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 1 ? aana 1 mu? arrixatu 1
yuwnaaniyyatu l mu?arrixaati t ta9baanaati 1
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
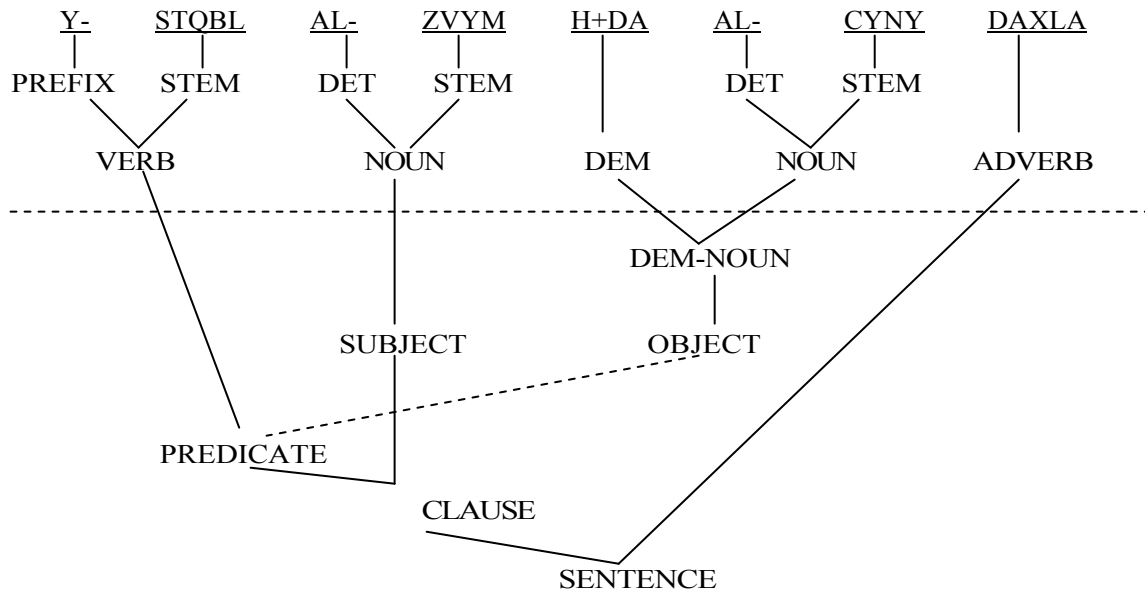
Fig. XIII-2. Illustration of a simplified two-dimensional grammatical
analysis of an input sentence as produced by the recognition
subprogram R. "The leader meets this Chinese inside."

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 sub-
routines 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 con-
struction 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 construc- tions 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 con- struction 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 constitu- ents.   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

---

<div style="border:1px solid black;">

obligatory rule

    MODIFIED-NOUN = ADJECTIVE+NOUN

multiple-choice rule set

    NOUN = BOY
    NOUN = GIRL
    NOUN = AGENT

</div>

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 sub- program 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 iden- tifiable through reference to the input sentence and its analysis, together com- posing 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.

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

The rule above may be interpreted as follows.   If the sequence of letters $\underline{X}$ $\underline{A}$ $\underline{C}$ 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 $\underline{+T}$ $\underline{B}$ $\underline{Y}$ $\underline{B}$ which has been interpreted as a noun N, then the multiple-choice rule ADJECTIVE = $\underline{PERSONAL}$ in subprogram E is specified.   In other environments the letter-string $\underline{X}$ $\underline{A}$ $\underline{C}$ 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 $\underline{A+SXAC}$ /? aṡxaaṣ/ 'persons'.

<div align="right">A. C. Satterthwait</div>

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.