[Mechanical Translation and Computational Linguistics, vol. 8, No. 2, February 1965]

Sentence-For-Sentence Translation: An Example*

by Arnold C. Satterthwait, Computing Center, Washington State University

A *computer program for the mechanical translation into English of an infinite subset of the set of all Arabic sentences has been written and tested. This program is patterned after Victor H. Yngve's framework for syntactic translation. The paper presents a generalized technique for thorough syntactic parsing of sentences by the immediate constituent method, a generalized structural transfer routine, and a consideration of the elements which must be included in a statement of structural equivalence with examples drawn from such a statement and the accompanying bilingual dictionary. Yngve's mechanism for the production of sentences is expanded by the introduction of a stimulator which brings stimuli external to the mechanism into effective participation in the construction of specifiers for the production of sentences. The paper includes a discussion of the requirement that a basic vocabulary for the output sentence be selected in the mechanical translation process before the specifier of that sentence is constructed. The procedure for the morphological parsing of Arabic words is also presented. The paper ends with a brief discussion of ambiguity.*

Introduction

The research discussed in this paper has resulted in the preparation of a working computer program which is the first example of sentence-for-sentence mechanical translation applying Victor Yngve's process. Of this process Yngve has written,

> Translation is conceived of as a three-step process: recognition of the structure of the incoming text in terms of a structural specifier; transfer of this specifier into a structural specifier in the other language; and construction to order of the output text specified.[1]

Yngve's process requires a grammar of the input language and a recognition routine, a statement of structural equivalence between the two languages and a structural transfer routine, and finally a grammar of the output language and a construction routine.

The present program causes the computer to prepare in the English sentence-construction subroutine sets of orders which direct the execution of the rules of an English sentence-construction grammar. The computer produces that specific sentence which is equivalent to any Arabic sentence selected from an infinite subset of the set of all Arabic sentences and submitted to the computer for translation.

Before the production of the sets of orders for the construction of the output sentence, the computer under control of the recognition subroutine makes a thorough morphological and syntactic analysis of any Arabic sentence selected from the subset. This analysis is compared with the rules in the statement of structural equivalence. As a result of this comparison and subsequent operations, the specific orders which will produce the English sentence equivalent to the Arabic are selected.

Yngve's theory[2] develops a context-free phrase-structure grammar which provides for the production of discontinuous constituents in the sentence-construction grammar and for their recognition in the sentence-recognition grammar. Details of the theory for the sentence-construction grammar as developed for the mechanical translation program presented here, the structure of the rules and so on are fully discussed in my first report.[3]

The sentences which the computer under control of the current program will translate are drawn from the subset of Arabic sentences which the Arabic sentence-construction grammar described previously is capable of producing.[3] The procedure by which a sampling of these computer-constructed sentences were tested for grammaticality is discussed at some length in "Computational Research in Arabic".[3a]

The computer will also translate any sentence composed by a human under restrictions of the rules following. These rules are in terms of traditional Arabic grammar and are not to be considered a linguistic description of the power of the translation program. 1) The sentence must be a simple statement, verbal (i.e. a *jumlah fi'līyah),* limited to one singly-transitive verb and one mark of punctuation, the period. 2) Grammatical categories set the following restrictions, a) Forms which include number category must be either singular or plural. (The program does not yet recognize duals.) b) Only imperfect, indicative, active forms of the verb may occur. c) Noun phrases may not contain constructs *(idāfāt)* or pronominal suffixes.

* This work was supported in part by the National Science Foundation: in part by the U.S. Army, the Air Force Office of Scientific Research, and the Office of Naval Research; and in part by the Research Laboratory of Electronics, Massachusetts Institute of Technology.

Research has been undertaken to explore problems dealing with syntactic and morphological structures rather than with problems of vocabulary. For this reason emphasis has been placed on a proliferation of structures which the program will translate rather than on the amassing of vocabulary. The vocabulary which the program recognizes is, therefore, small and limited to the items shown on pages 16 and 17.

The vocabulary was selected so that problems involving points of morphological analysis in Arabic, morphological and syntactic constructions in English, multiple meanings, idioms, orthography, etc. might be investigated. The program has translated over 200 sentences exemplified by the following:

Composed by an Arab:

اليوم يزور هذ ه الحرمه ز لك المحامى الكبير هنا .

'That big lawyer visits this woman here today.'

Constructed by computer:

يخون الا ن الحريم خارجا الجهال الثوريون هوءلا .

'These revolutionary children betray the women outside now.'

In Yngve's process the two grammars of the mechanical translation program with their routines are presented as units each of which may be operated independently of the other and of the structural transfer routine. While the present program does not maintain this autonomy between the three sub-programs, it is strongly indicated that such autonomy is both practically attainable and economically desirable. It is our intention, therefore, to make the changes in the program necessary to effect this independence.

Independence of the three subprograms has a number of implications. The input sentence remains intact, in order and form, as it does in the present program. The only changes which are made are in the form of added elements making grammatical information explicit. As the analysis is completely independent of the target language, the sentence-recognition grammar is expected to be usable for translation from the source language into any target language. The program which incorporates the sentence-construction grammar of the target language is written independent of reference to any source language. This portion of the program should, therefore, be usable for translation from any source language into the target language. The structural transfer section, due to its role as interpreter of two specific languages, must be rewritten for each pair of languages to be translated.

The Input

Modern Arabic is written with an alphabet of twenty-eight letters, punctuation marks and a set of diacritics. The diacritics symbolize vowels, mark length of vowels and consonants, and indicate elision. These marks rarely appear in journals and newspapers. The system of transliteration used in the program and the remainder of this paper is presented in my first report. As the diacritics are not represented in this system, the orthography is composed solely of consonants and marks of punctuation.

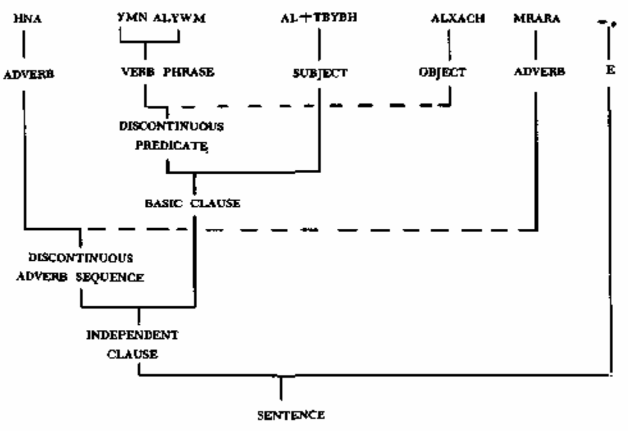


FIGURE 1.
Guide to the complete mechanical syntactic analysis of the sentence /hunaa yamunnu 1 yawma t tabiybatu 1 xaassata miraaran./ (cf. Figure 2). Word-for-word translation: Here he-weakens today the-physician-(feminine) the-special-officials-(masculine) at-times. Computer translation: The physician weakens the special officials here at times today.

While, at present, material intended for mechanical translation is punched on cards, economy will finally demand that most material be read automatically. The major problem in the automatic reading of Arabic will be the mechanical determination of word-division. The present program operates on the assumption that this problem has been solved.

In Arabic printing the letters of a word are characteristically joined and as in English handwriting the last letter of a word is not joined to the first letter of the following word. Unlike English, however, several letters in Arabic printing are not joined to following letters even within the same word. A break between two letters, the first of which is one of these “separate letters,” does not in itself constitute an indication of word-division. In careful handwriting intervals of two different lengths between unjoined letters are frequently observed. The longer interval indicates word-division. This distinction in the length of the interval is often, however, not observed in handwriting and sometimes is not observed even in printed matter. The magnitude of the problem that failure to identify word-division by spacing will present to automatic reading will require further investigation. It appears quite possible at the present time, however, that word-division may have to be determined morphologically rather than orthographically.

Verbs

guide	يَدُلّ	betray	يَخُون	hate	يَكْرَه
cheat	يَغُشّ	adorn	يَزِين	raise	يُرَبِّب
raise	يَرُبّ	disturb	يَرِيب	hear	يَسْمَع
ill-treat	يَزُرّ	fear	يَخَاف	awaken	يُحَسِّس
slaughter	يَحُسّ	visit	يَزُور	reproach	يُمَنِّن
hide	يَكُنّ	insult	يَحُوس	meet	يَسْتَقْبِل
weaken	يَمُنّ	pamper	يُدَلِّل	know	يَعْرِف
guard	يَصُون	cheat	يُغَشِّش	treat	يُعَايِن
provide	يَمُون	correspond with	يُكَاتِب	hide	يُكَنِّن

Demonstratives

	Masculine singular	*Feminine singular*	*Plural*
near	هٰذَا	هٰذِهِ	هٰؤُلَاءِ
distant	ذٰلِكَ	تِلْكَ	أُولٰئِك *or* أُولَائِكَ

Third-Person Pronominal Suffixes

	Singular	*Dual*	*Plural*
masculine	هُ		هُمْ
common		هُمَا	
feminine	هَا		هُنَّ

Adverbs

now	أَلْآنَ	here	هُنَا	a little	قَلِيلًا
today	أَلْيَوْمَ	there	هُنَاكَ	outside	خَارِجًا
inside	دَاخِلًا	a lot	كَثِيرًا	at times	مِرَارًا

Nouns

girl	بِنْت	man	رَجُل	woman	حُرْمَة
guide	دَلِيل	merchant	تَاجِر	physician	طَبِيب
clerk	كِتَاب	notable	عَيْن	unionist	إِتِّحَادِي
major general	لِوَاء	agent	وَكِيل	hermit	نَاسِك
manager	مُدِير	boy	وَلَد	upholsterer	نَجَّار
actor	مُمَثِّل	minister	وَزِير	seaman	بَحَّار
teacher	مُعَلِّم	leader	زَعِيم	usurer	مُرَابٍ
lawyer	مُحَامٍ	person	شَخْص	historian	مُؤَرِّخ

Adjectives

American	أَمْرِيكِيّ	famous	مَشْهُور	special	خَاصّ
sincere	صَادِق	Egyptian	مِصْرِيّ	Greek	يُونَانِيّ
little	صَغِير	Chinese	صِينِيّ	good	طَيِّب
ignorant	جَاهِل	psychic	نَفْسَانِيّ	thin	هَزِيل
handsome	جَمِيل	revolutionary	ثَوْرِيّ	fat	سَمِين
erudite	عَلَّامَة	murdered	قَتِيل	tall	طَوِيل
big	كَبِير	tired	تَعْبَان	short	قَصِير
lazy	كَسْلَان	visionary	تَخَيُّلِيّ	happy	فَرْحَان
idiotic	مَعْتُوه	just	عَادِل	sad	حَزْنَان

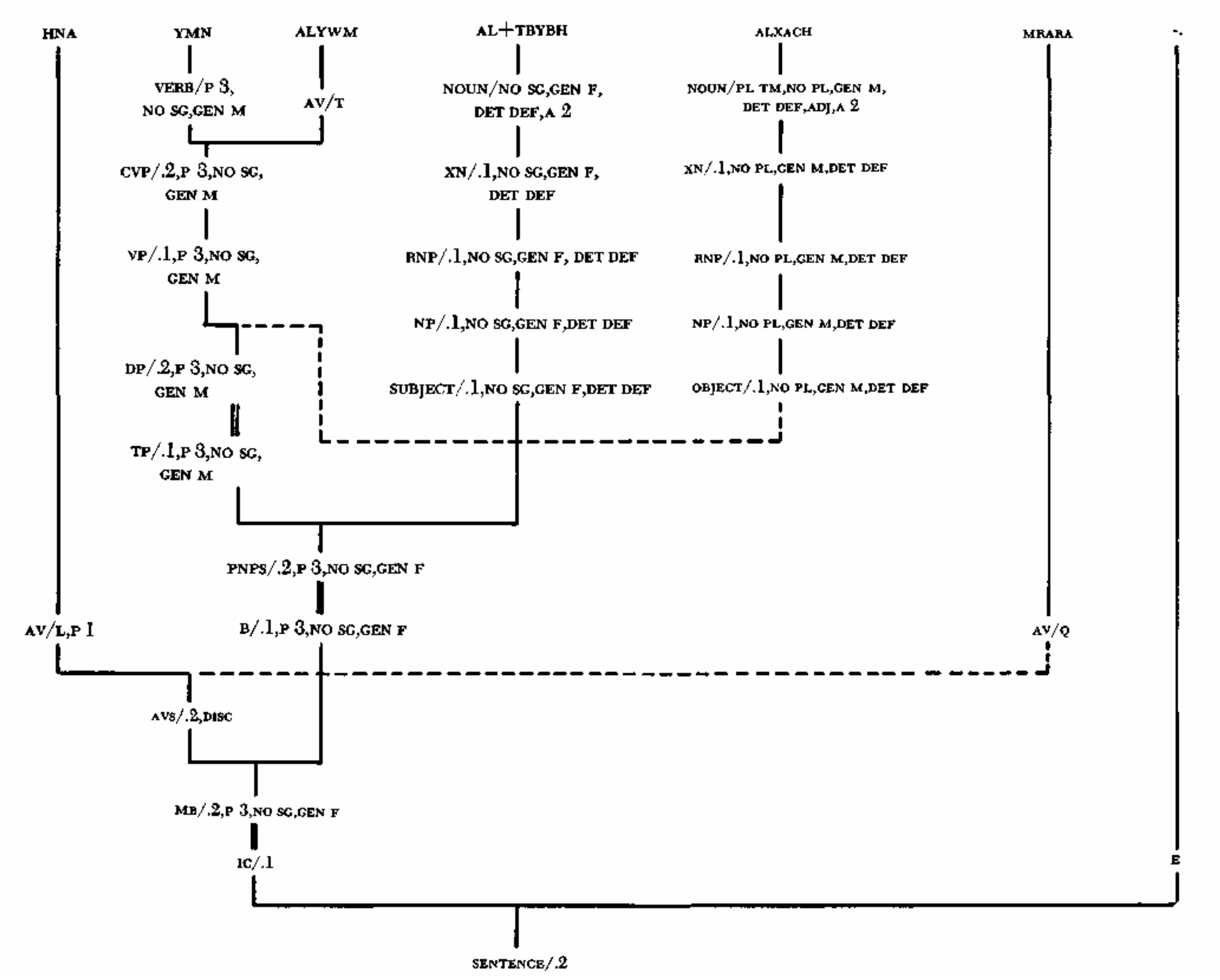

HNA
YMN
ALYWM
AL+TBYBH
ALXACH
MRARA
-.
VERB/P 3, NO SG,GEN M
AV/T
NOUN/NO SG,GEN F, DET DEF,A 2
NOUN/PL TM,NO PL,GEN M, DET DEF,ADJ,A 2
CVP/.2,P 3,NO SG, GEN M
XN/.1,NO SG,GEN F, DET DEF
XN/.1,NO PL,GEN M,DET DEF
VP/.1,P 3,NO SG, GEN M
RNP/.1,NO SG,GEN F, DET DEF
RNP/.1,NO PL,GEN M,DET DEF
NP/.1,NO SG,GEN F,DET DEF
NP/.1,NO PL,GEN M,DET DEF
DP/.2,P 3,NO SG, GEN M
SUBJECT/.1,NO SG,GEN F,DET DEF
OBJECT/.1,NO PL,GEN M,DET DEF
TP/.1,P 3,NO SG, GEN M
PNPS/.2,P 3,NO SG,GEN F
B/.1,P 3,NO SG,GEN F
AV/L,P 1
AV/Q
AVS/.2,DISC
MB/.2,P 3,NO SG,GEN F
IC/.1
E
SENTENCE/.2

FIGURE 2.
Tree-structure illustrating the complete syntactic mechanical analysis outlined in Figure 1.

Each Arabic letter has several forms. The particular form selected in any given instance is determined by the preceding and following letters. In general, therefore, in view of this redundancy only one computer symbol is assigned to a letter. For example, منهم /minhum/ 'from them' is transliterated MNHM without distinguishing the initial مـ M from the final م M.

The Sentence-Recognition Grammar

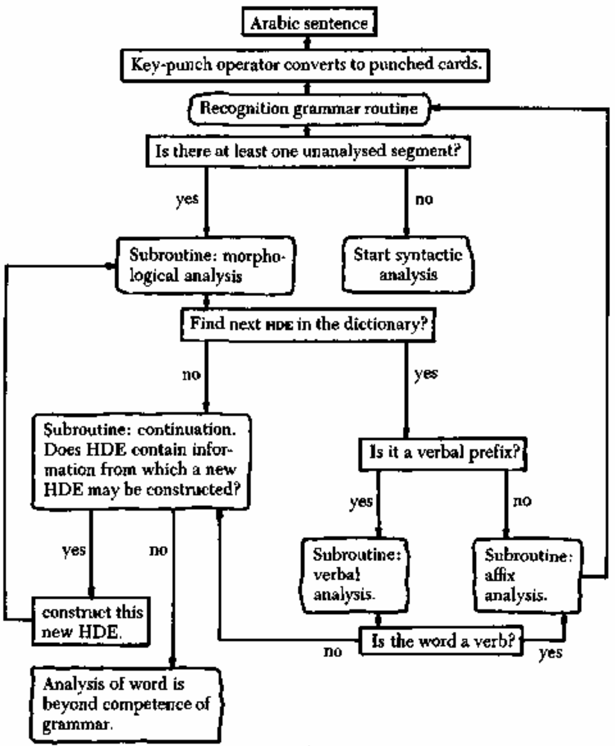
The computer parses the input sentence under control of two major subroutines, the morphological and the syntactic. The morphological subroutine identifies the lexical units of which each word is composed and makes the grammatical information derived from the analysis explicit. This grammatical information is added to the input in the form of a number of items named constitutes.

The syntactic subroutine associates groups of constitutes according to the rules of the grammar into increasingly general constructions also identified by constitutes to which further grammatical information is added as it is accumulated. If the input is grammatical, the whole sequence is identified as a sentence defined by the sum-total of the grammatical information derived from the analysis. If the sequence is ungrammatical or beyond the competence of the grammar, the analysis is carried as far as possible and then left incomplete. In such a case, no translation is attempted. In Arabic a fairly large number of morphemes may be grouped together to form a single word. While the present grammar is not comprehensive enough to parse the ten-letter orthographic word WSYFHMWNKH /wa sa yufahhimuwnakahu/ 'and they will explain, it to you', the word does illustrate the morphological problems which must be met by a complete sentence-recognition grammar of Arabic. This word is divisible into the following eight graphemes: W- 'and', S- 'will', Y- 'third person subject', FHM 'explain', -W 'masculine plural subject', -N 'indicative mode', -K 'you', -H 'it'.

The problem of the recognition of broken plural constructions was felt to be of sufficient interest to warrant the writing of rules to enable their identification as words derived from singular forms listed in the dictionary. Broken plural constructions are those which have as one constituent a plural prefix, infix, or a discontinuous affix or a suffix with a concomitant substantive stem the allograph of which differs from that of the singular stem. Singular and plural pairs illustrating the various types of plural affix follow. The singular noun is followed by the plural separated from it by a slash. RJL/A-RJL 'foot', RJL/RJ-A-L 'man', WZYR/WZR-AO 'minister', WLD/A-WL-A-D 'boy', LWAO/A-LWY-H 'major general', and TVB-AN/TV-A-B-Y 'tired'.

The Morphological Analysis

The subroutine for morphological analysis is broadly outlined in Flow Chart 1. The subroutine "morphological analysis" identifies the lexical items and morphemes in each word and makes explicit the grammatical information to be derived from them without reference to syntactic relations. The identification involves recognition of words and stems, prefixes, infixes and suffixes as well as various types of discontinuous morphemes. Distinctions are made between affixes on the one hand and identical sequences of letters which form parts of stems rather than affixes on the other hand. In addition, the grammar recognizes morphological ambiguities and keeps track of the alternates for possible solution by syntactic analysis.


Arabic sentence
Key-punch operator converts to punched cards.
Recognition grammar routine
Is there at least one unanalysed segment?
yes
no
Subroutine: morphological analysis
Start syntactic analysis
Find next HDE in the dictionary?
no
yes
Subroutine: continuation. Does HDE contain information from which a new HDE may be constructed?
Is it a verbal prefix?
yes
no
yes
no
Subroutine: verbal analysis.
Subroutine: affix analysis.
construct this new HDE.
Is the word a verb?
no
yes
Analysis of word is beyond competence of grammar.

FLOW CHART 1.
Subroutine for the morphological analysis of words in Arabic.

The analysis of YMNH and ALWYH illustrates in detail the computer subroutine for morphological analysis. YMNH (Figure 3) represents an *unanalyzed segment* (fourth box in Flow Chart 1), defined as any group of letters under immediate study. In the morphological analysis the word is assumed to be the first *hypothetical dictionary entry,* abbreviated to HDE. The HDE, YMNH, is looked up in the dictionary and not found.

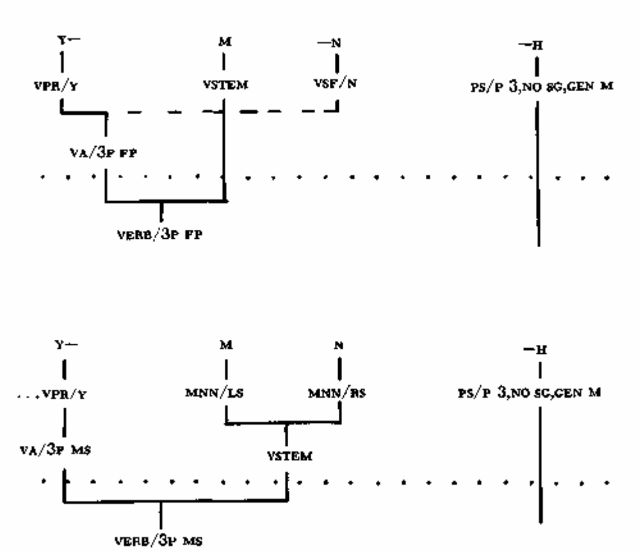
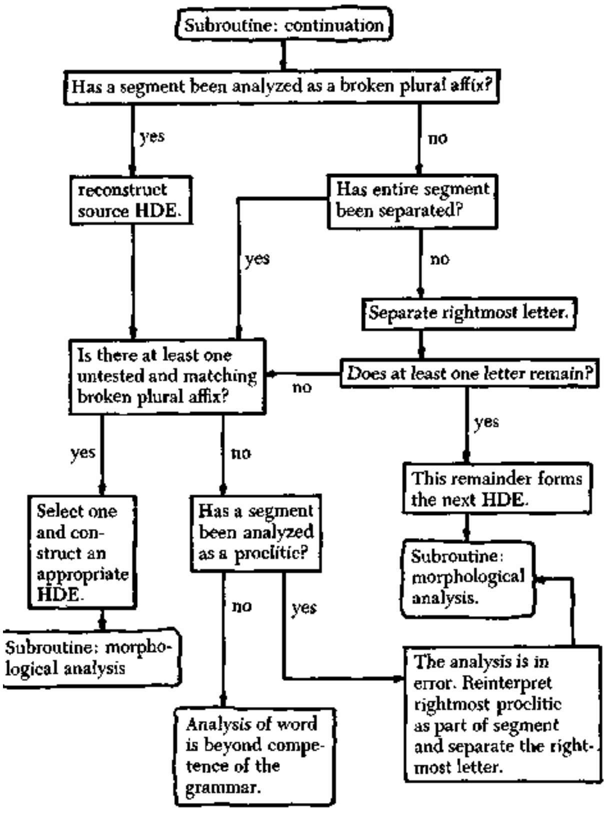


FIGURE 3.
The morphological analysis of the ambiguous word YMNH /yamunnahu/ 'they provide it' and /yamunnuhu/ 'he weakens it'.

Subroutine continuation is therefore entered. *Separation* (box 3 of subroutine continuation, p. 20) is a process which involves the splitting off of the rightmost letter of the current segment to form a new segment shorter than the preceding one. This process will form successively the new segments YMN, YM and Y from the original segment YMNH. The process does not involve deletion as the separate letters are preserved for further analysis.

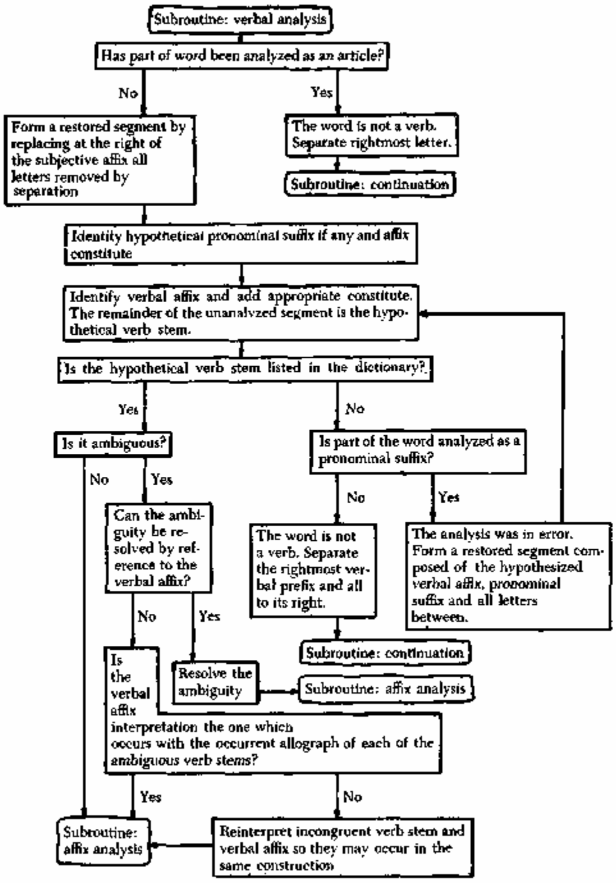
The segment YMN forms the next HDE. The process described as operating on YMNH is repeated until the final segment Y of YMNH is found in the dictionary and identified as a verbal affix. The subroutine verbal analysis is next entered (page 20).

The restored segment YMNH is formed. The H is now identified as the third person, masculine singular pronominal suffix, PS/P 3, NO SG, GEN M. The next step tentatively identifies the two letters Y and N of YMN as the two members of the third person feminine plural discontinuous verbal affix VA/3P FP. This leaves the unanalyzed segment M, which is found to be a dictionary entry. The dictionary lists M as an allograph of the stem MWN and the left side of an allograph of the


Subroutine: continuation
Has a segment been analyzed as a broken plural affix?
yes
no
reconstruct source HDE.
Has entire segment been separated?
yes
no
Separate rightmost letter.
Is there at least one untested and matching broken plural affix?
no
Does at least one letter remain?
yes
yes
no
This remainder forms the next HDE.
Select one and construct an appropriate HDE.
Has a segment been analyzed as a proclitic?
Subroutine: morphological analysis.
Subroutine: morphological analysis
no
yes
The analysis is in error. Reinterpret rightmost proclitic as part of segment and separate the rightmost letter.
Analysis of word is beyond competence of the grammar.

FLOW CHART 1.
Subroutine continuation.

stem MNN. The segment M is therefore ambiguous, and the ambiguity cannot be resolved by reference to the verbal affix. The computer next examines the fitness of the hypothesized verbal affix to occur in construction with the allograph of each of the ambiguous verb stems found in the word. Reference to the rules of the grammar incorporated in the program assures that M is the allograph of MWN which occurs in construction with VA/3P FP. Letters Y and N which constituted the hypothesized verbal affix VA/3P FP are now reanalyzed by the computer. The Y is reinterpreted as the third person masculine singular VA/3P MS and the N as the right side of the allograph MN of the verb stem MNN. The analysis of the two interpretations has reached the level of the dotted lines in the double analysis in Figure 3. The allograph MN of the verb stem MNN and the verbal affix may now occur in the same construction. Entrance is next made into the subroutine affix analysis. All sequences of letters have been identified, but three tree stems remain. Reference to the grammar rules directs the computer to associate the constitutes VA and VSTEM in the construction VERB. This constitute with information regarding the inflectional categories of gender, number and person are added to the analysis. The pronominal suffix is not treated as part of the word in the morphological analysis, and therefore the analysis is completed in this case


Subroutine: verbal analysis
Has part of word been analyzed as an article?
No
Yes
Form a restored segment by replacing at the right of the subjective affix all letters removed by separation
The word is not a verb. Separate rightmost letter.
Subroutine: continuation
Identity hypothetical pronominal suffix if any and affix constitute
Identify verbal affix and add appropriate constitute. The remainder of the unanalvzed segment is the hypothetical verb stem.
Is the hypothetical verb stem listed in the dictionary?
Yes
No
Is it ambiguous?
Is part of the word analyzed as a pronominal suffix?
No
Yes
No
Yes
Can the ambiguity be resolved by reference to the verbal affix?
The word is not a verb. Separate the rightmost verbal prefix and all to its right.
The analysis was in error. Form a restored segment composed of the hypothesized verbal affix, pronominal suffix and all letters between.
No
Yes
Subroutine: continuation
Is the verbal affix interpretation the one which occurs with the occurrent allograph of each of the ambiguous verb stems?
Resolve the ambiguity
Subroutine: affix analysis
Yes
No
Subroutine: affix analysis
Reinterpret incongruent verb stem and verbal affix so they may occur in the same construction

FLOW CHART 1.
Subroutine: verbal analysis.

with two tree stems. One of the alternate analyses of YMNH is placed in the pushdown store and the next word is processed for syntactic analysis.

The word ALWYH (Figure 4) is not listed in the dictionary and consequently is separated to AL which is identified as the article, DEF. The subroutine affix analysis is entered. DEF is a proclitic and therefore WYH forms the next HDE. The process is repeated until W is found in the dictionary listed as the proclitic conjunc-

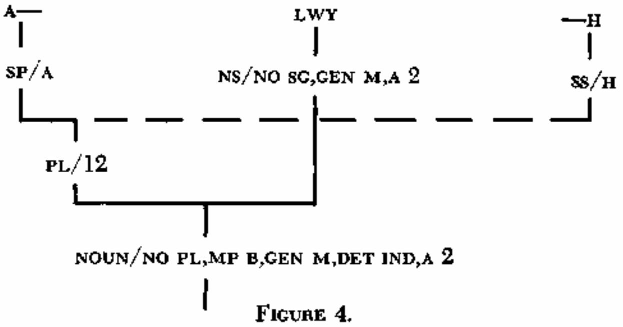
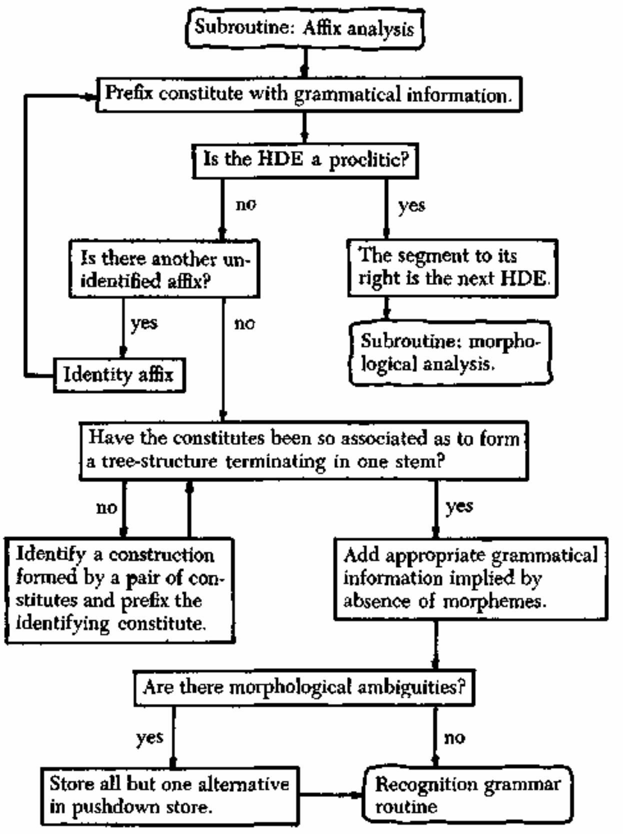


FIGURE 4.
The computer analysis of ALWYH 'major generals'.


Subroutine: Affix analysis
Prefix constitute with grammatical information.
Is the HDE a proclitic?
no
yes
Is there another unidentified affix?
The segment to its right is the next HDE.
yes
no
Identity affix
Subroutine: morphological analysis.
Have the constitutes been so associated as to form a tree-structure terminating in one stem?
no
yes
Identify a construction formed by a pair of constitutes and prefix the identifying constitute.
Add appropriate grammatical information implied by absence of morphemes.
Are there morphological ambiguities?
yes
no
Store all but one alternative in pushdown store.
Recognition grammar routine

FLOW CHART 1.
Subroutine: affix analysis.

tion 'and'. YH is constituted the next HDE. Y is found in the dictionary to be a potential verbal prefix and the subroutine verbal analysis is entered. Here it is found that AL has been analyzed as an article, and the analysis of YH as a possible verb is rejected. Subroutine continuation is now entered. At this point the entire word has been separated. No untested broken plural affix is recognized in the sequence YH. Two segments, the article AL and the conjunction W, are found to have been analyzed as proclitics. The interpretation of W as a proclitic is rejected, and its separation leaves the entire segment separated. Subroutine morphological analysis is reentered. Since there is no segment remaining to form an HDE to be looked up in the dictionary, subroutine continuation is immediately entered. No untested broken plural affix is recognized in the sequence WYH, but there is still the proclitic AL. The interpretation of AL as a proclitic is rejected, and the letter L is separated before reentering the subroutine morphological analysis.

The new HDE A is found in the dictionary and identified as a potential verbal prefix. At this point, no part of the word is analyzed as the article. The restored segment ALWYH is formed and the H is identified as the third person masculine singular pronominal suffix. The A is confirmed as the first person singular verbal affix and the hypothetical verb stem LWY is looked up in the dictionary where it is not listed. The hypothesis that the H was a pronominal suffix was in error. The restored segment ALWYH is then examined, and again the first person singular verbal affix A is confirmed. This time the hypothesized verb stem is LWYH, which also proves not to be listed in the dictionary. The analysis of ALWYH as a verb is consequently rejected.

Subroutine continuation is now entered. The entire segment has been separated. The untested broken plural affix A + . . . + H is now identified and the HDE, LWAO, is constructed from the unanalyzed segment LWY by application of the grammar rules. LWAO is listed in the dictionary and the subroutine affix analysis is entered. The constitute noun stem NS with the appropriate grammatical information is added to the analysis. At this point all elements of the input word have been identified, but the constitutes have not been associated to form a tree structure terminating in one stem. Reference to the grammar rules instructs the computer that the two constitutes PL and NS are associated in the construction NOUN. This constitute is added to the analysis. As there is no article in the word, the further grammatical information that the word is indefinite is added and the analysis is completed.

In the process of analysis the computer has considered the following six interpretations and rejected all but the last: 1. AL-W-Y-H 'the and he (verb stem)'; 2. AL-W-YH 'the and (plural substantive)'; 3. AL-WYH 'the (plural substantive)'; 4. A-LWY-H 'I (verb stem) it'; 5. A-LWYH 'I (verb stem)'; and 6. A-LWY-H 'major generals'.

The fifth alternative ALWYH 'I twist it' is rejected only because the stem LWY is not listed currently in the dictionary. If it were, the morphological analysis would remain ambiguous and await resolution in the syntactic analysis.

A characteristic feature of Arabic is the occurrence of discontinuous allomorphs, the presence of which is reflected in the orthography. The grammar contains rules which enable the computer to recognize such discontinuities in the formation of substantives and verbs.

The substantive plural affix manifests a number of discontinuous allomorphs. In the present grammar these plural allomorphs are described in terms of their component letters and the number of letters occurring to their left. The recognition of the stem allograph and the plural allograph occurs simultaneously by reference to a single grammar rule.

The rule for the recognition of the allograph PL/12 of the plural morpheme which occurs in the word ALWYH illustrates the procedure. The rule is A32LH=PL/12+SP/A+A—+32AO+LWY+SS/H+—H.
Three events are sought simultaneously on the left of

the equation: 1) a segment with an initial A, 2) any three letters to the right of the A, and 3) an H to their right. The right side of the rule then identifies the plural allograph PL/12 and its two constituents by simultaneously prefixing the constitutes SP/A and SS/H to the two members and the constitute PL/12 to the construction formed by them. In addition it identifies the three letters found to the left of the fifth letter H as the plural allograph of a hypothetical dictionary entry 32AO, interpreted as LWAO. The single rule thus results in three primary identifications, the identification of two constructions and the formation of a new HDE.

The Dictionary

The dictionary furnishes the sentence-recognition grammar with the grammatical information derivable from each lexical entry. The *lexical entry* may be a prefix, a stem or a portion of a stem, a proclitic or a word and is listed as the left side of a dictionary rule. The right side of the dictionary rule is composed of a constitute, which makes the grammatical information implied by the lexical entry explicit, and a repetition of the lexical entry. Generally a lexical subscript is attached to this repetition.

The *lexical subscript* consists of the term ARB and a subsubscript identical with the dictionary form of the item with which the lexical subscript is associated. The subsubscript identifies the vocabulary rule-set in the bilingual dictionary (Figure 7) by which is determined the output vocabulary subscript pertinent to the item with which the lexical subscript is associated. ALWYH/ARB LWAO derives its output vocabulary subscript from the vocabulary rule set LWAO.

```
A = VPR/A+A
B+HAR=NS/PL TM,NO  SG,GEN  M,A 1+B+HAR/ARB B+HAR
LWAO=NS/NO SG,GEN M,A 2+LWAO/ARB LWAO
M=VSTEM+MWN/ARB  MWN+VSTEM+MNN/ARB   MN
MNN=VSTEM+MNN/ARB MNN
MWN=VSTEM+MWN/ARB  MWN
Y=VPR/Y+Y
```

FIGURE 5
Examples of dictionary rules.

The seven lexical entries in Figure 5 fall into four grammatical classes. The ambiguity of lexical entry M is indicated by the occurrence of two pairs of items on the right side of that rule.

Stripping

In the actual computer program the aim has been to initiate the syntactic analysis with a single constitute per word. Where more than one constitute has been added in the course of the morphological analysis, the analysis of the word is stripped. The stripping process places a space to the left of each pronominal suffix and then deletes from the analysis of each word all but its single base constitute. A *base constitute* is a constitute which has not yet been identified as a constituent of a construction. The stripped morphological analysis of the Arabic sentence

هناك
يستقبل الوزير الصينى هوءلاء التجار المصريون •

follows: ADV/LOC, P 2 + HNAK/ARB HNAK + VERB/P 3, NO SG, GEN M+YSTQBL/ARB STQBL+NOUN/NO SG, GEN M, DET DEF, A 1 + ALWZYR/ARB WZYR+ADJ/NO SG, GEN M, DET DEF, A 1+ALCYNY/ARB CYNY+DEM/NO PL, P 1+H+WLAO/ARB H+WLAO+NOUN/MPB, NO PL, GEN M, DET DEF, A 1+ALTJAR/ARB TAJR+ADJ/NO PL, GEN M, DET DEF, C N,A 2+ALMCRYWN/-ARB MCRY+E+-. A word-for-word translation is 'there he-meets the-minister the-Chinese these the-merchants the-Egyptian.' After syntactic analysis the computer translation reads 'these Egyptian merchants meet the Chinese minister there.'

The Syntactic Analysis

The syntactic analysis of the input sentence is approached through the "immediate constituent" method. This method first identifies the most deeply nested structures and proceeds by building the tree-structure from the inside out. Immediate constituent analysis, therefore, is distinct from "predictive analysis," "analysis by synthesis" and the "dependency connection" approaches.[4]

The input to the syntactic analysis portion of the program is composed of the stripped morphological analysis of the input sentence. The input thus consists of any number of pairs of items each composed of a constitute and a word or pronominal suffix.

In essence, the program operates by searching in turn for each possible structure in the language starting with the most deeply nested one and proceeding structure by structure to the recognition of the final one, SENTENCE. Having selected a structure the identification of which is to be made, the computer seeks the constituent(s) required to form the construction and identifies it, wherever it occurs, through the addition of the appropriate constitute. This process is repeated until all constructions of the type sought are identified, and then the process is repeated with the next most deeply nested structure.

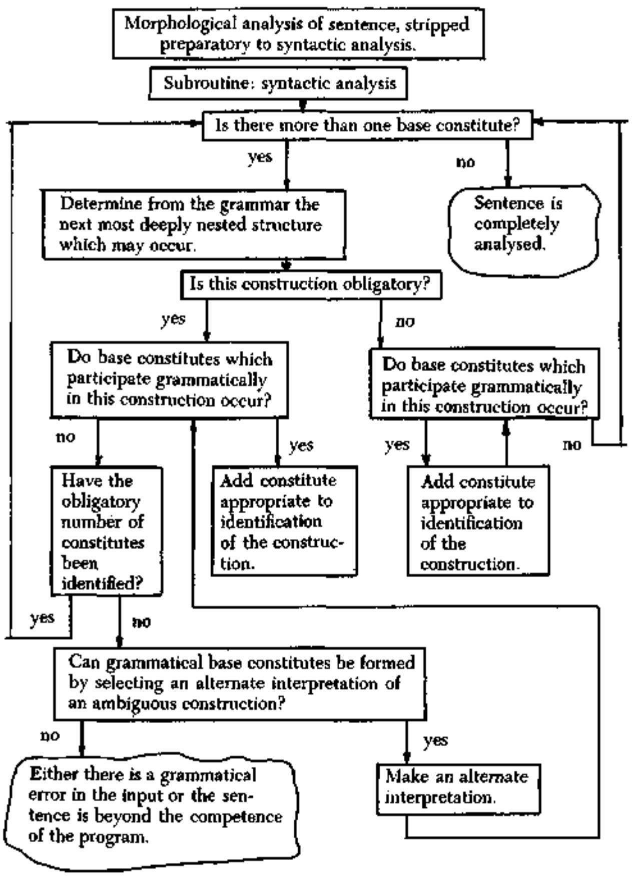
Under guidance of the program the computer identifies discontinuous as well as continuous dyadic and monadic constructions. It resolves cases of grammatical ambiguity when they are grammatically resolvable within the limits of the sentence and selects one of the alternates when the ambiguities are not resolvable. Some problems of agreement and concord are also solved by the computer.

The syntactic analysis program produces tree structures of the type found in Figure 2. The analysis of this sentence illustrates in some detail the steps taken by the computer in carrying out the syntactic analysis. The stripped morphological analysis to which the syntactic analysis is applied follows: AV/L, P 1 + HNA/ARB HNA + VERB/P 3, NO PL,GEN F + YMN/ARB MWN+AV/T+ALYWM/ARB ALYWM+NOUN/NO SG, GEN F, DET DEF, A 2 + AL+TBYBH/ARB +TBYB + NOUN/PL TM, NO PL, GEN M, DET DEF, ADJ, A 2 + ALXACH/ARB XAC +AV/Q+ MRARA/ARB MRARA + E+-. It will be noted that the constitute of YMN is not, at this stage, the same as that in the final stage exhibited in Figure 2.

The "immediate-constituent" recognition grammar must contain implicitly or explicitly a listing of constructions in order of nesting from the most deeply to the least deeply nested. In the present grammar the AJS construction consisting of a pair of adjectives is the most deeply nested construction.

Referring to Flow Chart 2, AJS is not obligatory, and no base constitutes which participate in this construction are found in the sentence above.

The first construction which the computer identifies in the sentence is the non-obligatory, monadic extended noun XN. The program adds the appropriate constitute and scans the analysis in an attempt to identify another such construction, which it does. The same process is followed in identifying the RNP and NP constructions.



Morphological analysis of sentence, stripped preparatory to syntactic analysis.
Subroutine: syntactic analysis
Is there more than one base constitute?
yes
no
Determine from the grammar the next most deeply nested structure which may occur.
Sentence is completely analysed.
Is this construction obligatory?
yes
no
Do base constitutes which participate grammatically in this construction occur?
Do base constitutes which participate grammatically in this construction occur?
no
yes
yes
no
Have the obligatory number of constitutes been identified?
Add constitute appropriate to identification of the construction.
Add constitute appropriate to identification of the construction.
yes
no
Can grammatical base constitutes be formed by selecting an alternate interpretation of an ambiguous construction?
no
yes
Either there is a grammatical error in the input or the sentence is beyond the competence of the program.
Make an alternate interpretation.

FLOW CHART 2.
Subroutine for the syntactic analysis of sentences by the immediate constituent method.

Next the adverbial sequence AVS is sought to the right of the verb. This construction may be either continuous or discontinuous and consists of two adverbs AV or an AV to the left of an adverb sequence AVS.

In accordance with Yngve's theory of grammar a discontinuous construction consists of two constituents separated by a single intervening construction. In a sentence-recognition grammar this intervening construction must be correctly and completely identified before the constituents of the enclosing discontinuous construction can be recognized in turn as members of a grammatical construction. This requirement imposed by the occurrence of discontinuous constructions in the syntactic analysis of natural languages is one reason which makes the ordering of search for the various substructures in the sentence so important.[5]

In Figure 2 the AV/L, P 1 and the AV/Q are two constituents of the discontinuous construction AVS/DISC. At the beginning of the syntactic analysis four base constitutes intervene between the two AV. Before these AV can be identified as constituents of the construction AVS/DISC, the four intervening constitutes must be identified as constituents of the basic clause construction B.

The program now directs the computer to seek to the right of the verb for two constituents of the construction AVS. It first locates a rightmost AV, in this case AV/Q. It fails to find to its immediate left the AV required to form a continuous AVS construction. Next it looks for an AV somewhere to the left of the first one and finds AV/T. The next step must determine whether the two may form a discontinuous AVS construction. The computer finds two base constitutes NP between the two AV. In the present grammar there is no construction which consists of two NP constitutes. Because of the requirement that one and only one base constitute may occur between the two constituents of a discontinuous construction, the computer rejects these two AV as candidates for a discontinuous AVS construction. The AV to the left of the verb is not considered as a constituent of an AVS construction until after the obligatory basic clause B has been identified.

Next the non-obligatory dyadic continuous verb phrase construction CVP is identified and the appropriate constitute is added by the same process used in identifying the XN. This CVP is then identified as a verb phrase, VP.

The program now directs the computer to identify the object of the VP and the subject if any. The first construction it seeks is the non-obligatory predicate with pronominal suffix PPS, such as YMNH, and does not find it. Then it attempts to identify the possible occurrence of a total predicate TP as a constituent of a

PNPS, predicate with noun-phrase subject. The two noun phrases make this an obligatory construction. The computer examines the VP to determine whether it is a base constitute which may participate in the PNPS construction. It is analyzed as third person feminine plural containing the constituent /yamunna/ 'they provide' derived from the stem MWN. Since no plural verb may participate in a PNPS construction, the alternate interpretation of YMN, VERB/3P MS derived from the stem MNN 'he weakens' is substituted from the pushdown store for the original interpretation. This interpretation of the verb may participate as a base constitute of the construction PNPS.

The next problem involves the identification of the obligatory monadic OBJECT and SUBJECT constructions. First a base constitute NP with case either accusative or oblique-accusative is sought. This is not found. Next a base constitute NP with case either nominative or nominative-oblique is sought. Such a NP would be identified as the SUBJECT, and the other NP as the OBJECT by elimination. No case distinctions are found and therefore the solution of the problem in this direction fails.

Gender concord between the verb and the hypothetical subject is the next possible means of solution. If the verb is contiguous with the subject noun phrase, concord in gender does occur, otherwise it need not. This means of solution also fails since the verb and NP are not contiguous.

The final solution is based upon word-order. In the normal Arabic word-order the object occurs to the right of the subject. The computer, therefore, identifies the righthand NP as object and the appropriate constitute is prefixed. The lefthand NP is next identified as the SUBJECT.

The computer now seeks a discontinuous predicate construction DP. Only one base constitute is found between the VP and the object, which may therefore form the two immediate constituents of DP. The dyadic PNPS construction is sought and identified immediately after the identification of the total predicate TP.

After PNPS has been identified as the monadic basic clause construction B, the computer examines the analysis to determine whether another AVS construction with the AV to the left of B as one constituent may be formed. It seeks an AV to the right of the substructure B. It does find AV/Q and associates the two AV in the discontinuous adverbial sequence construction AVS/DISC with one base constitute B intervening. The constituent AVS/DISC and B are next identified as the modified basic clause MB, and the analysis of the sentence is concluded.

The Structural Transfer Routine and the Statement of Structural Equivalence

The mechanism for the production of output sentences in the mechanical translation program is an adaptation of the one invented by Yngve. This mechanism is best described in his own words.

> The mechanism gives precise meaning to the set of rules by providing explicitly the conventions for their application. . . . It is an idealized computer and is physically realizable. It consists of four cooperating parts. There is an output device that prints the output symbols one at a time in left-to-right fashion on an output tape. There is a computing register capable of holding one symbol at a time. There is a permanent memory in which the grammar rules are stored, and there is a temporary memory, in the form of a tape, on which intermediate results are stored.[2]

Once Yngve's mechanism has been activated, it produces sentences randomly under control of the program, without external stimulus. In this respect Yngve's model does not attempt to simulate the human as a sentence-producer since the human speaker is stimulated not only to produce sentences but to produce specific sentences by events both outside and within his own body. The stimuli from without are received through various senses such as sight, hearing, pain, *etc.* Events within his body which affect the production of specific sentences will certainly include the effects of memory, habit and physiological state.

The mechanical translation program discussed here still falls short of a model of human speech behavior, however the production of sentences is determined by the perception of stimuli external to the mechanism in the form of the input sentence with its grammatical analysis.

A fifth cooperating part called the stimulator has been added to the four found in Yngve's mechanism. The stimulator is a device in which a simulation of certain events external to the mechanism may be placed. These events are those which influence speech-production. The simulation of these events is in a form which can be recognized, examined and analyzed in various ways by the mechanism. In effect, the stimulator is a model of an interesting part of that portion of the universe which effects and stimulates the human speaker's speech. To the present time the stimulator has contained only the output of the sentence-recognition pro-

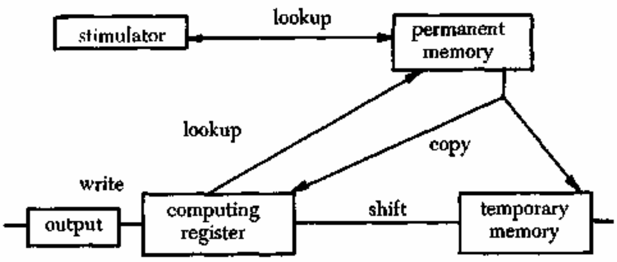


FIGURE 6.
Organization of the mechanism for the structural transfer and sentence-production routines.

gram. With some adaptation it is possible to imagine the stimulator as containing information which might simulate more generally visual, aural and other forms of perception.

At the time this research was undertaken I had not decided where in the mechanical translation process the specifier for the output sentence should be formed. As a result part of it is formed during the analysis of the input sentence, another part during the actual production of the output sentence and still another part between the two.

I now believe that no part of the output sentence specifier should be formed until the analysis of the input sentence has been completed. Decisions on the formation of the output specifier made during the analysis of the input sentence are so premature that many changes in it may be required after the analysis has been completed.

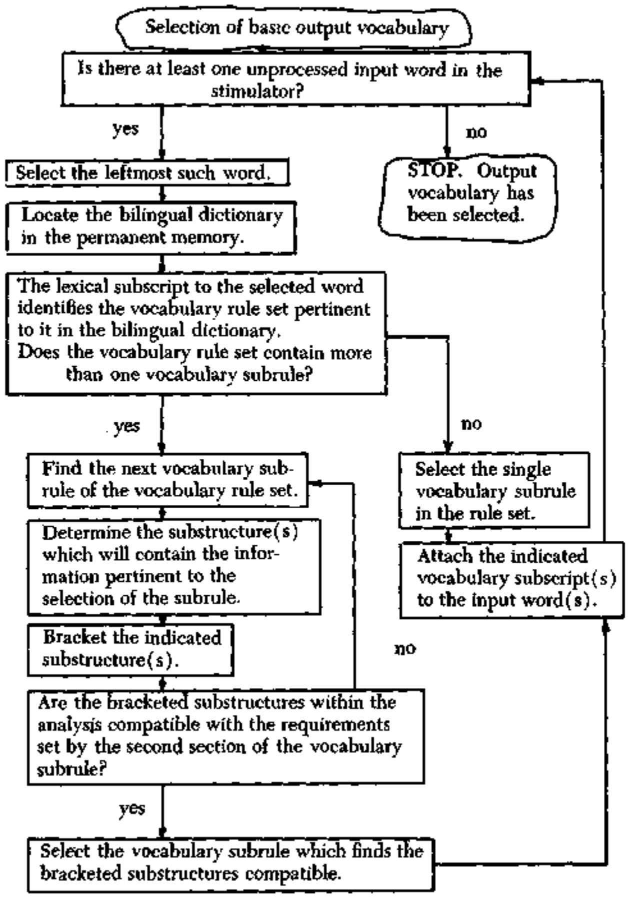
A more serious question is raised when one asks whether the specifier should be formed before or concurrently with the production of the output sentence. The answer to this question is at least partially dependent on the theory of sentence-construction grammar used. The current grammar is the one presented in my first report.[6] This grammar is written in accord with Yngve's model for language structure[2] which makes use of rule-sets composed of one or more subrules. The specifier consists of instructions for the selection of a number of rule-sets, the subrule to be selected in execution of each rule-set and the order in which they are to be executed. I now consider it most satisfactory to construct the output sentence specifier concurrently with the construction of the output sentence. The selection of the specific subrule to be executed is to be made immediately before the expansion of the constituent for which the subrule has been selected. It appears, however, that it will be convenient or even necessary to specify the selection of certain subrules before the production of the output sentence. The only subrules so specified at present are those which select the output vocabulary. The reason for the differentiation in the selection of these rules will be discussed below.

Yngve's mechanism operates under the control of two generalized programs specially designed for mechanical translation. The first operates before the production of the output sentence and is designed to select the basic output vocabulary. This program is presented in Flow Chart 3, which contains several new terms and two new operations.

The *bilingual dictionary* consists of that part of the statement of structural equivalence composed of the vocabulary rule sets. A *vocabulary rule set* consists of its name located at the head of the set and the vocabulary subrules which compose the set, listed below the name. A *vocabulary subrule* is composed of three parts. The first part is found in the lefthand column of Figure 7. Here is listed the constitute of the input analysis

		ALYWM
M/ARB ALYWM	- - -	M/ARB,ALYWM,TMPADV TODAY/-$
		HNAK
M/ARB HNAK	- - -	M/ARB HNAK,LOCADV THERE/-$
		JAHL
NOUN	M/ARB JAHL	M/ARB JAHL,NOUN CHILD/F
M/ARB JAHL	- - -	M/ARB JAHL,ADJ/ZAVJ IGNORANT/-$
		JMYL
NP	NP/GEN M	M/ARB JMYL,ADJ/ZAJEXC HANDSOME/-$
M/ARB JMYL	- - -	M/ARB JMYL,ADJ/ZAJEXC BEAUTIFUL/-$
		MVLM
M/ARB MVLM	- - -	M/ARB MVLM,NOUN TEACHER/A
		STQBL
M/ARB STQBL	- - -	M/ARB STQBL,VERB/T MEET/SSUF S
		XAC
NP	M/ARB +TBYB	M/ARB XAC,ADJ/ZAJA PERSONAL/-$
NP	M/ARB MVLM	M/ARB MVLM,NOUN TUTOR/A
AJ	AJ/C N	M/ARB XAC,ADJ/ZAJA SPECIAL/-$
AJ	AJ/C AOB	M/ARB XAC,ADJ/ZAJA SPECIAL/-$
NOUN	M/ARB XAC	M/ARB XAC,NOUN OFFICIAL/A,ADJ/ZAJA SPECIAL/-$
MBDL	M/ARB XAC AJ MBDL	M/ARB XAC,NOUN OFFICIAL/A,ADJ/ZAJA SPECIAL/-$
AJ/NOM	M/ARB XAC	M/ARB XAC,NOUN OFFICIAL,ADJ/ZAJA SPECIAL/-$
M/ARB XAC	- - -	M/ARB XAC,ADJ/ZAJA SPECIAL/-$

FIGURE 7.
Excerpt from the bilingual dictionary.


Selection of basic output vocabulary
Is there at least one unprocessed input word in the stimulator?
yes
no
Select the leftmost such word.
STOP. Output vocabulary has been selected.
Locate the bilingual dictionary in the permanent memory.
The lexical subscript to the selected word identifies the vocabulary rule set pertinent to it in the bilingual dictionary. Does the vocabulary rule set contain more than one vocabulary subrule?
yes
no
Find the next vocabulary subrule of the vocabulary rule set.
Select the single vocabulary subrule in the rule set.
Determine the substructure(s) which will contain the information pertinent to the selection of the subrule.
Attach the indicated vocabulary subscript(s) to the input word(s).
Bracket the indicated substructure(s).
no
Are the bracketed substructures within the analysis compatible with the requirements set by the second section of the vocabulary subrule?
yes
Select the vocabulary subrule which finds the bracketed substructures compatible.

Flow Chart 3.
Subroutine for the selection of the basic output vocabulary.

which defines the substructure of the input sentence in which is contained the information needed to determine whether or not the subrule is to be selected. In the first subrule in the vocabulary rule set XAC (Figure 7), NP indicates that all pertinent information for the selection of the subrule will be found in the substructure noun phrase which contains the constituent M/ARB XAC. M is a variable representing any sequence of letters. In this case, for example, it may represent XAC, XACA, ALXACWN, *etc.*

The second part of the vocabulary subrule is found in the central column. In the first subrule under XAC this section is represented by M/ARB +TBYB. This section defines the features of the environment which must be found in the substructure indicated by the first section if the vocabulary subscripts in the third part are pertinent. M/ARB +TBYB indicates that some form of the lexical item +TBYB must occur in the substructure NP if this subrule is to be executed. For example, the sentence-recognition portion of the program will identify AL+TBYBH ALXACH as a noun phrase NP. ALXACH will have the lexical subscript ARB XAC and AL+TBYBH will have ARB +TBYB. The first subrule under XAC will be found compatible with this substructure and will be selected. Eventually, as a result, ALXACH will be translated 'personal' to form the output phrase 'personal physician'.

The output *vocabulary subscript* identifies a subrule of a rule in the sentence-construction grammar of the output language. For purposes of this discussion its term will be the left side of the monadic output grammar rule and the subsubscript will be the right side of the same rule. In the first subrule under XAC (Figure 7) ADJ/ZAJA PERSONAL/-$ is a vocabulary subscript added to the word with the translation subscript ARB XAC. The subrule which this vocabulary subscript identifies is ADJ/ZAJA= PERSONAL/-$.

The *basic output vocabulary* is the set of all subsubscripts of the vocabulary subscripts in the bilingual dictionary. PERSONAL is an example of a member of this set. The basic output vocabulary is not the total vocabularly in the output grammar since the basic vocabulary does not include a number of function words.

The following eight sentences contain constructions which are compatible with each of the eight vocabulary subrules of the rule set XAC.

1. AVRF ALA + TBAO ALXACYN.
   I know the *personal* physicians.
2. AVRF ALMVLMAT ALXACH.
   I know *the tutors.*
3. YVRFH ALXACWN.
   The *special ones* know him.
4. AVRF ALWKLAO ALXACYN.
   I know the *special* agents.
5. AVRF ALXACH.
   I know the *special officials.*
6. AVRF ALXAC ALM + SHWR.
   I know the famous, *special official.*
7. AVRF ALXAC.
   I know the *special official.*
8. AVRF ALM + SHWR ALXAC.
   I know the famous, *special* one.

The term *bracketing* used in Flow Chart 3 applies to a process by which the substructure or substructures pertinent to an operation are isolated from the remainder of the analysis. The bracketed material contains the analysis of the substructure including the identifying constitute.

In the first sentence, ALXACYN/ARB XAC is found to be a constituent of the substructure NP, ALA+TBAO ALXACYN. This substructure is bracketed under direction of the program. The substructure NP does contain ALA+TBAO/ARB +TBYB which matches with the second section of the subrule. The bracketed substructure is thus compatible with the subrule and the vocabulary subscript is attached to ALXACYN.

In the second sentence ALXACH/ARB XAC is a constituent of the substructure NP, ALMVLMAT ALXACH. This substructure does not contain any constituent M/ARB +TBYB, but it does contain MVLMAT/ARB

MVLM. The substructure is compatible with the second subrule and the subscript NOUN TUTOR is attached to form ALMVLMAT/ARB MVLM, NOUN TUTOR replacing the earlier NOUN TEACHER. No vocabulary subscript is attached to ALXACH/ARB XAC with the result that no discrete output is produced as its equivalent.

The various forms of XAC in the third through the seventh sentences illustrate an interesting problem. The masculine XAC as the nucleus of a noun phrase with the distinctive plural XACH is to be translated 'special official'. In sentences three and four the occurrence of the plural suffixes —WN (nominative) and —YN (accusative-oblique) rather than —H prevent the translation of XAC as 'special official'. The substructure required to identify the translation in this case is restricted to the morphological constitute AJ with its telltale case.

In sentence five ALXACH is identified morphologically as a NOUN. The third section indicates that the vocabulary subscripts NOUN OFFICIAL and ADJ/ZAJA SPECIAL should be added to ALXACH. This subrule illustrates the selection of vocabulary when a single lexical item is to be translated by more than one output item.

In the sixth sentence ALXAC is found to be neither a constituent of a NP construction nor a constituent of an AJ/C N or an AJ/C AOB construction nor of a NOUN construction. It is included in a modified nominal construction with an adjective nucleus, MBDL. The environment required for the translation to 'special official' of a form of XAC as a constituent in a MBDL is of some complexity as is indicated by the form taken by the second section of this subrule. For such a translation the notation in the second section indicates that the form of XAC must be the nucleus of the MBDL. The substructure MBDL derived from ALXAC ALM+SHWR is given in Figure 8. The substructure is thus compatible with the requirements set by the vocabulary subrule. On the other hand, while ALXAC/ARB XAC in sentence eight is a constituent of a MBDL it is not compatible with the requirements set forth by the sixth subrule and so the phrase of which it is a constituent is translated 'the famous, special one' (Figure 9). ALXAC in this last sentence is compatible with none of the requirements of the first seven subrules. Any such form will be translated 'special' by default.

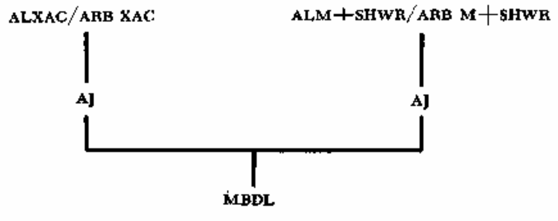


FIGURE 8.
Substructure of a MBDL with a form of XAC as the nucleus.

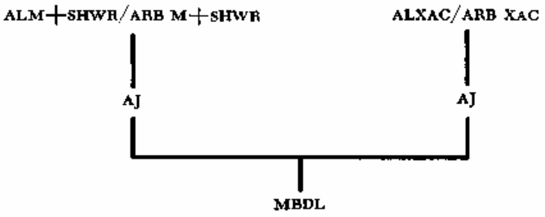


FIGURE 9.
Substructure of a MBDL with a form of XAC as an attribute.

An application of the structural transfer routine and the statement of structural equivalence to the analysis presented in Figures 10 and 12 to produce the output sentence in Figures 11 and 13 will illustrate this phase of the mechanical translation program and serve as a basis for a discussion of some of the problems involved.

Before the production of the output sentence is initiated the basic output vocabulary must be selected through the execution of the program in Flow Chart 3 applied to the pertinent vocabulary rule sets (Figure 7). The stimulator contains the mechanical analysis of the input sentence (Figure 12).

In initiating the subroutine for the selection of the output basic vocabulary (Flow Chart 3), the first word to be examined is HNAK/ARB HNAK. The vocabulary rule set HNAK (Figure 7) contains only one subrule. The vocabulary subscript LOCADV THERE is subscripted to it and the next word is sought. This process is repeated until the vocabulary subscript NOUN TEACHER/A has been added to the word ALMVLMH/ARB MVLM. The next word is ALXACH/ARB XAC. To be compatible the first vocabulary subrule of the rule set XAC requires some form of +TBYB. The second subrule requires some form of MVLM in the noun phrase of which ALXACH is a constituent. ALMVLMH meets the requirement, and the second subrule is compatible with the substructure. The subscript NOUN TUTOR/A replaces the subscript NOUN TEACHER/A attached to ALMVLMH. The fact that no vocabulary subscript is attached to ALXACH is a positive result of its processing by this portion of the program.

The next word ALJAHLH/ARB JAHL is not a constituent of a NOUN construction so the first subrule (Figure 7) is incompatible. The second is compatible and the subscript ADJ/ZAVJ IGNORANT is attached to ALJAHLH.

JMYL may be translated as 'handsome' when attribute to a substantive referring to a male. Otherwise it is translated as 'beautiful'. If the form of JMYL is itself the nucleus of a noun phrase and refers to a male, it is translated as 'handsome one,' otherwise as 'beautiful one.' In the present grammar all substantival references are to persons and so this classification is not specified.

In Arabic the gender of the adjective attribute is not generally in itself indicative of the gender of its

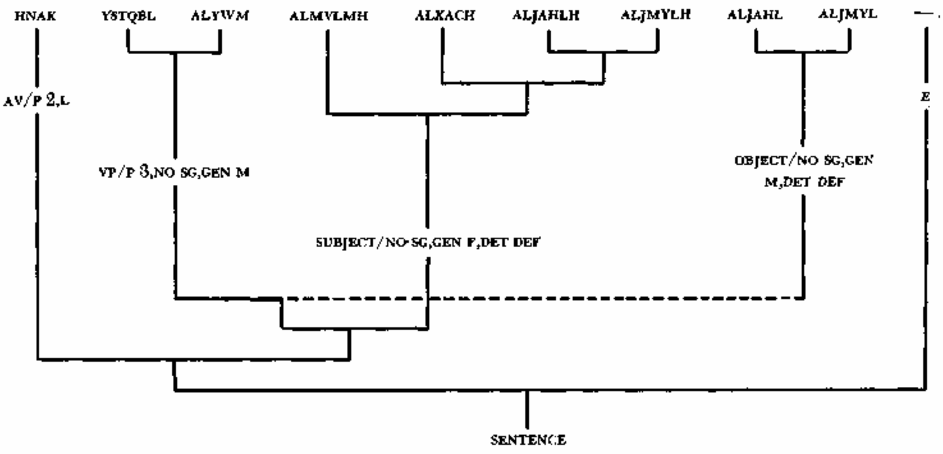


Figure 10.
Outline of mechanical analysis of input sentence /hunaaka yastaqbilu l yawma l mu9allimatu l xaassatu l jaahilatu l jamiylatu¨ l jaahila l jamiyla./

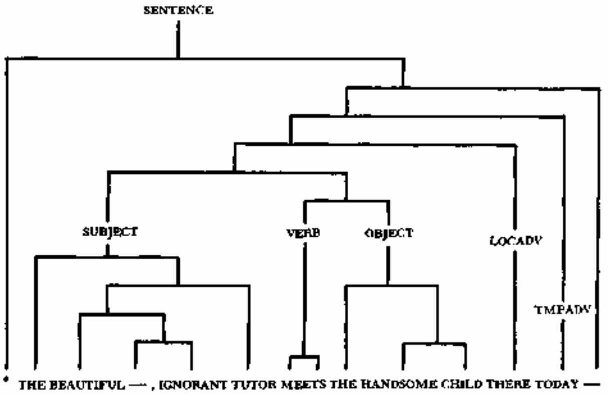

SENTENCE
SUBJECT
VERB
OBJECT
LOCADV
TMPADV
• THE BEAUTIFUL — , IGNORANT TUTOR MEETS THE HANDSOME CHILD THERE TODAY —

Figure 11.
Outline of mechanical construction of output sentence which translates sentence in Figures 10 and 12.

substantive. The feminine singular form of the attribute may occur in conjunction with a large class of masculine plural nouns as well as with both feminine plural and singular nouns, for example ALAWLAD ALJMYLH 'the handsome boys', ALBNAT ALJMYLH 'the beautiful girls' in addition to ALBNT ALJMYLH 'the beautiful girl'.

In the present grammar if the *noun phrase* of which the form of JMYL is a constituent is masculine the subscript ADJ/ZAJEXC HANDSOME is attached to the form of JMYL. Otherwise the subscript ADJ/ZAJEXC BEAUTIFUL is attached to it. The noun phrase of which ALJMYLH is a constituent is not masculine and so the first subrule is incompatible. By the second subrule the subscript ADJ/ZAJEXC BEAUTIFUL is added to the word. The last two words are processed as the others with the subscript NOUN CHILD and ADJ/ZAJEXC HANDSOME being added to each respectively. The selection of the subsubscripts IGNORANT and CHILD for JAHLH and JAHL, respectively, and of the subsubscripts BEAUTIFUL for JMYLH and HANDSOME for JMYL illustrates the capacity of the process to utilize rather subtle contextual differences.

At this phase of the translation, the input words with their subscripts appear in the stimulator as follows and furnish the skeletal word-for-word translation 'there meet today tutor ignorant beautiful child handsome.'

HNAK/ARB HNAK,LOCADV THERE+YSTQBL/ARB STQBL, VERB/T MEET/SSUF S+ALYWM/ARB ALYWM,TMPADV TODAY+ALMVLMH/ARB MVLM,NOUN TUTOR/A+ ALXACH/ARB XAC+ALJAHLH/ARB JAHL,ADJ/ZAVJ IGNORANT+ALJMYLH/ARB JMYL,ADJ/ZAJEXC BEAUTIFUL+ ALJAHL/ARB JAHL,NOUN CHILD+ALJMYL/ARB JMYL,ADJ/ ZAJEXC HANDSOME

After the basic output vocabulary has been selected through the application of the program in Flow Chart 3, the specific output sentence which translates the input sentence is produced by the concurrent application of the remainder of the structural transfer routine and the sentence-construction routine. These routines are applied to the statement of structural equivalence and the sentence-construction grammar of the output language in the permanent memory and the analysis of the input sentence in the stimulator. The two routines are combined in Flow Chart 4 which is an adaptation of the one in my first report.[7] This routine in turn is adapted from Yngve's.[2] Step IV contains the only significant change from the original routine. In the original step IV, subrules to be executed in the construction of a sentence were selected randomly. In the current routine the selection of the subrules is determined by the statement of structural equivalence and the analysis of the input sentence.

If one wishes to consider the program as a restricted example of the production of sentences stimulated by events occurring outside the mechanism, the statement of structural equivalence may be equated with a portion of general knowledge while the contents of the stimulator may be equated with one class of external stimuli.

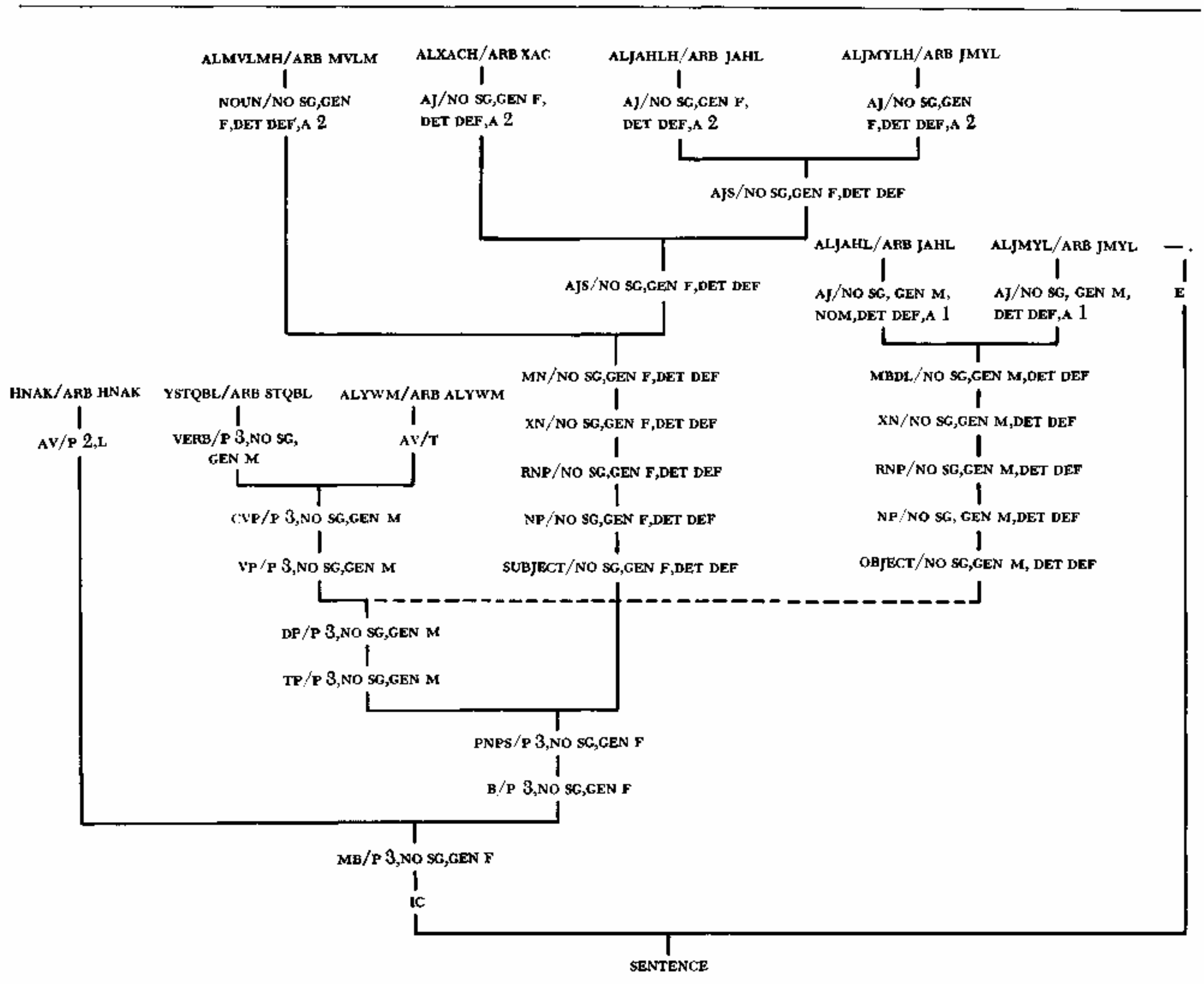


FIGURE 12.
Complete mechanical syntactic analysis of input sentence (Figure 10).

The structural transfer rule sets located in the permanent memory of the mechanism may be illustrated by two examples, one with a single subrule SENT and one with two subrules ART/NO SG,SUBJ (Figure 14).

SENT	SENT	------	INDCL+E
ART/NO SG,SUBJ	SUBJECT	NP/DET DEF	THE/-$
ART/NO SG,SUBJ	SUBJECT	NP/DET IND	AN/-$

FIGURE 14.
Two structural transfer rule sets.

The item in the first column is the lefthand side of the sentence-construction grammar rule with which the contents of the computing register match. The item in the second column identifies a specific substructure or substructures in the analysis of the input sentence located in the stimulator. These substructures will contain the information pertinent to the selection of the sentence-construction grammar subrule and are, therefore, to be bracketed.

Delimitation of the structure(s) which contain the pertinent information for the selection of the subrule is necessary since all non-pertinent input recurrences must be excluded. For example, in the present grammar the substructure noun phrase may occur in both the subject and the predicate. Most features of the noun phrase in the subject may also be reproduced in the predicate. If there were no delimiting operation there would be no means of identifying the source from which the information for the construction of the noun phrase might be unambiguously drawn. Bracketing is one means of making this identification possible.

The item(s) in the third column (Figure 14) are the items which must match constituents found in the bracketed substructures of the analysis of the input sentence if the ST subrule is to be compatible. The fourth column contains the righthand side of the sub-

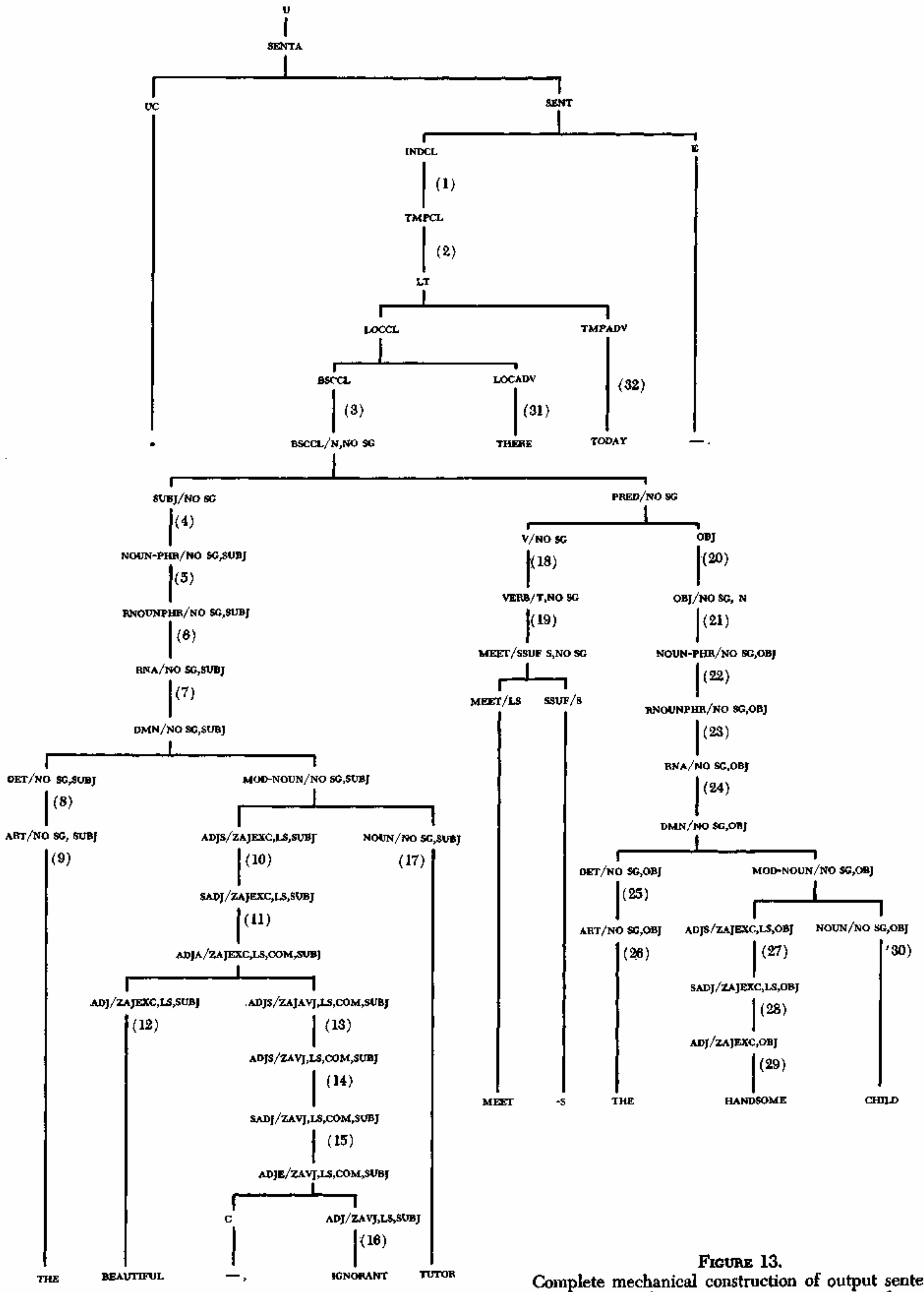
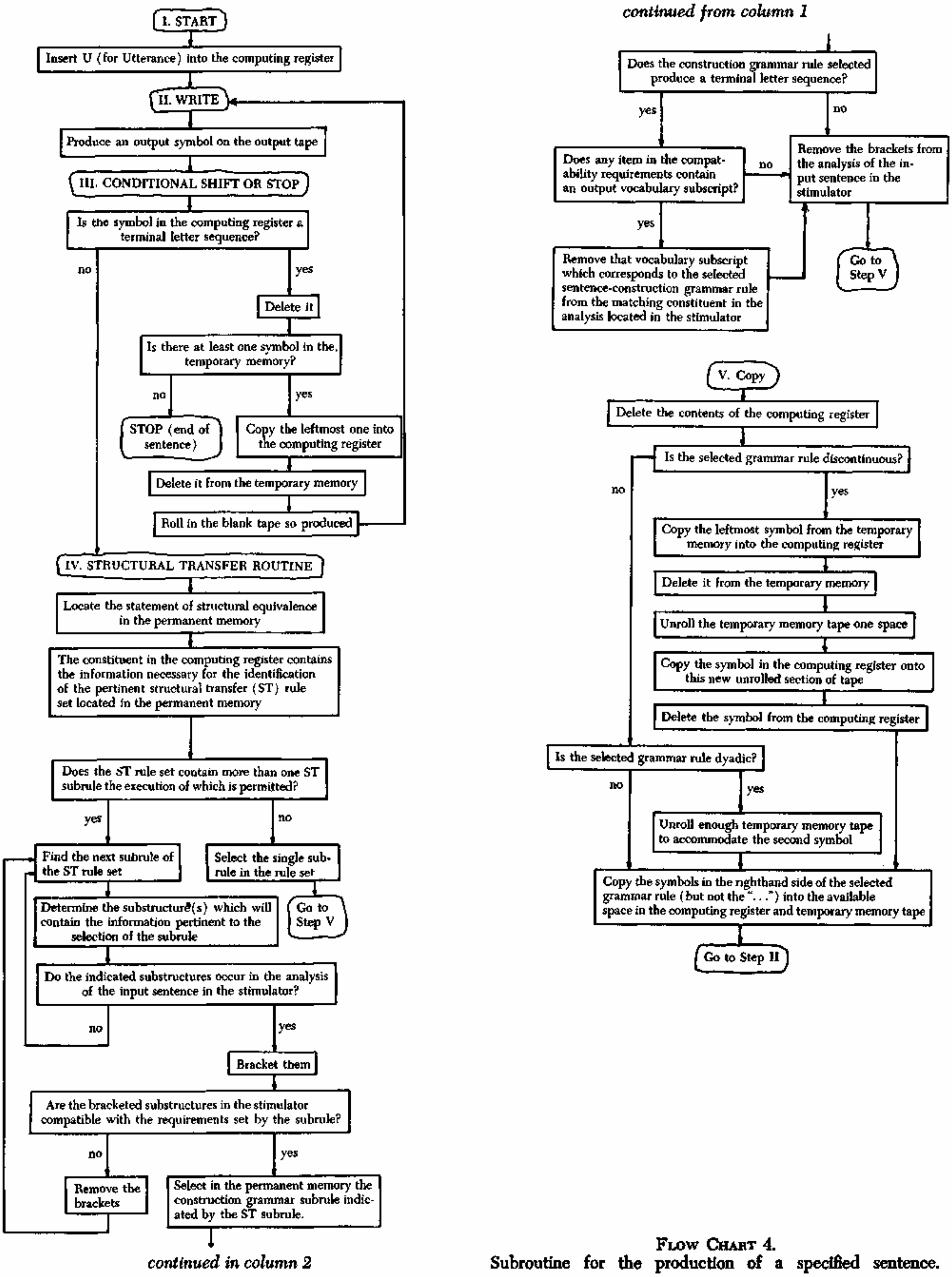


FIGURE 13.
Complete mechanical construction of output sentence which translates sentence in Figures 10 and 12.


I. START
Insert U (for Utterance) into the computing register
II. WRITE
Produce an output symbol on the output tape
III. CONDITIONAL SHIFT OR STOP
Is the symbol in the computing register a terminal letter sequence?
no
yes
Delete it
Is there at least one symbol in the, temporary memory?
no
yes
STOP (end of sentence)
Copy the leftmost one into the computing register
Delete it from the temporary memory
Roll in the blank tape so produced
IV. STRUCTURAL TRANSFER ROUTINE
Locate the statement of structural equivalence in the permanent memory
The constituent in the computing register contains the information necessary for the identification of the pertinent structural transfer (ST) rule set located in the permanent memory
Does the ST rule set contain more than one ST subrule the execution of which is permitted?
yes
no
Find the next subrule of the ST rule set
Select the single sub-rule in the rule set
Determine the substructure(s) which will contain the information pertinent to the selection of the subrule
Go to Step V
Do the indicated substructures occur in the analysis of the input sentence in the stimulator?
no
yes
Bracket them
Are the bracketed substructures in the stimulator compatible with the requirements set by the subrule?
no
yes
Remove the brackets
Select in the permanent memory the construction grammar subrule indic-ated by the ST subrule.
continued in column 2
continued from column 1
Does the construction grammar rule selected produce a terminal letter sequence?
yes
no
Does any item in the compat-ability requirements contain an output vocabulary subscript?
no
Remove the brackets from the analysis of the in-put sentence in the stimulator
yes
Go to Step V
Remove that vocabulary subscript which corresponds to the selected sentence-construction grammar rule from the matching constituent in the analysis located in the stimulator
V. Copy
Delete the contents of the computing register
Is the selected grammar rule discontinuous?
no
yes
Copy the leftmost symbol from the temporary memory into the computing register
Delete it from the temporary memory
Unroll the temporary memory tape one space
Copy the symbol in the computing register onto this new unrolled section of tape
Delete the symbol from the computing register
Is the selected grammar rule dyadic?
no
yes
Unroll enough temporary memory tape to accommodate the second symbol
Copy the symbols in the righthand side of the selected grammar rule (but not the "...") into the available space in the computing register and temporary memory tape
Go to Step II

Flow Chart 4.
Subroutine for the production of a specified sentence.

rule of the sentence-construction grammar which is to be executed if the ST subrule is found compatible with the analysis in the stimulator.

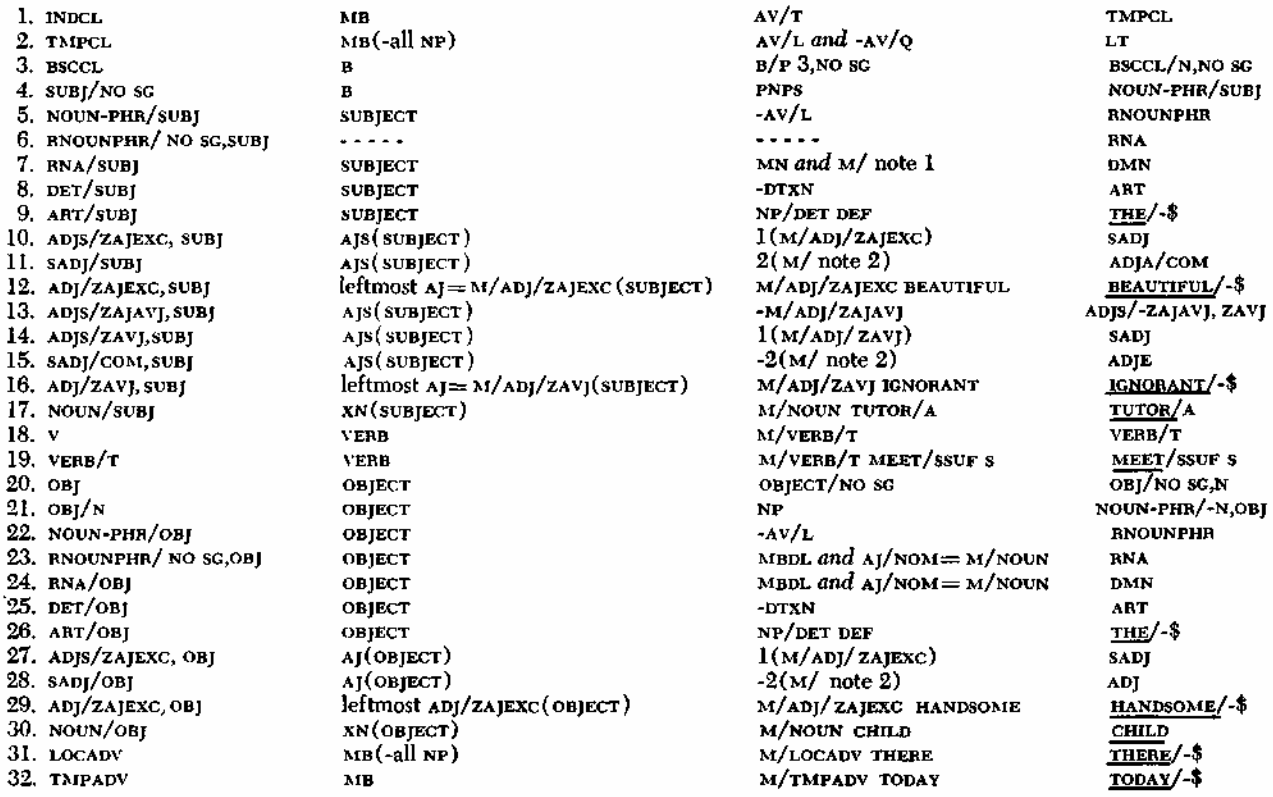
The first ST subrule (Figure 14) indicates that the rule SENT=INDCL+E is to be executed if the contents of the computing register match with SENT. In this case there is no choice. The second ST subrule indicates that the rule ART/NO SG,SUBJ=THE/-$ is to be executed if the analysis contains a substructure identified by the constitute SUBJECT and that substructure contains a definite noun phrase, NP/DET DEF.

We may now turn to an application of the routine (Flow Chart 4) to the translation of the input sentence in Figures 10 and 12. The complete production of the output sentence is presented in Figure 13 and in outline in Figure 11. Since fifty-one rules are executed in the production of this sentence, thirty-two of which are selected by structural transfer rule sets with more than one subrule, it is impractical to list all the subrules of all the structural transfer rule sets considered in the production of this sentence. The compatible subrules of the structural transfer rule sets which contain more than one subrule are presented in Figure 15 and a discussion of their more interesting features follows. Sole attention in the following is given to the operation of step IV (Flow Chart 4). Explanation and exemplification of the other steps are presented fully in my first report.

The stimulator (Figure 6) contains the total analysis of the input sentence (Figure 12) with the addition of the vocabulary subscripts (page 28, col. 2). The permanent memory contains the bilingual dictionary, the statement of structural equivalence and the sentence-construction grammar of the output language.[8] The numbers in parentheses (Figure 13) match the numbers of the ST subrules (Figure 15).

The first constituent written in the computing register is u. The structural transfer rule set in the permanent memory applicable to u contains one subrule, U=SENTA. In Figure 13 the results of executions of rule sets with only one subrule are identifiable by lack of a parenthesized number.

The constituent INDCL introduces the first rule-set composed of more than one subrule. The analysis in the stimulator does contain a modified basic clause MB a constituent of which is a temporal adverb AV/T (Figure 15). The bracketed substructure in the stimulator is therefore compatible with the requirements set by subrule 1, and the construction grammar subrule INDCL=TMPCL is executed.



1.	INDCL	MB	AV/T	TMPCL
2.	TMPCL	MB(-all NP)	AV/L *and* -AV/Q	LT
3.	BSCCL	B	B/P 3,NO SG	BSCCL/N,NO SG
4.	SUBJ/NO SG	B	PNPS	NOUN-PHR/SUBJ
5.	NOUN-PHR/SUBJ	SUBJECT	-AV/L	RNOUNPHR
6.	RNOUNPHR/ NO SG,SUBJ	- - - - -	- - - - -	RNA
7.	RNA/SUBJ	SUBJECT	MN *and* M/ note 1	DMN
8.	DET/SUBJ	SUBJECT	-DTXN	ART
9.	ART/SUBJ	SUBJECT	NP/DET DEF	THE/-$
10.	ADJS/ZAJEXC, SUBJ	AJS(SUBJECT)	1(M/ADJ/ZAJEXC)	SADJ
11.	SADJ/SUBJ	AJS(SUBJECT)	2(M/ note 2)	ADJA/COM
12.	ADJ/ZAJEXC,SUBJ	leftmost AJ=M/ADJ/ZAJEXC (SUBJECT)	M/ADJ/ZAJEXC BEAUTIFUL	BEAUTIFUL/-$
13.	ADJS/ZAJAVJ,SUBJ	AJS(SUBJECT)	-M/ADJ/ZAJAVJ	ADJS/-ZAJAVJ, ZAVJ
14.	ADJS/ZAVJ,SUBJ	AJS(SUBJECT)	1(M/ADJ/ ZAVJ)	SADJ
15.	SADJ/COM,SUBJ	AJS(SUBJECT)	-2(M/ note 2)	ADJE
16.	ADJ/ZAVJ, SUBJ	leftmost AJ= M/ADJ/ZAVJ(SUBJECT)	M/ADJ/ZAVJ IGNORANT	IGNORANT/-$
17.	NOUN/SUBJ	XN(SUBJECT)	M/NOUN TUTOR/A	TUTOR/A
18.	V	VERB	M/VERB/T	VERB/T
19.	VERB/T	VERB	M/VERB/T MEET/SSUF S	MEET/SSUF S
20.	OBJ	OBJECT	OBJECT/NO SG	OBJ/NO SG,N
21.	OBJ/N	OBJECT	NP	NOUN-PHR/-N,OBJ
22.	NOUN-PHR/OBJ	OBJECT	-AV/L	RNOUNPHR
23.	RNOUNPHR/ NO SG,OBJ	OBJECT	MBDL *and* AJ/NOM= M/NOUN	RNA
24.	RNA/OBJ	OBJECT	MBDL *and* AJ/NOM = M/NOUN	DMN
25.	DET/OBJ	OBJECT	-DTXN	ART
26.	ART/OBJ	OBJECT	NP/DET DEF	THE/-$
27.	ADJS/ZAJEXC, OBJ	AJ(OBJECT)	1(M/ADJ/ ZAJEXC)	SADJ
28.	SADJ/OBJ	AJ(OBJECT)	-2(M/ note 2)	ADJ
29.	ADJ/ZAJEXC, OBJ	leftmost ADJ/ZAJEXC(OBJECT)	M/ADJ/ ZAJEXC HANDSOME	HANDSOME/-$
30.	NOUN/OBJ	XN(OBJECT)	M/NOUN CHILD	CHILD
31.	LOCADV	MB(-all NP)	M/LOCADV THERE	THERE/-$
32.	TMPADV	MB	M/TMPADV TODAY	TODAY/-$

Note 1: an output vocabulary subscript equivalent to a rule to produce one of the classes of ADJ.
Note 2: an output vocabulary subscript.

FIGURE 15.
ST subrules compatible with the analysis of input sentence (Figure 12).

TMPCL in turn finds an ST rule set with several subrules. The substructure indicated by the applicable subrule is an MB. The entry in the second column, MB (-all NP) indicates that no information in any noun phrase occurring in any portion of the substructure MB may be used to determine the selection of the construction grammar subrule, and the NP'S are excluded from the bracketed material. The requirements set by the subrule in the third column are the presence of a locative adverb AV/L and the absence of any quantitative adverb AV/Q. The reason for the exclusion of the NP constructions must now be apparent. No locative adverb in a noun phrase construction is pertinent to the selection of the construction grammar subrule by this ST subrule. A locative adverb which is not the constituent of a NP does not occur in the analysis in the stimulator, and no quantitative adverb occurs. The substructure is compatible, and the rule TMPCL=LT is selected and executed.

ST subrule 4 adds the structural transfer subscript SUBJ to NOUN-PHR. The origin of each constitute must be kept distinct. If this were not done, ST rule 22, for example, might be selected instead of ST rule 5. In such a case, the Arabic object would be used to translate the English subject. Another way to discover the source of a constituent in the computing register is to search the constituents of the sentence so far produced. This search is impossible with the present mechanism. Eventually, however, for purposes of grammatical reference as well as for mechanical translation it will probably prove most economical to arrange for examination of these constituents.

When NOUN-PHR/NO SG,SUBJ is found in the computing register, ST subrule five brackets the substructure SUBJECT in the stimulator. The absence of a locative adverb AV/L in the NOUN-PHR construction is required by the third item of the subrule. The Arabic SUBJECT contains no AV/L, and the rule NOUN-PHR=RNOUNPHR is selected and executed.

When RNA/SUBJ, NO SG is found in the computing register, ST subrule 7 brackets the substructure SUBJECT. The symbol M/note 1 requires a word with an output vocabulary subscript which will be used to produce one of the classes of English ADJ. To meet the requirement of compatibility the third item in subrule 7 states that SUBJECT must contain both a modified noun MN and a word with one of the indicated output vocabulary subscripts. A search of the analysis finds that SUBJECT does include an MN and that two constituents of the MN contain the required vocabulary subscripts, ALJAHLH/ADJ/ZAVJ IGNORANT and ALJMYLH/ADJ/ZAJEXC BEAUTIFUL. The subrule is compatible and the rule RNA=DMN is selected and executed.

The occurrence of rules of the sort found here has forced me to program the selection of the basic output vocabulary (Flow Chart 3) before the initiation of the sentence-construction routine. If the Arabic construction MN had been derived from the phrase ALMVLMH ALXACH without further attributive adjectives and no prior regard had been paid to the output vocabulary, examination at this point would have to be made in order to determine whether the Arabic MN was to be translated by an English DN, 'the tutor' for example, or a DMN 'the special teacher.' To determine the choice of construction one might, at this point, examine the vocabulary required in the translation of the Arabic MN. I feel it is more economical to divide the structural transfer routine and the statement of structural equivalence into the two parts previously discussed.

The DTXN construction, the absence of which is required for the compatibility of ST subrule 8, contains as one constituent a demonstrative adjective. If it had occurred, the subrule DET=DEM would have been selected rather than the subrule DET=ART. The selection of the subrule DET=ART cannot be made on the basis that the nucleus of the Arabic construction is definite since it is definite in construction both with and without a demonstrative adjective.

With ART/NO SG, SUBJ in the computing register the bracketed substructure SUBJECT in the stimulator does have NP/DET DEF as a constituent. Rule nine is, therefore, compatible and the subrule ART=THE/-$ is selected. THE is the first terminal letter sequence produced. The item in the third column, in which the compatibility requirements are stated, is NP/DET DEF. This item does not contain an output vocabulary subscript. The brackets are removed from the stimulator and the selected sentence-construction grammar rule is executed.

The substructure indicated by the second item in subrule ten AJS(SUBJECT), is to be read “an adjective sequence which occurs as a constituent of the SUBJECT construction.” This substructure does occur in the analysis of the input sentence. The third item 1(M/ADJ/ZAJEXC) is to be read “one and only one word with an output vocabulary subscript the term of which is ADJ/ZAJEXC must occur in the bracketed substructure.” This is the first use of the number one, which is to be read “one and only one.” Compatibility does occur, and the construction grammar subrule ADJS/ZAJEXC=SADJ is executed.

SADJ/ZAJEXC, SUBJ, LS is next found in the computing register, and AJS(SUBJECT) is bracketed. The compatibility requirement in the third column 2 (M/note 2) is read “at least two words with an output vocabulary subscript must occur in the bracketed substructure.” The compatibility requirement is met and the rule SADJ=ADJA/COM is selected and executed.

When ADJ/ZAJEXC, SUBJ, LS is found in the computing register, the structural transfer subrule twelve indicates that the pertinent substructure is the leftmost adjective construction AJ which contains a word with an attached vocabulary subscript the term of which is ADJ/ZAJEXC. The compatibility requirement is met since the vocabulary subscript of this word is ADJ/ZAJEXC BEAUTIFUL (p. 28). The grammar subrule ADJ/

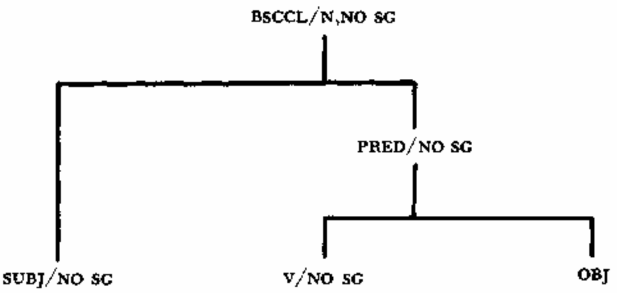


Figure 16.
Production of number concord between subject and verb.

ZAJEXC=BEAUTIFUL/-$ is, therefore, selected. This subrule produces a terminal letter sequence, BEAUTIFUL, and the item in the compatibility requirements, M/ADJ/ZAJEXC BEAUTIFUL/-$ contains an output vocabulary subscript. This vocabulary subscript corresponds to the selected grammar construction rule ADJ/ZAJEXC =BEAUTIFUL/-$. Therefore, the vocabulary subscript is deleted from the matching constituent ALJMYLH/ARB JMYL, ADJ/ZAJEXC BEAUTIFUL.

The compatibility requirement in ST subrule 15, -2(M/note 2), is read "less than two words with a translation subscript must occur in the bracketed substructure."

Subrule 23 is located when the computing register contains RNOUNPHR/NO SG, OBJ. The substructure pertinent to the selection of the construction grammar subrule has OBJECT as its constitute and must contain a MBDL constituent to meet one of its requirements. The second requirement, AJ/NOM=M/NOUN, states that it also must contain an adjective nucleus construction AJ/ NOM which contains a word with an output vocabulary subscript the term of which is NOUN. This requirement is met by ALJAHL/ARB JAHL, NOUN CHILD (page 28). The adjective JAHL furnishes an example of an input language adjective which, when nucleus of a noun phrase, is translated as an output language noun. The remaining steps in the execution of the structural transfer and sentence-construction routines parallel steps already discussed.

One problem illustrated by the translation of this sentence involves the treatment of difference in the handling of inflection in the input and output languages. In the theory of grammar applied in the present program, inflectional categories are produced by the attachment of subscripts to grammatical symbols. These subscripts are carried from the constitute to its constituents unless specifically deleted. Concord between subject and verb is produced by the attachment of the pertinent inflectional subscripts to a constitute common to both the subject and the verb (Figure 16).

The Arabic grammar distinguishes between singular and plural first and second person verbs. As a result an Arabic first person singular, for example, will be translated by the English pronoun 'I', classed as grammatically plural in the grammar used in the current mechanical translation program. If the problem ended here, one would be able to select the construction grammar subrule by reference to the person and number of the Arabic verb. The situation is, however, further complicated. If the Arabic sentence contains a SUBJECT, *fā'ilu-l-fi'l,* the number of the verb is always singular, for example YVRFH ALWLD 'the boy knows him' and YVRFH ALAWLAD 'the boys know him'. As a result of these syntactic peculiarities, the number information pertinent to the production of the English sentence is only completely gathered with the identification of the Arabic basic clause constitute B. If the Arabic sentence contains a SUBJECT, the constitute B derives gender and number from it and person from the verb, and only otherwise does it derive gender and number as well as person from the verb. As subrule 3 (Figure 15) indicates, if the constituent B is third person singular, a singular basic clause BSCCL/NO SG must be produced in English, otherwise a plural BSCCL is produced. The discussion shows that an analysis of the Arabic input at a high syntactic level is required to translate the Arabic verb.

The treatment of difference of structure in the two grammars may be illustrated by reference to ST subrule four (Figure 15). The limited English grammar in the program always produces a subject, either a noun phrase or a subjective pronoun SP (Figures 18 and 20). The Arabic may produce a basic clause B with or without a subject (Figures 17 and 19).

If the Arabic analysis of B contains a predicate with a noun phrase subject, PNPS, then the English subrule SUBJ=NOUN-PHR is selected, otherwise the English subrule SUBJ = PS is selected and executed (Figure 15, subrule 4).

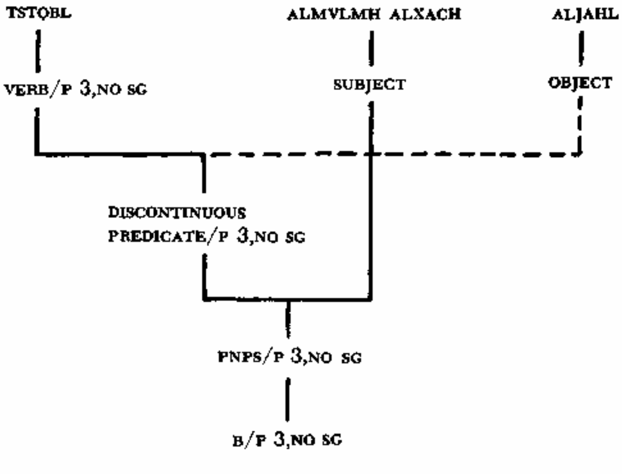


Figure 17.
The Arabic basic clause with a subject: 'the tutor meets the child.'

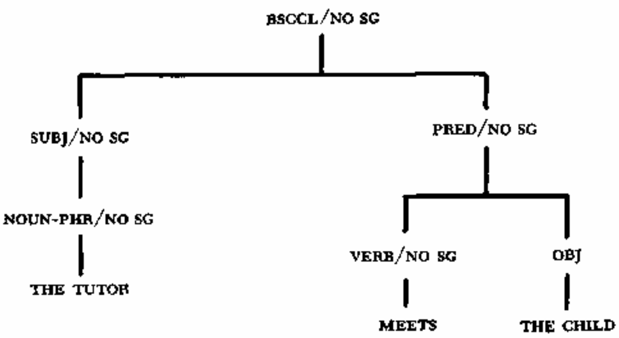


FIGURE 18.
The English basic clause with a noun-phrase subject.

The portion of the statement of structural equivalence which refers to the noun phrases offers the richest variety of differences in structure between the two languages furnished by the present program. A noun phrase composed of an adjective nucleus AJ/NOM may be translated as a determined adjective or a determined modified noun DMN: ALYWNANYWN 'the Greeks' but ALXACWN 'the special ones'. A NOUN may be translated as a determined noun DN or a DMN: ALMVLMH 'the teacher' but ALXACH 'the special officials'. A modified noun MN may be translated as a DN or a DMN: ALMVLMH ALXACH 'THE TUTOR' BUT ALMVLMH ALJAHLH 'the ignorant teacher'. A demonstrative pronoun is translated as a determined noun or a demonstrative pronoun: H+DA 'this one' but H+WLAO 'these'.

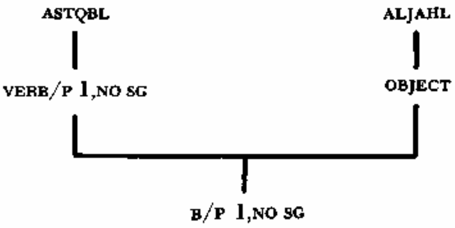


FIGURE 19.
The Arabic basic clause without a subject: 'I meet the child.'

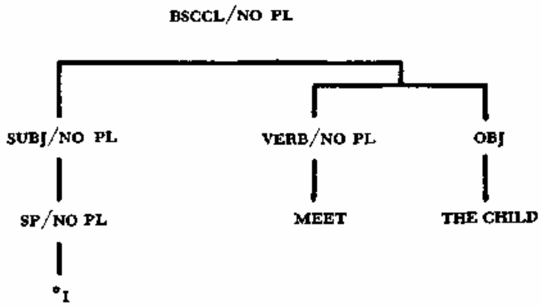


FIGURE 20.
The English basic clause with a pronoun subject.

One more example will illustrate further the capability of the system to handle the translation of a single structure into two quite different constructions. The following two sentences have exactly the same structure in Arabic, but the object of the second is translated by the English subject: Y+HB AL+HRMH 'he likes the woman' and YVJB AL+HRMH 'the woman likes him' or more literally 'he-is-pleasing-to the woman.'

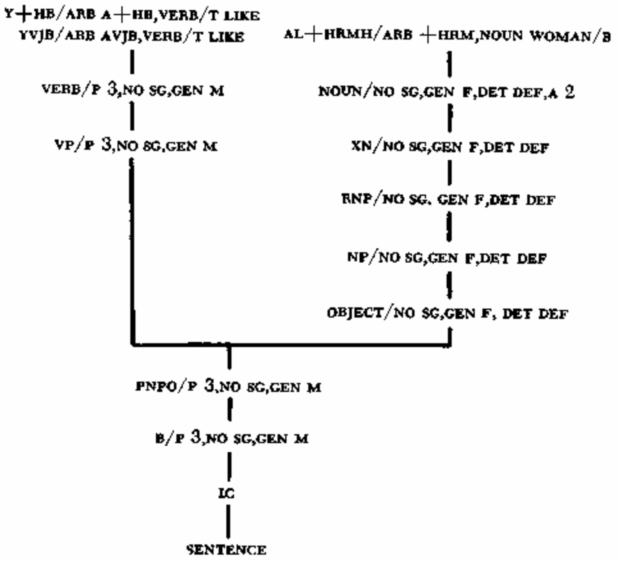
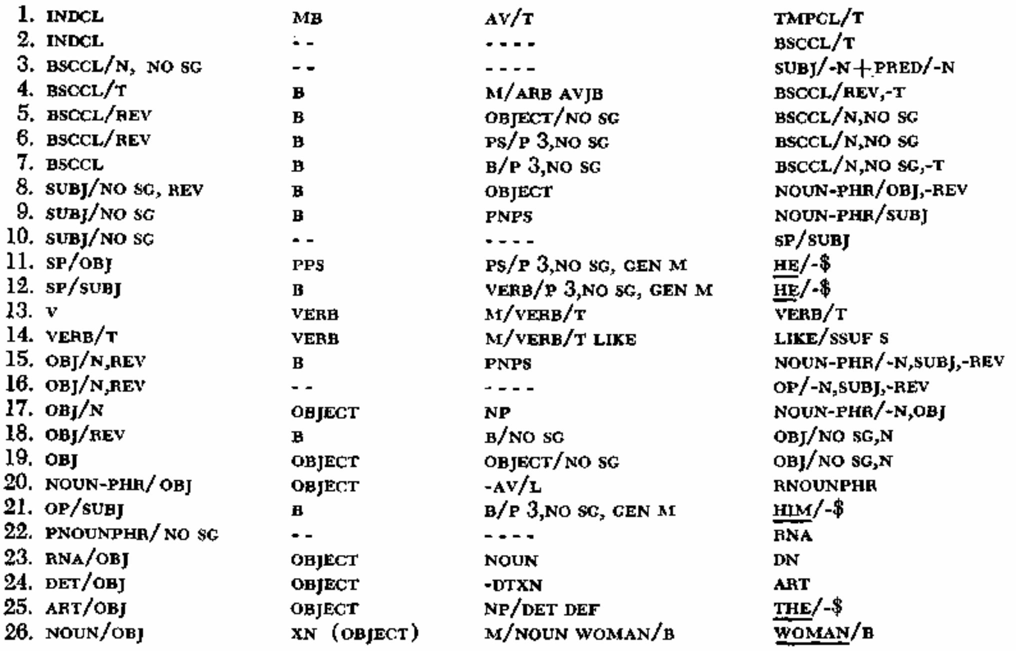


FIGURE 21.
Analysis of the two sentences Y+HB AL+HRMH and YVJB AL+HRMH.

The translation of the first sentence can be called parallel to its input sentence in that the subject is translated by the subject and the object by the object. The translation of the second sentence, however, must be carried out by translating the subjective affix into the objective pronoun and the object as subject.

The construction of the first sentence (Figure 23) proceeds as the constructions discussed previously. The second input sentence differs from the first only in the choice of the verb, YVJB contrasted with Y+HB. The difference in construction of the two output sentences is fixed with the expansion of BSCCL/T after the construction subrule has been selected by the structural transfer rule 4 (Figure 22). This rule brackets the construction identified by the constituent B in the analysis in the stimulator (Figure 21). One of the constituents of this substructure is the word YVJB with the subscript ARB AVJB. The subrule adds the structural transfer subscript REV. After the constituent BSCCL/REV is copied into the computing register, the constitute B in the analysis is again bracketed by ST subrule 5 and a singular object is identified. The subrule is compatible and BSCCL/NO SG is produced. If subrule 5 were found



1. INDCL	MB	AV/T	TMPCL/T
2. INDCL	- -	- - - -	BSCCL/T
3. BSCCL/N, NO SG	- -	- - - -	SUBJ/-N+PRED/-N
4. BSCCL/T	B	M/ARB AVJB	BSCCL/REV,-T
5. BSCCL/REV	B	OBJECT/NO SG	BSCCL/N,NO SG
6. BSCCL/REV	B	PS/P 3,NO SG	BSCCL/N,NO SG
7. BSCCL	B	B/P 3,NO SG	BSCCL/N,NO SG,-T
8. SUBJ/NO SG, REV	B	OBJECT	NOUN-PHR/OBJ,-REV
9. SUBJ/NO SG	B	PNPS	NOUN-PHR/SUBJ
10. SUBJ/NO SG	- -	- - - -	SP/SUBJ
11. SP/OBJ	PPS	PS/P 3,NO SG, GEN M	HE/-$
12. SP/SUBJ	B	VERB/P 3,NO SG, GEN M	HE/-$
13. V	VERB	M/VERB/T	VERB/T
14. VERB/T	VERB	M/VERB/T LIKE	LIKE/SSUF S
15. OBJ/N,REV	B	PNPS	NOUN-PHR/-N,SUBJ,-REV
16. OBJ/N,REV	- -	- - - -	OP/-N,SUBJ,-REV
17. OBJ/N	OBJECT	NP	NOUN-PHR/-N,OBJ
18. OBJ/REV	B	B/NO SG	OBJ/NO SG,N
19. OBJ	OBJECT	OBJECT/NO SG	OBJ/NO SG,N
20. NOUN-PHR/OBJ	OBJECT	-AV/L	RNOUNPHR
21. OP/SUBJ	B	B/P 3,NO SG, GEN M	HIM/-$
22. PNOUNPHR/NO SG	- -	- - - -	RNA
23. RNA/OBJ	OBJECT	NOUN	DN
24. DET/OBJ	OBJECT	-DTXN	ART
25. ART/OBJ	OBJECT	NP/DET DEF	THE/-$
26. NOUN/OBJ	XN (OBJECT)	M/NOUN WOMAN/B	WOMAN/B

FIGURE 22.
Structural transfer rules for sentences in Figure 21.

to be incompatible, then subrule 6 would have found a pronominal suffix PS in the substructure B. If this were third person singular, the BSCCL/NO SG would have been produced. These rules determine the source from which the information concerning the number of the English basic clause BSCCL must be drawn and assign the proper inflectional number. If a verb like YVJB is found in the input, then the information is drawn from the object of that verb. Otherwise, it is drawn from the subject. In a larger program other sources of information might have to be examined.

Ambiguity

The occurrence of ambiguity presents one of the more serious problems in mechanical translation. Ambiguity, in the context of translation, occurs in any situation in which an expression in one language, the *ambiguous expression,* may be rendered by two or more equivalent expressions with different meanings, the *discriminating expressions,* in the other. For example, English 'you meet him' is equivalent to any one of the following Arabic words depending upon the number of people addressed and their sexes: TSTQBLH, TSTQBLYNH, TSTQBLANH, TSTQBLWNH and TSTQBLNH.

If the ambiguous expression is in the output language, the precise meaning of the discriminating expression may be left unexpressed. All of the Arabic words in the example above may be translated indiscriminately 'you meet him'. The problem is handled this way in the present program. On the other hand, ambiguities of this sort may be resolved either grossly by adding some grammatical indication to the output or more subtly by a circumlocution at a suitable point in the total translation.

If the ambiguous expression is in the input language, resolution of the ambiguity is dependent upon the context available for examination. Given a sufficiently expanded context it is probable that many if not most ambiguities can be solved. If in English, considered as an input language, the context is restricted to 'flying planes can be dangerous', the clause is ambiguous with regard to the category, verbal noun or adjective, to which 'flying' is to be assigned. If the context is expanded so that the entire sentence 'flying planes can be dangerous but it is profitable' is available for inspection, 'flying' must be categorized as a verbal noun.

The size of the context required for the resolution of an ambiguity will depend upon the constituents of which the ambiguity is composed, but the limits are exceedingly broad. The ambiguous interpretation of a morpheme may be resolvable within the word of which it is a constituent. On the other hand, the examination of a book-length text may be unable to resolve other ambiguities. For example, Arabic MDYR may be translated 'principal (of a school)', 'director (of a company)' or even 'the person who pushes (a coffee-wagon)'. It is possible to imagine that the resolution of the ambiguity may be possible only through reference to a proper name. ALMDYR +HSN may be translated 'principal Hasan', ALMDYR QASM 'director Qasim', and ALMDYR ABRAHYM 'coffee-boy Ibrahim'. Knowledge of the actual occupation of each individual

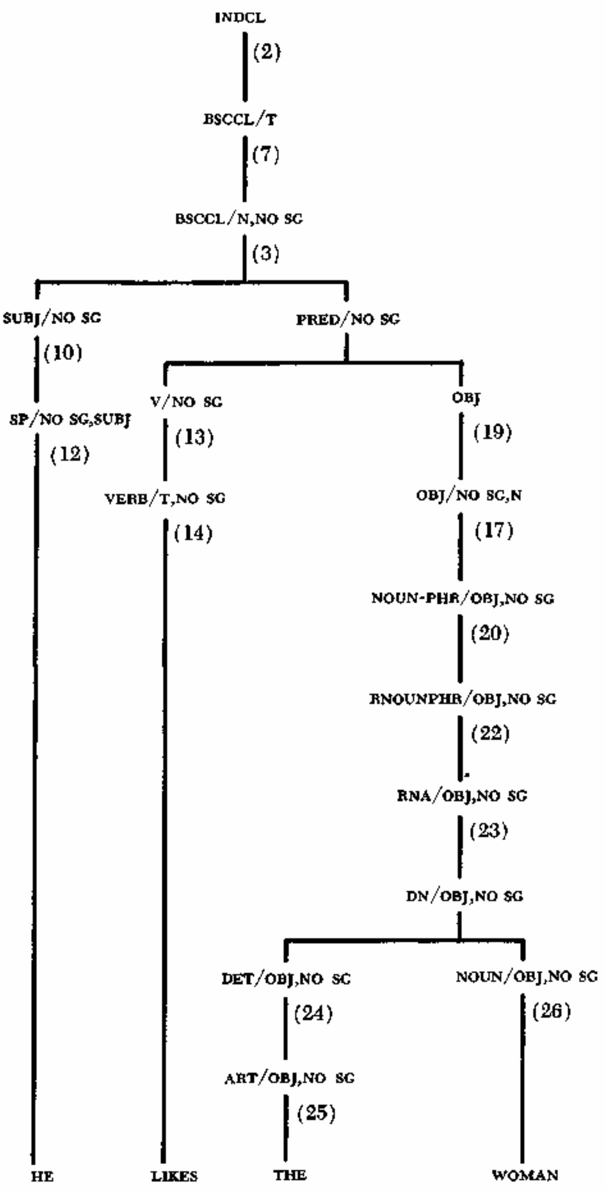


Figure 23.
Construction of the mechanical translation of
Y+HB AL+HRMH.

at the time referred to may be the only means of solving the ambiguities represented by such phrases. It is indicated that the contexts within which at least some ambiguities may be solved must include features of general knowledge. It is conceivable that the mechanism be given reference to an encyclopedia to aid in the solution of such problems. It is furthermore conceivable that the computer be able to add to this generalized knowledge through information derived from the text to be translated. The practical and general achievement of solutions by these means, however, appears to be beyond our present capacities.

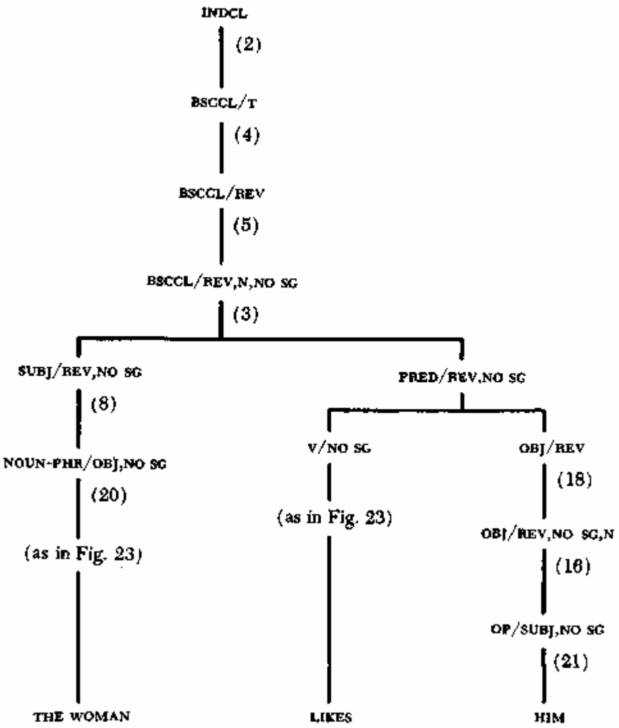


Figure 24.
Construction of the mechanical translation of
YVJB AL + HRMH.

The current program has undertaken to resolve those ambiguities the information for the resolution of which may be found within the limits of the sentence. For example, TSTQBL may be translated as either 'meets', 'meet', 'you meet' or 'she meets'. This potential ambiguity is resolved in the program by reference to context as in the sentence TSTQBL ALBNT ALWLD, 'the girl meets the boy.' On the other hand, context does not give sufficient information to resolve completely the ambiguity in TSTQBL ALBNT, where TSTQBL may be translated either as 'you meet' or 'she meets'. Currently the program selects one of the translations at random and produces that one.

Projected Research

Considerable thought has been given to the utilization of the sporadically occurring diacritics in the Arabic orthography for the resolution of ambiguity. It is now felt that such utilization is possible through enabling the computer first to parse the words without diacritics and then to consider the compatibility of any occurrent diacritics with the alternate interpretations of the original parsings. Parsings incompatible with the diacritics can be eliminated and further information derivable from the occurrent diacritics can be added to the various parsings. By this method WLD would be parsed as follows:

/walad-/	singular noun	'boy' *or*
	1st measure, past active verb	'begot'
/wald-/	singular noun	'birth'
/wuld-/	plural noun	'boys'
/wulid-/	1st measure, past passive verb	'was born'
/wallad-/	2nd measure, past active verb	'generated'
/wullid-/	2nd measure, past passive verb	'was generated'

The occurrence of one or more diacritical marks would appreciably reduce the number of parsings compatible with the input text. Addition of a subroutine of this sort should not be overly difficult and would add to the efficiency of the sentence-recognition grammar.

[1] Yngve, Victor H., “A Framework for Syntactic Translation,” *Mechanical Translation 4,* December, 1957, p. 59.

[2] Yngve, Victor H., “A Model and an Hypothesis for Language Structure,” *Proceedings of the American Philosophical Society* 104, October, 1960, pp. 444-446.

[3] Satterthwait, Arnold C., *Parallel Sentence-Construction Grammars of Arabic and English,* Massachusetts Institute of Technology, Research Laboratory of Electronics, 1962, pp. 18-37, 61-68,

[3a] Satterthwait, Arnold C., “Computational Research in Arabic,” *Mechanical Translation 7,* August, 1963 pp. 62-70.

[4] Knowlton, Kenneth C., *Sentence Parsing with a Self-Organizing Heuristic Program,* Massachusetts Institute of Technology, Research Laboratory of Electronics, 1963, pp. 1-5.

[5] Garvin, Paul L., “Syntactic Retrieval,” *Proceedings of the National Symposium on Machine Translation,* H. P. Edmundson, ed., Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961, p. 290.

[6] Satterthwait, Arnold C., *Parallel Sentence-Construction Grammars of Arabic and English,* pp. 191-218, 262-270.

[7] *Ibid.,* 22-24.

[8] *Ibid.,* 191-218, 262-270.