English words from an analysis of the written form. This determination depends upon the ability to determine the number of graphic syllables in the word. It is natural, then, to speculate as to the nature of graphic syllabification and the relation of this phenomenon to the practice of word-breaking in dictionaries and style manuals.

It is not at all clear at the start that dictionary word-breaking is subject to any fixed structure. In fact, certain forms cannot be broken uniquely in isolation since the dictionary provides different forms depending upon whether the word is used as a noun or a verb. However, it is shown in this paper that letter strings can be decomposed into 3 sets of roughly the same size in the following manner: in the first, strings are never broken in English words; in the second, the strings are always broken in English words; and in the third, both situations occur. Rules for breaking vowel strings are obtained by a study of the CVC forms. Breaks involving consonants can be determined by noting whether or not the consonant string occurs in penultimate position with the final *c.* The final *e* in compounds also serves to identify the forms that are generally split off from the rest of the word.

A thorough analysis is made of the accuracy of the rules given when applied to the 12,000 words of the Government Printing Office Style Manual Supplement on word-breaking. Comparisons are also drawn between this source and several American dictionaries on the basis of a random sample of 500 words.

Writing of Chinese Recognition Grammar for Machine Translation

Ching-yi Dougherty

University of California, Berkeley

Our approach to this problem is based on the stratificational grammar outlined and the procedures proposed by Dr. Sydney Lamb. How the theory and the procedures can be applied to written Chinese is briefly discussed. For the time being our research is limited to the particular kind of written Chinese found in chemical and biochemical journals. First the Chinese lexes are classified by detailed syntactical analysis, then binary grammar rules are constructed for joining two primary or constitute classes. How a more and more refined classification can eliminate one by one the ambiguity resulting from all possible constructions arising from juxtaposition of two distributional classes is discussed in detail.

The Behavior of English Articles

H. P. Edmundson

Thompson Ramo Wooldridge Inc.

Machine translation has often been conceived as consisting of three steps: analysis of source-language sentence, transformation of analyzed pieces, and synthesis of target-language sentence. This paper is concerned with one aspect of the last step, namely, the rules of behavior of English articles. Since the classical definitions of definite and indefinite articles are operationally imprecise, proper mechanistic rules must be formulated in order to permit the automatic insertion or non-insertion of English articles. The rules discussed are of syntactic origin; however, note is also taken of their semantic aspects. This paper describes the methods used to derive these rules and offers ideas for further research.

On Representing Syntactic Structure

E. R. Gammon

Lockheed Missiles and Space Company

The idea of sentence depth of Yngve (A Model and an Hypothesis for Language Structure, *Proc. Am. Phil. Soc*., Vol. 104, No. 5, Oct. 1960) is extended to the notion of "distance" between constituents of a construction. The distance between constituents is defined as a weighted sum of the number of IC cuts separating them. Yngve's depth is then a maximum distance from a sentence to any of its words.

Various systems of weighting cuts are investigated. For example, in endocentric structures we may require that the distance from an attribute to the structure exceeds the distance from the head to the structure, and in exocentric structures that the distances from each constituent to the structure are equal.

Representations of constructions are considered which preserve the distance between constituents. It is shown that it is impossible to represent some sentences in Euclidean space with exact distances, but a representation may be found if only relative order is preserved. If more general spaces are used then exact distances may be represented. It follows that for a wide class of sentence types, there is a weighting, and a space, in which the distance preserving representations are identical with the diagrams of traditional grammar.

La Traduction Automatique et l'Enseignement du Russe

Yves Gentilhomme

Centre National de la Recherche Scientifique, Paris

Les recherches effectuées depuis quelques années en vue de la Traduction Automatique ont conduit à des méthodes de travail et à des résultats intéressant la pédagogie des langues.

Une expérience d'enseignement du russe a l'usage des scientifiques fondée sur ces données a été poursuivie pendant deux ans à Paris (Centre National de la Recherche Scientifique et Faculté des Sciences), et a abouti à la publication d'un manuel.

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.