Le present compte-rendu a pour objet de préciser les principes généraux utilisés, la réaction des étudiants et le rendement pédagogique obtenu.

1. *Graphes morphologiques:* Les mots d'une même famille. Notion de base. La double ramification. Les graphes abstraits. Les néologismes scientifiques.
2. *Graphes syntaxiques:* La double structure d'une phrase. Multiplicité des modèles. Point de vue psychologique. Notion de fonction. Continuité et discontinuité.
3. *Les séparateurs:* La segmentation d'une phrase. Le vocabulaire prioritaire.
4. Théorie de la valence: macro et microcontexte. Qu'est-ce-que "connaître un mot"?
5. Point de vue de l'étudiant; point de vue du traducteur humain; et point de vue de l'Enseignant.

Word and Context Association by Means of Linear Networks

Vincent E. Giuliano

Arthur D. Little, Inc.

This paper is concerned with the use of electrical networks for the automatic recognition of statistical associations among words and contexts present in written text. A general mathematical theory is proposed for association by means of linear transformations, and it is shown that this theory can be realized through use of passive linear electrical networks. Several small-scale experimental associative networks have been built, and are briefly described in the paper; one such device will be demonstrated in the course of the oral presentation of the paper. Some of the devices generate measures of association among index terms used to characterize a document collection, and between the index terms and the documents themselves. Another uses syntactic proximity within sentences as a criterion for the generation of word association measures. Examples are given of associations produced by these network devices. It is conjectured that the network-produced association measures reflect two distinct types of linguistic association—"synonymy" association which reflects similarity of meaning, and "contiguity" association which reflects real-world relationships among designata.

A Study of the Combinatorial Properties of Russian Nouns

Kenneth E. Harper

Rand Corporation

A statistical study was made of the extent to which Russian nouns enter into certain kinds of syntactic combination. The basis of the study was a corpus of 180,000 running words of Russian physics text prepared for analysis by the Automatic Language Data Processing group at The Rand Corporation; for each sentence of text the syntactic dependency of each word had been previously coded. A data retrieval program was applied, showing for each noun in text the number of occurrences (a) with at least one genitive noun dependent, (b) with at least one adjective dependent, and (c) with either type of dependent. A listing of all nouns in text (64,026 occurrences of 2,993 nouns) was prepared, ordered by frequency, and showing counts for a, b, and c above. Separate listings were prepared, showing for each noun that occurred 50 times or more the probability P that it would be modified in each of these three ways; these listings were ordered on P.

The data suggests, among others, the following conclusions: there is statistical significance in the variability with which nouns enter into the given combinations; the partial interchangeability of adjective and genitive noun modification is supported; a general correspondence exists between combinatorial groupings of nouns and morphological or semantic groupings (concrete nouns have low P for genitive complementation, abstract nouns have high P, etc); the use of words in a given field of discourse can be determined empirically (*e.g.,* the use of deverbative nouns either to indicate a process or the result of a process). It is suggested that the distributional approach is a useful supplement to traditional syntactic and semantic classification schemes, and that it is of direct utility in automatic parsing programs.

Connectability Calculations, Syntactic Functions, and Russian Syntax

David G. Hays

*Common Research Center, EURATOM, Ispra**

A program for sentence-structure determination can be divided into routines for analysis of word order and for testing the grammatical connectability of pairs of sentence members. The present paper describes a connectability-test routine that uses the technique called *code matching.* This technique requires elaborate descriptions of individual items, say the words or morphemes listed in a dictionary, but it avoids the use of large tables or complicated programs for testing connectability. Development of the technique also leads to a certain clarification of the linguistic concepts of *function, exocentrism,* and *homography.*

In the present paper, a format for the description of Russian items is offered and a program for testing the connectability of pairs of Russian items is sketched. This system recognizes nine dominative functions: subjective; first, second, and third complementary; first, second, and third auxiliary; modifying; and predicative.

* On leave from The RAND Corporation, 1962-63. The work reported in this paper was accomplished in part at RAND and completed at EURATOM. A fuller account of the connectability-test routine for Russian dominative functions is to appear as a EURATOM report.

The nature of a program for testing connectability with respect to coordinative functions (coordination, apposition, etc.) is suggested.

Punctuation and Automatic Syntactic Analysis*

Lydia Hirschberg

University of Brussels

In this paper we discuss how algorithms for automatic analysis can take advantage of information carried by the punctuation marks.

We neglect stylistic aspects of punctuation because they lack universality of usage and we restrict ourselves to those rules which any punctuation must observe in order to be intelligible. This involves a concept we call "coherence" of punctuation. In order to define "coherence", we introduce two characteristics, which we prove to be mutually independent, namely "separating power" and "syntactic function".

The *separating power* is defined by three experimental laws expressing the fact that two punctuation marks of different separating power prevent to a different extent syntactic links from crossing them. These laws are defined independently of any particular grammatical character of the punctuation marks or of the attached grammatical syntagms.

On the other hand, whichever grammatical system we choose, we may assimilate the punctuation marks to the ordinary words, to the extent that we can assign to them a known *grammatical character and function,* well defined in any particular context. They differ however from the other words by their large number of homographs and synonyms i.e. by the fact that almost every punctuation mark can occur with almost every grammatical value in each particular case, and in quite similar contexts.

The syntactic functions, in general, and in particular those of the punctuation marks, *can be ordered* according to an arbitrary scale of decreasing "value" of syntactic links, where the "value" of a link is directly related to the number of syntactic conditions the links must satisfy.

The law of coherence, then, shows that in a given context, a particular punctuation mark cannot indistinctly represent all its homographs, so that a certain number of assumptions about its syntactic nature and function can be discarded. This law can be stated as follows: "When moving from a punctuation mark to its immediate (left or right) neighbor in any text, the separating power cannot increase if the value of the syntactic function increases and vice-versa".

In addition we review two related topics, namely the stylistic character of punctuation and the necessity and existence of intrinsic criteria of grammatically, i.e. independent of punctuation. We propose such a criterion, and suggest a formalism related to the parenthesis free notation of logic.

* This investigation was performed under EURATOM contract No. 018-61-5-CET.B.

Application of Decision Tables to Syntactic Analysis

Walter Hoffman, Amelia Janiotis, and Sidney Simon

Wayne State University

Decision tables have recently become an object of investigation as a possible means of improving problem formulation of data processing procedures. The initial emphasis for this new tool came from systems analysts who were primarily concerned with business data processing problems. The purpose of this paper is to investigate the suitability of decision tables as a means of expressing syntactic relations as an alternative to customary flow charting techniques. The history of decision tables will be briefly reviewed and several kinds of decision tables will be defined.

As an example, parts of the predicative blocking routine developed at Wayne State University will be presented as formulated with the aid of decision tables. The aim of the predicative blocking routine is to group a predicative form together with its modal and temporal auxiliaries, infinitive complements, and negative particle, if any of these exist. The object of the search is to define such a syntactic block, but it may turn out instead that an infinitive phrase is defined or that a possible predicative form turns out to be an adverb.

Simultaneous Computation of Lexical and Extralinguistic Information Measures in Dialogue

Joseph Jaffe, M.D.

College of Physicians and Surgeons, Columbia University

An approach to the study of information processing in verbal interaction is described. It compares patterns of two indices of dispersion in recorded dialogue. The lexical measure is the mean segmental type—token ratio, based on 25-word segments of the running conversation. It is computed from a key punched transcript of the dialogue without regard to the speaker of the words. The extralinguistic measure is the H statistic, computed from the temporal pattern of the interaction. The latter is prepared from a two-channel tape recording by a special analogue to digital converter (AVTA system) which key punches the state of the vocal transaction 200 times per minute. Probabilities of the four possible states (either A or B speaking, neither speaking, both speaking) are the basis for the computation. All analyses are done on the IBM 7090. The methodology is part of an investigation of information processing in dyadic systems, aimed toward the reclassification of pathological communication.