Sentence Structure Diagrams

Susumu Kuno

Computation Laboratory, Harvard University

A system for automatically producing a sentence structure diagram for each analysis of a given sentence has been added to the program of the multiple-path syntactic analyzer. A structure code, consisting of a series of structure symbols or phrase markers that identify the successive higher-order structures to which the word in question belongs, is assigned to each word of the sentence. The set of structure codes for the words of a given sentence is equivalent to an explicit tree diagram of the sentence structure, but more compact and easier to lay out on conventional printers.

The diagramming system makes some experimental assumptions about the dependencies of certain structures upon higher-level structures. All the major syntactic components of a sentence (i.e., subject, verb, object, complement, period, or question mark) are represented in the current system as occurring on the same level, all being dependent on the topmost level, "sentence". A floating structure such as a prepositional phrase or adverbial phrase or clause, whose dependency is not determined in the analyzer, is represented as depending upon the nearest preceding structure modifiable by such a floating structure. Different assumptions as to structural dependencies would yield different diagrams without requiring modification on the main flow of the diagramming program.

The diagrams thus obtained contribute greatly to the rapid and accurate evaluation of the analysis results, and they are also useful for obtaining basic syntactic patterns of analyzed structures, and for detecting the head of each identified structure.

Linguistic Structure and Machine Translation

Sydney M. Lamb

University of California, Berkeley

If one understands the nature of linguistic structure, one will know what design features an adequate machine translation system must have. To put it the other way around, it is futile to attempt the construction of a machine translation system without a knowledge of what the structure of language is like. This principle means that if someone wants to construct a machine translation system, the most important thing he must do is to understand the structure of language.

Any MT system, whether by conscious intention on the part of its creators or not, is based upon some view of the nature of linguistic structure. By making explicit the underlying theory for various MT systems which have been proposed we can determine whether or not they are adequate. Similarly, by observing linguistic phenomena we can determine what properties an adequate theory of language must have, and such determination will show what features an MT system must have in order to be adequate.

It can be shown that some of the approaches to MT now being pursued must necessarily fail because their underlying linguistic theories are inadequate to account for various well-known linguistic phenomena.

On Redundancy in Artificial Languages

W. P. Lehmann

Linguistics Research Center, The University of Texas

Artificial languages are one concern of work in computational linguistics, if only as a mnemonic device for interlinguas which will be developed. Even if it does not gain wider use, the structure of an artificial language is of general interest.

In contrast to the artificial languages which have been widely proposed, linguistic principles underlying a well-designed artificial language and its usefulness are well-established, particularly through Trubetzkoy's article, TCLP 8.5-21. which indicates phonological limitations for such a language. Since Trubetzkoy's specifications yield a total of approximately 11,000 morphemes, if an artificial language incorporated the degree of redundancy found in natural languages it would be severely handicapped by the size of its lexicon. The paper discusses the problem particularly with regard to suprasegmentals, which Trubetzkoy almost entirely ignored.

A Procedure for Automatic Sentence Structure Analysis

D. Lieberman

IBM Thomas }. Watson Research Center

The two main considerations in the design of this procedure were the economical recognition and representation of multiple readings of syntactically ambiguous sentences, and general applicability to "all" languages (English, Russian, Chinese). The following features will be discussed: types of structural descriptions, form of linguistic rules, use of linguistic heuristics to achieve economical multiple analyses, application to linguistic research and application to production MT systems. Also, the relation between this procedure and other existing sentence analysis procedures will be discussed.

An Algorithm for the Translation of Russian Inorganic-Chemistry Terms

L. R. Micklesen and P. H. Smith, Jr.

IBM Thomas J. Watson Research Center

An algorithm has been devised, and a computer program written, to translate certain recurring types of inorganic-chemistry terms from Russian to English. The terms arc all noun-phrases, and several different types of such phrases have been included in the program. Examples are: