

# The Passive in Arabic, Hebrew and English and Machine Translation\*

JUDITH ROSENHOUSE and ARIEL MORDECHAI COHEN†  
The Technion, IIT, Haifa, Israel

## Abstract

In order to enable the generation of valid syntactic rules for automatic translation from English into Hebrew-Arabic, basic morphological features of the passive as well as its semantic functions, syntactic structures, and distribution restrictions are examined here. This study has led to several basically syntactic translation rules. Some of the rules which deal with the English-Arabic translation have been tested in a small framework of real computer programming environment, and examples of the translation are presented.

## Introduction

In the field of syntax, the passive has been occupying a very important place for a very long time. The elusive links between passive and active forms, which are morphologically marked by various devices in many languages, have intrigued many linguists and yielded many approaches to this issue in the literature.

Some of the recent approaches, which are naturally based on the previous views in their deep-set basis, were studied as a preliminary stage of this research. It was hoped that an analysis of these issues may bring out underlying linguistic structures in each of these languages and serve as a basis for defining rules for automatic translation.

Siewierska (1984) analyses the facts along lines of comparative linguistics and from the general point of view of universal linguistics, though, as she notes, not all languages attest the passive. Also Shibatani (1985) describes the issue from a general linguistic point of view.

For the English passive, the recent comprehensive book by Quirk *et al.* (1985) has also been consulted.

Saad (1982) analyses the passive in modern literary Arabic by the case grammar method. (As a matter of fact, this approach has some common features as that of the classical Arabic from over 1,000 years ago, cf. Wright, 1967).

Cantarino (1974-5) provides important material on the passive in modern Arabic prose. Though methodologically unlike Saad, it presents the facts mainly in the 'conventional' syntactic approach, it also provides many details about the distribution and roles of the passive in Arabic.

For Modern Hebrew passive, Rubinstein (1971), Berman (1973), Chayen and Dror (1976), and Tsadqa (1981)

\* This paper is a revised and expanded version of a paper read at the 15th International Conference on Literary and Linguistic Computing held in Jerusalem, June 1988.

† Ariel Mordechai Cohen was at the time an undergraduate student at the Faculty of Computer Sciences, The Technion, IIT, Haifa, Israel.

Correspondence: Dr. J. Rosenhouse, Department of General Studies, The Technion, IIT, Haifa, 32000 Israel.

Literary and Linguistic Computing, Vol. 5, No. 1, 1990

have been consulted. Not much has been written concerning this problem in Hebrew syntax, but as these works represent the transformative-generative approach, they deal also with the passive in Hebrew.

In the following we briefly review the morphology of the passive in the three languages under study (Section 1), the types of the passive in linguistic literature (Section 2), and the occurrence of the passive forms and their selections (Section 3). After these sections will follow our tentative translation rules (Section 4) and a section on an actual automatic translation experiment from English into Arabic (Section 5).

## 1. Morphological comparison

English has the least morphological complexity among these three languages: its only method of transformation is an auxiliary verb ('be' and rarely 'get') + the past participle form of the main verb (e.g. The food was cooked by the maid; the food got cooked).

Hebrew and Arabic have basically similar verb systems, which include two ways of passive 'transformations' (see Appendix I):

1. An 'internal passive', which transforms an active verb-form into a passive one by a fixed change of the vowel patterns. Thus, in Arabic, for example: *kataba*<sup>1</sup> (he wrote) vs. *kutiba* (it was written); *kaataba* (he wrote to someone) vs. *kuutiba* (he was written to); *'aktaba* (he made someone write) vs. *'uktaba* (he was made to write); etc. And in Hebrew, too: *sipper* (he told a story) vs. *suppar* (it was told); *hixtiv* (he dictated) vs. *huxtav* (it was dictated). But in contrast with the productivity of this passivization process in Arabic, where ten verb measures may theoretically be passivized by this device, Hebrew has only the above two forms of passive of the parallel active verb patterns (measures).
2. A special pattern or 'measure' which denotes the passive through its special prefix: *'infa'ala* in Arabic and *nif'al* in Hebrew. As a matter of fact, Saad (1982, ch. 6) shows that *'infa'ala* in Arabic is not a real passive but a reflexive. This observation is true in part, at least, also concerning the use of *nif'al* in Hebrew, for though *nif'al* is usually the passive of verb forms in the basic measure (*pa'al*, or *qaal*), it is sometimes used with non-passive meanings, e.g. *nifgash* (meet), *ne'elam* (disappear).

As Arabic has more measures than Hebrew, it has more passive forms to deal with than Hebrew, which situation seems to require more rules for translation from English into Arabic than into Hebrew. In addition, the conditions for the use of each of the passive forms are more numerous in Arabic than in Hebrew. Thus,

© Oxford University Press 1990

despite the common basic structural similarity between Hebrew and Arabic, the translation rules need to be language-specific.

## 2. Types of passives

A main feature of the passive is that it is usually contrasted with the active. Linguists have noted that passivization is linked with 'object promotion' and 'subject demotion', and much debate is going on concerning which of these processes is more crucial for the passive (cf. Shibatani, 1985). Some linguists claim that there is no basic link between the active and the passive structures, and the link between them is formed on the lexical or derivational level (Siewierska, 1984: 75). Siewierska thinks that both processes are important, but 'demotion' is significant especially for the impersonal passive.

Siewierska (ibid.) classifies the passive into four types which are attested in many languages: personal, impersonal, periphrastic, and reflexive passive as well as 'exceptions' (see Appendix II).

Hopper and Thompson (1980) analyse the passive as part of the contrasts of transitivity and of topicalization, i.e. transposing the object to the agent position, as part of other high and low transitivity features, which are actually semantic features. This approach is interesting, but seems difficult to apply to automatic translation rules because of its heavy semantic load.

Most of these types of passives exist in Arabic and Hebrew in varying distributions and forms. To the personal passive, which is the most prevalent type of passive, Saad (1982) devotes a whole chapter, in which he classifies smaller groups of passivizable and non-passivizable verbs. In the previous group he includes Implemental verbs (e.g. open, hit, suck, kick, etc.); Physical Perception verbs (which may be stative, inchoative, and active); Reciprocate verbs (in measure III); Resultative verbs (e.g. solve, destroy, build, stop); Emotive and Cognitive verbs (hate, envy, have pity, know, ignore, understand, be sad, etc.); Verbs of Transforming (appoint, make, name, etc.); Verbs of Acquiring (hold, buy, adopt, etc.); Causative verbs (make melt, make unhappy, etc.); Verbs of Certainty and Doubt (consider, see, know, find, claim, etc.); and Miscellanea (Expose verbs, Declare verbs, Remove verbs, Surpass verbs).

Among the non-passivizable verbs Saad includes one-place verbs (stative, inchoative, or active-intransitive), defective verbs and 'be', and verbs of the type of 'cost' and 'resemble'. (Some of these groups are described already in classical Arabic grammar, cf. Wright, 1967, vol. 2.) In any case, 'the deep structure of Arabic sentences . . . can be better described in terms of ergativity than in terms of transitivity, that is in terms of cause and effect' (Saad, 1982: 90). Saad then classifies Arabic verb classes into four groups:

- morphologically non-derived transitive verbs are transitive;
- morphologically derived transitive verbs are causative;
- morphologically non-derived intransitive verbs are intransitive;
- and morphologically derived intransitive verbs are reflexive.

The transitive and reflexive verbs have the feature { -middle}, while intransitive and causative verbs have the feature { -middle}. Passive verbs are of the { +middle} type, and may be transitive and causative verbs (Saad, 1982: 93-4).

In Arabic, due to morphological rules, theoretically any verb may be arbitrarily transformed to the internal passive even if it is not 'logical'. But in fact, passivizability is decided by the verb's case frame and not by 'general logic', and not all verbs are passivized (see Saad's list above).

In Arabic the passive has a subject-agent which need not be overt, nor the patient. Saad therefore advocates to distinguish between subjectivization and passivization, and formulates the following rule: 'The closest noun phrase to the verb becomes subject provided the noun phrase and the verb are not separated by a preposition' (Saad, 1982: 64).

Impersonal passives occur in Arabic, as well as in English. Often, as Siewierska (1984) describes, impersonal passives are derived of intransitive verbs, and many of them are also locatives or time expressions. Cf. 'niima fi al-bayti' ((it) was slept in the house), 'siira 'ilaa al-qudsi' ((it) was gone to Jerusalem). Saad considers such forms possible as a result of a Targetization transformation. However, the verb in such cases must be in the 'unmarked' (impersonal) masculine 3rd person singular.

In English the dummy subject 'it' must appear in such cases, thus making the verb formally similar to the Arabic pattern, i.e. in the 3rd person singular. It should be noted, however, that as in English, such Arabic examples hardly occur in modern Standard Arabic, although they are grammatically legitimate. Secondly, such passives in Arabic occur with intransitive verbs which govern 'internal objects' and only with them is this pattern possible (e.g. niima . . . 'was slept' siira . . . 'was gone'). Impersonal passives such as 'qiila' (it was said) or 'ulima' (it was known, it was reported) occur often as they are not transformed of intransitive verbs.

In Hebrew similar structures exist, but we do not know that any case grammar analysis has been done yet, to analyse the conditions of their usage. It should be mentioned, that in Biblical Hebrew there occur impersonal passive forms in the 3rd person singular (cf. Rabin, 1963, 1971). But in Modern Hebrew such 'Biblical' impersonal passive forms do not occur and when such passivization is required, it is usually done by the impersonal active in the 3rd person plural.

## 3. Distribution problems

We have seen that in contrast with the one basic form of the passive in English, Hebrew and Arabic present numerous potential forms of passive. For automatic translation this seems to raise the question of defining which passive form to use in each situation, since this is not at all clear from the standpoint of the English structures. In relatively many cases, there may be some passive forms to one active verb form, each with its special semantic nuance. It seems logical, then, to start independently with each of Hebrew, Arabic, and English, and then compare results. A bilingual dictionary

may then help in solving problems of pattern selection, by automatic reference to the correct passive sense of each meaning of the active root (e.g. I or VII in Arabic, Nif'al or Pu'al in Hebrew).

But also distribution problems exist among these languages. Elsewhere (Rosenhouse, 1988) we find that for different types of written discourse (literary, scientific, and journalistic texts) there are differences in the rate of occurrence of the passive. English appears to have the largest rate of passive occurrence, Hebrew follows, and Arabic lags far behind. The difference between the English and the Hebrew rate of use of the passive is not large but solid. Arabic, however, tends to use the passive only when it is unavoidable.

The basic approach in modern Arabic prose (cf. Cantarino, 1975) is that in Arabic every finite verb has a subject, and when the subject-agent is known (even if not explicit), the verb has to be active rather than passive. The use of passive for predicate topicalization which is often applied in English is not used in Arabic, for the 'cleft sentence' structure is preferred for this problem. This reduces the rate of passive occurrence in Arabic compared with English passives, though also most of the passive forms in English (about two-thirds in fact—cf. Svartvik, 1966) are agentless transformations of transitive verbs. In fact, however, passives do occur in Arabic even when the agent is specified by the 'by phrase'. (Saad (1982: 36) also notes that such structures are present already in the Qur'an, i.e. since classical Arabic.) But the above 'rule' probably represents an inherent dislike of the passive in Arabic grammar. These tendencies complicate the issues for an automatic translation program.

In addition, a well-liked syntactic structure in Arabic (and to some extent also in Hebrew), which often translates English impersonal passives, is the impersonal active which retains 'impersonality' not by voice but by the use of a 'vague' subject in the 3rd person plural. (Impersonal passives are related to, but should not be confused with, impersonal active structures, which exist also in English.) On the other hand, sometimes an Arabic text may use the passive when no passive occurs in an equivalent English text.

Hebrew passives resemble the English ones in rate of use more than the Arabic passives do. Still, also this area does not show in the translation of a one-to-one conformance with the English source. The main categories that are relevant for Hebrew are (according to Berman, 1973): transitivity, reflexivity, middle, and passivity. Also the perfective and durative aspects of the verb forms (with and without time element), are important as they affect the occurrence of the passive participle. Following Saad's analysis, ergativity should be considered also in Hebrew: though there are less 'measures', similar principles seem to be operating here, too, and we may try to apply his rules for the Hebrew use of the patterns *pu'al* and *huf'al* (real passive of *pi'el* and *hif'il*, respectively) vs. *nif'al* (real passive of 'qal', but often non-passive at all) vs. *hitpa'el/nitpa'el* (reflexive, sometimes passive). Such information, however, would be included more easily in the bi-lingual dictionary rather than in the syntactic rules of translation.

#### 4. Tentative translation rules from English into Arabic and Hebrew

The tentative rules below are limited to passive cases as they occur in the English source-text and will refer only to 'standard' Hebrew and Arabic texts (i.e. without any consideration of different passive rules in any Arabic dialects, e.g. Retsö, 1983, Rosenhouse, to appear). In all the examples quoted with the rules, the English source sentence is translated first into Arabic then into Hebrew, both in phonetic transcription.

Our approach to the subject was considerably influenced by the 'case grammar' or 'relational grammar'. According to these theories, the relations between words in any sentence are based on deep semantic relationships which are termed agent, patient, goal, object, recipient, benefactive, locative, temporal, etc. These relations also govern the linguistic expressions which distinguish between the active and the passive (Saad, 1982). In fact, many recent studies led to the conclusion that semantic considerations operate on all the levels of linguistic structure. Thus, it is only logical to incorporate them somehow also in any automatic translation.

However, this understanding had to remain here implicit, since in the present framework our tools were insufficient for a complete semantic description of the Arabic or Hebrew semantic systems.

Due to these pragmatic limitations, the rules presented here are simplified as much as possible by referring to the most basic cases, with the least possible reference to semantically special verb or noun subgroups. The style of the rules is also rather syntactic than semantic, due to the same problem.

Still, some semantic features which are essential to each item are given in the dictionary, e.g. human or animate agent required by a verb, governed preposition, gender and number specification. These were, naturally, only few, in comparison with the semantic features underlying any human linguistic communication and which would be required in a project on a larger scale.

*Rule I.* If there is in the English source a 'by-phrase' denoting the agent of an action specified by a passive verb, then translate the sentence into Arabic while transforming it into an active structure. Translation into Hebrew will retain the passive structure as in the original English text, if no other restrictions prevent it. Cf.:

The apple was eaten by the boy  
 → 'akala al-waladu al-tuffaahata (active)  
 → ha-tapuah ne' exal 'al-yedey ha-yeled (passive), but  
 The apple was eaten → 'ukilat al-tuffaahatu/ha-tapuah ne'exal.

*Rule II.* If the active equivalent of the passive form in English is a transitive verb, to be translated in the dictionary into a morphologically non-derived Arabic verb form, or if the active equivalent of the passive form in English is a causative verb, which is translated in the dictionary into a morphologically derived transitive, then keep the translated structure in the passive in Hebrew and in Arabic (if rule I does not prevail):

The book was written → kutiba al-kitaabu ha-sefer nixtav.

The book was hidden at once → 'uxbi'a al-kitaabu fii al-haali/ha-sefer hubba miyad.

The picture was given to him → 'u' (iyathu al-suuratu/ha-tmuna nitna lo.

He was given the picture → huwa 'u' tiya al-suurata (\*hu nitna lo/nitan 'et ha-tmuna).

This rule appears to be valid also for cases where in English the auxiliary verb is 'get'. (This rule reflects Saad's classification of verbs and transitivity/intransitivity).

**Rule III.** If in the English source text there is a passive form (often of the group of verbs of 'saying') preceded by the dummy agent 'it' (thus being actually an impersonal passive), the passive form will be retained also in the translation into Arabic and Hebrew, while the dummy-subject 'it' is deleted. For example:

It is said that ... → yuqaalu 'inna ... /ne'emar she ... (There are, of course, other translation options, and some verbs require different options because they cannot appear in the passive at all but as noted, these are only the basic cases.)

**Rule IV.** If the passive verb in English governs a prepositional phrase, in the translation the verb will be transformed into an active sentence (in both Hebrew and Arabic), with the verb in the past tense and the 3rd person plural.

This rule is defined thus for the sake of simplicity, for in some cases it is possible to keep the passive verb form, in both Hebrew and Arabic. But this will cost in modifications of the sentence structure rules and the verb type specification in the bilingual dictionary. The following examples demonstrate the difficulties involved in the translation:

The bill is paid for → Arabic: al-hisaabu madfuu 'un. (Different lexeme, using the passive participle form.)

dufi'a al-hisaabu. (The same lexeme as above, but in the past tense.)

Hebrew: ha-heshbon meshulam. (Different lexeme using the passive participle form.)

ha-heshbon shulam. (The same lexeme as above, but in the past tense.)

The required translation, according to the rule would yield:

The bill is paid for → Arabic: dafa'uu al-hisaaba.

Hebrew: shilmu 'et ha-heshbon. (With the object marker 'et' introduced in Hebrew.)

**Rule V.** If the passive verb belongs to the group of verbs of feelings and affections (love, hope, hate, etc.), the translation will tend to use the passive participle form with the copula 'to be' indicating the tense, rather than the passive verb form in the finite tense conjugations (past/non-past). For example:

He was loved by all → kaana mahbuuban 'ala al-kulli 'he was beloved on all' (\*'uhibba min qibali al-kulli).

hu haya 'ahuv 'al ha-kol 'he was beloved on the-all' (\*hu ne'chav 'al-yedey kullam).

(This modification somewhat reduces the dynamic as-

pect of the verb phrase, which is normally present in English passive forms. In this example, however, the dynamic aspect is absent also from the English structure because of the verb semantic features.)

**Rule VI.** If in the English text the passive follows a modal verb (e.g. 'it can/may/should be done'), or if there is an adjective which ends with the suffix '-able/-ible', we actually deal with semantico-syntactic notions of procedure and probability of occurrence. The translation into Arabic and Hebrew will require different paths. Our rule is phrased in two parts as follows:

(1) If in English the passive verb form follows a modal auxiliary, in the translation into Arabic this modal verb will be omitted, and the passive verb form (participle) will be translated into the passive imperfect form in 3rd person singular (Monteil (1960) considers this structure a major characteristic of modern Arabic style. Cf. also Blohm *et al.* (1981), vol 2/II), for example:

It can be done easily → haqa yu' malu bi-suhuulatin.

The house can be built in a year → yubnaa al-baytu fii sanatin.

This problem cannot be forgotten → laa tunsaa haqdihi al-qadiyyatu.

(2) If in English the main verb phrase is the verb 'be' which is followed by an adjective which ends in the suffix '-able/-ible', then in the translation into Arabic the auxiliary verb 'be' will be omitted and the adjective will be transformed into a passive verb form in the imperfect tense (3rd person singular). Thus:

This matter is unforgivable → laa yusaamaahu haaqa al-'amru.

This food is not edible → laa yu' kalu haaqa al-ta'aamu.

This story is unbelievable → laa tušaddaqu haaqihi al-hikaayatu.

In Hebrew the rule must be defined separately. We suggest the following tentative phrasing:

(3) If in English the passive verb form follows a modal auxiliary, in the translation into Hebrew this modal verb will be kept while using the passive verbal noun if it is of the measure *nif'al* (e.g. yaxol le-he' asot). Cf.:

It can be done easily → ze yaxol le-he' asot be-qallot.

The house can be built in a year → ha-bayit yaxol le-hibanot be-shana.

This problem cannot be forgotten → ha-be'aya ha-zot 'eyna yexola (lo tuxal) le-hishaxaf.

(4) If in English the main verb phrase is the verb 'be' which is followed by an adjective which ends in the suffix '-able/-ible', then in the translation into Hebrew the whole verb phrase will have to change into a phrase, e.g. 'nitan l(e) + active verbal noun', or 'elshar l(e) + active verbal noun' (+ accusative marker, and/or the negative particle 'i', if required), e.g.:

The thing is unforgivable → ha-davar 'eyno nitan li-sliha, 'the thing not—it given to—forgiving'.

This food is not edible → ha-mazon ha-ze 'eyno nitan la-'axila, 'the-food this not-it given to-eating'.

This story is unbelievable → 'i 'elshar le-hu'amin la-sipur ha-ze, 'not possible to believe to-the story this'.

In terms of occurrence rate, it is more likely that in Hebrew the structure with the passive verbal noun will be less common than in English, because of their general little occurrence (and cf. Berman, 1973). A separate issue is to define which verbs can take the passive infinitive and which cannot. As this requires semantic definition of the verb class (whatever its 'measure'), we do not deal with it here.

Sometimes, the English verbal phrase may be analysed in two ways, as a real passive verb and/or as an adjective following the copula (periphrastic structure). This ambiguity raises a translation problem. Ambiguity problems occur also in other areas of automatic (and non-automatic) translation. Let us, however, try to resolve here the problem of ambiguity of some passive structures. Quirk *et al.* (1985: 167) note that there is a 'passive gradient'. From the examples quoted there, it seems possible to distinguish between verbal and adjectival structures by the use of adjacent adverbial and prepositional phrases, i.e. by the context. Indeed, for a computer program such a scanning would require much more computing, but it seems useful for the long run, also for other sentence members and other syntactical elements, to build the syntactic part of the program to include contextual morphosyntactical scanning. For instance, if we compare the following sentence pair: 'Leonard was interested in linguistics' with 'Leonard was interested by his professor in linguistics' (Quirk *et al.*, 1985), we would tend to interpret 'interested' in the first case as an adjectival as the 'by phrase' does not occur there. To make the past participle form into a real passive structure, it seems necessary here to add the 'by phrase'.

Another example from Quirk *et al.* (1985: 167) is: 'The building is already demolished'. The parallel example: 'The building is already demolished by the workers' is not acceptable in English tense-wise; it should be either 'The building has been already demolished by the workers' or 'The building is being (already/now) demolished by the workers'. Thus, the tentative rule would be:

*Rule VII.* If the tense form is in conflict with the adverbial ('already'), then the past participle cannot be analysed as part of a passive verb structure, and cf.:

The house is already built → Arabic: al-baytu mabniyyun 'the-house built' (passive participle without translation of 'already').

or: al-baytu saara mabniyyan, 'the house became built'.

Hebrew: ha-bayit kvar hanuy, 'the house already built (passive participle).

The above rule seems for the time being a comfortable solution for many of these sentences. However, there are also 'semi-passives' and 'pseudo-passive' as Quirk *et al.* (1985: 168-70) call them. From analysing the examples there (e.g. 'I feel rather let down by his indifference; we were unimpressed by his indifference') we suggest the following rule, to distinguish between participle adjectives and real passives:

*Rule VIII.* If the participle is stative (rather than dynamic) and if the head noun of the 'by phrase' is not the agent of the participle form, then the participle form will

be considered adjectival rather than a real passive. We may know that the noun-head of 'by' is not the agent if, for example, in the bilingual dictionary the verb 'kill' is specified as one that requires an animate agent while, at the same time, 'river' does not get a marker which makes it part of the 'animate' group. The form 'tired' is to be marked 'adjective' in the dictionary, which will prevent such translation problems.

He was killed by the river → huwa qutila 'inda al-nahri/ hu neherag le-yad ha-nahar.

He was tired by the river → huwa kaana ta'baanan 'inda al-nahri/hu haya 'ayef le-yad ha-nahar.

Of course, there are more complex cases to tackle, e.g. 'we were all worried about the complication; I was a bit surprised at her behaviour; I was surprised that the food was so good; By the time she got there, her friend was gone; I'll soon be finished with this work', etc. (These examples, too, are from Quirk *et al.*, 1985.) But these often reflect 'idiomatic phrases' rather than simple passive forms. Therefore, it seems they may be separated from 'simple' passive translation rules, and dealt with in the dictionary on an 'individual' basis.

It seems preferable to tackle such ambiguities in the stage of the source sentence syntactic analysis. The translation procedure may then follow its normal course, with passive forms translated according to the relevant rules and lexical problems (such as phrases or governed prepositional phrases) solved on the dictionary level.

#### 5. Example: automatic translation of some passive sentences from English into Arabic

The above rules were applied for automatic translation from English into Arabic. The program was implemented in Common Lisp on a Symbolics Lisp Machine. It made use of PEG, a parser used at the Haifa IBM Science and Technology Center as a front end syntactic analyser.

The program scans a sentence and performs the required transformations and translation simultaneously. The algorithm imposes a definite order on the analysis.

First, indirect speech is looked for (rule III). If found, the transformation is carried out and the program is applied recursively to the rest of the sentence.

Secondly, an agentive 'by phrase' is searched (rules I, VIII). If found, the sentence is translated (using the rest of the rules) while being transformed into the active form; otherwise it remains in the passive form.

In both cases, the main verb appears first (before the subject), unless the sentence is a clause appearing after the particle 'inna' (as described in rule III). In this case, the subject precedes the verb (predicate). If there is no explicit subject, a dummy subject (3rd singular pronoun) is supplied and suffixed to the particle 'inna'. The rest of the sentence follows.

The program uses a dictionary. It includes the translations of English words into Arabic ones, with some details of their morphological forms, as necessary for the application of the translation rules, e.g. plural forms of nouns, past and future forms of the verbs, animate/human agent requirement (rule VIII) or whether a verb

belongs to the group of 'report' or 'saying' verbs (rule III).

When a compound English verb translates into a simple Arabic verb (e.g. 'was made up' translated by 'uxturi'a'), separate entries are kept for 'make' and 'make-up'. A more difficult case is the opposite one: when an English simple verb is translated into a compound verb phrase (e.g. 'caught'—'alqaa al-qabja 'ala'—literally, 'threw the grabbing on'). A simple dictionary look-up is not adequate, and a new object ('al qabja' in this example) which does not appear in the original source must be introduced for independent manipulation in the sentence.

See Appendix III which presents parts of the program, the dictionary and a set of test sentences, which were given in English and automatically translated into Arabic.

The example sentences represent the above rules which dealt only with the passive structure. However, when real translation is involved, morpho-syntactic rules had to be added, to deal with, for example, concord between a noun and its attribute (adjective), or a verb and its subject-head noun, etc.

In Arabic these morpho-syntactic elements are rather involved and therefore they were treated here for the example by 'ad hoc' rules. An issue which we left out is the case endings in Arabic which are suffixed to nouns and verb-forms according to certain morpho-syntactic conditions. As these are suffixes, they tend nowadays to be deleted in 'spoken literary Arabic' (i.e. literary Arabic used orally, not only for reading, in formal occasions), while colloquial Arabic has long discarded them. These could also be programmed, of course; but we did not consider them necessary for the problem at hand. This lack of the case-endings is seen in the difference between our above examples for the rules (e.g. 'yubnaa al-bayt fii sana' for 'yubnaa al-baytu fii sanatin'). But, as noted, our main purpose here was to start a translation procedure of passive verb forms from English into Arabic, and this goal has been achieved.

The translated sentences apply both the passive forms in Arabic and the transformation of the English passive into Arabic active verbs when the agent is specified. Also, the problem of verb+complement has been tackled, even in an example of the impersonal passive and some though not all concord problems which are, as mentioned, different in Arabic from English.

Translation from English into Hebrew is undertaken by the Haifa IBM Science and Technology Center (cf., for instance, Golan *et al.*, 1988), so we do not present examples for it.

## 6. Summary and conclusions

This paper presents a comparative analysis of the passive voice systems in three languages, English, Hebrew, and Arabic, in the aim of proposing rules for automatic translation from English into the other two languages. Due to their being Semitic languages there is between Hebrew and Arabic a basic similarity which goes beyond the morphological verb systems. But structural (semantic and syntactic) details required somewhat different end-rules for the automatic translation, although in many cases the general concepts seemed similar.

In the suggested rules we included some semantic information in the lexical database (the lexicon). These rules, however, have been simplified, for not all cases could be provided with such semantic information.

Some help in the analysis of the three systems was provided by the existing literature on the passive in other languages. Especially Saad's approach to the Arabic system was found useful, not only for Arabic but partly also applicable to Hebrew.

Naturally, these translation rules are tentative, and deal only with one syntactic issue. A fully automatic text translation requires much elaboration for application in combination with other syntactic elements besides full details of all the cases of passive forms. Ambiguity problems should also be tackled, and it seems preferable to do so at the source-language analysis stage, for otherwise the translation might become interpretation, which is not always considered desirable.

## Acknowledgements

We cordially thank the Haifa IBM Science and Technology Center and especially Mr M. Rimón for his permission to use their parsing system. The Faculty of Computer Sciences at the Technion, and especially Dr Y. Gold and Mr I. Dagan, are likewise gratefully acknowledged for their help.

This study is partly supported by the Technion Lawrence Deutsch Fund—VPR grant No. 2000074.

## Note

1. The Arabic transcription is the usually conventional one. Note the following items: 'sh' is as sh in English; 'x' is the velar unvoiced fricative (somewhat 'harder' than 'ch' in English 'loch'); 'q' is emphasized (velarized) 'd'; 't' is emphasized (velarized) 'l'; 'th' is as English th in 'think'; 'q' is as th in English 'this'; 'ʔ' is the glottal stop, while 'h' is the pharyngeal voiced stop; 'h' is the unvoiced pharyngeal (somewhat similar to 'ch' in English 'loch'); 'ʕ' is emphasized (velarized) s; 'q' is a velarized 'k'; there are only three vowels: 'a, i, u'; long vowels and geminated consonants are written twice (aa, ii, dd).

## References

- Berman, R. A. (1973). 'Verbal Nouns in Modern Hebrew'. Unpublished Ph.D. thesis. The Hebrew University of Jerusalem.
- Blohm, D., Reuschel, W., and Samarraia, A. (1981). *Lehrbuch des modernen Arabisch*, vol. II/2. Leipzig: Der Verlag Enzyklopaedie.
- Cantarino, V. (1974-5). *Modern Arabic Syntax*. Bloomington: Indiana University Press.
- Chayen, M. and Dror, Z. (1976). *Mavn Le-Diqduq Titsroni 'Ivri* (Introduction to Hebrew Transformational Grammar). Tel-Aviv.
- Golan, I., Lappin, Sh., and Rimón, M. (1988). An Active Bilingual Lexicon for Machine Translation, Technical Report 88.242 (July 1988). IBM Israel: The Science and Technology Center.
- Hopper, P. J. and Thompson, S. A. (1980). 'Transitivity in Grammar and Discourse', *Language*, 56, no. 1, 251-299.

Literary and Linguistic Computing, Vol. 5, No. 1, 1990

Monteil, V. (1960). *L'Arabe Moderne*. Paris.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., and Crystal, D. (1985). *A Comprehensive Grammar of the English Language*. London.

Rabin, Ch. (1963). *Tahbir Leshon Ha-Miqra* (Syntax of the Biblical Language). Jerusalem.

— (1971). *Mashma'uyoteihen shel Ha-Tsurat Ha-Diqduqiyot Bi-Leshon Ha-Miqra*. Jerusalem: Akademon.

Retsö, J. (1983). *The Finite Passive Voice in Modern Arabic Dialects*, Acta Universitatis Gothoburgensis.

Rosenhouse, J. (1988). 'Occurrence of the Passive in Different Types of Text in English, Hebrew and Arabic', *BABEL*, 34, no. 2, 90-103.

Rosenhouse, J. 'The Occurrence of the Passive in Some Bedouin Dialects', *Journal of AfroAsian Linguistics*, 3:1 (in press).

Rubinscin, E. (1971). *Ha-Tseruf Ha-Po'oli: Iyyunim Be-Tahbir Yameynu* (The Verbal Phrase: Studies in Modern Syntax). Tel-Aviv.

Saad, G. N. (1982). *Transitivity, Causation and Passivization*. London: Kegan Paul International.

Shibatani, M. (1985). 'Passives and Related Constructions: A Prototype Analysis', *Language*, 61, no. 4, 821-48.

Siewierska, A. (1984). *The Passive. A Comparative Linguistic Analysis*. London.

Svartvik, J. (1966). *On Voice in the English Verb*. The Hague.

Tsadqa, (1981). *Tahbir Ha'Ivrit Be-Yameynu* (Modern Hebrew Syntax). Jerusalem.

Wright, W. (ed. and translator) (1967). *A Grammar of the Arabic Language*, 3rd edn. Cambridge.

**Appendix I**

*Morphological comparison of the three languages*

English	Hebrew	Arabic
be	Nif'al	fu'ila
+ Past participle	Pu'al	fu'ila fu'ila u'ila
get	Huf'al	(tufu'ila?) tafa'ala
	(Hitpa'el)	(unfu'ila?) infa'ala
	(Nitpa'el)	ufu'ila (ifna'ala)
by	al-yedey	min qibali
	bi-ydey	'ala yadi

**Appendix II**

*Passive types after Siewierska, 1984*

Personal	Impersonal	Periphrastic	Reflexive
Overt passive	No real subject	Stative sense	Subject is affected
Corresponding active construction	(Possible 'dummy' subjects)	With various auxiliary verbs	Subject = object
Subject of the passive = object of the active			May be impersonal

*Examples for these passive types*

Impersonal	Periphrastic	Reflexive
—Es wurde getanzt (it was danced)	—Six men got arrested	—Si notano subito (are noted quickly le belle donne the beautiful women)
—niima fii al-layli (it was slept at night)	—S. got lost	—S. got three of her paintings accepted for an exhibition
—siira saa'atayni (it was walked)	—The glass is broken	—Es tanzt sich (it is danced gut hier well here)

*Special symbols as transcription of the Arabic letters in the translated examples*

- ʔ—Glottal unvoiced stop.
  - C—Pharyngeal voiced fricative (ʕ).
  - H—Pharyngeal unvoiced fricative (ħ).
  - x—Velar unvoiced fricative (like Scottish 'ch').
  - ?—Denti-alveolar voiced fricative (like th in 'this').
  - !—Intidental unvoiced fricative (like English th in 'thin').
  - \$—Palatal unvoiced fricative sibilant (like sh in 'Shake').
  - \$—Velarized denti-alveolar unvoiced fricative sibilant (emphatic 'š').
  - T—Velarized dental unvoiced stop (emphatic 't').
  - D—Velarized dental voiced stop (emphatic 'd').
  - q—Velarized velar unvoiced stop (retracted or 'back' English k).
  - y—Palatal semi-vowel (like English y in 'year').
  - w—Labial semi-vowel (like English w in 'win').
- Long vowels are indicated by two vowel-letters: aa, ii, uu.
- The dash indicates the boundary of preclitic particles, namely the definite article ('al-) and some prepositions (bi-, li-).

### Appendix III

#### Parts of the program, dictionary, and translation

(a.1)

```
-- Default-character-style: (:fix :roman :large) --

;; Translate a verb
(DEFUN TRAN-VERB
  (TREE VERB TENSE P-NUM PASSIVE GENDER &AUX FORM BASE PP TRANSLATION PREP TR-PREP)
  ; determine the correct form of the verb
  (COND
    ((EQ TENSE 'PAST) (COND
      (PASSIVE (SETQ FORM 'PAST-PASSIVE))
      (T (SETQ FORM 'PAST-ACTIVE))))

    (T (COND
      (PASSIVE (SETQ FORM 'FUTURE-PASSIVE))
      (T (SETQ FORM 'FUTURE-ACTIVE))))))
  (SETQ BASE (GET-ATTR VERB 'BASE))
  (SETQ *PREP* "")
  (COND
    ((AND
      (SETQ PP (NODE-POST-BROTHER VERB))
      (EQ (NODE-TYPE PP) 'PP) ; if after the verb there comes a PP
      (ONEP (LENGTH (NODE-SONS PP))) ; of length one
      (COND ; then if the verb plus the PP is a known expression
        ; which exists in the dictionary, translate it as a unit
        ((SETQ TRANSLATION
          (LOOKUP (INTERN (CONCATENATE
            'STRING
            (STRING BASE)
            (STRING (GET-ATTR (SETQ PREP (CAR (NODE-SONS PP))) 'BASE))))
            FORM GENDER P-NUM TENSE)))
          ; or if it's a required preposition of the verb, get its
          ; translation from the verb's entry in the dictionary
          ((SETQ TR-PREP (GET BASE (GET-ATTR PREP 'BASE)))
            (SETQ TRANSLATION (LOOKUP BASE FORM GENDER P-NUM TENSE))
            (SETQ *AUX-SUBJECT* T) ; an auxiliary subject is needed
            (SETQ *PREP* TR-PREP))))
      (ERASE PP TREE))
    ; otherwise translate the verb normally
    (T
      (COND ; if the verb contains a preposition, use it
        ((SETQ TR-PREP (GET BASE 'PREP)) (SETQ *PREP* TR-PREP)))
      (SETQ TRANSLATION (LOOKUP BASE FORM GENDER P-NUM TENSE))))

  TRANSLATION)
```

(a.2)

```
-- Default-character-style: (:fix :roman :large) --

;; Translate an English sentence's tree.
(DEFUN TRANS (TREE VERB-FIRST) ; VERB-FIRST is a flag, signifying whether the verb should go
  ; first in the translation
  (COND
    ((GET-ATTR TREE 'PASSIVE) ; only passive sentences are translated
      ; find the first NP that is also a subject
      (SETQ *AUX-SUBJECT* NIL)
      (SETQ *SUBJECT* (FIND-TYPE TREE 'NP (GET-ATTR *NODE* 'ISASUBJECT)))
      (SETQ *VERB* (ADD-NEG
        TREE
        (TRAN-VERB
          TREE
          (FIND-VERB TREE)
          (GET-ATTR TREE 'TENSE)
          (GET-ATTR TREE 'P-NUM)
          (GET-ATTR TREE 'PASSIVE)
          (GET (GET-ATTR *SUBJECT* 'BASE) 'GENDER))))))
    (APPEND
      (OR ; try three options in this order:
        (INDIRECT TREE) ; check for indirect speech
        (WITH-AGENT TREE VERB-FIRST) ; check for the existence of an agent
        (NORMAL TREE VERB-FIRST))) ; continue normally
      (T ('(UNABLE TO TRANSLATE))))))
```

Fig. 1. Part of the translation program.



(b.1)

```

--*-- DEFAULT-CHARACTER-STYLE: ((FIX :ROMAN :LARGE) --*--
;;;
;;; A dictionary for the program
;;;

;; put VALUE at in property list of SYMBOL under PROPERTY
(DEFMACRO PUT (SYMBOL PROPERTY VALUE) '(SETF (GET (QUOTE ,SYMBOL) (QUOTE ,(EVAL PROPERTY))) ,VALUE))

```

;; Verbs

```

(PUT EAT 'PAST-ACTIVE "akala")
(PUT EAT 'PAST-PASSIVE "ukila")
(PUT EAT 'FUTURE-ACTIVE "ya'kulu")
(PUT EAT 'FUTURE-PASSIVE "yukalu")
(PUT EAT 'AGENT 'ANIM)
(PUT BE 'PAST-ACTIVE "kaana")
(PUT BE 'FUTURE-ACTIVE "yakuunu")
(PUT SLEEP 'PAST-ACTIVE "naama")
(PUT SLEEP 'PAST-PASSIVE "niima")
(PUT SLEEP 'FUTURE-ACTIVE "yanaamu")
(PUT SLEEP 'FUTURE-PASSIVE "yunaamu")
(PUT SLEEP 'IN "fi")
(PUT SLEEP 'AGENT 'ANIM)
(PUT GIVE 'PAST-ACTIVE "aCTaa")
(PUT GIVE 'PAST-PASSIVE "uCTiya")
(PUT GIVE 'FUTURE-ACTIVE "yuCTaa")
(PUT GIVE 'FUTURE-PASSIVE "yuCTii")
(PUT GIVE 'AGENT 'ANIM)
(PUT SAY 'PAST-ACTIVE "qaala")
(PUT SAY 'PAST-PASSIVE "qilla")
(PUT SAY 'FUTURE-ACTIVE "yaquulu")
(PUT SAY 'FUTURE-PASSIVE "yuqaalu")
(PUT SAY 'AGENT 'HUM)
(PUT SAY 'REPORT-VERB T)
(PUT BARK 'PAST-ACTIVE "nabaHa")
(PUT BARK 'PAST-PASSIVE "nubiHa")
(PUT BARK 'FUTURE-ACTIVE "yambaHu")
(PUT BARK 'FUTURE-PASSIVE "yunbaHu")
(PUT BARK 'AT "bi-")
(PUT BARK 'AGENT 'ANIM)
(PUT MAKEUP 'PAST-ACTIVE "ixtaraCa")
(PUT MAKEUP 'PAST-PASSIVE "uxturiCa")
(PUT MAKEUP 'FUTURE-ACTIVE "yaxtariCu")
(PUT MAKEUP 'FUTURE-PASSIVE "yuxtaraCu")
(PUT MAKEUP 'AGENT 'HUM)

```

(b.2)

```

(PUT DO 'PAST-ACTIVE "Camila")
(PUT DO 'PAST-PASSIVE "Cumila")
(PUT DO 'FUTURE-ACTIVE "yaCmalu")
(PUT DO 'FUTURE-PASSIVE "yuCmalu")
(PUT FORGIVE 'PAST-ACTIVE "saamaHa")
(PUT FORGIVE 'PAST-PASSIVE "suu(Ha)")
(PUT FORGIVE 'FUTURE-ACTIVE "yusaamiHu")
(PUT FORGIVE 'FUTURE-PASSIVE "yusaamaHu")
(PUT FORGIVE 'AGENT 'HUM)
(PUT BUILD 'PAST-ACTIVE "banaa")
(PUT BUILD 'PAST-PASSIVE "buniya")
(PUT BUILD 'FUTURE-ACTIVE "yabnii")
(PUT BUILD 'FUTURE-PASSIVE "yubnaa")
(PUT BUILD 'AGENT 'ANIM)
(PUT FORGET 'PAST-ACTIVE "naaiya")
(PUT FORGET 'PAST-PASSIVE "nusiya")
(PUT FORGET 'FUTURE-ACTIVE "yanai")
(PUT FORGET 'FUTURE-PASSIVE "yunsa")
(PUT FORGET 'AGENT 'ANIM)
(PUT CATCH 'PAST-ACTIVE "algaa")
(PUT CATCH 'PAST-PASSIVE "ulqiya")
(PUT CATCH 'FUTURE-ACTIVE "yulqli")
(PUT CATCH 'FUTURE-PASSIVE "yulqaa")
(PUT CATCH 'PREP "Calaa")
(PUT CATCH 'ACTIVE-RECIPIENT "'al-qabDa")

```

(b.2-continued)

```

(PUT CATCH 'PASSIVE-RECIPIENT "'al-qabDu")
(PUT CATCH 'AGENT 'ANIM)
(PUT DANCE 'PAST-ACTIVE "raqaSa")
(PUT DANCE 'PAST-PASSIVE "ruqiSa")
(PUT DANCE 'FUTURE-ACTIVE "yarquSu")
(PUT DANCE 'FUTURE-PASSIVE "yurqaSu")
(PUT DANCE 'AGENT 'HUM)

```

;; Nouns

```

(PUT APPLE 'SING "tuffaaH")
(PUT APPLE 'PLUR "tuffaaHaat")
(PUT APPLE 'GENDER 'MALE)
(PUT BOY 'SING "walad")
(PUT BOY 'GENDER 'MALE)
(PUT POSTMAN 'SING "muwazziC al-bariidi")
(PUT POSTMAN 'GENDER 'MALE)
(PUT POSTMAN 'CONSTRUCT-CASE T)
(PUT STORY 'SING "Hikaaya")
(PUT STORY 'GENDER 'FEMALE)
(PUT IT 'SING "haaTa")
(PUT IT 'GENDER 'MALE)

```

(b.3)

```

(PUT HOUSE 'SING "bayt")
(PUT HOUSE 'GENDER 'MALE)
(PUT PROBLEM 'SING "muSkila")
(PUT PROBLEM 'GENDER 'FEMALE)
(PUT YEAR 'SING "sana")
(PUT YEAR 'GENDER 'FEMALE)
(PUT THIEF 'SING "saariq")
(PUT THIEF 'GENDER 'MALE)
(PUT DANCE 'SING "raqSa")
(PUT DANCE 'PLUR "raqSaat")
(PUT DANCE 'GENDER 'FEMALE)
(PUT EVENING 'SING "masaa")
(PUT EVENING 'GENDER 'MALE)
(PUT BED 'SING "sariri")
(PUT BED 'GENDER 'MALE)

```

;; Adverbs

```

(PUT QUICKLY 'ANY-GENDER "bisurCatin")

```

;; Adjectives

```

(PUT BIG 'MALE "kabiir")

```

;;; Determiners

```

(PUT THIS 'MALE "haaTa")
(PUT THIS 'FEMALE "haa?ihi")
(PUT THAT 'MALE "?aalika")

```

;; Quantifiers

```

(PUT THREE 'MALE "ialaa!atu")
(PUT THREE 'FEMALE "ialaa!u")

```

;; Prepositions

```

(PUT AT 'ANY-GENDER "bi-")
(PUT IN 'ANY-GENDER "fi")
(PUT TO 'ANY-GENDER "li-")

```

Fig. 2. Part of the dictionary used with the translation program.

(c.1)

-- Default-character-style: (:fix :roman :large) --

```
THE BIG APPLE WAS EATEN >>> 'ukila 'al-tuffaaH 'al-kabiir
THE BIG APPLE WAS EATEN BY THE BOY >>> 'akila 'al-walad 'al-tuffaaH 'al-kabiir
THE BOY WAS GIVEN THE APPLE >>> 'uCTiya 'al-walad 'al-tuffaaH
IT IS SAID THAT THE APPLE WAS EATEN >>> yuqaalu 'inna 'al-tuffaaH 'ukila
THE POSTMAN WAS BARKED AT >>> nubiiHa bi-muwazziC al-bariidi
THE STORY WAS MADE UP >>> 'uxturiCat 'al-Hikaaya
IT CAN BE DONE QUICKLY >>> yuCaalu haa?a bisurCatin
IT CAN'T BE FORGIVEN >>> laa yusaamaHu haa?a
THE HOUSE CAN BE BUILT IN A YEAR >>> yubnaa 'al-bayt fii sana
THIS PROBLEM CANNOT BE FORGOTTEN >>> laa tunaaa haa?ihi 'al-muSkila
THE THIEF WAS CAUGHT IN THE HOUSE >>> 'ulqiya 'al-qabDu Calaa 'al-seariq fii 'al-bayt
THREE DANCES WERE DANCED THAT EVENING >>> ruqiSat lalaalu raqSaat ?aalika 'al-maaa'
THE BED WAS NOT SLEPT IN >>> maa niima fii 'al-sariiri
```

THANK-YOU

(c.2)

-- Default-character-style: (:fix :roman :large) --

```
THREE APPLES WERE NOT GIVEN TO THE BOY >>> maa 'uCTiyat lalaatu tuffaaHaat li-'al-walad
THE THIEF WAS CAUGHT BY THE POSTMAN >>> 'alqaa muwazziC al-bariidi 'al-qabDu Calaa 'al-seariq
THE BOY CAN'T BE FORGIVEN >>> laa yusaamaHu 'al-walad
THE STORY CAN BE MADE UP QUICKLY >>> tuxtaaruCu 'al-Hikaaya bisurCatin
IT IS SAID THAT THE BED WAS NOT SLEPT IN >>> yuqaalu 'innahu maa niima fii 'al-sariiri
```

THANK-YOU

Fig. 3. Examples of translated sentences.