DOI: 10.14746/linpo.2025.67.2.2

Morphological and semantic opacity as factors of linguistic change: A study of Form VIII verbs in Arabic*

Ahmed Ech-Charfi

Mohammed V University in Rabat
a.echcharfi@um5r.ac.ma | ORCID: 0000-0002-4946-1380

Abstract: This paper deals with the role of morphology in the reconstruction of lexical meaning. It focuses on the case of the Arabic Form VIII verbs in order to illustrate the challenge that morpho-semantics presents for historical lexicographers assuming the unity of a language throughout a long period of its use. In this connection, the paper attempts to show that, although Form VIII verbs have been in use since the early stages of Arabic, it is likely that users assigned them different meanings according to whether Form VIII morpho-semantics was transparent or opaque. Three factors have been identified that increase the opacity of this category: allomorphy, polysemy and frequency of the derivation base. 529 items were culled from a bilingual dictionary for the purposes of the study, and allomorphy was found to contribute about 12% to morphological opacity, and more than 70% of the verbs had a non-prototypical sense. Many of the extended senses seem to have lost all kinds of semantic relation to the prototypical sense, thus resulting in less transparency in the semantics of the derived forms. The study also argues that the less frequent the base of the derivation is, the more opaque Form VIII will be. The paper concludes that, given the lack of rich data from the early stages of Arabic, it is likely that a satisfactory reconstruction of the meaning of derived forms will probably never be achieved.

Keywords: morphology, semantic change, word meaning, Arabic, Form VIII verbs

1. Introduction

Reconstructing lexical meanings of morphologically complex words can be particularly challenging and rather imprecise. The challenge, as will be argued in this paper, is posed mainly by the intricate polysemy networks that derivational morphemes develop over time as well as the phonological changes these morphemes undergo. When these two factors are combined, the relation between meaning and form can become extremely opaque so much

* Some changes have been made to an earlier version of this paper, including the addition of some references, as a response to remarks made by an anonymous reviewer, which are thankfully acknowledged.

so that native speakers fail to recover it from use, a fact that calls for reanalysis. In the case of languages like written Arabic, which has been in use for around a millennium and a half, it is a risky mission to determine whether and when such a reanalysis happened and what its consequences on the grammar of the language were. The study will focus on Arabic verbal morphology, as represented by Form VIII verbs.

Arabic verbal morphology is very elaborate. It relies on affixes (essentially prefixes and infixes) to derive new verb forms with various meanings. For example, the augmented form *ʔafʕal* is derived from the basic form *faʕal* to convert an intransitive verb into a transitive one, e.g. *xarağ* 'to get out' vs. *ʔaxrağ* 'to put s.o/s.th out'. Similarly, *ftaʕal* is augmented with a *-t-* infix that expresses the notion of reflexivity and related meanings, e.g. *bāʕ* 'to sell' vs. *btāʕ* 'to buy'. As is often noted, derivational morphemes tend to be restricted in productivity but highly polysemous (Bauer 2001, Kotowski & Plag 2023). The infix *-t-*, for example, is limited to a small set of verbs, e.g. **ʕtamal,* from *ʕamal* 'to do', is not attested, though conceptually possible with the meaning 'to do s.th for oneself'. In turn, the infix can express a variety of meanings, besides that of reflexivity, as will be explained later. In diachronic studies on Arabic, the semantic content of morphological forms tends to be overlooked, probably because of their polysemous nature that often escapes satisfactory characterization. This problem becomes even more acute when there is no community of native speakers to whom researchers can turn for nuanced judgments about meaning and use. For example, the online Doha Historical Dictionary of the Arabic Language (*Muʕğam*) provides the following allegedly original meaning of the verb *ḥtaram:* "*waqqara-hu wa rāʕā mā yağibu min ḥaqqi-hi*" ('to respect s.o and to acknowledge his due'). This definition, however, seems to reflect modern use rather than the old use that goes as far back as pre-Islamic times, as this study will show.

The most problematic aspect of this reconstructed meaning is the total disregard of the morphological structure and its contribution to the sense of the verb. As it stands, *ḥtaram* is an augmented Form VIII verb that is derived by the insertion of the affix *-t-* after the first root consonant of its Form I equivalent. The basic form from which it is derived should be *ḥaram* (to deny s.o s.th) or *ḥarum* (to be forbidden), assuming that the *t* after the first consonant is an infix. Given the reflexive meaning of the affix *-t-* to be discussed later, *ḥtaram* should mean something like 'to deny oneself s.th' . The root 'Ḥ-R-M' also refers to the sanctity of the thing forbidden, probably because of its association with the Kaaba called *al-bayt al-ḥarām* (literally, the forbidden house) and, later on, with the Islamic notion of *ḥarām* (sin, wrongdoing, etc.). This shade of meaning can be detected in other related words like *ḥaram* (inviolable, anything that must be defended with arms) and *ḥarīm* (one's wives and children), probably because the family was regarded as the most sacred thing in the Arab society of the time. When this nuance is taken into account, *ḥtaram* should mean something like: 'to withhold the desire to cause harm because of deep religious or similar feelings'. This nuance has completely disappeared in modern use, however, and it is not clear when exactly that happened. In fact, even its earliest attested uses cannot be claimed with certainty to have had that meaning.

Meaning change does not involve lexis only, but it can also affect sub-lexical components, and for that matter, morphology must also be taken into account when studying language change. It has been argued by many linguists, particularly those who work within the

cognitive linguistic framework, that morphemes as well as the schematic form of derived words usually express meanings, though these meanings are characteristically more abstract than lexical meanings (cf. Bybee 1985, Langacker 1987, 1991, Lakoff 1987, Talmy 2000, among many others). Like lexical meaning, abstract grammatical meaning is also subject to change through extension, shift, re-analysis, or even loss (cf. Fortson 2003, Traugott 2000). The morphological system of the Arabic verb, for instance, can express a variety of grammatical meanings that often determine their argument structure and, thus, have an interface with syntax. Consequently, any change that affects this kind of morphological system can also affect lexical meaning as well as syntactic structure. In this paper, I will illustrate this situation by focusing on the change in the meaning of Form VIII verbs as Classical Arabic (CA) developed into Modern Standard Arabic (MSA).

The paper is constituted of two main sections in addition to this introduction and a conclusion. Section 2 will present the verbal system in different Arabic varieties with the view of setting the scene for later discussions. In this section, we will explain that the focus of this study is not on the development of verbal morphology from CA to the modern colloquial varieties, but rather on the changes that have affected this morphological system during different stages of the standard variety itself, i.e. CA and MSA. Section 3, which constitutes the main contribution of this study, will provide a detailed discussion of the morphological change in Form VIII, which resulted from phonological and/or semantic factors.

2. The verbal system of CA, MSA and the colloquial varieties

Although historical linguists are interested in the study of the mechanisms of language change, they have not always considered seriously the sociolinguistic status of language varieties in their historical development (cf. Romaine 1982). Any diachronic study of the Arabic verbal system, however, cannot afford overlooking the different paths in the development of the Arabic language. This is particularly so because this language is well known for being diglossic and it is not clear when this situation exactly began; some scholars argue that it developed during the early periods of the Muslim empire (e.g. Versteegh 1984: Chap. II) while others claim that it was already characteristic of the pre-Islamic period (e.g. Zwettler 1978: 101). Because of Arabic diglossia, the verbal system of CA followed two paths: one path into the modern dialects, which differ from one variety to another across and beyond the Arab World, and a second path into MSA, a primarily written variety that is spoken natively by no one. Most linguists would show more interest in the first path probably because it is more "natural", but the Doha Historical Dictionary of the Arabic Language is rather concerned with the second. Like other lexicographers, the compilers of this historical dictionary target users of the standard variety in which most learned culture is written. Given that the Arabic learned culture spans over more than a millennium in what is considered to be one and the same language standard, namely *al-Fuṣḥā* (i.e. the pure language), it is no surprise that Arab lexicographers show interest in change in the standard rather than the colloquial.

Arabic, like other Semitic languages, has a basically non-concatenative morphology that weaves roots and patterns into pronounceable words. The tri-consonantal root *K-T-B*, for example, is not a word in itself although it denotes the semantic field of writing; it can be realized as a word only when combined with a (usually) vocalic template, as in *katab* 'to write', *kutib* 'to be written', *kātib* 'writer', *kitāb* 'book', etc. The affixes are quite regular in both form and meaning when combining with other roots to the extent that traditional Arab grammarians refer to them by using the root *F-ʕ-L* 'to do' because of its schematic meaning. Thus, the pattern *faʕal* refers to the perfective form of any tri-consonantal verb, *fuʕil* to the passive of the perfective form of such verbs, etc. In comparison, Western scholars use capital Cs instead to refer to root consonants for practically the same purpose (e.g. McCarthy 1979).

In CA, as in MSA, there are a number of verbal patterns with different schematic meanings. These are called verb forms in the Western tradition of Arabic studies. The most basic one is Form I *CaCaC* for tri-consonantal roots and *CaCCaC* for quadri-literals, both are generally assumed to be non-derived by traditional Arab grammarians, though there are cases that are possibly derived from nouns (e.g. *talfan* 'to give a phone call' from the loan word *tilifūn* 'telephone'). All the other verb forms are augmented by the addition of an affix, a fact which indicates that they are derived. Traditional Arab grammarians generally assume that augmented forms are derived either directly from Form I or indirectly via other augmented forms. For example, Form II CaC_2C_2aC is derived from the basic form by geminating the second root consonant while Form V $taCaC_2C_2aC$ is derived from Form II by prefixing *ta-* to it. If this prefix is attached to the basic form, it would be an inflectional rather than a derivational affix (viz. *taffʕal* 'you do/she does').

The number of augmented forms in Arabic varieties can vary significantly. In CA, for example, 14 augmented forms are usually identified (cf. Wright 1896: I, 29-46) whereas MSA uses only 10 (cf. Ryding 2005: 434). This is clear indication that at least four forms have fallen into disuse as CA developed into MSA. The other forms may not also have the same frequency in the two varieties, nor even the same use for that matter; hence, the significance of scrutinizing their diachronic development. As to the colloquial varieties, the number of verb forms can be extremely reduced. In Moroccan Arabic, for instance, there are only two derived forms: Form II (corresponding to CA Form II) and Form III (corresponding to CA Form V), while the others have completely disappeared because of sound change, or are represented only by a small group of verbs and, consequently, are unlikely to be open to new innovative forms (cf. Harrell 1962: 29-34). In comparison, Gulf Arabic seems to have preserved most of the augmented forms of MSA except form IV, which was lost as a result of the loss of the glottal stop from its phonological system (cf. Qafisheh 1977: Chap. 6). Like other Bedouin dialects, Gulf Arabic seems to be more conservative than those descending from dialects of settled communities (cf. Versteegh 2014: Section 10.3). It is likely, however, that not all the verb patterns are as productive in such Bedouin varieties as they are in MSA or CA.

Regarding the semantics and syntax of verb patterns, it is worth noting that each form is usually associated with a set of senses that determine the argument structure of the clause. Thus, while the basic form verbs can be intransitive, monotransitive or ditransitive, augmented verbs are often associated with a specific clause type given that they are syntactic

and semantic operations on the base form. Form IV, for example, is morphologically derived from Form I by the prefixation of *ʔa-*, and the result of such a derivation is usually the causative form of the Form I verb, as a comparison of *māt* 'to die' and *ʔamāt* 'to kill' shows. The basic form *māt* is intransitive but its augmented equivalent *ʔamāt* is transitive; but when the basic form is transitive, as is the case with *ʕalim* 'to know', the causative is ditransitive; viz. *ʔaʕlam* 'to inform' or *ʕallam* 'to teach'.

Similarly, Form VIII, which is the focus of this study, has its specific syntactic and semantic properties. Form VIII verbs are generally derived from the basic form by the infixation of *-t-* immediately after the first root consonant; e.g. *bāʕ* 'to sell' vs. *btāʕ*[1] 'to buy'. According to traditional Arab grammarians, Form VIII verbs can have up to six different senses (cf. Ibn ʕUṣfūr 1987, II 192-194, Ibn Yaʕīš 2001, IV: 441)[2]. The first meaning can be called "resultative", as exemplified by *rtafaʕ* 'to rise, to climb, to soar' from *rafaʕ* 'to raise, to lift'; it is as if the situation expressed by the derived form of the verb results from that expressed by its basic form. The second meaning that Form VIII verbs can express is the use of some object in some way understood through encyclopedic knowledge, as in *ḥtabas* 'to imprison', that is to take someone as a prisoner. Some Form VIII verbs can also have a reciprocal meaning illustrated by *qtatal* 'to kill each other' derived from *qatal* 'to kill'. Others can refer to the effort made by the subject during the action denoted by the basic form of the verb. For example, the difference between *kasab* 'to earn' and *ktasab* 'to earn' is that the second stresses the role of the agent in the action and, thus, the difference is not truth-functional. As to the remaining two senses, it seems that the tradition cannot distinguish them clearly from the senses of other forms, as is the case of *btasam* and *tabassam* 'to smile', and *xaṭaf* and *xtaṭaf* 'to snatch'. In the first pair, no semantic distinction is detected between the Form VIII and Form V verbs, and the second pair also indicates that the Form I and Form VIII verbs are apparently synonymous. Generally, however, no attempt is made within this tradition to identify the relation between the different senses of a verb form or to explain why they should be expressed by one and the same verb pattern.

By contrast, polysemy in morphology is one of the issues that modern researchers are fascinated with. For these researchers, affixes, just like independent words, usually express different, but related, senses (cf. Copestake & Briscoe 1995, Lehrer 2003, Lieber 2004, Rainer 2014, Rainer et al. 2014, Schulte 2015, among others). Within the cognitive linguistic framework, polysemy is usually explained by the extension of a prototypical sense through metaphor, metonymy or some similar cognitive process. Since verbal affixes express grammatical meanings, they generally operate on the syntactic structure of the clause and, thus, affect the argument structure of the verb. Their syntactic role, however, is rarely uniform precisely because of polysemy, as the discussion of the infix *-t-* of Form VIII in Arabic will illustrate.

[1] This form is pronounced as [ibtāʕ]; the initial i vowel is epenthetic to avoid complex syllable onsets, and will be ignored hereafter.

[2] I would like to thank an anonymous reviewer who drew my attention to al-Tūnī (1997), a study dedicated to Form VIII in the Quran. Though more detailed than what is generally found in Arabic writings on the topic, the study, however, remains faithful to the spirit of the tradition by identifying the same senses mentioned in reference grammars.

The role of the *-t-* infix in Form VIII is primarily to demote the subject and promote the object. A comparison of the examples under (1) illustrates this idea:

(1) a. *ʕazala al-raʔīs-u al-wazīr-a*
'The president dismissed the minister.'
b. *ʕuzila al-wazīr-u*
'The minister was dismissed.'
c. *ʕazala al-wazīr-u nafs-a-hu*
'The minister dismissed himself.'
d. *iʕtazala al-wazīr-u*
'The minister resigned.'

The verb *ʕazal* 'to dismiss' is transitive because it describes an action with an agent and a patient. In (1a), the agent is referred to by the subject NP (marked for nominative case) while the object NP (marked for the accusative) refers to the patient. In (1b), however, the verb is in the passive form, as indicated by the *-u-i-* vocalic melody, and passivization involves the promotion of the NP referring to the patient to the subject position to indicate that the entity is focused. Nonetheless, the agent, though not mentioned, remains in the background as an entity that is distinct from the patient. In comparison, the reflexive pronoun *nafs-a-hu* 'himself' in (1c) indicates that the agent and the patient roles are performed by one and the same entity. Despite that, the basic form of the verb *ʕazal*, by virtue of its semantic content, describes a two-participant event and, consequently, (1c) frames the agent and the patient as conceptually distinct entities. In comparison, the Form VIII *ʕtazal* in (1d) frames the event as a one-participant event, and the action is conceptualized as being performed on the self. In this sense, Form VIII is similar in many respects to the middle voice (cf. Kemmer 1993, Klaiman 1991, Zúñiga & Kittilä 2019).

Like most morphological patterns, Form VIII has uses other than the one discussed in the previous paragraph. Glanville (2018: Chap. 4), one of the rare works on Arabic verbal patterns[3], cites three major uses of this form; these are: subject as beneficiary, actions on the self, and symmetry; which of these is intended will usually be determined by the semantic content of the verb. They are illustrated by the following respective examples:

(2) a. *iqtaṭaʕa ʔarḍ-an*
'He cut out a piece of land for himself.'
b. *irtamā fī al-māʔ*
'He threw himself into the water.'
c. *iǧtamaʕa maʕa ʔaṣḥāb-i-hi*
'He met with his companions.'

In (2a), the verb is derived from the basic *qaṭaʕ* 'to cut', but the result is not an intransitive verb, as is the case in the examples cited in the preceding paragraph. Rather, the derived verb is still transitive and the affix *-t-* adds the meaning that the outcome of cutting will benefit to

[3] An anonymous reviewer thankfully pointed out that Fleisch (1978) and Larcher (2003) also discuss the semantics of Form VIII verbs.

the subject. In comparison, (2b) is a clear case of the middle use already discussed. As to (2c), the verb is derived from *ǧamaʕ* 'to gather' and the resultant meaning is that the subject and the object referents came together. Syntactically, the derived verb is intransitive but the object must be mentioned in the oblique. Glanville (2018: 64-66) also points out that Form VIII verbs can be derived from nouns, given that the form has become established in the grammar as a schema with a set of meanings, though these meanings can be various but related. For example, *ʕtanaq* 'to embrace' comes most probably from *ʕunuq* 'neck' since embracing involves taking and being taken by the neck. The verb is also used to denote adoption of ideas, religions, or theories; but this meaning is apparently an extension from 'to embrace' for reasons that seem obvious. This example is reminiscent of verbs like *ǧtamaʕ* 'to meet' in that the action involves the subject and one or more participants. More will be said later about the various uses of Form VIII and how this polysemy has contributed to semantic change.

In the remainder of this paper, the claim that the meaning and use of Form VIII has undergone some change will be discussed and illustrated on the basis of a list of verbs culled from Baalbaki's (1995) Arabic-English dictionary. This large dictionary was chosen essentially because it is organized on the basis of words rather than roots, as is the usual practice in Arabic lexicography. Since the words follow the alphabetical order, it was much easier to search for verbs with the form *ftaʕal* than would have been the case if a root-based dictionary were used instead. The final list included 529 Form VIII verbs, which were put in a spreadsheet in order to facilitate their arrangement and re-arrangement according to different criteria.

3. Derivational opacity, reanalysis and meaning change

As was mentioned at the outset, this study aims at showing that Arabic words change their meaning in the course of time partly because of opacity in derivational morphology. The relation between morphology and semantics is not a straightforward one and varies according to the theory adopted (see Anderson 2015 and Chapters 19-23 in Part V of Himmisley & Stump 2016). In this paper, however, we will focus more on data than on their theoretical implications and, therefore, we will do our best to couch the description and the argumentation in theory-neutral terms. The issue of opacity/transparency will occupy a central position in the discussion because of its role in semantic change. This section will be divided into three subsections: one on allomorphy, the second on polysemy, and the third on the frequency of the basic form.

3.1. Allomorphy

Allomorphy constitutes a type of morphological change, albeit a small one with non-dramatic consequences. The English simple past marker *-ed*, for example, is usually realized as [d], but can also be realized as [t] in cases like *looked* or as [ɪd] in cases like *wanted*, depending on the preceding consonant. In such a situation, the learner must first make a con-

nection between the three realizations and infer, based on linguistic and contextual clues, that they are allomorphs of the same morpheme, namely the suffix marker of the past tense. However, when the connection between the different allomorphs is no more transparent, this can become a major change leading to the decay or reanalysis of the morpheme. It can result also in the change of the whole paradigm of which that morpheme is an element.

This seems to be the case with the *-t-* of Form VIII. In cases like *manaʕ* 'to prevent' vs. *mtanaʕ* 'to refrain, to abstain', the learner can easily make the connection between the base and the derived forms essentially because the infix surfaces under a form that is identical with the underlying one. When the first root consonant is emphatic, however, the *-t-* gains emphasis by assimilation, as in *ḍarab* 'to hit' vs. *ḍṭarab* 'to be confused'. This kind of allomorphy can be confusing especially that this assimilation excludes the phoneme /r/, which is also emphatic in Arabic, except when followed by the vowel /i/, viz. *ramā* 'to throw' vs. *rtamā* 'to throw oneself'. Similarly, the *-t-* can be realized as [d] when adjacent to /z/, /d/ or /ð/ only, as in *zād* 'to add' vs. *zdād* 'to add to oneself/to be born'. This is apparently a case of voice assimilation although it is restricted to the context of non-emphatic voiced dental sounds. In both emphasis and voice assimilation, the connection between the base and the derived forms can become opaque, thus constituting a challenge for the learner. In yet a third situation, allomorphy can affect not only the affix, but also the stem, making the derivation even more opaque. *ttaxað* 'to assume', for instance, is derived from *ʔaxað* "to take", but the glottal stop assimilates completely to the affix *-t-*. Under this category, we can also include cases of verbs with an initial *t* as a root consonant such as *ittabaʕ* 'to follow', especially that geminates are represented graphically in Arabic by a single letter; viz. اتبع. Similarly, *ddakar* 'to recall' derives from *ðakar* 'to mention', but the affix *-t-* assimilates first to the interdental fricative and the geminate [ðð] is strengthened after that to yield [dd]. (Actually, both *ððakar* and *ddakar* are attested in the language as free or dialectal variants.) Obviously, these morpho-phonological changes obscure the derivational relation between the base and the derived form and, thus, make the learning process much more challenging than would be the case with less opaque derivations. This is true for all learners, but it is more so for L2 learners such as learners of MSA.

The effect of allomorphy on the derivational system will partly depend on its frequency. If only a small set of verbs exhibit differences between the base and the derived forms, allomorphy will probably not have any significant consequences on the morphological category despite its high level of opacity. But when a large number of verbs take a form of the morpheme that is not identical with its underlying representation, the weaker the connection between the different allomorphs is, the more likely the morphological category will be lost. In the case of Form VIII verbs in Arabic, allomorphy seems to have contributed to the opacity of the derivation.

There are 64 cases in which the *-t-* affix occurs under a different form in the list of Form VIII verbs compiled for this study. This figure represents 12.10% of the list; they are divided into five categories exhibited in Table 1.

Table 1. Frequency according to allomorphy

Allomorph/ Grapheme	Frequency	Examples
Geminate /t/	20	*ttabaʕ* 'to follow' *ttaxað* 'to assume' *ttaṣal* 'to be connected to'
Geminate /d/	6	*ddakar* 'to recollect' *ddaxar* 'to save'
Single /d/	11	*zdahar* 'to prosper'
Single /ṭ/	24	*ṣṭafa* 'to choose'
Geminate /ṭ/	3	*ṭṭalaʕ* 'to examine'

The first line shows cases in which the affix *-t-* occurs as part of a geminate, either because the first root consonant is /t/ or because it is a glottal stop or /w/ that assimilates regressively to the affix. In either case, and because the geminate is represented graphically by a single letter, the affix is likely to be opaque to the learner. The same remark holds for the second and the last lines in which the affix surfaces as part of a 'd' or a 'ṭ' geminate, respectively. In the remaining two lines, the affix is represented graphically by a separate letter, but because of progressive assimilation, the letter is different from that found in regular cases, namely ـدـ in the third line and ـط in the fourth. Although 12% does not seem to be a very large percentage, it is large enough to introduce opacity in the derivation and, by consequence, cause confusion to the learner who is initiated to the language typically through the writing mode.

But although allomorphy has a share of responsibility in reducing the transparency of Form VIII derivation, polysemy certainly plays a more crucial role in the semantic opacity. Obviously, when the two factors are combined, the consequences can be drastic, as the discussion below will show.

3.2. Polysemy

Traditionally, polysemy was assumed to be a characteristic of lexical items, and affixes were treated as part of polysemous items. For this reason, the phenomenon of polysemy, and semantic change in general, was rarely discussed in relation to morphology. More recently, however, many researchers have turned to the semantic contribution that affixes bring to the meaning of words (cf. Rainer 2014). In some theoretical frameworks such as Construction Grammar (cf. Booij 2013), for example, an affix is represented as forming a schema together with the grammatical category to which it is attached. The schema has a semantic content, just like any other lexical item in the language. For instance, $[V \_ er]_N$ stands for the combination of a verb and the affix *-er* to form what is called an "agent" noun in English. This schema can denote the agent of some activity, e.g. *writer*, but it could also denote an instrument, e.g. *blender*, or even a theme of an activity, e.g. *bestseller*, etc. This is clear indication that the $[V\_er]_N$ is polysemous.

Similarly, Form VIII is polysemous as a schema. Previously, we pointed out that the function of the *-t-* infix is to turn a transitive verb into an intransitive one by demoting the subject and promoting the object; compare *manaʕ* 'to prevent' and *mtanaʕ* 'to refrain, to abstain' discussed earlier. When a verb is ditransitive, insertion of the infix turns the verb into a monotransitive one, as in the following pair of examples:

(3) a. *bāʕ-a al-tāğir-u al-rağul-a biḍāʕat-an*
'The trader sold the man a merchandise.'
b. *ibtāʕ-a al-rağul-u biḍāʕat-an*
'The man bought a merchandise.'

In (3a), the verb takes a direct and an indirect object, just like its English equivalent. Semantically, the subject is an Agent, the direct object a Theme and the indirect object a Beneficiary. In comparison, the Beneficiary in (3b) is promoted to the subject position while the Agent is backgrounded. While this de-focusing operation is performed lexically in English through the selection of a different verb, it is performed in Arabic morphologically through infixation. In both monotransitive verbs like *manaʕ* and ditransitive verbs like *bāʕ*, the corresponding Form VIII focuses attention on the Patient or the Beneficiary as the initiator of the action. In this sense, the two could be said to express the same meaning, although there is a slight difference between them.

In other cases, however, infixation of *-t-* does not result in any syntactic operation. Such is the case of *btadaʔ* 'to begin' illustrated in these examples:

(4) a. *badaʔ-a al-ʕāmil-u šuγl-a-hu*
'The worker began his work.'
b. *ibtadaʔ-a al-ʕāmil-u šuγl-a-hu*
'The worker began his work.'

In both examples, the verb has the same argument structure, thus indicating that there has been no change in syntactic structure. Semantically, the infix does not seem to bring much to the meaning of the sentence since the activity of beginning work must involve some effort on the part of the agent anyway. Baalbaki (1995) does not provide any equivalent for *btadaʔ* but merely refers the user to the *badaʔ* entry, implying that the two are equivalent. This use of Form VIII is a clear deviation from the one illustrated in the preceding paragraph and, therefore, it must have been a semantic change introduced sometime in the history of the Arabic language.

We have been able to identify eight different uses of Form VIII in the corpus collected, though these should not be treated as tight categories. In addition to the two cases just discussed, which can be considered as the two ends of a continuum, there is a third class of verbs whose Form I and Form VIII are clearly related, though not synonymous. An example that illustrates this class is the pair *raʔā* 'to see, to think' and *rtaʔā* 'to consider'. The fourth category of Form VIII verbs includes verbs for which no corresponding basic form can be identified. As a case in point, *btahal* 'to supplicate' does not seem to be derived from any Form I verb since no such form as *bahal* can be found in the bilingual dictionary from which

the list was culled nor in any other monolingual dictionary of CA. A similar class includes verbs for which there is a corresponding basic verb but with an apparently unrelated meaning. For instance, while both *rāḥ* and *rtāḥ* are attested, the first means 'to leave' and the second 'to rest'. Apparently, the Form VIII verb is derived from the noun *rāḥa* 'rest' rather than from any basic verb form. A sixth class includes Form VIII verbs cited in the bilingual dictionary but for which no entry is cited in monolingual dictionaries of CA or MSA. For instance, Baalbaki (1995) cites *ḥtaðar* with a note of reference to the adjective *ḥaðir* 'cautious', implying that the verb probably means 'to be cautious'. This verb, however, is not cited by dictionaries of CA. The seventh class includes Form VIII verbs that express reciprocity and are, therefore, synonymous with the corresponding Form VI verbs. An example of such verbs is *xtaṣam*; Baalbaki (1995) does not provide any explanation for this verb but merely refers to Form VI *taxāṣam* 'to dispute', implying that the two are synonymous. The last class includes verbs expressing the intensification of an activity. For instance, while the basic form *ḥafā* means 'to welcome', *ḥtafā* means 'to welcome heartily'. As illustrated by the examples, these classes provide clear indication that Form VIII is polysemous.

The eight classes do not all have the same type frequency. By type frequency, we mean the number of verbs in each class, not the frequency of a verb within a given corpus of texts, which is usually called *token* frequency. Table 2 exhibits the frequency of each class.

Table 2. Type frequency of verb classes

Verb class	Example	Frequency	Percentage
Class 1	*mtanaʕ* 'to refrain'	141	26.65
Class 2	*btadaʔ* 'begin'	195	36.86
Class 3	*rtaʔā* 'to suggest'	58	10.96
Class 4	*btahal* 'to supplicate'	51	9.64
Class 5	*rtāḥ* 'to rest'	45	8.50
Class 6	*ḥtaðar* 'to be cautious'	19	3.59
Class 7	*xtaṣam* 'to quarrel'	16	3.02
Class 8	*ḥtafā* 'to welcome heartily'	4	0.75

As can be noticed, the second class ranks first with 195 verbs, covering almost 37% of the list, followed by the first class with 141 verbs. The smallest class includes only 4 verbs while the remaining classes range between 16 and 58. But despite the varying frequencies, the fact that some incompatible classes have more or less the same frequency indicates that the semantics of Form VIII has become very loose. A clear illustration of this point can be provided by a comparison of the first and the second classes. As was explained earlier, while *-t-* affixation in cases like *mtanaʕ* 'to refrain' affects the syntax-semantics of the verb, it does not change much in cases like *btadaʔ* 'to begin'. The question that should be raised in this connection is: how can a language learner connect between the different uses of the infix *-t-*?

In treatments of polysemy, most researchers argue that there is usually a core or "prototypical" sense from which the other senses are derived by metaphor or metonymy or some

other cognitive process (cf. Brugman 1988, Lakoff 1987, Tsohatzidis 1990, Evans 2009, among many others). Regarding the case under study, it seems that Class 1 exemplified by *mtanaʕ* 'to refrain' is the prototype of Form VIII in Arabic. Although this is no place to develop the argument, the fact that many languages have a middle voice through which the subject is defocused and the object focused can be an indication that this is the initial function of this morphological derivational (cf. Kemmer 1990). In other words, all the cases in which no syntactic transformation results from the affixation of *-t-* must have been developed from the prototypical use by extension. From the reflexive use of *-t-* in *mtanaʕ*, for example, the learner may focus on the effort made by the experiencer to refrain from doing something instead of the number of participants in an activity. This attention is transferred later to a verb like *ḥtafā* from Class 8 to intensify the warmth of welcome without any effect on the argument structure of the verb. As long as native speakers are able to infer the right semantics of the non-prototypical uses and their connection to the prototype of an affix, the morphological operation can be claimed to be productive. But when such a connection starts to wane, there is some probability that some change has occurred in morphological derivation as well as in the meaning of the derived forms.

The frequencies in Table 2 above can be interpreted as an indication in this direction. Although Class 1 seems to represent the prototypical use of Class VIII, as argued above, the number of verbs constituting this class is lower than the number of verbs in the second class, which deviates from the prototype. This class constitutes almost 37% of the corpus. Of course, there are several factors that determine the prototypicality of a sense, but frequency is usually regarded as being one of them (cf. Fenk-Oczlon & Fenk 2010). Therefore, the fact that the second class of verbs outnumbers the first could be considered as an indication that Form VIII has changed its prototypical sense or, perhaps more accurately, that it no longer forms a homogenous category from a semantic perspective. This should come as no surprise given that Standard Arabic has been in use as a High diglossic variety for more than a millennium, a time span long enough to account for the reported change. Although there are no studies to my knowledge on the acquisition of the Arabic verbal system by Arab learners (see Zalami 2007 for a review), it seems that even at an advanced level of proficiency, learners are unable to link the various uses of Form VIII, for example, in such a way that these uses form a web of interrelated meanings. They are simply too heterogeneous, both syntactically and semantically, to be included in a single category.

The issue to be considered now is the following: what meaning do language users assign to a Form VIII verb if they are unable to relate it to a basic form through a derivational process? For example, given that there is no attested basic form *bahal*, how do they interpret and store the derived form *btahal* 'to supplicate'? Similarly, how can they link the meaning of *ḍṭarab* 'to be confused' to that of *ḍarab* 'to hit' despite the apparent lack of relation between the two meanings? An attempt to answer these questions will be made after the frequency of the basic form is discussed in the following subsection.

3.3. Frequency of the base form

For a derivation to be productive, both the base and the derived forms must be accessible to the learner and the relation between the two must be transparent to a significant degree. For example, both *manaʕ* 'to prevent' and *mtanaʕ* 'to refrain' are frequent enough in MSA to enable the language learner to establish a certain connection between the two. In comparison, although both *ḍarab* 'to hit' and *ḍṭarab* 'to be confused' are frequent in MSA[4], the fact that the affix has changed into *ṭ* and the meanings of the two verbs are not clearly related has made the two forms unrelated for many speakers. For these reasons, the derivation in the first case will be qualified as transparent while in the second case it is opaque. The more factors there are that intervene to remove the derived form from its base, the more opaque the derivation will be. For this reason, transparency and opacity tend to form a continuum rather than discrete categories.

The intervening factors considered so far are allomorphy and polysemy, and a third factor is the relative frequency of the base. Obviously, if a language user is unable to connect a derived form with its base, it is not clear how that form can be considered derived for him/her. Consider again the example of *btahal* 'to supplicate' from Class 4, for which no corresponding *bahal*[5] can be identified, nor any other base whatsoever. For speakers of MSA at least, there seems to be no sense in claiming that this verb is constituted of a base and an affix, and that its meaning is the result of combining the meanings of its constituents. Therefore, it is very likely that cases like this are learned as non-derived verbs and that the *t* is reanalyzed as a root consonant rather than an affix. Class 5 may also be behaving in the same way. As explained above, this class includes verbs whose Form I and Form VIII do not share the same meaning, at least in MSA usage, though they share the same root consonants. For instance, *btaɣā* is derived from *baɣā*, both of which mean 'to seek' in CA; but in MSA, *baɣā* is usually used with the meaning of 'to assault'[6]. Therefore, speakers of MSA are unlikely to consider the second as the base of derivation for the first and, consequently, the *t* of *btaɣā* is perhaps not treated as an affix. Given that the two classes of verbs include 96 cases and form more than 18% of the corpus (see Table 2 above), we can easily imagine the impact such cases must have had on the internal consistency of Form VIII as a morphological category.

Frequency does not involve only these two classes but cuts through all the others as well. Even Class 1 includes cases in which the corresponding Form I verb is of low frequency in MSA and may not be familiar to a large number of speakers. For instance, *ntaʕaš* 'to become refreshed' should correspond to *naʕaš*, a form that does exist in CA with the meaning of 'to

[4] *Ḍarab* occurs around 32 thousand times in ArabiCorpus while *ḍṭarab* occurs about one thousand times.

[5] While *btahal* occurs 145 times in ArabiCorpus, a search for *bahal* does return some 9 instances. A close scrutiny, however, indicates that these are colloquial forms from Levantine Arabic that are combinations of the preposition *b-* 'with' and the demonstrative *hal* 'this'/'these'.

[6] This is particularly the case in parts of the Arab World where *bɣa* is used in the colloquial variety with the meaning of 'to want' such as North Africa. Apparently, MSA users avoid such forms in order not to be suspected of wrong use of the language. Where this is not the case, however, as in the Middle East, *baɣā* is used with same meaning, as an inspection of the 907 instances in ArabiCorpus shows.

raise' but which is almost inexistent in MSA; the ArabiCorpus, for example, does not include any instance of this form. Baalbaki (1995) does cite *naʕaš* but merely refers to Form IV *ʔanʕaš* for explanation, implying that the two have more or less the same meaning. For MSA speakers, *ntaʕaš* is more linked to *ʔanʕaš* than to any other form, but it is not clear how one could be derived from the other. In other cases, the base form may be familiar to MSA speakers but with a different meaning. As a case in point, both *nahā* 'to prohibit' and *ntahā* 'to finish' are quite frequent in modern usage, but the first is linked to the action noun *nahy* 'prohibition' and the second to *nihāya* 'end, termination' or *ntihāʔ* 'completion'. In CA, however, all these seem to be connected; specifically, *ntahā* used to denote restraining oneself from doing or enjoying something, and the meaning of putting an end to an activity is a mere extension of this sense by implication. In comparison, the two senses are separate in MSA. Therefore, we must conclude that the organization of the lexicon of MSA must be different from that of CA, a conclusion that is not surprising given that the second was spoken by native speakers while the first is used by second language learners only (cf. Khamis-Dakwar & Froud 2019, Froud & Khamis-Dakwar 2021). Some words are often represented in the mental lexicon of second language learners as simplex even when they are complex (cf. Milton 2009: 103-105). Although it was not possible to measure the frequency of the base form of all the verbs in the corpus due to ambiguities in the classification, a significant number of cases seem to lack a transparent connection between the base and the derived forms.

Obviously, the more opaque the relation is between the base and the derived forms, the more likely the derivational process will be obscured and, ultimately, lost. Semantically, once the two forms are no more morphologically linked, each will develop its own meaning separately from the other. To take the examples of *btahal* 'to supplicate' and *ntahā* 'to finish' once again, each of them has developed a separate meaning. In particular, the first stands now on its own since its base was lost in the course of language development and, consequently, can be argued to be synchronically underived. As to the second, it stands somewhere between transparent verbs and completely opaque verbs. On the one hand, it seems to be unrelated to *nahā* 'to forbid' for many speakers of MSA, but on the other, it is still connected to the noun *nihāya* 'end', which does not include the affix *-t-*. The fact that *ntahā* and *nihāya* are semantically related is likely to encourage the analysis of the first as a Form VIII verb, possibly derived from the noun form instead of the putative base *nahā*. (It should be recalled that many Form VIII verbs are derived from nouns, not necessarily Form I verbs, as explained in Section 2 above.) If this is indeed the case, the meaning of *ntahā* in MSA should not include any reference to self-restraint, which apparently it used to have in CA by virtue of its derivation from *nahā*. This remark holds for most cases exhibiting some degree of opacity due to allomorphy, polysemy or infrequency of the basic form.

One final caveat, however, is in order. Speakers of Arabic today are of varying degrees of proficiency and may have different experiences with written Arabic. They may even hold different perceptions and attitudes toward the various styles and usages, which could manifest in their prescriptive views. Therefore, unless deep investigation has been carried out into the mental lexicon of different categories of speakers, our understanding of the degree of semantic change in the verbal morphology of the language will remain imprecise. Some

change has occurred, nonetheless, though it needs to be studied in language use rather than on the basis of individual intuitions only.

4. Conclusion

It has been argued in this paper that Form VIII verbs show various degrees of transparency/opacity depending on at least three factors: allomorphy, polysemy, and frequency of the base. If transparency/opacity can be measured against the intuitions of modern day speakers of the language, no access is possible to past speakers. One consequence of this fact is that we may never know whether and to what extent a morphological derivation was productive in the past and, if some change happened in this regard, when it happened exactly. To take the example of *ḥtaram* 'to respect' again as cited in the introduction, we will probably never know whether the first recorded occurrence of this verb was treated as a form derived from *ḥarum* 'to be forbidden' or not, and if yes, when it stopped to be so as is the case for modern speakers. This being said, the study of texts from a given period could turn out to be helpful in determining the frequency of the derived forms and their putative bases. As argued in this paper, when a base is frequent with a given sense, it is more likely that part of that sense will be maintained in the derived form than when it is not; compare *manaʕ* vs. *mtanaʕ* and *mahan* 'to serve' vs. *mtahan* 'to practice a profession'. Unlike *manaʕ*, *mahan* is very archaic and is, therefore, unknown to most users MSA. Thus, those who use *mtahan* today are unlikely to think of it as derived from *mahan* and, consequently, would not include 'service' as a component of its meaning much like they would exclude 'forbidding' from the meaning of *ḥtaram*.

References

Anderson, S. 2015. Morphological change. In Bowern, C. & Evans, B. (eds.), *The Routledge handbook of historical linguistics*, 264-285. London – New York: Routledge.

Baalbaki, R. 1995. *Al-Mawrid: A modern Arabic-English dictionary*. Beirut: Dār al-ʕIlm li-l-Malāyīn.

Bahloul, M. 2008. *Structure and function of the Arabic verb*. London – New York: Routledge.

Bauer, L. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.

Booij, G. E. 2013. Morphology in Construction Grammar. In Hoffmann, T. & Trousdale, G. (eds.), *The Oxford handbook of Construction Grammar*, 255-273. Oxford: Oxford University Press.

Brugman, C. 1988. *The story of 'over': Polysemy, semantics and the structure of the lexicon*. New York: Garland Press.

Bybee, J. L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam – Philadelphia: Benjamins.

Copestake, A. & Briscoe, T. 1995. Semi-productive polysemy and sense extension. *Journal of Semantics* 12. 15-67.

Evans, V. 2009. *How words mean: Lexical concepts, cognitive models and meaning construction*. Oxford: Oxford University Press.

Fenk-Oczlon, G. & Fenk, A. 2010. Frequency effects on the emergence of polysemy and homophony. *International Journal "Information Technologies and Knowledge*" 4(2). 103-109.

Fleisch, F. 1979. *Traité de philologie arabe*. Vol. II: *Pronoms, morphologie verbale, particules*. Beyrouth: Dar El-Machreq.

Fortson, B. W. 2003. An approach to semantic change. In Joseph, D. & Janda, R. (eds.), *The handbook of historical linguistics*, 648-666. Malden etc.: Blackwell.

Froud, K. & Khamis-Dakwar, R. 2021. The study of Arabic language acquisition: A critical review. In Ryding, K. C. & Wilmsen, D. (eds.), *The Cambridge handbook of Arabic linguistics*, 48-82. Cambridge: Cambridge University Press.

Glanville, P. J. 2018. *The lexical semantics of the Arabic verb.* Oxford: Oxford University Press.

Harrell, R. 1962. *A short reference grammar of Moroccan Arabic*. Washington: Georgetown University Press.

Himmisley, A. & Stump, G. 2016. *The Cambridge handbook of morphology*. Cambridge: Cambridge University Press.

Ibn ʕUṣfūr, A. 1987. *Al-Mumtiʕ fī al-taṣrīf*. Ed. by F. Qabāwa. Bayrūt: Dar al-Maʕrifa.

Ibn Yaʕīš, M. 2001. *Šarḥ al-Mufaṣṣal*, vol. 4. Dār al-Kutub al-ʕIlmiyya.

Kemmer, S. 1993. *The middle voice*. Amsterdam – Philadelphia: John Benjamins Publishing Company.

Khamis-Dakwar, R. & Froud, K. 2019. Diglossia and language development. In Al-Wer, E. & Horesh, U. (eds.), *The Routledge handbook of Arabic sociolinguistics*, 300-313. Abingdon – New York: Routledge.

Klaiman, M. H. 1991. *Grammatical voice*. Cambridge: Cambridge University Press.

Kotowski, S. & Plag, I. (eds.). 2023. *The semantics of derivational morphology: Theory, methods, evidence*. Berlin – Boston: De Gruyter.

Lakoff, G. 1987. *Women, fire and dangerous things: What categories can reveal about the mind*. Chicago: Chicago University Press.

Langacker, R. 1987. *Foundations of cognitive grammar I: Theoretical pre-requisites*. Stanford: Stanford University Press.

Langacker, R. 1991. *Foundations of cognitive grammar II: Descriptive application*. Stanford: Stanford University Press.

Larcher, P. 2003. *Le système verbal de l'arabe classique*. Aix-en-Provence: Publications de l'Université de Provence.

Lehrer, A. 2003. Polysemy in derivational affixes. In Nerlich, B. & Todd, Z. & Herman, V. & Clarke, D. D. (eds.), *Polysemy: Flexible patterns of meaning in mind and language*, 217-232. Berlin – New York: Mouton de Gruyter.

Lieber, R. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

McCarthy, J. 1979. *Formal problems in Semitic phonology and morphology*. Cambridge, Massachusetts: MIT. (Doctoral dissertation.)

Milton, J. 2009. *Measuring second language vocabulary acquisition.* Bristol – Buffalo – Toronto: Multilingual Matters.

Muʕğam = Muʕğam al-Dawḥa al-Tārīxī li-l-Luɣa al-ʕArabiyya. (https://www.dohadictionary.org/) (Accessed 2023-05-10.)

Qafisheh, H. 1977. *A short reference grammar of Gulf Arabic*. Tucson: University of Arizona Press.

Rainer, F. 2014. Polysemy in derivation. In Lieber, R & Štekauer, P. (eds.), *The Oxford handbook of derivational morphology*, 338-353. Oxford: Oxford University Press.

Rainer, F. & Dressler, W. U. & Gardani, F. & Luschützky, H. C. 2014. Morphology and meaning: An overview. In Rainer, F. & Gardani, F. & Luschützky, H. C. & Dressler, W. U. (eds.), *Morphology and meaning*, 3-46. Amsterdam: Benjamins.

Romaine, S. 1982. *Socio-historical linguistics: Its status and methodology*. Cambridge University Press.

Ryding, K. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.

Saad, G. N. 1982. *Transitivity, causation and passivization: A semantic-syntactic study of the verb in Classical Arabic*. London – New York: Routledge.

Schulte, M. 2015. *The semantics of derivational morphology: A synchronic and diachronic investigation of the suffixes* -age *and* -ery *in English*. Tübingen: Narr.

Talmy, L. 2000. *Toward a cognitive semantics*. 2 vols. Cambridge – London: The MIT Press.

Traugott, E. C. 2000. Semantic change: An overview. In Chen, L. & Sybesma, R. (eds.), *The First Glot International State-of-the-Article Book: The latest in linguistics*, 385-406. Berlin: Mouton de Gruyter.

Tsohatzidis, S. L. (ed.). 1990. *Meanings and prototypes: Studies in linguistic categorization*. London – New York: Routledge.

al-Tūnī, M. Z. Ḥ. (1997). Ṣīγat *iftaʕal* fī al-Qurʔān al-Karīm: Dirāsa luγawiyya. *Mağallat Kulliyyat al-Tarbiya (al-Qism al-ʔAdabī)* 3(3). 1-120.

Versteegh, K. 1984. *Pidginization and creolization: The case of Arabic*. Amsterdam – Philadelphia: John Benjamins.
Versteegh, K. 2014. *The Arabic language*. Edinburgh: Edinburgh University Press.
Wright, W. 1896. *A grammar of the Arabic language*, vol. 1. Cambridge: Cambridge University Press.
Zalami, F. B. 2007. First language acquisition. In Versteegh, K. & Eid, M. & Elgibali, A. & Woidich, M. & Zaborski, A. (eds.), *Encyclopedia of Arabic language and linguistics*, vol. 2, 96-104. Leiden – Boston: Brill.
Zúñiga, F. & Kittilä, S. 2019. *Grammatical voice*. Cambridge: Cambridge University Press.
Zwettler, M. 1978. *The oral tradition of Classical Arabic poetry: Its character and implications*. Columbus: Ohio State University Press.