XIII. MECHANICAL TRANSLATION*

V. H. Yngve
Carol M. Bosche
Elinor K. Charney
Ann Congleton
J. L. Darlington
D. A. Dinneen
G. H. Harman
Renate I. Heil
Phyllis Hirshfang
Muriel S. Kannel
E. S. Klima
K. C. Knowlton
J. D. McCawley
T. More, Jr.
W. K. Percival
A. C. Satterthwait
Rosemarie Sträussnigg
M. R. Quillian

A. SENTENCE PARSING WITH A SELF-ORGANIZING HEURISTIC PROGRAM

Self-organizing heuristic search procedures for automatic sentence parsing have been studied. Experimental results demonstrate that a statistically guided program can produce correct parsings even with a grammar involving extreme overgeneralization, that strategies exist for producing these parsings efficiently, and that these strategies can be developed automatically from a training sequence of sentences and their correct parsings.

A computer program has been developed which learns to parse correctly and efficiently on the basis of a training sequence of sentences and their correct parsings. Its strategy of search for the grammatical structure of a sentence is based upon a set of treelike patterns representing local configurations of nodes and words in correctly parsed sentences. Two numbers are associated with each pattern, one showing how many times the indicated configuration has been observed in correct parsings, and the other showing how many times the configuration has been constructed during search. The ratio of the first number to the second is taken as the estimated probability, according to the pattern, that the connection resulting in the indicated configuration is correct. At any point in the search tree, the strongest such probability estimate determines the next connection for the partial structure corresponding to that point; hence it determines the next move from that point in the tree. The strategy used for determining the point from which to proceed is always to generate that partial structure with the highest average of probability estimates for all of its connections. This strategy is in effect a method of repeated reconsideration as to whether the last move made leads toward the right answer, or whether an alternate move from some other point in the search tree may be better.

The parsing program, coded in the COMIT language for the IBM 7090 digital computer, was used in an experiment with 300 sentences of Basic English. The machine attempted to parse each of these sentences on the basis of experience gained from all previous attempts and their hand-parsed correct answers. The grammar used could have allowed a very large number of incorrect parsings, because of the overgeneralizations implied by its rules, but only the first parsing found was taken as the machine's answer.

---

*This work was supported in part by the National Science Foundation (Grant G-24047).

The results indicate that a simple grammar can be used effectively for parsing, that an efficient search procedure can be based upon a set of treelike patterns and associated parameters, and that such a set of patterns and parameters can be developed automatically from a training sequence of sentences and parsings. In the experiment with 300 sentences, few nonsensical parsings resulted; an incorrect parsing almost always represented either an ambiguity of English which is very difficult to resolve on the basis of syntax even with a "good" grammar, or a difficulty which could have been resolved by a simple mechanized procedure that does not have to fight the battle of a long combinatorial search. Even better performance is to be expected from a slightly modified search strategy and from an extended set of patterns; modifications are also suggested for the self-organization process.

The results of the study suggest a design for a practical parsing mechanism. It consists essentially of: (a) a linear array of patterns ordered according to the estimated probability that they do the right thing if they apply to a partial construction that is right thus far, and (b) a strategy for selecting from the already produced partial structures one that will probably lead to the correct parsing because the next step of the construction maintains a low average "information content" per connection.

Techniques similar to those developed may be useful in other structure-building areas in which problems consist of completing structures in "typical" ways, and for which the number of possible structures is so large that useful statistics must be statistics not on structures but on their properties.

K. C. Knowlton

B. GERMAN WORD-ORDER RULES

Most German sentences are matched by other sentences in which the same words are arranged in a different order but in such a way that the meaning remains the same except for stylistic overtones. Thus the sentence Ich habe das Buch gelesen 'I have read the book' is matched by Das Buch habe ich gelesen, and Gelesen habe ich das Buch.

I propose to handle these sentences by choosing one permutation to represent the whole set of sentences, generate this permutation in the normal way, and then generate the others by means of special transformational rules operating on this basic type. These special transformational rules will have to proceed by first moving up some portion of the sentence into first position, and then permuting the subject noun phrase and the finite verb. In order to carry out the first part of the operation, however, the rules will have to specify what portions of the sentence are preposable. In the example quoted above there are two such preposable portions: the word gelesen and the phrase das Buch. Notice that in the basic sentence these two portions are contiguous, do not overlap, and make up all of the sentence except the subject and finite verb. Facts of this kind have

prompted grammarians to assume that all sentence constituents other than the subject and the finite verb are preposable. This is unfortunately not the case.

The restrictions on the operation of the prepose rule are in fact quite extensive, some of them of a trivial nature, others not. If we take, for example, the sentence Ich habe meinen Bruder schon zweimal gesehen, 'I have seen my brother twice', we discover that there are four preposable portions: meinen Bruder, schon zweimal, zweimal, and gesehen. We note that one of these portions (zweimal) is contained in one of the others (schon zweimal), and one word in the basic sentence is not preposable, viz. schon. In the sentence Ich weiss es 'I know it', there are no preposable portions, although it is usually possible to prepose an object pronoun (e.g., Ich kenne ihn 'I know him' - Ihn kenne ich). Other sentence constituents occur only preposed. Thus the so in So bin ich überzeugt, dass er da ist. 'So I'm convinced that he's there', must appear in first position. In such a case we have no basic sentence from which to derive the sentence with the preposed so, or alternatively we must regard the prepose rule in some cases as obligatory, a solution that is clearly counterintuitive. One restriction on the operation of the prepose rule is of an unusual kind and will be described at some length. The argument will necessitate a digression into the area of semantics.

There are certain ambiguous noun phrases in German, and in other European languages, exemplified by the English phrase my sister. If we compare the sentence My sister has gone to New York with the sentence That woman over there is my sister, we notice the following fact. The phrase my sister in the first sentence refers to a very specific person, whereas in the second sentence one obvious interpretation is to make the phrase my sister merely refer descriptively to a characteristic of the 'woman over there', or more exactly to a relationship between her and the speaker of the utterance. I shall say, therefore, that the phrase my sister can be used either referentially or descriptively; in the first sentence, My sister has gone to New York, we have an instance of its referential use, and in the second sentence, That woman over there is my sister, an instance of its descriptive use.

One other type of ambiguity is relevant to our problem, and this is exemplified in the English sentence Thursday is the fifth day of the week. There are two rather obvious interpretations of this sentence. In the first the sentence would supply information about Thursday, and in the second it would supply information about the fifth day of the week. Or to put the matter a little differently, we can imagine this sentence as an answer to the question What is the fifth day of the week? or as an answer to the question What is Thursday? If we reverse the order of the noun phrases, i.e., if we say The fifth day of the week is Thursday, a similar ambiguity appears.

We shall now consider analogous sentences in German, and observe their behavior under the prepose rules. For this purpose let us use the following:

(1) Meine Schwester ist das tüchtigste Mädchen in der Klasse.

(2) Das tüchtigste Mädchen in der Klasse ist meine Schwester.

We shall number the interpretations in each case a and b, and adopt the convention of saying that there are as many sentences as there are interpretations. Thus we are dealing with four sentences, not two.

Let us run through the four sentences and recall their interpretations. Sentence (1a) Meine Schwester ist das tüchtigste Mädchen in der Klasse, is an answer to the question Wer ist das tüchtigste Mädchen in der Klasse?, whereas sentence (1b) of the same shape answers the question Wie ist es mit deiner Schwester? or the like. Sentence (2a) Das tüchtigste Mädchen in der Klasse ist meine Schwester is, as (1a) is, an answer to the question Wer ist das tüchtigste Mädchen in der Klasse?, sentence (2b) being, on the other hand, an answer to the question Wie ist es mit dem tüchtigsten Mädchen in der Klasse? or some such similar question. The situation, in other words, is entirely parallel to the English examples that we discussed.

But now let us examine the phrasal ambiguity of meine Schwester. The descriptive use of the phrase occurs only once, namely in (2b), whereas in the other three sentences the phrase is used referentially. It is also the case that while the referential phrase occurs both as initial noun phrase and as noninitial noun phrase (initially in (1a) and (1b), and noninitially in (2a)), the descriptive phrase occurs only in the noninitial position.

Let us now examine the operation of the prepose rules on these sentences, concentrating for this purpose on (1b) and (2b). Sentence (1b) appears after the transformation has been applied as Das tüchtigste Mädchen in der Klasse ist meine Schwester. This sentence has the same phonemic shape as (2), but differs in that the sentence stress occurs on Klasse. (In other words, the sentence stress appears on the same word in the transformed sentence as in the basic sentence.) Let us recall that this sentence conveys information about a particular person designated by the phrase meine Schwester. In other words, the phrase meine Schwester is being used referentially. Notice, moreover, that the referential phrase appears in the noninitial position. However, this fact is not in any way surprising, since we were already aware of the fact that the referential phrase can equally well occupy initial or noninitial position. We recall that it was the descriptive phrase that, thus far, has only appeared in one position, namely noninitially. But now the following problem arises. If we were to take (2b) and apply the same transformation, would we not end up with a sentence in which the descriptive phrase meine Schwester appears in initial position? This would in fact be the case, if such a sentence existed. However, it appears that no such sentence occurs. There is no sentence, that is, of the shape Meine Schwester ist das tüchtigste Mädchen in der Klasse such that the phrase meine Schwester is used descriptively and the sentence as a whole is an answer to the question Wie ist es mit dem tüchtigsten Mädchen in der Klasse?

There is, therefore, a rule in German which requires that descriptively employed noun phrases occur only in noninitial position in the sentence, and that noun phrases occurring in initial position should be interpreted referentially. Moreover, this rule has priority over the type of transformational rules that we proposed to establish to take care of the preposing of sentence elements. Can it legitimately be said, therefore, that the semantic rule acts as a restriction on the transformational rules? Clearly, that would be the case only if we had some way of referring to the restriction purely in terms of the symbols in the string to be transformed and their derivational history. But, as far as is known at present, semantic distinctions of the kind with which we are concerned here would not be likely to show up in the derivational history, nor would they be represented in the string of symbols constituting the structural description of the sentences in question.

W. K. Percival

C. ARABIC TO ENGLISH TRANSLATION

Program AE has been written for the mechanical translation of a limited corpus of Arabic into English. This program is in the final stages of debugging. The Arabic corpus that the IBM 7090 computer is capable of translating under control of the program is defined by a restricted sentence-construction grammar of Arabic[1] according to the theory of grammar proposed by V. H. Yngve.[2]

1. Sentence-Construction Grammar

Program A which includes the sentence-construction grammar of Arabic enables the computer to compose a series of Arabic sentences described by the grammar. These sentences are used to test the validity of the grammar and the capability of the translation program AE.

The sentences constructed by the computer under control of program A are always verbal, declarative statements, each limited to one singly transitive, imperfect, indicative, active verb. The noun phrases contain no constructs or pronominal suffixes. All nouns are animate, referring only to persons. The verbs are either sound, hollow or doubled.

The Arabic produced by program A and translated by program AE is represented in a COMIT-symbol transliteration of the strictly consonantal Arabic orthography without indication of vowels or other diacritical marks. Figure XIII-1 furnishes an example of such a sentence produced by program A followed by a phonemic transcription and the translation which is to be expected from program AE.

YKATB QLYLA ALAN ALM+WRXH ALYWNANYH ALM+WRXAT ALTVBANAT ALKBYRH ALSMYNAT TLK HNAK.

/yukaatibu qaliylani l ?aana l mu?arrixatu l yuwnaaniyyatu l mu?arrixaati t ta9baanaati l kabiyrata s samiynaati tilka hunaak./

The Greek historian corresponds with those big, fat, tired historians there a little now.

Fig. XIII-1. A computer-produced Arabic sentence with phonemic interpretation and English translation.

2. Mechanical Translation Program

The mechanical translation program is an application of Yngve's framework for syntactic translation.[2] Ideally, the translation procedure involves three stages: analysis or recognition, structural transfer, and synthesis or construction.

a. Analysis

The first stage is realized through the application of a subprogram R of program AE to any input sentence that has been produced by program A. Under the guidance of this subprogram, the computer produces a grammatical analysis of the input sentence. Figure XIII-2 furnishes a very much abbreviated example of such an analysis rewritten in two-dimensional form.

Subprogram R contains the sentence-construction grammar expressed in terms of a set of expected forms and structures. It directs the computer to compare the items of the input sentence with the predicted forms. These forms are identified as grammatical elements. The grammatical elements, in turn, are compared with expected constructions. Sets of elements which match predicted constructions are identified as grammatical elements of higher levels.

The recognition procedure is divided into two steps, morphological (above the dotted line in Fig. XIII-2) and syntactic (below the line). In the morphological phase each word is analyzed in turn during a single left-to-right sweep.

Three major steps may be taken in the analysis of a word. (i) If the word is found to coincide with a dictionary entry, it is given a grammatical tag and the morphological analysis is complete. In Fig. XIII-2 H+DA = DEM (demonstrative) and DAXLA = ADVERB furnish examples of morphological analysis completed at this step. (ii) If the item does not coincide with any dictionary entry, the rightmost letter is deleted and the remainder is found to coincide with a stem or prefix listed in the dictionary. A series of subroutines then identifies the remainder as a set of suffixes and/or a stem. If the

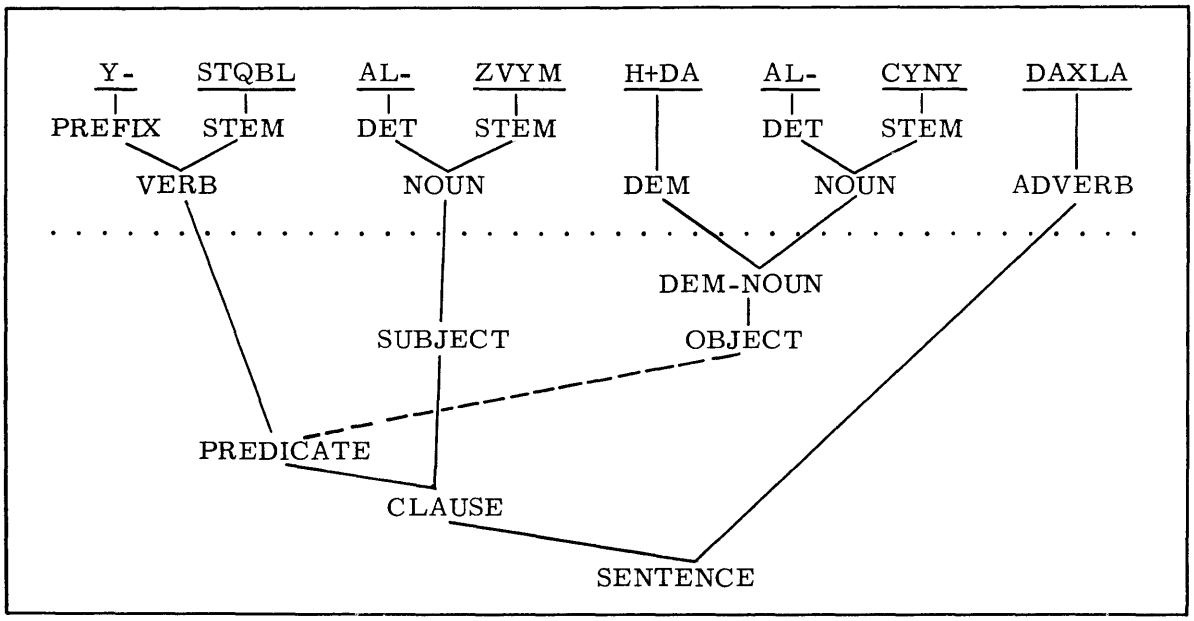

Fig. XIII-2. Illustration of a simplified two-dimensional grammatical analysis of an input sentence as produced by the recognition subprogram R. "The leader meets this Chinese inside."

remainder does not coincide with any entry in the dictionary, step (ii) is repeated until the recognition is achieved. In Fig. XIII-2 the prefixes Y- and AL- with the stems STQBL and ZVYM furnish examples of successful identification at this stage. (iii) If the word remains unidentified by either step (i) or step (ii), a subroutine seeks any one of a number of broken plural substantive affixes. If such an affix is identified, the possible canonical forms of substantive stems are hypothesized and formed. The subroutines for the identification of suffixes are activated after one or more such stems are found to coincide with dictionary entries.

The syntactic recognition phase is based upon a general characteristic of the sentence-construction grammar. The sentence is described as a set of constructions arranged in specified relations to each other. Each construction is composed of one or two constituents. A constituent may be terminal such as STQBL or nonterminal such as VERB (Fig. XIII-2). If the constituent is nonterminal, it identifies a construction. For example, in Fig. XIII-2 CLAUSE is a constituent of the construction identified by SENTENCE. In turn it identifies the construction PREDICATE+SUBJECT.

Constructions may be described as included within other constructions. For example, the construction identified by PREDICATE may be said to be included within the construction identified by CLAUSE. All constructions are included within the construction identified by SENTENCE. One construction may be described as "more included" than another. For example, CLAUSE is included within SENTENCE and PREDICATE is included within both CLAUSE and SENTENCE. PREDICATE is,

therefore, described as being "more included" than CLAUSE.

Not all constructions bear the relation of inclusion to each other. For example, within the scope of the present grammar, the construction identified by VERB is neither included within nor does it include the construction identified by SUBJECT. The sentence-construction grammar indicates whether or not any two constructions have the relation of inclusion and if so what that relation is.

Syntactically most-included constructions are defined as those syntactic constructions that include no other syntactic constructions. In Fig. XIII-2 DEM-NOUN identifies a situation that illustrates what is meant by a most-included construction. This construction consists of DEM+NOUN. DEM and NOUN are both words, and so identify no further syntactic constructions.

The syntactic recognition program identifies the most-included constructions first. It then identifies the next most-included constructions in turn until a sentence has been identified.

In Fig. XIII-2 adjectives as attributes of nouns are first sought. These are not found. Then demonstrative adjectives are sought. H+DA is found and identified as a grammatically possible member of a DEM-NOUN construction. The constituents of a clause are next sought. VERB+NOUN+DEM-NOUN are recognized as possible constituents. Case, number, gender, person and relative position within the sentence are then examined to determine a grammatically compatible interpretation of the three items as constituents of a CLAUSE. CLAUSE+ADVERB are recognized as two constituents of a sentence and the analysis is complete.

b. Synthesis

The third stage of program AE consists of a subprogram E which includes a sentence-construction grammar of English. Program E is similar to program A in that it enables the computer to compose series of English sentences described by the grammar.

The grammar is composed of obligatory rules and multiple-choice rule sets (Fig. XIII-3). The multiple-choice rule set is composed of all the rules in the grammar

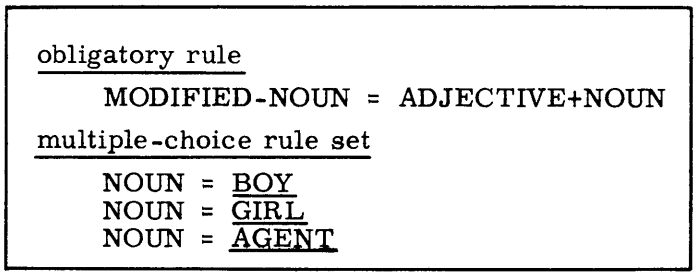

Fig. XIII-3. Examples of an obligatory rule and a multiple-choice rule set.

the left sides of which are identical. The expansion of a constituent identical with the left side of a multiple-choice rule set is effected by the selection and expansion of one of the rules of the set. Sentences produced by the sentence-construction grammar vary only as the selection of alternative multiple-choice rules varies.

C. Structural Transfer

Subprogram ATE includes a set of structural-transfer rules called the structural-transfer grammar. Ideally, this subprogram should be located between subprograms A and E in the mechanical translation program AE. At present the rules of the executive routine for subprogram ATE and the structural-transfer grammar are distributed through the other two subprograms.

The function of the subprogram ATE is the identification and arrangement in order of application of the specific multiple-choice rules with which subprogram E may produce the output sentence most nearly equivalent in meaning to the input sentence.

The specific output sentence that is equivalent to the input sentence is identifiable through reference to the input sentence and its analysis, together composing the specifier of the input sentence. This output sentence may be defined by reference to the sentence-construction grammar that produces it and an ordered list of multiple-choice rules. The obligatory rules are implied by the grammar and need not be mentioned.

The specification of the multiple-choice rules is effected by searching the specifier of the input sentence for features determined to be significant for the definition of the output sentence equivalent to it. When a significant feature is found, the multiple-choice rules that this feature specifies are selected and arranged in order. An example of a structural-transfer rule follows.

$$\text{MN(N(+TBYB);AJS(AJ(XAC)))} \longrightarrow \text{ADJECTIVE=PERSONAL}$$

The rule above may be interpreted as follows. If the sequence of letters X A C has been interpreted as an adjective AJ and if it is a constituent of an adjective string AJS, which in turn is a constituent of a modified noun phrase MN another of the constituents of which is a string of letters +T B Y B which has been interpreted as a noun N, then the multiple-choice rule ADJECTIVE=PERSONAL in subprogram E is specified. In other environments the letter-string X A C may be equivalent to the English adjective 'special', the English modified noun 'special official' or it may be only part of the stem of the Arabic noun A+SXAC /ʔaš̆xaaṣ/ 'persons'.

A. C. Satterthwait

References

1. A. C. Satterthwait, Parallel Sentence-Construction Grammars of Arabic and English, Mechanical Translation Group Note, Research Laboratory of Electronics, M.I.T., May 1962.

2. V. H. Yngve, A framework for syntactic translation, Mechanical Translation, Vol. 4, No. 3, pp. 59-65, 1957.