

**Linguistic Problems  
In Computer Aided Translation  
From English to Arabic**  
*Wilson, B. Bishai, Harvard University*

The idea of computers aiding in or performing translation from one language to another has been a tantalizing one since it was first conceived. Linguists and software programmers have always pursued high hopes for producing computer-aided translations that would be reliable tools for academic and commercial purposes. At the present time, however, the state of the art falls short of this goal of perfection. And this is precisely why we are meeting here today in this workshop.

I am aware of the fact that there are several research projects in the field of mechanical translation, which have claimed "success". However, this "success" must be measured by the goals set forth by these projects in the first place. If we set the goal of computer-aided translation as merely the preparation of a look-up table of certain specialized terms, and then store in a bilingual computer these terms in one language with their respective equivalents in the other language, then we may claim success only as far as we are able to type a word or an expression in one language and get its translation in the other. We may then expand our goals to include the preparation of a number of look-up tables for highly specialized terms or even idioms, and claim "success" to the extent that we are able to accomplish these goals. But, is the real goal of machine translation merely the preparation of look-up tables? Is this the topic we have come here to discuss? Certainly not! I believe that we have come here to discuss the problems of preparing adequate computer programs capable of producing reliable and acceptable translations.

As a matter of fact, the minimal goal for an acceptable computer-aided translation project should be nothing less than the preparation of a program by which the user would type in sentences in a source language, English for example, and receive a grammatically correct translation of these sentences in the target language, Arabic for example, without the need for any post-editing that would require the services of a human translator. This means that if any computer-aided translation program put out a piece of translation which would require a human translator to consult the original text and correct the output translation, then such a translation program should be judged as inadequate; since the human translator would prefer to do his own translation without the help of a computer.

However, if a computer translation program is capable of putting out a piece of translation which is grammatically correct and understandable, requiring only stylistic modifications without the need to check the source text, then such a translation program may be judged as adequate and acceptable.

The question begs itself at this point: Is the preparation of such a translation program feasible? My answer to this question is definitely yes. However, before embarking on such a project for the purpose of producing acceptable and correct machine translation we must define the pertinent problems and plan their solutions at every level of the three levels of machine translation, namely, Analysis, Transfer, and Synthesis.

In this light, the goals of computer-aided translation from English to Arabic may be summarized as following

1. Producing a device to analyze English syntax as it relates to Arabic syntax.
2. Preparing a coded dictionary capable of storing certain grammatical and semantic specifications of nouns, verbs, and particles.
3. Producing a synthesis device by which Arabic morphological and syntactic grammatical rules could be stored in the computer and used to render the Arabic output of the computer translation grammatically correct.

Let me now elaborate on these three basic goals:

1. On the analysis level, the goal is to analyze the basic types of English sentences in such a way as to enable the computer to correctly transfer their syntactic units (syntagmems) into Arabic syntactic units. So far, specialists working on machine translation from English to other languages have concentrated on analyzing English syntactic features without much attention given to target languages. In other words, they have tried to produce computer software programs capable of analyzing English sentences purely for the sake of English. With regards to machine translation, these specialists have faced major obstacles, chief of which is the obstacle of analyzing complex English syntactic combinations. They came to realize that such analysis requires the extra-linguistic human dimensions of intelligence and common sense.

Be this as it may, I do not believe that we really need this kind of analysis of English ambiguous and complex sentences; since complex English sentences could be rewritten in a simpler way at the input level whenever there is a need for pre-editing.

I am therefore recommending that the operation of analysis of English syntax be restricted to the basic and most common sentence types, which the computer can reasonably and easily analyze. Other sentence type beyond the capability of the computer should be isolated and then rewritten. This means that once we isolate about twenty or twenty-five English sentence types that can be easily analyzed, a computer program would be made to flag out other sentence types which it does not recognize. During the phase of typing in the text to be translated, the English typist, who is not required or even expected to know any Arabic, would be asked by the computer itself to rewrite those sentences in a simpler way. Otherwise, the computer would be programed to render a literal translation of these sentences with a marker indicating the existence of a problem with these sentences. In any case, I expect the number of such problematic sentences to be minimal, and a way to avoid them could be found.

Once we isolate the analyzable English sentence types, we must make sure that their analysis be made in comparison with Arabic syntax. I shall present here an illustration of what I mean. We know that English infinitive phrases such as "to write", "to eat", and "to study" are rendered into Arabic in four different ways:

(1) After verbs of beginning or verbs of existence, an English infinitive phrase is always translated into Arabic as an imperfect verb فعل مضارع, such as "He began to write" which appears in Arabic as بدأ يكتب, and "He continued to write" rendered into Arabic as استمر يكتب.

(2) After verbs of desire or request, the English infinitive is translated into Arabic by using the particle أن plus a subjunctive verb, such as, "He wanted to write" which is translated into Arabic as أراد أن يكتب.

(3) After other verbs, the English infinitive is translated into Arabic by the particle كي plus a subjunctive verb, such as, "He came to write", which appears in Arabic as جاء كي يكتب.

(4) When the English infinitive is not preceded by a main verb, it is translated into Arabic by a noun form known as a masdar مصدر, such as, "To write is good" rendered into Arabic as الكتابة حسنة.

It is evident from the above illustration that any analysis of the English infinitive done for the sake of translating from English to Arabic must include these four syntactic variations dictated upon us by a comparative study of English syntax and Arabic syntax.

(Other illustrations of this feature are the Arabic relative clauses with their resumptive pronouns and the participial forms which function as adverbs).

II. The second basic level in machine translation is the level of Transfer. For the purpose of computer-aided translation from English to Arabic, this level must include a coded dictionary with code numbers for certain basic grammatical specifications of nouns, verbs, and particles. These coded specifications can be determined after a complete plan of the translation project is drawn and finalized.

It is generally believed that an ideal dictionary as a medium of transfer in machine translation should include a minimal number of coded symbols with only one meaning given for one entry. Linguists specializing in machine translation from English to Arabic are faced with various problems, the most significant of which are as follows:

(1) a need to determine the gender of all nouns as required by rules of Arabic grammar.

(2) a need to determine whether a certain plural noun is human or non-human, since non-human plurals in Arabic grammar are treated as feminine singulars.

(3) a need to determine variations in verb meanings due to prepositions which follow, such as "to go" meaning "to depart" and "to go on" meaning "to continue".

(4) a need to determine variations in meaning based on whether the object of the verb is human or otherwise, such as to "receive a gift" translated into Arabic as *استلم هدية* and "to receive a guest" translated as *استقبل ضيفا*.

(5) a need to determine the general semantic specifications of verbs in terms of whether they are verbs of beginning, verbs of existence, verbs of declaration or belief, verbs of desire or request, or other verbs. These determinations are made necessary by rules of Arabic grammar.

(6) a need to determine that certain nouns can also function as adjectives, such as the word "Arabic" itself, which may mean "the Arabic language" or "Arabic" as an adjective. The reason for this need is to enable the computer to distinguish between noun-noun phrases and noun-adjective phrases. This means that nouns designated as n+a (nouns and adjectives) will be considered only as adjectives whenever they are followed by other nouns.

(7) a need to distinguish between governing particles and non-governing particles. The reason for this is to enable the computer at the synthesis level to choose the correct verbs and nouns after governing particles. Examples of this are the subjunctive and jussive forms of verbs as well as the accusative and genitive forms of nouns.

(8) a need to list multiple meanings of certain words individually on separate lines with numerical designations, allowing only one meaning to be assigned to one entry. As example of this is the English word "race" which should be listed twice as follows:

- |                       |      |
|-----------------------|------|
| 1 - race (running)    | سباق |
| 2 - race (human type) | جنس  |

As the computer processes a certain text that includes the word "race", it would request from the English typist, who does not have to know Arabic, to choose either (1) or (2) depending upon the two meanings indicated between parentheses.

A coded dictionary is considered the main device in machine translation at the Transfer level. However, such a dictionary must be accompanied by an elaborate program of Transfer Rules. These rules would combine the Analysis rules of Level I with the Coded Dictionary of level II in order to convert English Syntagmems (syntactic units) into Arabic Syntagmemes, and then re-order the latter to fit Arabic normal word order.

III - The third basic level in machine translation is the level of Synthesis. This is perhaps the most important of all operations, because a grammatically correct translation depends mainly on a good and reliable synthesis device which includes an interactive grammar of the Arabic language.

Following are some examples of the functions which a computerized interactive Arabic grammar is expected to perform:

1. Arabic syntactic rules require the verb to precede the subject, and require the object to appear in the accusative case and to follow the subject in word order. Accordingly, this interactive Arabic grammar is expected to re-arrange the Arabic syntactic units to conform to the normal Arabic word order, and at the same time add the accusative case ending to the object of the verb.
2. If the object of the verb is a pronoun, Arabic grammar rules require that such a pronoun should appear as a suffix attached to the verb which usually appears before the subject.
3. Accusative particles such as **لكن** "but" and **لأن** "because" require that the noun subject should precede the verb. Clauses introduced by these particles force the verb to follow the subject. In this case, according to Arabic grammar, the verb must agree with the subject in number as well as gender; whereas, in the normal word order of verbal sentences, the verb would remain singular.

4. Another syntactic complexity in Arabic is the resumptive pronoun in relative clauses. An English sentence such as "The boy, whom I saw, came", should be transferred into Arabic, according to Arabic grammar, as "The boy, who I saw him, came" جاء الولد الذي رأيته. There is no doubt that the synthesis device in any translation project from English to Arabic should allow for this special feature of Arabic grammar.

5. This interactive Arabic grammar is expected to rewrite adjectives to agree with nouns in number, gender, case, and determination, and also to adjust adverbs of manner and circumstance to appear in the accusative case and to agree with nouns to which they refer in number and gender.

These are only a few illustrations to show the importance of a computerized device in any machine translation program designed to translate from English to Arabic.

At this point, I would like to add that what I have mentioned above is not by any means a complete plan for a project to produce computer-aided translation from English to Arabic. It is only an outline of the types of problems that are encountered. If and when such a project is undertaken, a complete plan of operation should be prepared, which would explain in details all the problems to be considered as well as specific ways and devices to solve them.

In closing I would like to emphasize that, in the computer business, a good programmer is expected to be a master of the science as well as the art of programming. Accordingly, in the final analysis, it might be said that the secret of success of any machine translation project is a happy combination of a group of creative linguists and a group of imaginative programmers.

Thank you, والسلام عليكم ورحمة الله