Issues in the Morphological Analysis of the Arabic Passive Verb

Khaled F. Shaalan
Institute of Informatics
The British University in Dubai
P. O. Box 502216, Dubai,UAE
Honorary Fellow, School of Informatics, University of Edinburgh
khaled.shaalan@buid.ac.ae

Eman Othman
Computer Science Dept., Institute of Statistical Studies and Research (ISSR), Cairo Univ.
5 Tharwat St., Orman, Giza, Egypt
emy_othman@yahoo.com

Abstract. Arabic is a strongly structured and highly derivational language. Arabic morphology and syntax provide the ability to add a large number of affixes to each word which makes combinatorial increment of possible words. In Arabic, passive voice is used as a writing style when: 1) the subject is unknown, 2) the subject is unimportant enough to be mentioned, or 3) the author wants to highlight the object. In this paper, the issues related to the recognition of the Arabic passive verbs which impact the automated understanding of Arabic sentences were addressed. An experiment using the Buckwalter Arabic morphological analyzers, one of the mature Arabic morphological analyzer, were conducted in order to highlight the limitations in the analysis of Arabic passive verbs. Results indicated that there exists a need for handling the problems related to the morphological analysis of passive verbs in order to improve the recognition accuracy of Arabic words.

1 Introduction

Arabic is a strongly structured and highly derivational language (Kiraz, 2001). Understanding Arabic requires the treatment of the language constituents at all levels: morphology, syntax, and semantics. Each component requires extensive study and exploitation of the associated linguistic characteristics (Black, 2004).

Arabic morphology and syntax provide the ability to add a large number of affixes to each word which makes combinatorial increment of possible words (Ditters, 2001; Jaccarini, 2001; Habash, 2004). This is the reason that most of the researches on Arabic natural language processing are mainly concentrated in the field of morphological analysis (Farghaly; 1987; Rafea et al., 1993; Al-Shalabi et al., 1998; Beesley, 2000; Freeman, 2001; Darwish, 2002; Soudi et al., 2003). For a comprehensive survey of Arabic morphological analysis further reference can be made to Al-Sughalyer et al. (2004).

In Arabic, passive voice is used as a writing style when: 1) the subject is unknown, 2) the subject is unimportant enough to be mentioned, or 3) the author wants to highlight the object. In this paper, the issues related to the recognition of the Arabic passive verbs which impact the automated

understanding of Arabic sentences were addressed. An experiment using the Buckwalter Arabic morphological analyzers, one of the mature Arabic morphological analyzer, were conducted in order to highlight the limitations in the analysis of Arabic passive verbs. Results indicated that there exists a need for handling the problems related to the morphological analysis of passive verbs in order to improve the recognition accuracy of Arabic words.

The paper is structured as follows. In Section 2, the relevant aspects of Arabic passive verbs were described. Special attention was given to the Arabic verb form when it changes from active voice to passive voice. In Section 3, an experiment conducted to analyze a test suite of Arabic passive verbs using the Buckwalter Arabic morphological analyzer was presented. Section 4 the analysis of the results of this experiment was discussed. In Section 5, some concluding remarks were given. An Appendix was used to list a sample output from the test suite.

2 Linguistic Aspects of Arabic Passive Verbs

Arabic words are divided into three types: noun, verb, and particle (Shaalan, 2005). The verb is any word that indicates the occurrence of an action. The Arabic verb is mainly classified according to the following: tense (perfect, imperfect and future), object (intransitive, transitive), structure (sound, weak), syntax (declinable, indeclinable), and voice (active, passive).

Figure 1 shows a partial classification hierarchy of Arabic verbs into strong and weak verbs. For each verb, the conjugations of perfect and imperfect forms in both the active and passive voices were shown. For the purposes of this paper, the Buckwalter transliteration convention[1] was followed.

The correct morphological analysis of passive verbs plays a very important role in the automated understanding of Arabic sentences. As an example, consider the following unvocalized verbal sentence (Othman et al., 2004):

أكل الطعام
The food was eaten ("أُكِلَ" >ukila passive voice)
The food ate ("أَكَلَ" >akala active voice) *wrong analysis

The recognition of this verb just as active voice ("أَكَلَ" >akala) would lead to either incorrect understanding of the Arabic sentence or failing to parse this sentence.

3 An Experiment

In this experiment, the Buckwalter Arabic morphological analyzer[2] was used to get the analysis of a test suite of Arabic passive verbs. The aim of this experiment was to highlight the gap, even in the mature Arabic

[1] See Buckwalter's transliteration convention http://www.qamus.org/transliteration.htm
[2] See the description of the Arabic morphological analysis using the Buckwalter Arabic morphological analyze http://www.qamus.org/morphology.htm

morphological analyzer, in the morphological processing of Arabic passive verbs. It is worth noting that the objective of this experiment was not to evaluate the Buckwalter Arabic morphological analyzer.

The approach that was followed in this experiment was to: 1) select a test suite text comprising both classic and modern standard Arabic (MSA) sentences, 2) extract the Arabic passive verbs from these sentences, 3) run these verbs on a widely available and recognized Arabic morphological analyzer, and 4) discuss the obtained results.

The MSA sentences are randomly selected from newswire and technical brochures. The classic Arabic sentences include all verses of the last (thirteenth) part of the Holy Qur'an. We have extracted 280 passive verbs from this test suite: 50 verbs from the agricultural domain, 106 from newswire, and 124 from the Holy Qur'an. This is shown in Table 1.

Table 1 Test suite

	Tech. brochure	**Newswire**	**Qur'an**	**Total**
Strong Verbs	38	20	92	150
Weak Verbs	12	86	32	130
Total	50	106	124	280

4 Analysis of Results

The test suite of 280 passive verbs was fed into the Buckwalter Arabic morphological analyzer. In the Appendix, a sample output from this experiment was given. Table 2 shows the percentage of the recognized verbs for both strong and weak passive verbs. Table 3 shows the percentage of the recognized perfect and imperfect passive verbs. Table 4 shows the percentage of the recognized verbs in the different cases. In particular, it shows the cases of strong passive verbs as well as the different types of weak passive verbs. In each case, the perfect tense as well as the imperfect tense were explored.

Table 2 Percentage of recognized passive verbs (Strong vs. weak)

Passive Verbs	**Passive recognized**	**%**
Strong Verbs	**No**	62%
	Yes	38%
Weak Verbs	**No**	55%
	Yes	45%

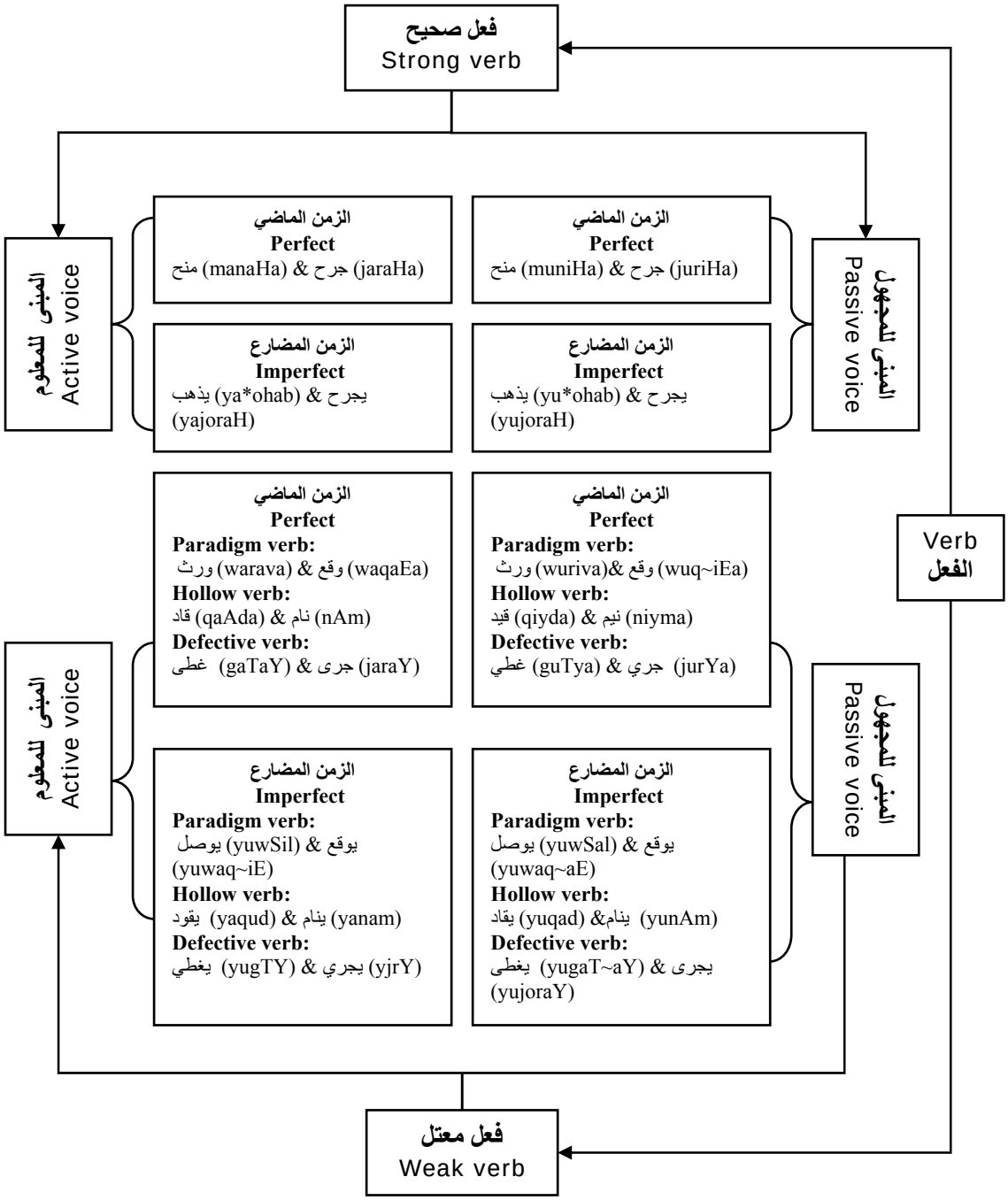

Figure 1 A partial classification hierarchy of Arabic verbs.

Table 3 Percentage of recognized passive verbs (perfect vs. imperfect)

Passive Verbs	Passive recognized	Perfect Tense	Imperfect Tense
Strong Verbs	**No**	70%	53%
	Yes	30%	47%
Weak Verbs	**No**	86%	21%
	Yes	14%	79%

Table 4 Percentage of recognized passive verbs (strong vs weak, perfect vs imperfect, and considering different types of weak verbs)

Passive Verbs		Passive recognized	Perfect Tense	Imperfect Tense
Strong Verbs		**No**	70%	53%
		Yes	30%	47%
Weak Verbs	**Paradigm**	**No**	90%	36%
		Yes	10%	64%
	Hollow	**No**	81%	33%
		Yes	19%	67%
	Defective	**No**	88%	9%
		Yes	12%	91%

The morphological analysis of the 280 input passive verbs has shown that 164 (59%) passive verbs were not recognized by the Buckwalter Arabic morphological analyzer. The number of the unrecognized strong passive verbs is 93. The number of the unrecognized weak passive verbs is 71. In the following we classify the problems related to the unrecognized passive verbs:

- *Unrecognized passive verbs due to the change in weak letter.* This category includes defective and hollow (with Alef) passive verbs. There are 104 input passive verbs of this category. The results indicate that the number of the unrecognized verbs of this category is 55 (53%). They have their weak letter changed in its passive voice. For example, the weak letter "ا" (Alef) in the verb "تَاجَرَ " (tajara- perfect active voice) is changed to "و" (Waw) in its passive voice "تُوجِرَ" (tujira- perfect passive voice). In particular, 35 verbs out of these 55 verbs (65%) are failed to be recognized in the perfect tense although it is recognized in the imperfect tense. The possible

cause of this problem is that the weak letter in the passive imperfect verb is the same as the active perfect verb such that the verb is recognized in the passive imperfect tense more often than in the passive perfect tense. For example, the defective active perfect verb "أَجْرَى" (>ajoraY) is not recognized in its passive perfect tense " أُجْرِيَ " (>ujoriya) while it is recognized in the passive imperfect "يُجْرَى" (yujoraY). Also, the hollow (with Alef) active perfect verb "قَالَ" (qaAla) is not recognized in its passive perfect tense "قِيْلَ" (qiyla) while it is recognized in the passive imperfect "يُقَالُ" (yuqAlu). We suggest that the morphological analyzer detects the change in the weak letter in order to handle this problem.

- *Unrecognized passive verbs due to the change in the diacritic signs.* This category includes strong, paradigm, and hollow (other than Alef) passive verbs. There are 176 input passive verbs of this category. The results indicate that the number of the unrecognized verbs of this category is 109 (60%). 85% of these verbs are strong verbs and 15% are weak verbs. The possible cause of this problem is that both the unvocalized active and passive voices are identical. For example, the strong active verb "أَكَلَ" (>akala) looks like its passive voice "أُكِلَ" (>ukila), the paradigm active verb "وَصَلَ" (waSala) looks like its passive voice "وُصِلَ" (wuSila), the hollow (other than Alef) active verb "سَوَّرَ" (saw~ara) looks like its passive voice "سُورَ" (suw~ira). We suggest that the morphological analyzer gives the two possible voices as output and leaves the distinction to the subsequent analysis stages.

5 Conclusion

As Arabic is a highly inflected language, there will always be possibilities for improving its natural language processing approaches and tools. In this paper, issues related to the Arabic morphology of passive verbs were discussed. The changes occurred on the strong and weak verbs when they come in the passive voice were explained. The findings of the conducted experiment with a test suite that include both MSA and classic Arabic passive verb forms seem to show advantages that may arise from handling the gaps in the recognition of Arabic passive verbs. To distinguish between active and passive voices for Arabic strong verbs and some forms of weak verbs, we need to recognize the diacritic signs which are usually unavailable. In this case, both alternative features should be produced from the analysis of the verb. Afterwards, during the sentence analysis the recognition of a proagent may also be used to select the right passive verb. Moreover, in certain forms of weak verbs, the passive voice can be distinguished using the changes in the weak letter(s). The correct morphological analysis of Arabic passive verbs will improve the word recognition accuracy which plays very important role in the automated understanding of Arabic sentences.

References

1. Riyad Alshalabi and Martha Evens. A Computational Morphology System for Arabic. Workshop on Semitic Language Processing. In the Proceedings of the COLING-ACL-98, University of Montreal, Montreal, PQ, Canada, pp. 66-72, 1998.
2. Imad A. Al-Sughaiyer and Ibrahim A. Al-Kharashi, Arabic Morphological Techniques: A Comprehensive Survery, Journal of the American Society for Information Science and Technology, 55(3):189-213, John Wiley & Sons, U.K., 2004.
3. Kenneth R. Beesley and Lauri Karttunen, Finite-State Non-Concatenative Morphotactics, In the Proceedings of the 38th Annual Meeting of the Association for Computational inguistics, Hong Kong, pp.191-198, 2000.
4. William J. Black and Sabri El-Kateb, A Prototype English-Arabic Dictionary Based on WordNet. In the Proceedings of the Second International WordNet Conference— GWC 2004, Brno, Czech Republic, 2004.
5. Kareem Darwish, Building a Shallow Arabic Morphological Analyzer in One Day, In the Proceedings of the Computational Approaches to Semitic Languages, A workshop affiliated with ACL-2002, University of Pennsylvania, 2002.
6. Everhard Ditters, A Formal Grammar for the Description of Sentence Structure in Modern Standard Arabic, In the proceeding of Arabic NLP Workshop at ACL/EACL, 2001.
7. Ali Farghaly, Three-level Morphology, In Proceeding of the Arabic Morphology Workshop, Stanford University, 1987.
8. Andrew Freeman, Brill's POS Tagger and a Morphology parser for Arabic, In the proceedings of Arabic NLP Workshop at ACL/EACL, 2001.
9. Nizar Habash, Large scale lexeme based Arabic morphological generation. In Proceedings of Traitement Automatique du Langage Naturel (TALN-04). Fez, Morocco, 2004.
10. André Jaccarini, A modifiable structural editor of grammars for Arabic processing, In the proceeding of Arabic NLP Workshop at ACL/EACL, 2001.
11. George Kiraz, Computational Nonlinear Morphology: with Emphasis on Semitic Languages, Cambridge University Press, UK, 2001.
12. Eman Othman, Khaled Shaalan, and Ahmed Rafea, Towards Resolving Ambiguity in Understanding Arabic Sentence, In the Proceedings of the International Conference on Arabic Language Resources and Tools, NEMLAR, PP. 118-122, 22nd–23rd Sept., Egypt, , 2004.
13. Ahmed Rafea, and Khaled Shaalan, Lexical Analysis of Inflected Arabic words using Exhaustive Search of an Augmented Transition

Network, Software Practice and Experience, Vol. 23(6), pp. 567-588, John Wiley & sons, U.K., 1993.
14. Shaalan K. Arabic GramCheck: A Grammar Checker for Arabic, Software Practice and Experience, John Wiley & sons Ltd., UK, 35(7):643-665, June 2005.
15. Abdelhadi Soudi and Violetta Cavalli-Sforsa, Interfacing an Arabic Morphology and Sentence Generation with an English-to-Arabic knowledge-based Machine Translation System, In the Proceedings of the workshop on Information Technology, Rabat, Morocco, March 17-19, 2003.

Comments	Perfect Verbs	imperfect Verbs
Passive voice of a strong verb is produced in imperfect tense but not included in the perfect tense	INPUT STRING: ذهب LOOK-UP WORD: *hb SOLUTION 1: (*ahaba) [*ahab-a_1] *ahab/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + go/depart + he/it <verb> SOLUTION 2: (*ahaba) [*ahab-a_2] *ahab/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + take (with) + he/it <verb> SOLUTION 3: (*ah~aba) [*ah~ab_1] *ah~ab/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + gild + he/it <verb> SOLUTION 4: (*ahab) [*ahab_1] *ahab/NOUN (GLOSS): + gold + SOLUTION 5: (*ahab) [*ahab_2] *ahab/NOUN_PROP (GLOSS): + Dhahab/Zahab +	INPUT STRING: يذهب LOOK-UP WORD: y*hb SOLUTION 1: (ya*ohab) [*ahab-a_1] ya/IV3MS+*ohab/VERB_IMPERFECT (GLOSS): he/it + go/depart + SOLUTION 2: (ya*ohab) [*ahab-a_2] ya/IV3MS+*ohab/VERB_IMPERFECT (GLOSS): he/it + take (with) + SOLUTION 3: (yu*ah~ib) [*ah~ab_1] yu/IV3MS+*ah~ib/VERB_IMPERFECT (GLOSS): he/it + gild + SOLUTION 4: (yu*ohib) [>a*ohab_1] yu/IV3MS+*ohib/VERB_IMPERFECT (GLOSS): he/it + remove/eliminate + SOLUTION 5: (yu*ohab) [>a*ohab_1] yu/IV3MS+*ohab/VERB_IMPERFECT (GLOSS): he/it + be removed/be eliminated +
Passive voice of a strong verb is produced in both perfect and imperfect tenses	INPUT STRING: منح LOOK-UP WORD: mnH SOLUTION 1: (manaHa) [manaH-a_1] manaH/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + grant/award/bestow + he/it <verb> SOLUTION 2: (muniHa) [manaH-a_1] muniH/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + be granted/be awarded/be bestowed + he/it <verb> SOLUTION 3: (manoH) [manoH_1] manoH/NOUN	INPUT STRING: يمنح LOOK-UP WORD: ymnH SOLUTION 1: (yamonaH) [manaH-a_1] ya/IV3MS+monaH/VERB_IMPERFECT (GLOSS): he/it + grant/award/bestow + SOLUTION 2: (yumonaH) [manaH-a_1] yu/IV3MS+monaH/VERB_IMPERFECT

	(GLOSS): + granting/bestowal/awarding + SOLUTION 4: (minaH) [minoHap_1] minaH/NOUN (GLOSS): + grants/scholarships/gifts +	(GLOSS): he/it + be granted/be awarded/be bestowed +
Passive voice of a strong verb is neither produced in perfect nor imperfect tenses	INPUT STRING: لعب LOOK-UP WORD: lEb SOLUTION 1: (laEiba) [laEib-a_1] laEib/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + play + he/it <verb> SOLUTION 2: (laEob) [laEob_1] laEob/NOUN (GLOSS): + game/sport + SOLUTION 3: (liEab~) [Eab~_1] li/PREP+Eab~/NOUN (GLOSS): for/to + drinking/quaffing + SOLUTION 4: (laEab~) [Eab~_1] la/EMPHATIC_PARTICLE+Eab~/NOUN (GLOSS): indeed/truly + drinking/quaffing + SOLUTION 5: (laEab~a) [Eab~-u_1] la/RESULT_CLAUSE_PARTICLE+Eab~/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): would have + drink/quaff + he/it <verb>	INPUT STRING: يلعب LOOK-UP WORD: ylEb SOLUTION 1: (yaloEab) [laEib-a_1] ya/IV3MS+loEab/VERB_IMPERFECT (GLOSS): he/it + play +
Passive voice of a weak verb is produced in both perfect and imperfect tenses	INPUT STRING: وقع LOOK-UP WORD: wqE SOLUTION 1: (waqaEa) [waqaE-a_1] waqaE/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + fall down/take place/be located + he/it <verb> SOLUTION 2: (waq~aEa) [waq~aE_1] waq~aE/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + sign + he/it <verb> SOLUTION 3: (wuq~iEa) [wuq~iE_1] wuq~iE/VERB PERFECT+a/PVSUFF SUBJ:3MS (GLOSS): + be signed + he/it <verb> SOLUTION 4: (waqoE) [waqoE_1] waqoE/NOUN (GLOSS): + impression + SOLUTION 5: (waqoE) [waqoE_2] waqoE/NOUN (GLOSS): + falling + SOLUTION 6: (waqoE) [waqoE_3] waqoE/NOUN	INPUT STRING: يوقع LOOK-UP WORD: ywqE SOLUTION 1: (yuwaq~iE) [waq~aE_1] yu/IV3MS+waq~iE/VERB_IMPERFECT (GLOSS): he/it + sign + SOLUTION 2: (yuwaq~aE) [wuq~iE_1] yu/IV3MS+waq~aE/VERB_IMPERFECT (GLOSS): he/it + be signed + SOLUTION 3: (yuwqiE) [>awoqaE_1] yu/IV3MS+wqiE/VERB_IMPERFECT (GLOSS): he/it + inflict/bring about + SOLUTION 4: (yuwqaE)

	(GLOSS): + beat (music) +	[>awoqaE_1] yu/IV3MS+wqaE/VERB_IMPERFECT (GLOSS): he/it + be inflicted/be brought about +
Passive voice of a weak verb is neither produced in perfect nor imperfect tenses	INPUT STRING: **يـئس** LOOK-UP WORD: y}s SOLUTION 1: (ya}isa) [ya}is-ai_1] ya}is/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + despair/be hopeless + he/it <verb>	INPUT STRING: **يـيـئس** LOOK-UP WORD: yy}s SOLUTION 1: (yayo}is) [ya}is-ai_1] ya/IV3MS+yo}is/VERB_IMPERFECT (GLOSS): he/it + despair/be hopeless + SOLUTION 2: (yuyo}is) [>ayo>as_1] yu/IV3MS+yo}is/VERB_IMPERFECT (GLOSS): he/it + make despair +
Imperfect weak verb is not recognized either as passive or active voices	INPUT STRING: **وزن** LOOK-UP WORD: wzn SOLUTION 1: (wazana) [wazan-i_1] wazan/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): + weight/equilibrate + he/it <verb> SOLUTION 2: (wazan~a) [wazan-i_1] wazan/VERB_PERFECT+na/PVSUFF_SUBJ:3FP (GLOSS): + weight/equilibrate + they [fem.pl.] <verb> SOLUTION 3: (wazon) [wazon_1] wazon/NOUN (GLOSS): + weight/weighing + SOLUTION 4: (wazan~a) [zan~-u_1] wa/CONJ+zan~/VERB_PERFECT+a/PVSUFF_SUBJ:3MS (GLOSS): and + drone/buzz + he/it <verb> SOLUTION 5: (wazan~) [zan~_1] wa/CONJ+zan~/NOUN (GLOSS): and + droning/buzzing + SOLUTION 6: (wazin~a) [zAn-i_1] wa/CONJ+zin/VERB_PERFECT+na/PVSUFF_SUBJ:3FP (GLOSS): and + decorate/adorn + they [fem.pl.] <verb>	INPUT STRING: **يـوزن** LOOK-UP WORD: ywzn Comment: ywzn NOT FOUND
Perfect	INPUT STRING: **تـوجر**	INPUT STRING: **يـتـاجر**

weak verb is neither recognized as active nor passive voice	LOOK-UP WORD: twjr Comment: twjr NOT FOUND	LOOK-UP WORD: ytAjr SOLUTION 1: (yutAjir) [tAjar_1] yu/IV3MS+tAjir/VERB_IMPERFECT (GLOSS): he/it + deal with/do business +
Passive voice of a weak verb is produced in imperfect tense but not included in the perfect tense	INPUT STRING: عصي LOOK-UP WORD: ESy SOLUTION 1: (EaSiy~) [EaSiy~_1] EaSiy~/ADJ (GLOSS): + rebel/mutineer + SOLUTION 2: (EaS~iy) [EaS~_1] EaS~/NOUN+iy/POSS_PRON_1S (GLOSS): + hardening + my	INPUT STRING: يعصى LOOK-UP WORD: yESY SOLUTION 1: (yuEoSaY) [EaSaY-i_1] yu/IV3MS+EoSaY/VERB_IMPERFECT (GLOSS): he/it + be resisted/be refused +