[*Mechanical Translation*, Vol.7, no.2, August 1963]

Computational Research in Arabic

by Arnold C. Satterthwait*, Research Laboratory of Electronics,
Massachusetts Institute of Technology

In the preparation of an Arabic to English sentence-for-sentence mechanical translation program, a computer has been applied to the testing of statements concerning various phases of the morphological and syntactic structure of Arabic and structural equivalences between Arabic and English, This paper discusses briefly the testing procedure used, the grammatical statements tested in the form of an Arabic sentence-construction grammar programmed for a computer and some results of the testing procedure applied to sentences randomly composed in Arabic by the computer.

The research in Arabic at M.I.T. has been strongly influenced by an original interest in the preparation of a computer program for the translation of Arabic to English. It was initially recognized that it would be quite impossible to prepare, in one step, a mechanical translation program which would be capable of translating *any* modern Arabic text into English. It was, therefore, decided to make a general study of the problems of mechanical translation by attempting to produce a workable computer program which would translate a limited corpus of Arabic. The corpus which the program would be capable of translating was to be defined by a restricted sentence-construction grammar which, when programmed for a computer, produced random sentences in Arabic. The sentences which the resulting program is potentially able to produce represent only a very limited number of all the sentences capable of being composed in Arabic. However, the corpus described by the grammar, limited even to the extent to which this one is, is much too extensive to be listed.

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

As a first attempt in writing an Arabic sentence-construction grammar of this type, the rules were so written as to produce only strictly grammatical sentences. Since one of the purposes of the computer program was to facilitate the study of the extreme limits of grammaticalness, it was soon realized that the grammar would have to be rewritten to allow the production of unusual and rare constructions. It was agreed that this program might, on occasion, produce sentences beyond the limits accepted even after an extremely tolerant definition of grammaticalness had been admitted. As a result of this rewriting of the program, the computer has produced a satisfying number of sentences which do push to the borders of grammaticalness and perhaps trespass beyond. As will be seen from the examples quoted in the following pages, the computer has produced sentences and sequences which could never have been imagined by the writer of this paper nor perhaps by any native-speaker of Arabic. Many of these sentences can be only termed nonsense sentences. Such sentences do, however, inspire numerous questions which in turn serve as stimuli for further research. Such being one purpose of the program, it is to be hoped that the sentences presented in this paper will be read with this point kept constantly in mind.

As a result of the various freedoms and restrictions built into the program, the grammar produces sentences of the following type.

> /?a 1 yawma yastaqbilu 1 ?awlaadu 1 kibaaru haa?ulaa?i tilka 1 mu9allimaati 1 xaaşşaati 1 jamiylata hunaaka./
> 'These big boys will meet those beautiful tutors there today.'

Since the primary goal of all this research is the production of a program for the translation of contemporary literary Arabic into English, the Arabic, whether gencrated or to be translated, is represented in the strictly consonantal orthography without indication of vowels or other diacritical marks. In addition, no distinction is made between the *hâ'* and *tâ' marbûtah* nor between the final *yâ'* and the *alif bi-şûrati-l-yâ'*. On the other hand, distinctions are made between the *hamzah* without a *kursi* and the *hamzahs* with *wâw* and *yâ'* serving as their supports.

As all material must be key-punched for submission to the computer, a special one-to-one transliteration of the Arabic is made for this purpose.

Fig. 1 illustrates the types of ambiguities which the transliteration causes.

In the first two examples the *dammatayn* /-un/ and

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

Key-punch transliteration	*Consonantal orthography*	*Full orthography*	*Transcription*
1. WLD	ولد	وَلَدٌ	/waladun/
2. WLD	ولد	وَلَدٍ	/waladin/
3. H+DA	هذا	هٰذَا	/haaḏaa/
4. WLDA	ولدا	وَلَدًا	/waladan/
5. +HRMH	حرمه	حُرْمَةٌ	/ḥurmatan/
6. YKRH	يكره	يَكْرَهُ	/yakrahu/
7. KSLY	كسلى	كَسْلَى	/kaslaa/
8. ALCYNY	الصين	الصِّينِيُّ	/ʔa ṣ ṣiyniyyu/
9. ALAN	الان	الْآنَ	/ʔa l ʔaana/
10. LWA+WKM	لواؤكم	لِوَاؤُكُمْ	/liwaaʔukum/
11. LWA+YKM	لوائكم	لِوَائِكُمْ	/liwaaʔikum/
12. LWAOKM	لواءكم	لِوَاءُكُمْ	/liwaaʔakum/

Figure 1

the *kasratayn* /-in/ are not distinguished. The final *alif* with the *fathatayn* /-an/ is not distinguishable from any other final *alif,* as illustrated by items three and four. The only indication of the *fathatayn* occurs when the final *alif* is *written.* Their occurrence and omission are illustrated in items four and five. The *tâ marbṭah* /-at-/ and the final *hâ'* /-hu/ are not differentiated in the fifth and sixth words. The *alif bi-şûrati-l-yâ'* /-aa/ and the *yâ'* /-iyyu/ are not differentiated in the seventh and eighth words, and the *maddah* /-aa-/ is not indicated in the ninth. The last three words illustrate grammatical differences, phonemically vocalic, but adequately identified in the strictly consonantal orthography.

The writer's research in Arabic at M.I.T. may be divided into four main areas: the sentence-construction and recognition grammars of Arabic, the sentence-construction grammar of English and the statement of structural equivalences between the two languages.

As has been noted above, the sentence-construction grammar of Arabic defines the corpus which the mechanical translation program is designed to translate. One test of the translation program is the requirement that it be able to translate any sentences produced by the computer under control of this sentence-construction program.

The statement of structural equivalences is composed of a set of rules which indicate the English equivalences of the Arabic. For example, the sequence of letters WLD, once they have been identified as a noun, are equated with the English noun, 'boy'. We cannot equate the letters WLD with 'boy' until the grammatical identity 'noun' has been made, either explicitly or implicitly. In computer research, the identification must be explicit. If grammatical identifications were not first made, we might find, for example, the letters XAC in A+SXAC /ʔasxaaş/ 'persons' equated with XAC /xaaşş/ 'special' rather than with the plural of +SXC /saxş/.

Furthermore, the statement of structural equivalences is not restricted to stating equivalences between words or portions of words. If each grammatical construction in a source language were exactly parallel to an equivalent construction in a target language, it might be possible to limit the statement of structural equivalences to the vocabulary. This parallelism, however, rarely occurs exactly. Constructions which are not parallel may be equivalent. An example may be given here to indicate the use of the word 'parallel' in this context.

The Arabic prepositional phrase construction as exemplified by */fî l bayti/* may be described as approximately parallel to the English prepositional phrase construction 'in the house'. In both languages the construction 'prepositional phrase' may be partially described as 'preposition plus noun-phrase'. These phrases are also structurally equivalent in that an English prepositional phrase may be used to translate an. Arabic prepositional phrase. On the other hand, the Arabic modified noun phrase construction /ʔa l mu9allimu l xaaşşu/ (literally 'the special teacher') is equivalent to the English noun 'tutor' without a descriptive adjective attribute, thus failing to parallel the Arabic construction. While these two expressions are termed equivalent, they are not parallel.

The third area of research in Arabic deals with the construction of a recognition grammar of Arabic. A grammar of this type has been programmed to direct the computer in the analysis of any sentence produced

by the sentence-construction grammar of Arabic or by a speaker of Arabic, keeping himself within bounds defined by rules which describe the limits of the grammar's capability.

A sentence such as

هَنَاكَ يَخُونُ ٱلتُّجَّارَ ٱلْمُضْرِبُونَ ٱلْأَعْيَانَ ٱلصِّينِيِّينَ .

is punched out on a card and submitted to the computer under control of the program. The computer produces a one-dimensional analysis in V. H. Yngve's adaptation of the Polish parenthesis-free algebraic notation.[1] The one-dimensional analysis may be mechanically transformed to a two-dimensional tree-structure. Fig. 2 shows the tree-structure which is equivalent to the one-dimensional analysis produced by the computer. In the computer analysis each node has at least one construction name with subscripted grammatical information, but for the sake of brevity, only a minimum of the construction names devoid of subscripted information is indicated here.

The recognition grammar of the source language is necessary for mechanical translation. As has been indicated in the discussion of structural equivalences, no successful translation at even a very low level of style can be obtained solely by word substitution.

Several examples of some of the mechanical translation problems which have been solved by a thorough grammatical analysis of the sentences may be illustrated by some of the noun phrases in Fig. 3. In the first two phrases, the number and gender of the phrase may be derived from the nuclei CYNY 'Chinese (masc. sg.)' and CYNYAT 'Chinese (fem, pl.) .'

In the English of the first two phrases, the number of the nucleus 'Chinese' is indeterminate. The number of the English noun phrases is made explicit by the number of the demonstrative adjectives, 'this' and 'these'. The number of the English demonstratives, however, cannot be determined by reference to the number of the

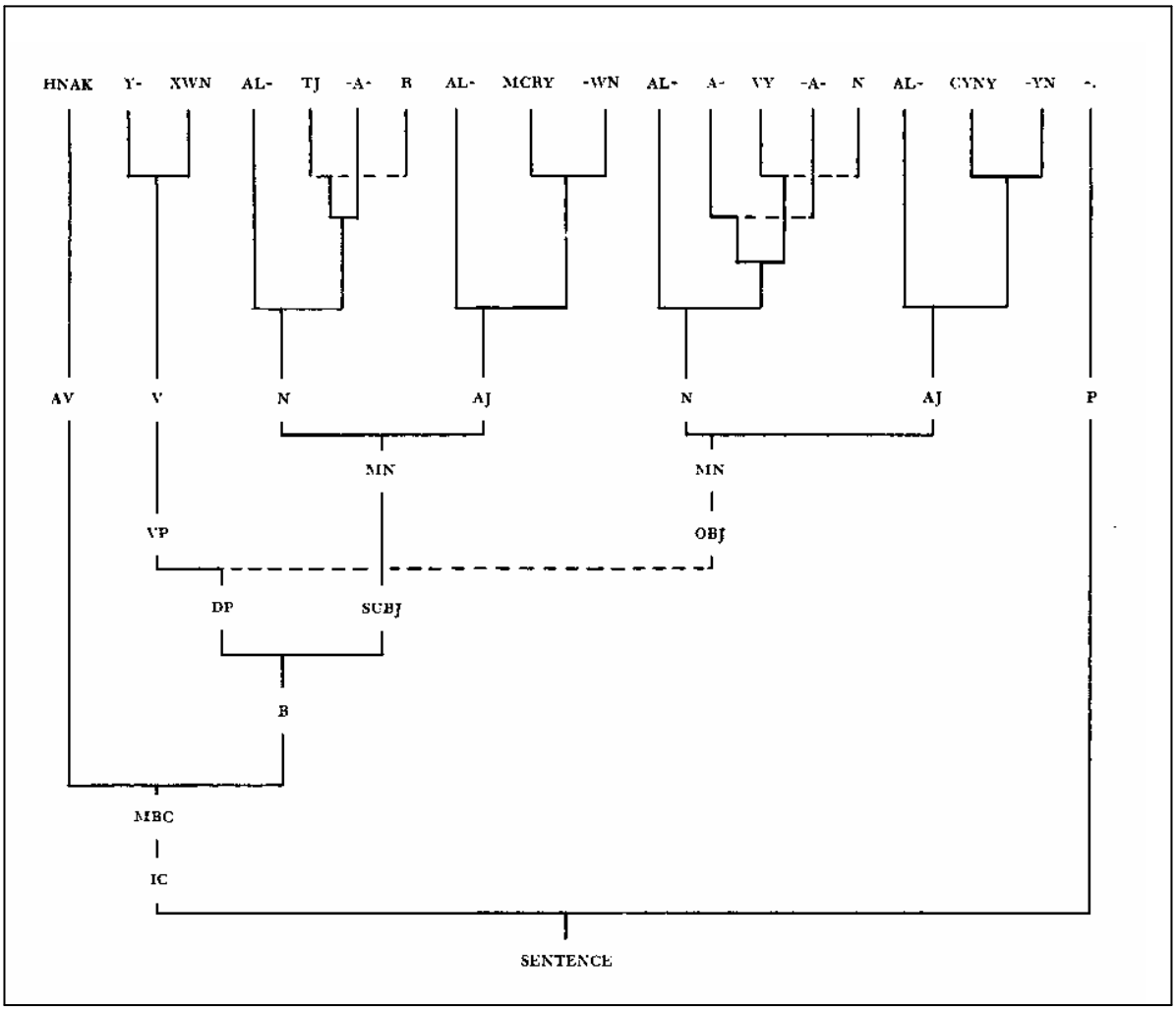

Figure 2

H + DA ALCYNY ALJMYL	this handsome Chinese
H+DH ALCYNYAT ALJMYLH	these beautiful Chinese
KANT ALVRB JMYLH	the Arabs were handsome
KANT ALVHBYAT JMYLH	the carts were beautiful
BYT W+HDYQH ALWZYR	the minister's house and garden
BYT ALWZYR W + HDYQTH	the minister's house and garden

/haaða ş şiyniyyu 1 jamiylu/
/haaðihi ş şiyniyyaatu 1 jamiylatu/
/kaanati 1 9arabu jamiylatan./
/kaanati 1 9arabiyyaatu jamiylatan./
/baytu wa hadiyqatu 1 waziyri/
/baytu 1 waziyri wa hadiyqatuhu/

Figure 3

Arabic demonstratives, which are both singular. It is the number of the Arabic *noun phrase* which determines the number of the *English* noun phrase. The selection, then, of the appropriate forms of the noun and the demonstrative is determined by the number of the *English* noun phrase.

The second pair of sentences illustrates the need for a thorough grammatical analysis in order that the correct translation of the word JMYLH 'handsome' or 'beautiful' may be selected. Only reference to the subject in these cases will enable us to decide whether the predicate complement should be translated as 'handsome' or 'beautiful'. A complete syntactic analysis seems to be needed, first to identify the subject and second to determine its gender and animation. If the noun phrase of which the feminine adjective JMYLH is a constituent or to which it is a predicate complement is masculine *and* animate, then JMYLH will be translated as 'handsome'. Otherwise, it will be translated as 'beautiful'.

At least a partial grammatical analysis of the source language is generally recognized as essential for mechanical translation. Partial analyses of several languages are currently being used experimentally. The research being carried on at M.I.T. supplies a fairly complete grammatical analysis of a restricted segment of Arabic grammar. The question as to whether such a complete analysis is essential for mechanical translation cannot yet be answered.

No matter how this question is finally answered, however, a thorough mechanical recognition grammar of a language will be a valuable tool for linguistic research. Entire texts may be analyzed by computer. Such analyses without further work might be of small value. The electronic computer allows us, however, to ask innumerable questions of these analyses. For example, we might learn whether or not the construction of the type represented by the fifth item in Fig. 3 in contrast with the construction of the type represented by the sixth item occurs, and if so what the comparative rates of occurrence of the two types of construction are. We might wish to learn the comparative frequency of occurrence of nominal sentences compared with the frequency of occurrence of verbal sentences. Or, all the relative clauses in a text might be gathered and listed in relation to their various environments, and so on almost indefinitely. Once the analyses are made, they will be permanently available for linguistic research. So far, research of the type suggested by the use of a computer has been extremely limited; and when undertaken, only carried out at the expense of great human labor.

The fourth area of research is concerned with the preparation of a sentence-construction grammar of English, equivalent to the sentence-construction grammar of Arabic. 'Equivalence' in the sense used here means that any sentence produced by the grammar of one language may be matched with at least one sentence produced by the grammar of the other, in such a way that one will be accepted as a translation of the other.

The remainder of the paper will discuss the sentence-construction grammar of Arabic briefly, and will examine some of the sequences produced by the computer under the control of the program in which the grammar is incorporated. The problem is to write a set of rules which may be manipulated so as to select a set of allomorphs from a list of allomorphs such as that found in Fig. 4, and then to arrange them in grammatical and only grammatical sequences. An example of such a sequence would be YSTQBL ALBNT AL+HRMH, /yastaqbilu 1 binta 1 hurmatu/, 'the woman meets the girl.' A full discussion of one type of grammar designed to solve this problem is furnished in the writer's report, *Parallel Sentence-Construction Grammars of Arabic and English.*[2]

+HRM	'woman'	AL-	'the'
WLD	'boy'	BNT	'girl'
TLK	'that'	H+DH	'this'
Y-	'he'	T-	'she'
VRF	'know'	STQBL	'meet'
VAYN	'treat'	-Y-	(plural)
-H	*tâ' murbûṭah*	-AT	(plural)

Figure 4

The theory of the grammar is based on that developed by V. H. Yngve in *A Model and an Hypothesis for Language Structure.*[1] This theory furnishes a mechanism for the construction of sentences in any language for which a grammar is supplied. The grammar upon which the mechanism operates consists of an unordered set of immediate constituent rules.

The current sentence-construction grammar of Arabic produces sentences with and without noun phrase subjects; for example, YVRFWN ALWZYR, /ya9rifuwna 1 waziyra/, 'They know the minister' and YVRF ALWKLAO ALWZYR, /ya9rifu 1 wukalaa?u 1 waziyra/, 'The agents know the minister'. The object may occur either to the right or the left of the subject: YSTQBL ALWLD BNTA, /yastaqbilu 1 waladu bintan/, or YSTQBL BNTA ALWLD,

/yastaqbilu bintani 1 waladu/, both translated as 'the boy meets a girl.' The subject may be first, second or third person, masculine or feminine, singular or plural. The verb-stems are sound, hollow or doubled and some end in N, a letter which assimilates with the feminine plural subjective-suffixes. Examples are /yu9aayinna/, the feminine plural of /yu9aayinu/, both spelled YVAYN. Nouns may be masculine or feminine. The feminine plural nouns and the masculine broken plural nouns may participate in constructions containing both singular and plural attributes. The nouns may be definite or indefinite. Broken plurals, that is, plurals with either discontinuous or infixed number affixes, are formed from singular stems in the sentence-construction grammar. The recognition grammar identifies broken plurals in terms of singular stems and plural affixes.

A sentence-construction grammar programmed to enable a computer to produce sentences in a given language represents a testable hypothesis. The hypothesis may be stated in a generalized form as follows: An appropriate computer under control of an identified program will produce grammatical sequences in a designated language. This generalized hypothesis can be restated with specific application to the sentence-construction program discussed in this paper. The hypothesis will then read: An IBM 709 or 7090 under control of the Arabic sentence-construction program MIT-l-A will produce grammatical sequences in Arabic.

This hypothesis can be tested in at least two ways. 1) The sequences produced by the computer can be examined to determine whether or not they represent grammatical sequences in the designated language. 2) The program can be examined to determine whether or not it can direct the computer to produce grammatical sequences identical with those composed by a tester.

The first test may be exemplified by an examination of sentence 4004. YDLHA AWL+YK ALM+HAMWN ALCYNYWN ALHZLY HNAK, /yadulluhaa ?ulaa?ika 1 muhaamuwana ş şiyniyyuwna 1 hazlaa hunaaka./ 'Those thin Chinese lawyers there guide her.'

If space permitted to illustrate an application of the second test, it could be shown that sentence 1005 composed by a native-speaker of Arabic may be reproduced by the computer. 1005 reads HNAK YSTQBL ALWZYR ALCYNY H+WLAO ALTJAR ALMCRYWN, /hunaaka yastaqbilu 1 waziyra ş şiyniyya haa?ulaa?i t tujjaaru 1 miṣriyyuwna./, 'These Egyptian merchants will meet the Chinese minister there.'

There are numerous values to be gained from the application of the computer to linguistic research. In the first place, complete explicitness is demanded. To date, our program does not expect the computer to do any interpolating.

The testability of the program demands that all grammatical statements be mutually consistent. If this consistency were not maintained, the sequences produced by the computer would prove, on examination, to be ungrammatical.

The random sequences produced by the computer suggest improvements in the grammar. For example, the present grammar contains three classes of adverbs, temporal, locative and quantitative. Members of these classes may be produced before or after the verb, the subject and the object. It was, of course, immediately observed that this range of occurrence must be narrowed. No adverb may occur between the pronominal suffix object and its verb. Consider the following three sequences:

HNA YKATB ALYWM ALWLD H + DA ALMVLM MRARA.
/hunaa yukaatibu 1 yawma 1 walada haada 1 mu9allimu miraaran./

HNA YKATB ALYWMH H + DA ALMVLM MRARA.
/hunaa yukaatibu 1 yawmahu haaḍa 1 mu9allimu miraaran./

HNA YKATBH ALYWM H + DA ALMVLM MRARA.

/hunaa yukaatibuhu 1 yawma haaḍa 1 mu9allimu miraaran./

The first sentence, 'this teacher will write to the boy here at times today' is acceptable, but the second is ungrammatical. The adverb ALYWM may not intervene between the pronominal suffix object -H and the verb. This latter sentence must be rewritten in the form of the third sentence, 'this teacher will write to him here at times today.'

Another example of computer-produced sequences which may suggest questions leading to more precise grammatical statements is furnished by adjective-strings attributive to the nuclei of noun phrases.

In English we may say 'the big, gray, American skyscraper' with normal intonation. We do not, however, say 'the American, gray, big skyscraper', with the same intonation. Observations of this nature lead us to divide English attributive adjectives into positional subclasses.

Our computer produces sequences of adjectives in Arabic noun phrases such as the following.

2100: ALBNAT ALJMYLAT ALKSLANH ALPWRWYH ALFRA + HY ALSMYNAT ALFRA+HY, /?a 1 banaatu 1 jamiylaatu 1 kaslaanatu *θ* *θ*awrawiyyatu 1 faraaha s samiynaatu 1 faraahaa/, 'the beautiful, lazy, short, fat, happy, happy girls'. Such a sequence causes us to ask a number of questions about our grammar. 1) Should Arabic attributive adjectives be assigned to positional classes of a nature similar to those found in English? 2) May the same adjective occur more than once in the same adjective string, for example, ALFRA+HY? 3) If a feminine plural nucleus of a noun phrase occurs in a construction with several adjective attributes some singular and others plural, is the order of occurrence of these attributes determined by the number category into which they fall or is it random? The phrase above, in which the plural ALJMYLAT is followed by two singulars in turn followed by three more plurals illustrates the possibilities. 4) May the contemporaneous sound-form of diptote

adjectives ending in -AN /-aan/ occur in the same adjective string or even in the same sentence with the older broken form? This question is suggested by the occurrence of ALKSLANH with ALFRA + HY above.

These questions will illustrate the stimulus to further refinement furnished by the study of sequences produced by computers under the control of programs such as the one discussed here.

The remainder of this paper will consist of an examination of the sequences produced by the computer. These sentences were compared with those composed by a native-speaker of Arabic. In general, it can be said that the human sentences were more conservative than the computer sentences.

The freedom of occurrence of the three adverb classes may exceed that actually found in Arabic. At the best a number of sentences composed by the computer will be termed stylistically poor because of the position of the adverbs.

> It was realized from the beginning that this situation would result from composing the grammar with this freedom built into it; but since all positions of occurrence seem individually possible, it was deemed best to allow them all. This decision will render the grammar more serviceable in the course of further research.[2]

The computer has produced about fifty different combinations of adverb classes. The first group of computer-constructed sentences which follow, parallel, in general, the adverb distribution found in human sentences.

2018. ALAN Y+HSWNHA. /?a 1 ?aana yahussuwnahaa./ 'Now they are feeling it.'

2003. YSTQBLWNHM MRARA. /yastaqbiluwnahum miraaran./ 'They meet them at times.'

2005. ALYWM KPYRA HNA AMNHM. /?a 1 yawma ka*θ*iyran hunaa ?amunnuhum./ 'I am weakening them here a lot today.'

2064. MRARA TCNH ALYWM DAXLA. /miraaran taşunnahu 1 yawma daaxilan./ 'You are guarding him inside at times today.'

The next sentence illustrates the production of adverbs immediately after the verbs. The human sentences furnished only one example of this type of ordering. A computer-produced sentence follows.

2073. TSTQBL KPYRA TLK ALMMPLH ALYWM. /tastaqbilu ka*θ*iyran tilka 1 muma*θθ*ilata 1 yawma./ 'You will meet that actress a lot today.'

The basic clause is composed of a verb, a pronominal suffix or noun phrase object, and an optional noun phrase subject. The sentences are all verbal. The noun phrase subject may precede or follow the noun phrase object. The following sentences illustrate the various orders of occurrence produced by the computer. The subject is underlined once and the object twice.

2006. YKRH +DLK ALTVBAN HNAK ALBNAT ALMVTWHH H+WLAO HNA. /yakrahu daalika t ta9baanu hunaaka 1 banaati 1 ma9tuwhata haa?ulaa?i hunaa./ 'That tired one there hates these idiotic girls here.'

2048. YXWN ALAN AL + HRYM XARJA ALJHAL ALPWRWYWN H + WLAO. /yaxuwnu 1 ?aana 1 hariyma xaarijani 1 juhhaalu *θ θ*awrawiyyuwna haa?ulaa?i./ 'These revolutionary children are betraying the women outside now.'

ALPWRWYWN H+WLAO. /yaxuwnu 1 ?aana

The following sentence presents an interesting situation.

2011. Y + HSSHMA HNAK ALAN MRARA VLAMH +HZNANH HNAK. /yuhassisuhumaa hunaaka 1 ?aana miraaran 9allaamatun haznaanatun hunaaka./ 'A sad, erudite one there feels it there at times now.'

This last sentence contains a feminine noun-phrase subject with the common-gender adjective VLAMH 'erudite' as its nucleus modified by the feminine adjective + HZNANH 'sad'. This subject is separated from its verb, the masculine form of which the computer has correctly produced. As a result, the only means of determining the gender of the basic clause is by way of the attributive adjective +HZNANH.

The grammar produces the singular and plural forms of the imperfect, indicative, active verb. Sound, hollow and doubled verb-stems are included in the vocabulary. The orthographic problem involving verb-stems ending in final N is solved.

The definition of the sound verb-stem differs somewhat from the traditional one. In this grammar the sound verb-stem class includes any verb-stem which has only one orthographic allomorph in the imperfect indicative active.

Within the limits of the corpus of approximately one hundred fifty sentences so far produced by the computer, it has proved impossible to produce examples of each form of each verb-stem included in the vocabulary. A fair percentage of the forms were produced, however, and most of the conjugations can be filled in if several stems are used. In this way, the conjugation of the sound verb is completed by the use of one or another stem (Fig. 6). The hollow verb-stem has two allomorphs, one occurring in the second and third person plural forms, for example ZR in YZRN /yazurna/ 'they visit' and the other ZWR elsewhere, as in YZWR /yazuwru/ 'he visits'. The only feminine plural forms of the *hollow* verb which the computer has produced to date are from the

	singular		*plural*	
3 masc:	Y-RBB	/yurabbibu/	Y-STQBL-WN	/yastaqbiluwna/
3 fem:	T-VRF	/ta9rifu/	Y-G+S+S-N	/yuğaššišna/
2 masc:	T-RBB	/turabbibu/	T-STQBL-WX	/tastaqbiluwna/
2 fem:	T-RBB-YN	/turabbibiyna/	T-STQBL-N	/tastaqbilna/
1 common:	A-RBB	/?urabbibu/	N-VRF	/na9rifu/

Figure 6

verb YCWN /yaşuwnu/, with a stem-final N. This radical N assimilates with the N of the suffix yielding /yaşunna/ 'they guard' spelled YCN. In order to produce the allograph C of the canonical form CWN used in constructions with the second and third person feminine plural suffixes, the computer must be directed to delete both the W and the X of the canonical stem form. Happily the sentences produced by the computer furnish a complete plural conjugation of this verb (Fig. 7).

Y-CWN-WN	/yaşuwnuwna/	Y-C-N	/yaşunna/
T-CWN-WN	/taşuwnuwna,	T-C-N	/taşunna/
N-CWN	/naşuwnu/		

Figure 7

The nucleus of a noun phrase may be either a noun, an adjective or a demonstrative. The demonstrative pronoun may occur in construction with a locative adverb as in sentence 2062, ARB H+DA HNA HNA. /?arubbu haadaa hunaa hunaa./ 'I raise this one here here.' The first HNA is in construction with H + DA, and the second with the verb.

In the following discussion, examples with either noun or adjective nuclei will be used indifferently to illustrate the structure of the noun phrase.

The attributes of the nucleus of the noun phrase may be one demonstrative, an unlimited number of attributive adjectives and one locative adverb. The locative adverb appears only at the extreme right of the noun-phrase construction. The demonstrative may appear at the extreme left of the phrase or at the right of the attributive adjectives, immediately to the left of the locative.

Agreement in the noun phrase may occur in five inflectional categories; number, gender, case, definition and person. The following phrase, even in the consonantal orthography, illustrates agreement in number, person and definition.

2024. AWLA + YK ALTVBANYN ALVLAMAT AL+TWAL HNAK /?ulaa?ika t ta9baaniyna l 9allaamaati ţ ţiwaala hunaaka/ 'those tall, tired, erudite ones there'.

The demonstrative adjective AWLA+YK 'those' agrees in person with the locative adverb HNAK 'there'. Each word, with the exception of the locative which does not distinguish number, is plural. The nucleus and the attributive adjectives are all definite.

The following phrase exhibits consonantal agreement in gender and case.

2013. NSAKA +HZNANYN /nussaakan haznaaniyna/, 'sad hermits'.

In certain cases in the Arabic noun phrase a feminine singular attribute may occur in construction with a plural nucleus, for example, ALBNAT ALVLAMAT ALCQYRH /? l banaatu l 9allaamaatu l şaqiyratu/, 'the erudite, little girls'.

The sequences produced by the computer, however, raise some interesting questions. Before they can be stated, two points must be clarified. Plurals of the form -i-aa- are of common gender. For example we may say /banaatun kibaarun/ 'big girls' as well as /?awlaadun kibaarun/ 'big boys'. The grammar assumes that the plural adjectives, of this form may occur with common gender as nuclei of noun phrases. For examples, the computer might produce AL+TWAL ALMCRYWN /?a ţ ţiwaalu l mişriyyuwna/ as well as AL+TWAL ALMCRYAT /?a ţ ţiwaalu l misriyyaatu/ loosely translated as 'the tall Egyptians' masculine and feminine respectively. Adjectives which form their plural in this way are assumed also to have an alternate sound feminine plural form, for example AL+TWYLAT ALMCRYAT /?a ţ ţawiylaatu l mişriyyaatu/.

The grammar produces feminine singular attributes in the following situations. If the nucleus of the noun phrase is a masculine broken plural form or a feminine plural, then the demonstrative and attributive adjectives may be feminine singular. Limited by this general statement of distribution, the computer produced the following adjective sequence the likes of which the writer has never seen nor heard.

2089. ALM+WRXAT ALKBYRH ALSMYNAT TLK /?a l mu?arrixaatu l kabiyratu s samiynaatu tilka/ 'those big, fat, historians'

While it is recognized that a singular and a plural adjective may occur in a single noun-phrase construction, the repeated shift from singular to plural back to singular again was judged ungrammatical and the grammar has been changed to prevent this construction.

In the following phrase ALSMAN ALKSLANYN ALKBYRH TLK /?a s simaana l kaslaaniyna l kabiyrata tilka/ 'those big, fat, lazy ones', the masculine gender of KSLANYN is the only element to certify that the gender of the entire phrase is masculine. Morphologically the gender of the broken plural ALSMAN is indeterminate. Syntactically it is masculine. In accordance with the traditional rules of Arabic grammar* it occurs with feminine singular attributes. The noun phrase produced by the computer poses an interesting question. May a feminine singular attribute of a broken plural head of a noun phrase occur to the right of a masculine sound plural attribute such as ALKSLANYN? The question has been answered in the negative, and the statements of agreement in the grammar have been reframed.

Finally, a few words may be said about the morphology of the substantive.

The indefinite accusative suffix /-an/ represented by the *fathatayn* is not consistently indicated in the consonantal orthography. In most grammars this suffix is described from a more or less phonemic viewpoint with orthographic rules supplied by the way. A consonantal grammar must be restricted to a description of the occurrence of the *alif* suffix -A which only in some cases accompanies the *fathatayn.* As a result the consonantal grammar classifies the substantives rather differently from the normal grammars in this respect. The situation is rather complex and the details are treated in the writer's report, *Parallel Sentence-Construction Grammars of Arabic and English.*[2] The substantives are divided into two accusative classes, A and negative-A. Class A includes those substantives which may occur in a construction with the *alif* suffix. Class negative-A includes the remaining substantives. The accusative suffix occurs in construction with class A substantives only but not always if they are indefinite accusatives. The problem is to write the program in such a way that the accusative suffix will be produced only when no mutually exclusive constituents occur. Examples of computer-produced accusative class A substantives are given in Fig. 8.

The computer produces three singular suffixes; the *tâ' marbûṭtah* -H /-at/, -AN /-aan/, and the *alif bi-ṣûrati l-yâ'* -Y /-aa/. Examples are M + WRX /mu?arrixun/ 'historian (masc.)' M+WHX-H /mu?arrix-at-un/ (fem.), KSLAN /kaslaanu/ (masc.) KSLAN-H /kaslaan-at-un/, (fem.) the contemporary form for 'lazy', KSL-Y /kasl-aa/ the traditional form for 'lazy' (fem.), and the -AN /-aan/ of KSL-AN the traditional masculine singular form for 'lazy'

Three suffixes occur in constructions with stems that exhibit allomorphs different from the canonical singular form.

* For example, Wright states "The *pluralia fracta,* even when derived from a masc. sing, are construed with adjectives in the fem. sing. or plural *(sanus or factus)*."[3]

NSAK-A	/nussaak-an/	'hermits'
A + SXAC-A	/?ašxaaṣ-an/	'persons'
+HRYM-A	/hariym-an/	'women'
+ HRMH	/hurmat-an/	'woman'
KBAR-A	/kibaar-an/	'big'
AL + TWAL	/?a ṭ ṭiwaala/	'the tall'
B + HAR-A	/bahhaar-an/	'seaman'
B + HARH	/bahhaarat-an/	'seamen'
VLAMH	/9allaamat-an/	'erudite'
TVBAN	/ta9baan-a/	'tired'
TVBANYN	/ta9baaniyna/	'tired (plural)'
+ HZNANAT	/haznaanaati/	'sad (plural)'
+ HZANY	/hazaanaa/	'sad (plural)'
QTYL-A	/qatiyl-an/	'murdered (fem. sg.)'
QTLY	/qatlaa/	'murdered (plural)'

Figure 8. Accusative substantives produced by the computer

	singular	*plural*
'girl'	BNT /bint/	BN-AT /ban-aat/
'agent'	WKYL /wakiyl/	WKL-AO/wukal-aa?/
'murdered'	QTYL /qatiyl/	QTL-Y /qatl-aa/

One infix, -Y-, was produced in a construction with the stem +HRM which underwent no consonantal change: +HRMH /hurmat-/ 'woman', +HR-Y-M /hariym/ 'women'.

The program produced -A- infixes which occur in constructions with stems exhibiting allomorphs other than the canonical singular forms. For example:

	singular	*plural*
'fat'	SMYN /samiyn/	SMAN /simaan/
'hermit'	NASK /naasik/	NSAK /nussaak/

In these forms, the consonants -A- and -Y- of the canonical singular forms are deleted and the infix added.

Four discontinuous plural affixes have been produced by the computer. They are

'physician'	+TBYB /tabiyb/	A- + TB-AO /?a-tibb-aa?/
'major general'	LWAO /liwaa?/	A-LWY-H /?a-lwiy-at-/
'boy'	WLD /walad/	A-WL-A-D /?a-wl-aa-d/
'tired'	TVB-AN /ta9b-aan/	TV-A-B-Y /ta9-aa-b-aa/

In conclusion, it may be said that the grammar is basically satisfactory as far as it goes. The sequences it produces are, in general, grammatical. The major points which require further research involve

1. the selection and ordering of the adverbs, and
2. the random ordering of attributive adjectives in an adjective sequence.

These points may be reduced to a single question, Do the rules of the grammar permit a greater freedom of construction than observance of the natural language warrants?

References

1. Yngve, Victor H., "A Model and an Hypothesis for Language Structure," *Proceedings of the American Philosophical Society,* CIV (October, 1960), 444-466.
2. Satterthwait, Arnold C., "Parallel Sentence-Construction Grammars of Arabic and English," Doctoral Thesis, Harvard University, 1962.
3. Wright, W., A *Grammar of the Arabic Language* (3rd ed.; Cambridge: the University Press, 1955), II, 273.