 35, 58, 67,
 69, 75, 78
 Lubaszewski, W. 39
 Ludvikova, M. 13
 Lukaszewicz, W. 40
 Machová, S. 72
 Marciszewski, W. 56
 Mateeva, A. 59
 Mavlov, L. 11, 77
 Mikhailova, T. 85
 Misheva, A. 16, 19
 Mitsev, I. 66
 Mladenoff, T. 60
 Morin, J. Y. 61
 Naidenov, K. 20
 Neubert, G. 22
 Niculescu-Mizil, A. 81
 Nikolayev, B. 27
 Nikolov, V. 15
 Nořl, J. 50

Ovcharova, P. 78, 79

Pashova, T. 15

Petrov, A. 62

Picard, M. 61

Pitha, P. 47

Popova, M. 17, 18

Rainov, V. 11

Raychev, R. 75, 78

Rondeau, G. 24

Rüddiger, B. 41

Séljan, E. 37

Semeniuk-Polkowska, M. 31

Sgall, P. 44

Stamenova, I. 66

Steels, L. 49

Tilkov, D. 19

Tzoneva-Pencheva, L. 79

Uzunova, I. 71

Vulcheva, M. 79

Walter, H. 54

Wechler, W. 62



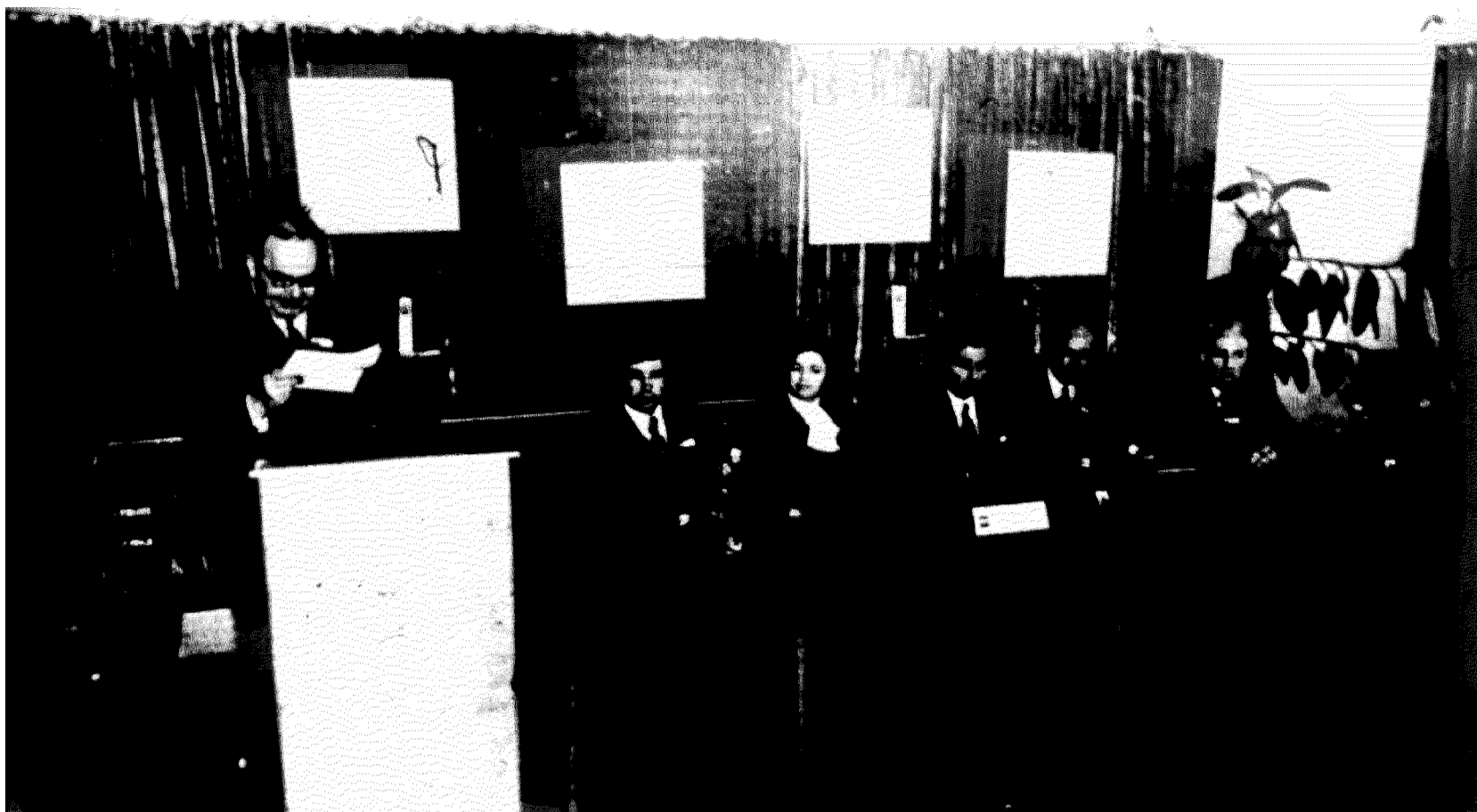
A. Ljudskanov.

Professor Dr. B. Sendov, Chancellor of the Sofia University "Kliment Ohridski"



**I-BA. НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МАЙ ВАРНА V 75**

The leader of the delegation from the USSR.



**I-BA НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МДЖ ВАРНА V 75**

The Organizer of the Conference; The Rector of the University of Sofia; the Lady; The Delegate from the USSR; The President of ICCL; the Director of the Institute of Mathematics.



**I-BA НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МДЖ ВАРНА V 75**

M. Untel; M. Vauquois; Hr. Karlgren; Pr. Ljudskanov.



**І-ВА НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МДЖ ВАРНА V 75**

Kay, Hays, and others listen to the opening remarks
of the Organizer.



**I-BA НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МАЖ ВАРНА V 75**

French delegates with the Rector: M. Deweze (son verre levé), M. Chauché (barbu), une irrésistible jeune femme r(o)usse.



**I-BA НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МДЖ ВАРНА V 75**

The Organizer of COLING 76, Professor Guy Rondeau,
with his associate Brian Harris.



**I-BA НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА-МДЖ ВАРНА V 75**

Academician Ilieff, Director of the Institute of
Mathematics and Mechanics, Sofia



**Г-ВА НАЦИОНАЛНА КОНФЕРЕНЦИЯ С МЕЖДУНАРОДНО
УЧАСТИЕ-КОМПЮТЪРНА ЛИНГВИСТИКА_МАШ РАДУА 1/78**

