ing morphosyntactic properties were analyzed within each language first and compared among the four languages:

1) the category of gender,
2) the category of animateness,
3) the category of case and number.

The product of this comparative analysis is a set of formation rules which embody a system for the identification of the inflected forms. The detailed result will be presented in an additional report.

Types of Language Hierarchy

E. D. Pendergraft

Linguistics Research Center, The University of Texas

Various relations lead to hierarchical systems of linguistic description. This paper considers briefly a typology of descriptive metalanguages based on such relations and sketches possible consequences for computational linguistics.

Its scope is accordingly limited to metalanguages having operational interpretations which specify individual linguistic processes and structural interpretations which specify language data of individual languages. Immediate-constituent, context-free metalanguages are used to illustrate hierarchical types.

Path Economization in Exhaustive Left-to-right Syntactic Analysis

Warren J. Plath

Computation Laboratory, Harvard University

In exhaustive left-to-right syntactic analysis using the predictive approach, each path of syntactic connection which originates at the beginning of a sentence must be followed until it is clear whether or not it will lead to the production of a well-formed analysis. The original scheme of following each path until it terminates either in an analysis or in a grammatical inconsistency has been considerably improved through the. incorporation of two path-testing techniques. Using the first technique, the program abandons a path as unproductive whenever a situation is detected where the prediction pool contains more predictions of a given type than can possibly be fulfilled by the remaining words in the sentence. Employment of the second technique, which is based on periodic comparison of the current prediction pool with pools formed on earlier productive paths, eliminates repeated analysis of identical right-hand segments which belong to distinct paths.

Taken together, the two path-testing procedures frequently enable the program to terminate the processing of a path well before its end has been reached. For most sentences, this means a considerable reduction in the total path length traversed, accompanied by a corresponding increase in the speed of analysis. Comparison of runs performed using both versions of the program indicates that employment of the new techniques reduces the average running time per sentence to less than one-fifth of its former value.

A Computer Representation for Semantic Information

Bertram Raphael

Computation Center, Massachusetts Institute of Technology

This paper deals with the problem of representing in a useful form, within a digital computer, the information content of statements in natural language. The model proposed consists of words and list-structure associations between words. Statements in simple English are thought of as describing relations between objects in the real world. Sentences are analyzed by matching them against members of a list of formats, each of which determines a unique relation. These relations are stored on description-lists associated with those words which denote objects (or sets of objects). A LISP computer program uses this model in the context of a simple question-answering system. Functions are provided which may grow, search, and modify this model. Formats and functions dealing with set-relations, part-whole and numeric relations, and left-to-right spatial relations have been included in the system, which is being expanded to handle other types of relations. All functions which operate on the model report information concerning their actions to the programmer, so that the applicability and limitations of this kind of model may more easily be evaluated.

Specifications for Generative Grammars Used in Language Data Processing

Robert Tabory

IBM Thomas]. Watson Research Center

It becomes more and more evident that successful pragmatics (i.e. automatic recognition and production procedures for sentences) cannot be performed without previously written generative grammars for the languages involved, using an underlying meta-theoretical framework proposed by the present school of mathematical linguistics. Two aspects of grammar writing are examined:

1. A taxonomy over the non-terminal vocabulary, using a subscripting system for signs and fitting into the more general string taxonomy of phrase structure components. The resulting more complex lexical organization is studied.

2. A command syntax for phrase structure components limiting the full, not necessarily needed generative power of these grammars. The proposed restrictions correspond to a priori linguistic intuition. Applicational order and location of the rules is studied.

Finally, the recognitional power and generative capacity of a computer are examined, the machine being structured according to a Newell-Shaw-Simon list system. It is well known that pushdown stores are particular cases of list structures, that context-free grammars

are particular cases of phrase structure grammars and that pushdown stores are the generative devices for context-free grammars.

Collecting Linguistic Data for the Grammar of a Language

Wayne Tosh

Linguistics Research Center, The University of Texas

Establishing the grammatical description of a language is one of the major tasks facing the technician in machine translation. Another is that of creating the system of programs with which to carry out the translation process. The Linguistics Research Center of The University of Texas recognizes the advantages in maintaining the specialties of linguistic research and computer programming as two separate areas of endeavor.

We regard the linguistic task as a problem in convergence. We do not expect ever to have a final description of a language (except theoretically for a given point in the history of that language). We do expect, however, to begin with almost immediate application of the very first grammatical description. We shall make repeated revisions of the grammar as we learn how to make it approximate better the language text fed into the computer.

The grammatical description of any one language is based primarily on specific text evidence. We are not attempting to describe "the language". We are, however, attempting to make descriptive decisions sufficiently general that new text evidence does not require extensive revision of earlier descriptions.

Corpora selected for description are chosen so as to have similar texts within the same scientific discipline for the several languages. Tree diagrams are drawn for each sentence in detail. The diagrams are inspected for consistency before corresponding phrase-structure rules are compiled in the computer. The grammar is then verified in the computer system and revised as necessary.

Derivational Suffixes in Russian General Vocabulary and in Chemical Nomenclature

John H. Wahlgren

University of California, Berkeley

A grammar based upon a conventional morphemic analysis of Russian will have a rather large inventory of derivational suffixes. A relatively small number of these recur with sufficient generality to acquire lexemic status (i.e., to be what is usually termed "productive"). Names of chemical substances in Russian may likewise be analyzed as combinations of roots or stems with derivational affixes, in particular, suffixes. The number of productive suffixes in the chemical nomenclature is considerably larger than in the general vocabulary. These suffixes derive from adoption into Russian of an international system of chemical nomenclature. A grammar of this system is basically independent of any grammar of Russian. It must, however, be consistently incorporated into the grammar and dictionary which are to serve in a machine translation system for texts in the source language containing chemical names.

Grammatical analysis of chemical suffixes and connected study of general Russian derivational suffixes has raised certain practical problems and theoretical questions concerning the nature of derivation. On the practical side, where a complex and highly productive system is involved, effective means of detecting and dealing with homography have required development. Theoretical consideration has been given to the question of grammaticality in chemical names and to problems of sememic analysis and classification of root and stem lexemes into tactic classes on the basis of co-occurrence with derivational suffixes.

On the Order of Clauses*

Victor H. Yngve

Department of Electrical Engineering and Research Laboratory of Electronics, Massachusetts Institute of Technology

We used to think that the output of a translation machine would be stylistically inelegant, but this would be tolerable if only the message got across. We now find that getting the message across accurately is difficult, but we may be able to have stylistic elegance in the output since much of style reflects depth phenomena and thus is systematic.

As an example, the order of the clauses in many two-clause sentences can be reversed without a change of meaning, but the same is not normally true of sentences with more than two clauses. The meaning usually changes when the clause order is changed. Equivalently, there appear to be severe restrictions on clause order for any given meaning. These restrictions appear to follow from depth considerations.

The idea is being investigated that there is a normal depth-related clause order and any deviations from this order must be signalled by special syntactic or semantic devices. The nature of these devices is being explored.

When translating multi-clause sentences, there may be trouble due to the fact that the clause types of the two languages are not exactly parallel. Therefore the list of allowed and preferred clause orders in the two languages will not be equivalent and the special syntactic and semantic devices available to signal deviations from the normal order will be different. Thus one would predict that multi-clause sentences in language A often have to be split into two or more sentences when translated into language B, while at the same time multi-clause sentences in language B will often have to be broken into two or more sentences when translating into language A.

* This work was supported in part by the National Science Foundation, in part by the U.S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research, and in part by the National Bureau of Standards.