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

In the following we briefly review the morphology and syntax 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).

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-sea basis, were studied at a preliminary stage of this research. It was hoped that an analysis of these issues may bring out underlying linguistic structures of each of these languages and serve as a basis for defining rules for automatic translation.

Siewierska (1964) analyses the facts along lines of comparative linguistics and as the general point of view of universal linguistics, though, as she notes, not all languages attest the passive. Also Shibatani (1965) describes the same 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 as an active verb form in a passive one by a fixed change of the word patterns. Thus, in Arabic, for example: 'kana aqada' (he wrote) vs. 'kana hada' (he was written); 'kana he wrote to someone) vs. 'kana he was written to'; 'kana he made someone write' vs. 'kana he was made to write'; etc. And in Hebrew too: 'sippa' (he told the story) vs. 'sippa' (it was told); 'hastin' (he dictated) vs. 'hastin' (it was dictated). But in contrast to the productivity of this passivization process in Arabic, in Hebrew the verb noun may be passivized by this device.

Hebrew and Arabic have basically similar verbal systems, which include two ways of passive 'transformations' (see Appendix 1):

1. An internal passive, which transforms an active verb-form into a passive one by a fixed change of the word patterns. Thus, in Arabic, for example: 'kana aqada' (he wrote) vs. 'kana hada' (he was written); 'kana he wrote to someone) vs. 'kana he was written to'; 'kana he made someone write' vs. 'kana he was made to write'; etc. And in Hebrew too: 'sippa' (he told the story) vs. 'sippa' (it was told); 'hastin' (he dictated) vs. 'hastin' (it was dictated).

2. A special pattern or 'measure' which denotes the passive through its special prefix: 'maful ala' in Arabic and 'maful' in Hebrew. As a matter of fact, Saad (1982, ch. 6) shows that 'maful ala in Arabic is not a real passive but a reflexive. This observation is true in part, at least, also concerning the use of 'naf' in Hebrew. For though 'naf' is a usually the passive of verb forms in the basic measure, 'naf' ala, 'naf' is sometimes used with non-passive meanings, e.g. 'naf akh (meet), 'naf' ala (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. Thsi,

* The paper is a revised and expanded version of a paper read at the 16th 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.
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 gone 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 (Sieverska, 1984: 75). Sieverska thinks that both processes are important, but "demotion" is significant especially for the impersonal passive.

Sieverska (1984) classifies the passive into four types which are attested in many languages: personal, imper- sonal, periphrastic, and reflexive passive as well as "exceptions" (see Appendix III).

Hopper and Thompson (1980) analyze the passive as part of the contrasts of transitivity and of topicalization, i.e., transporting 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 passive, inchoative, and active); Interpersonal verbs (in measure III); Resultative verbs (e.g. solve, destroy, build, stop); Emotive and Cognitive verbs (see, 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 Miscellaneous (Expose verbs, Decline verbs, Remove verbs, Surpass verbs).

Among the non-passivizable verbs Saad includes co- place verbs (tense, inchoative, or active-intransitive), defective verbs and "be", and verbs of the type of "cast" and "remove". (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 direct 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 Sieverska (1984) describes, impersonal passives are derived of intransitive verbs, and many of them are also locatives or time expressions. Cf. "nima fi al-bayt (direct: it was slept in the house), "nima li-ibad (it was gone to Jerusalem). Saad considers such forms as 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. nima... was slept) or 'unknown' (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. Rabbin, 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 possess 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 nuances. 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 II in Arabic, "the" or "which" 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) these are differences in the rate of occurrence of the passive. English appears to have the largest rate of passive occurrence, Hebrew follows, and Arabic lies 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 'it is' sentence structure is preferred for this purpose. This reduces the rate of passive occurrence in Arabic compared with English, though it is more use of the passive forms in English (about two-thirds in fact—cf. Sourvik, 1986) 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. among 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 includes '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 the 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 the translation of a one-to-one correspondence with the English source. The main categories that are relevant for Hebrew are (according to Berman, 1973): transitivity, reflexivity, middle, and passivity. Also the perceptive 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 "md/nd and hif/nd (real passive of "qal), but often non-passive at all) vs. hitpael 'emipta 'el (reflexive, sometimes passive). Such information, however, would be included more easily in the bilingual 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. Reiss, 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 at all the levels of linguistic structure. Thus, it is only logical to incorporate them somewhere 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 synthetic 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 a noun 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 the original English text, if no other restrictions prevent it. CI.

The apple was eaten by the boy

Ari al-waladu bi-l-taṣṣabita (active)

Ari al-waladu bi-l-taṣṣabi (passive), but

The apple was eaten by the boy

Ari al-waladu bi-l-taṣṣabita ha-tsabita (active)

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

Litawthu al-kitabu ha-seter mitzav.
The book was hidden at once-wa'adda al-kitabu fii al-bah/ha-tefer hubshu nida. 

The picture was given to him--'iyyatu al-quratu fii hakima mima a. 

He was given the picture---liwaa V'yahu al-suwaratu (*la' tina la'dinat et ha-tmuna). 

This rule appears to be valid also for cases where in English the auxiliary verb is 'get'. (This rule reflects Stand'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") rendered 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...---yuqatsu 'lama...hunemmar... (There are, of course, other translation options, and some verbs require different options because they cannot appear in the passive at all but as needed, 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-basasa mu'llad. (Different lexeme, using the passive participial form.)

Haaa al-basasa. (The same lexeme as above, but in the past tense.)

Hebrew: ha-bashon mishulam. (Different lexeme using the passive participle form.)

ba-basho 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: da'aa' wu al-basasa. Hebrew: shima' bi-ka-bashon. (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 participial form with the copula 'to be' indicating the tense, rather than the passive verb form in the tense tense combinations (past/non-passed). For example:

He was loved by all---kaama mabhuuban la' a-blkhul he was beloved on all (*'ishaa min qubul al-bakhul). 

hu haya 'shva al-ba-kul he was beloved on the all (*hu nena biyade'y kulfu.)

(This modification somewhat reduces the dynamic aspect 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 semantic-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 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 (Moncel 1969) considers this structure a major characteristic of modern Arabic style. Cf. also Blohm et al. (1984), vol 2/II, for example:

It can be done easily---haqda yu'mal bi-nahaddatin.

The house can be built in a year---yinbusa al-bayt u fisanin.

This problem cannot be forgotten---laa tunnasa haqda al-qaydyyatu.

(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. Thus:

This matter is unfor neglectable---laa yusa'muha haqq du al-lamru.

This food is not edible---laa yu' ala hu sada haqda al-ta'a'amu. This story is unbelievable---laa tuu'tadduqqu haqda al-bikaaya
tu.

In Hebrew the rule must be defined separately. We suggest the following intransitive 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 nj'al (e.g. yqos l-bhe'satis). Cf.:

It can be done easily---nj'al yid bhe'satis be-qolit.

The house can be built in a year---laa baya' yid bhe'satis be-nahaddathu.

(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. 'nitya qet) (+active verbal noun, or of 'nitya qet) (+active verbal noun') (++accusative marker, and/or the negative particle "i", if required, e.g.:

The thing is unfor neglectable---ha-bayy yqos ni'ta li-sili'tu, the thing not to be forgotten.

This food is not edible---ha-mazn a-bay yqos ni'ta la-asdu, 'the food this not to be eating'.

This story is unbelievable---ni'sha yinbusa la'shum la-ha, 'not possible to believe to-the story this'.

Library and Linguistic Computing, Vol 5, No. 1, 1980
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 lack of occurrence (cf. Herman, 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 structure of the passive of indirect 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 occupied. The parallel example: 'The building is already occupied by the workers' is not acceptable in English tense-wise; it should be either 'The building has been already occupied by the workers' or 'The building is being (already) occupied by the workers'. Thus, the intuitive 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 it:

The house is already built—Arabic: al-baytu mah-niyin 'the house built' (passive participle without translation of 'already');

or: al-baytu saara mah-liyyan 'the house became built'---Hebrew: ha-bayt kvar haruy, '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 'passivised' 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 unexpressed by his indifference') we suggest the following rule, to distinguish between participle adjectives and real passives:

**Rule VIII.** If the participial is passive (rather than dynamic) and if the head noun of the 'by phrase' is not the agent of the participial form, then the participial 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 'adjectival' in the dictionary, which will prevent such translation problems.

He was killed by the river—Arabic: 'inda al-nahhū bi neher bra-yad ha-nahhar.

He was tired by the river—Arabic: 'inda al-nahhū blur y-layl yad ha-nahhar.

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 IRM Science and Technology Center as a front end syntactic analyzer.

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 II). 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 'that' (as described in rule II). In such cases, 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 'that'. The rest of the sentence follows.

The program uses a dictionary. It includes the translations of English words into Arabic words, 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 ‘wax’), 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 com-

pound verb phrase (e.g. ‘caught’—’Aalaa al-qabda ‘la’—literally, ‘threw the grabbing out’). A simple dic-

tionary look-up is not adequate, and a new object (‘al qabda’ 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 that deal only with the passive structure. However, when real translation is involved, morpho-syntactic rules have 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 either derived and therefore they were treated here for the example by ‘ad loc 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 ordinary 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 differences between our above examples for the rules (e.g. ‘yehna al-bayti fii sina’ for ‘yehna al-bayti fii anattin’). 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 imperative 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 (etc., for instance, Golub et al., 1980), 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 preparing 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 (semant-

ic 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 Saade’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 translator 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. Rimon 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. 2000/74.

Note

1. The Arabic transcription is the usually conventional one.

Note the following examples: ‘al in Arabic; ‘it is the unvoiced fricative (somewhat ‘fatter’ than ‘sh’ in English ‘vetch’); ‘sh’ is emphasized (vestibular) ‘t’, ‘t’ is emphasized (vestibular) ‘l’; ‘th’ in an English th as in ‘think’; ‘v’, ‘i’ as in English this; ‘th’ is the glottal stop, while ‘th’ is the phonetical vowel step; ‘t’ is the unvoiced pharyngeal (somewhat similar to ‘th’ in English ‘teeth’); ‘t’ is emphasized (vestibular) ‘t’; ‘th’ is a voiced ‘t’; there are only three vowels: ‘a’, ‘i’, ‘u’; long vowels and geminated consonants are written twice (as ‘ii’, ‘vv’).

References


klopddie.


Clarys, M. and Droiz, Z. (1976), Mein Le-Disque Tetron Tet (Introduction to Hebrew Transformational Grammar), Tel-

Aviv.

Golub, I. Lappin, Sh., and Rimon, M. (1988), An Active Bilingual Lexicon for Machine Translation, Technical Re-


Appendix I

Morphological comparison of the three languages

<table>
<thead>
<tr>
<th>English</th>
<th>Hebrew</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>be</td>
<td>נָפָל</td>
<td>fiṭat</td>
</tr>
</tbody>
</table>
| Past participle | פָּזִיל | fiṣa ᵐα | fiṣa (“it was broken”)
| get     | חָפַל | ṣama ᵐα | ṣama (“it was walked”)
| by      | alayd | min qibbi | ala yaddi |

Appendix II

Passive types after Sinkiewicz, 1984

<table>
<thead>
<tr>
<th>Imperfective</th>
<th>Perfective</th>
<th>Reflexive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overt passive</td>
<td>No real subject</td>
<td>Passive sense</td>
</tr>
<tr>
<td>Corresponding construction</td>
<td>With various auxiliary verbs</td>
<td>Subject = object</td>
</tr>
</tbody>
</table>

Examples for these passive types

<table>
<thead>
<tr>
<th>Imperfective</th>
<th>Perfective</th>
<th>Reflexive</th>
</tr>
</thead>
<tbody>
<tr>
<td>fiṭat ᵐuṣṣal</td>
<td>fiṭat ᵐuṣṣal</td>
<td>fiṭat ᵐuṣṣal</td>
</tr>
<tr>
<td>(it was stepped on)</td>
<td>(it was stepped on)</td>
<td>(it was stepped on)</td>
</tr>
<tr>
<td>sink as an object</td>
<td>sink as an object</td>
<td>sink as an object</td>
</tr>
<tr>
<td>alayd ᵐaḥṣar</td>
<td>alayd ᵐaḥṣar</td>
<td>alayd ᵐaḥṣar</td>
</tr>
<tr>
<td>(it was walked)</td>
<td>(it was walked)</td>
<td>(it was walked)</td>
</tr>
<tr>
<td>nawāṣṣa ᵐaḥṣar</td>
<td>nawāṣṣa ᵐaḥṣar</td>
<td>nawāṣṣa ᵐaḥṣar</td>
</tr>
<tr>
<td>alayd ᵐaḥṣar</td>
<td>alayd ᵐaḥṣar</td>
<td>alayd ᵐaḥṣar</td>
</tr>
<tr>
<td>(it was broken)</td>
<td>(it was broken)</td>
<td>(it was broken)</td>
</tr>
<tr>
<td>sink as an object</td>
<td>sink as an object</td>
<td>sink as an object</td>
</tr>
</tbody>
</table>

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

- Glottal unvoiced stop.
- Pharyngeal voiced fricative (‘).
- Pharyngeal unvoiced fricative (b).
- Velar unvoiced fricative (like Scottish ‘ch’).
- Dental-alveolar voiced fricative (like th in ‘this’).
- Incidental unvoiced fricative (like English th in ‘thin’).
- Palatal unvoiced fricative sibilant (like sh in ‘shake’).
- Velarized dental unvoiced stop (emphatic ‘t’).
- Velarized dental unvoiced stop (emphatic ‘d’).
- Velarized velar unvoiced stop (retracted or ‘back’ English ‘k’).
- Labial semi-vowel (like English ‘y’ in ‘year’).
- Labial semi-vowel (like English ‘w’ in ‘win’).

Long vowels are indicated by two vowel-letter: au, ii, uu.

The dash indicates the boundary of predicative particles, namely the definite article (‘a-’) and some prepositions (bit-, li-).
Appendix III

Parts of the program, dictionary, and translation

(a.1) ;-- default-character-style: t;fix roman {large} ---

;;; Translate a verb

(setq tran-verb
  (lookup-base form (make-base name form base pp translation prep tr-thrs)
    ;; determine the correct form of the verb
    (cond
      ((eq these '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 (set-attr verb 'base))
    (setq 'prep "")
    (cond
      (and
        (setq pp (make-base-from-noun verb))
        (eq (mode-type pp) 'pp)
        (equal (length (mode-nouns pp)) 0)
        (cond
          ;; check if the verb plus the PP is a pronoun 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 (mode-nouns pp))) 'base))))
            (t
              ;; translation from the verb's entry in the dictionary
              (setq tran-verb (set base (get-attr prep 'base))
                (setq translation (lookup-base form (make-base name form base pp translation prep tr-thrs)))
                (cond
                  (aux-subject
                    (t
                      ;; an auxiliary subject is needed
                      (setq 'prep (set-prep 'tr-prep)))))
                (t tran-verb)))))
      (else
        ;; otherwise translate the verb normally
        (cond
          (prep (setq tran-prep (get base 'prep))
            (setq 'prep 'tr-prep))
          (t tran-verb)))))
    (t (setq translation (lookup-base form (make-base name form base pp translation prep tr-thrs))))

TRANSLATION

(a.2) ;-- default-character-style: t;fix roman {large} ---

;; Translate an English sentence's tree.

(setf tran (tree-verb-first) 'verb-first is a flag, signifying whether the verb should go first in the translation
  (cond
    (get attr tree 'passive)
      (cond
        (get attr tree 'aux-subject)
          (find-type tree 'np (get attr 'node') 'auxsubject))
        (get attr 'verb) (find-type tree 'np))
    (append
      (or
        (get attr tree 'verb-first)
        (get attr tree 'ayer)
        (get attr tree 'tr-prep)
        (get attr tree 'passive)
        (t)
        (translate)))

Fig. 1. Part of the translation program.
(c.1) Default-character-style: (fixed roman large) ***

THE BIG APPLE WAS Eaten
THE BIG APPLE WAS Eaten BY THE BOY
IT IS SAID THAT THE APPLE WAs Eaten
THE POSTMAN WAS BARKED AT
THE STORY WAS MADE UP
IT CAN'T BE DONE QUICKLY
IT CAN'T BE FORGIVEN
THE HOUSE CAN BE BUILT IN A YEAR
THIS PROBLEM CANNOT BE FORGOTTEN
THE TRIP WAS CAUGHT IN THE HOUSE
THREE DANCES WERE DANCED THAT EVENING
THE BED WAS NOT SLEPT IN

THANK YOU

(c.2) Default-character-style: (fixed roman large) ***

THREE APPLES WERE NOT GIVEN TO THE BOY
THE CHEF WAS CAUGHT BY THE POSTMAN
THE BOY CAN'T BE FORGIVEN
THE STORY CAN'T BE MADE UP QUICKLY
IT IS SAID THAT THE BED WAS NOT SLEPT IN

THANK YOU

Fig. 3. Examples of translated sentences.