Design of a Generalized Information System

Ronald W. Jonas

Linguistics Research Center, The University of Texas

While mechanical translation research involves the design of a computer system which simulates language processes, there is the associated problem of collecting the language data which are to be used in translation. Because large quantities of information will be needed, the computer may be useful for data accumulation and verification.

A generalized information system should be able to accept the many types of data which a linguist encodes. A suitable means of communication between the linguist and the system has to be established. This may be achieved with a central input, called Linguistic Requests, and a central output, called Information Displays. The requests should be coordinated so that all possible inputs to the system are compatible, and the displays should be composed by the system such that they are clearly understandable.

An information system should be interpretive of the linguist's needs by allowing him to program the data manipulation. The key to such a scheme is that the linguist be permitted to classify his data freely and to retrieve it as he chooses. He should have at his disposal selecting, sorting, and displaying functions with which he can verify data, select data for introduction to a mechanical translation system, and perform other activities necessary in his research.

Such an information system has been designed at the Linguistics Research Center of The University of Texas.

Some Experiments Performed with an Automatic Paraphraser

Sheldon Klein

System Development Corporation

The automatic paraphrasing system used in the experiments described herein consisted of a phrase structure, grammatically correct nonsense generator coupled with a monitoring system that required the dependency relations of the sentence in production to be in harmony with those of a source text. The output sentences also appeared to be logically consistent with the content of that source. Dependency was treated as a binary relation, transitive except across most verbs and prepositions.

Five experiments in paraphrasing were performed with this basic system. The first attempted to paraphrase without the operation of the dependency monitoring system, yielding grammatically correct nonsense. The second experiment included the operation of the monitoring system and yielded logically consistent paraphrases of the source text. The third and fourth experiments demanded that the monitoring system permit the production of only those sentences whose dependency relations were non-existent in the source text. While these latter outputs were seemingly nonsensical, they bore a special logical relationship to the source. The fifth experiment demanded that the monitoring system permit the production of sentences whose dependency relations were the converse of those in the source. This restriction was equivalent to turning the dependency tree of the source text upside down. The output of this experiment consisted only of kernel type sentences which, if read *backwards,* were logically consistent with the source.

The results of these experiments determine some formal properties of dependency and engender some comments about the role of dependency in phrase structure and transformational models of language.

Interlingual Correspondence at the Syntactic Level*

Edward S. Klima

Department of Modern Languages and Research Laboratory of Electronics, M.l.T.

The paper will investigate a few major construction types in several related European languages: relative clauses, attributive phrases, and certain instances of coordinate conjunction involving these constructions. In each of the languages independently, the constructions will be described as resulting from syntactic mechanisms further analyzable into chains of partially ordered operations on more basic structures. Pairs of sentences equivalent in two languages will be examined. Sentences will be considered equivalent if they are acceptable translations of one another. The examples used will, in fact, be drawn primarily from standard translations of scholarly and literary prose. Equivalence between whole sentences can be further analyzed, as will be shown, into general equivalence 1) between the chains of operations describing the constructions and 2) between certain elements (e.g., lexical items) in the more basic underlying structures. It will be seen that superficial differences in the ultimate shape of certain translation pairs can be accounted for as the result of minor differences in the particular operations involved or in the basic underlying structure. We shall examine two languages (e.g., French and German) in which attributive phrase formation and relative clause formation on the whole correspond and in which, in a more or less abstract way, the rules of relative clause formation are included as intermediate links in the chain of operations describing attributive phrases. The fact that in particular cases a relative clause in the one language corresponds to an attributive phrase in the other will be found to result from, e.g., differences in the choice of perfect auxiliary in the two languages.

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

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: