Abdullah Eisa

Marḥaban: Reconsidering the Criteria of an Arabic Phraseme

Abstract: This paper deals with the difficulties that face Arabic phraseology when the established criteria of phraseology as defined by Gries (2008) are applied. The paper focuses especially on the number of elements involved in a phraseme and here we introduce the concept of one-word + zero-element phrasemes in Arabic.

1 Introduction

Studies on Arabic phraseology focus on empirical applications of Arabic phrasemes. Scholars adopted the definition already established in the research to define an Arabic phraseme (Müller 1993, 2001; Ghariani Baccouche 2007). However, an Arabic phraseme challenges the established criteria for a phraseme. In order to illustrate this, we base our discussion of Arabic phraseology on the criteria for a phraseme as defined by Gries (2008: 6):

1. Natural elements are lexemes or lemmas (words)
2. The number of elements is two or more
3. Frequency of co-occurrence is greater than expected
4. The distance between elements is usually short (interrupted by just one word) or nonexistent
5. The flexibility of elements should not exceed more than one element
6. A phraseme should function as one semantic unit

The six parameters criteria “underlie most phraseological work” (Gries 2008: 5) and provide a precise definition that would help phraseologists and researchers from other fields to identify a phraseme in general and an Arabic phraseme in particular.

Gries suggested that his first criterion (the nature of the elements) included not only lexical items, but also grammatical patterns (Gries 2008: 5). He further argued that lexical items and lemmas should be accepted as phraseological (Gries 2008: 5).

As to the second criterion (the number of elements), a phraseme must be created from two or more elements. The minimum number of elements in the case of
Arabic should be the focus of more scholarly attention, since the morphological concepts *manḥūt* and *al-murakkab al-mazjī* – both of which originally contained two or more lexical items – are dealt with as single lexemes in dictionaries.

With regard to the third criterion (the number of occurrences), Gries claims that a phraseme can be identified as such “if its observed frequency of occurrence is larger than its expected one” (Gries 2008: 5). Although the strong tendency of two items to co-occur has been mentioned in most of the published definitions of phrasemes, such a method requires a well-established corpus, which does not exist for classical Arabic.

Regarding the fourth criterion (the permissible distance between the elements), Gries adopted a “widespread broader perspective” that allowed word coloclocations that contained discontinuous items to be identified as phrasemes (Gries 2008: 5). Arguments in favour of this criterion can be found in papers based on N-gram studies of natural language processes (Gries 2008: 5). However, applying this criterion to Arabic would tend to conflict with Arabic’s syntactic nature as a free-order language.

The fifth criterion (the degree of flexibility of the elements) revolves around the question of how flexible a phraseme ought to be. What tenses can it contain and still be considered a phraseme? What is the level of lexical flexibility for a phraseme? Completely inflexible forms, i.e. full-phrasemes are accepted, but the criterion also allows “relatively flexible patterns”, such as phrases that allow multiple tenses but exclude one particular tense (Gries 2008: 5). Also, the criterion includes “partially lexical-filled patterns”.

Lastly, the sixth criterion (semantic unity) is a semantic one, acting as the core of the definition of a phraseme: any word combination deemed a phraseme should function as one semantic unit (Gries 2008: 6). However, a debate has arisen over whether a phraseme should be semantically non-compositional. Gries argued that this was unnecessary, but advocated unity of meaning (Gries 2008: 6). The final definition of a phraseme he arrived at, based on the foregoing six criteria, was as follows:

[A] phraseologism is defined as the co-occurrence of a form or lemma of [a] lexical item and one or more additional linguistic elements of various kinds which function as one semantic

---

1 N-grams, bigrams, and trigrams are the extracted results of a study that statistically analyses “recurrent continuous sequences of two or more words”. Phraseological studies based on N-gram analysis have usually advocated the continuity of the items of a phraseme (Granger and Paquot 2008: 38–39).
unit in a clause or sentence and whose frequency of co-occurrence is larger than expected on the basis of chance.[1]

(Gries 2008: 6)

The six parameters and definitions discussed above provide us with six clear criteria for the definition of a phraseme. These criteria focus on three main concepts: the individual elements, the occurrence of the elements as a single unit, and the semantic unity of the phraseme. Although this definition provides a comprehensive definition of a phraseme within the frame of the European languages, it needs to be examined within Classical Arabic, the object of this study.

In this paper I will investigate the challenges that applying the criteria of a phraseme provides, aiming to redefine a phraseme within the context of Arabic.

2 Investigating the Criteria

2.1 The Nature of a Phraseme Element

According to the definition proposed by Gries and adopted in this paper, all elements of a phraseme should be words. Words, according to Gries, are “a form or lemma of lexical items and any kind of linguistic element” (Gries 2008: 5). The term ‘word’, however, requires further discussion. In Arabic tradition, ‘word’ is defined as a letter (ḥarf), a noun, or a verb (Ibn ʿAqīl 1980: 14). A noun is thus a word with an independent meaning but no tense; a verb is a word with an independent meaning and a tense; and a letter is a word with neither (Ibn ʿAqīl 1980: 15). Also, given that pronouns in Arabic are considered to be nouns, as they refer to a meaning by themselves and function as nouns grammatically (al-Nīlī 1999: 596; Ibn ʿAqīl 1980: 15), a suffix pronoun – e.g., kāf al-khiṭāb [second person singular] – is considered an independent element of a phraseme and can, with another lexeme, form a phraseme (Ibn ʿAqīl 1980: 31). As a result, any word of any word-class, whether a noun, pronoun, verb, or ḥarf, can form a phraseme under certain conditions as in ḥanānay-ka (your [dual] mercies) = be patient. The second element of the phraseme ḥanānay-ka is the second person pronoun kāf al-

---

2 In traditional Arabic grammar, conjunctions and determiners are included in the ḥarf word-class, while pronouns are included in the noun word-class (Ibn ʿAqīl 1980: 15).
khīṭāb [second person singular], which with the first element ḥanānayn forms the phraseme.  

Additionally, the Arabic definition of ‘word’ implies that there is no distinction between lexical items and grammatical patterns in terms of fulfilling the requirements of phraseme elements; i.e., the granularity level of the element can be either a lemma or a morphological form (Gries 2008: 15).

### 2.2 The Number of Elements

A phraseme, by definition, comprises a phrase. An English phrase, for instance, is defined as “any syntactic unit which includes more than one word and is not an entire sentence” (Matthews 1997: 255). Applying this criterion to the Arabic language calls for further investigation, however, due to the existence of what I will term ‘one + zero elements’ phrasemes. In the following, I will discuss how the word marḥaban is actually a phrase, in the deep structure, and a phraseme made up of one explicit element and a zero element.

Some Arabic phrasemes are made up of two elements, one explicit and the other implicit, i.e., understood from context. The words marḥaban [to be in a spacious place] = to be welcome, and ahlan [to be among one’s people] = to be welcome, are two good examples of this phenomenon. Marḥaban is a word used for greeting, and has the original meaning ‘wide’ (Ibn Manẓūr 2005: 1472–1473). Marḥaban is classified as a cognate object, or what is known in Arabic as mafʿūl muṭlaq. The cognate object is a verbal noun derived from the main verb (Taha 2011: 1), used after a verb to either describe or emphasize it (Ibn ʿAqīl 1930: 169). Given the grammatical class to which marḥaban belongs, we can surmise that the phrase has a missing element. That element can be defined as a zero element on both a syntactic and a semantic level. Syntactically, the accusative case (naṣb) requires a verb from which the cognate object is derived. Marḥaban is therefore in the accusative case as it is influenced al-ʿāmil (the governor) by taqdir. The concept of taqdir can be explained as follows:

The speaker ‘hides’ things in speech, and it is the grammarian’s task to reconstruct these hidden elements in order to explain the surface structure of the sentences. The most important aim of Arabic grammar is the explanation of the case endings (iʿrāb) in the sentences that are produced by the action (ʿāmil) of a visible element in the sentence. If no such

---

3 The -n at the end of ḥanānayn is dropped in the formation of the merging of the word and the pronoun.
element is available, the grammarian must have recourse to an underlying structure in which these elements are made explicit.

(Versteegh 2011: 1)

In the case of marḥaban, the implicit element is the ‘āmil, which is the verb arḥaba. It is crucial to mention here that the implicit element can be either a zero element or a semantic ellipsis. These two potential explanations are both considered below.

First, with regard to ellipsis, an elliptical phrase is one in which some elements are omitted, especially if its meaning is supplied by its context (Matthews 1997: 111). Linguists distinguish between different kinds of ellipsis. In Arabic, there are a number of linguistic phenomena considered to be ellipsis, including sluicing, verb-phrase (VP) ellipsis, and noun-phrase (NP) ellipsis (Mughazy 2011: 2). In sluicing, the omitted element is preceded by a wh-question tool, as in example (1a), where an omission can be understood from the antecedent, thereby allowing the phrase to be interpreted as arāda ‘Ali an yadhhaba ilā l-bayti in (1b). Sluicing therefore contrasts with NP ellipsis, illustrated in example (2), in which the missing element is not a phrase but a single noun. The quantifier can be tanwīn: the suffix n, or the prefix al-. In example (2), the omitted noun can be interpreted as al-muwaẓẓafīn, as in (2b). Lastly, in VP ellipsis, the omitted element is the head verb and its internal object the argument (Mughazy 2011: 2). This type of ellipsis only occurs after auxiliary verbs. No examples of VP ellipsis have been identified in classical Arabic, other than in ‘āmiyy-yah (colloquial), which is beyond the scope of this research (Mughazy 2011: 2).

(1a) ‘Ali arāda dh-dhahāba ilā l-bayti wa lā adrī limādhā.
‘Ali wanted to go home, and I don’t know why.’

(1b) \[
\text{[past., Sing., 3rd person]}[\text{arāda}]_\text{obj}\text{dh-dhahāba}_\text{conj}\text{wa}_\text{neg}_\text{limādhā}_\text{past., sing., 1st person}[
\text{adrī}][\text{pres., sing., 1st person}][\text{Δ}]\]

(2a) al-mudīrū qābala l-muwaẓẓafīn ila l-baʿḍ/ baʿḍa-n
‘The manager met all the employees except for a few.’

(2b) \[
\text{[past., Sing., 3rd person]}[\text{qābala}][\text{obj} \text{l-muwaẓẓafīn}][\text{excep} \text{illa}][\text{obj} \text{l-baʿḍ}][\text{npΔ}][\text{npΔ}][\text{npΔ}][\text{npΔ}][\text{npΔ}][\text{npΔ}][\text{npΔ}][\text{npΔ}]\]

The above examples, although they are not formulaic, refer to the syntactic sub-sentential level, and in the case of a one-element phrase like marḥaba-n, none

---

4 There is no agreement about the typology of ellipses. Nevertheless, VP ellipsis, NP ellipsis, and sluicing, albeit under various names, are widely acknowledged.

5 Mughazy (2011: 2) does not mention the suffix al- as a quantifier, although it can be used in al-baʿḍ for the same purpose, as in example (2).
of the ellipsis types can be applied. Linguists have proposed two rival explanations for this phenomenon. Haddar and Ben Hamadou (1998: 271) referred to it as “false ellipsis”, which can be understood without constructing the complete form. The same authors claimed that false ellipses “can be resolved at the lexical level” (Haddar and Ben Hamadou 1998: 271) and gave two examples of it: ʿida-n saʿīda-n (Happy New Year), and an-nāra n-nāra! (Fire, fire!). The elliptical element in the first example is the verb atamannā (I wish), and in the second, the verb iḥdhar (be careful). Although their examples indicate more than one lexeme, both demonstrate the concept of omitted ʿāmil (action) – the case with which we are specifically concerned. It may thus be claimed that ellipsis can be understood on a lexical level, but further investigation into the syntactic level is nevertheless required.

Stainton discussed two potential modes of analysing/explaining the phenomenon: a pragmatics-oriented approach, and semantic ellipsis (Stainton 2005: 386). The first requires that an utterance’s “face value” be the main focus of analysis, while its pragmatics – i.e., gestures and context – treated as the responsibility of the utterance’s receiver, who reconstructs missing elements and fills in gaps (Stainton 2005: 387). Crucially, however, it is not the non-sentential phrases that this approach intends to reconstruct; it has no interest in filling in linguistic gaps (Stainton 2005: 387). Rather, the non-linguistic context in which the elliptical phrase occurs fills the semantic gap in the utterance.

Analysing the phenomenon of one-word Arabic phrasemes using Stainton’s pragmatics-oriented approach therefore leads us to either a) accept or b) reject the idea that single words can be phrasemes. The first option, however, must be rejected as contradicting the definition of phraseology and its units: for a phraseme, by definition, is formed from a phrase, which cannot comprise fewer than two elements (in the case of the Arabic language, lexemes and pronouns). And in considering the second option, we cannot overlook the fact that one-word expressions function as phrasemes in Arabic, and syntactically reflect a missing element – the ʿāmil (action) – which changes their grammatical case from nominative to accusative. These cases are marked by case-endings: ḍammah [suffix u],

---

6 Stainton provides two “competing views” of how to explain how the gap is filled. The first view, advocated by Barton (2005: 386), “postulates (i) a sub-module of linguistic context, that operates exclusively on the sub-sentence uttered plus prior explicit discourse, (ii) a sub-module of conversational context, that takes the output of the first sub-module as input, and uses non-linguistic context […] to derive what the speaker meant to convey”. The other view, advocated by Stainton himself, is that while gap-filling does occur “via non-deductive inference”, there are no pragmatics modules at work, but rather “central system progresses, inferential processes not specific to language, [that are used] to bridge the gap”.

wāw [suffix ū], alif [suffix ā] and nūn [suffix n] for nominative, and fathāh [suffix a], yāʾ [suffix ʾ], alif [suffix ā], and hadhf [deletion] for accusative.

Semantic ellipsis occurs when a sentence is elliptic, but the ellipsis can be reconstructed by applying the syntactic rules of the language in the absence of an uttered antecedent. It differs from the previously described varieties of ellipses in that it could potentially explain one-word phrasemes. Although marḥab-ta, the omitted verb of marḥaba-n, does not exist (or at any rate has not been detected) in lexicons of the Arabic language, it must still be reconstructed – especially in combination with the cognate object marḥaba-n – to justify the accusative case ending (al-Farāhīdī 2005: 342). Stainton defended the pragmatics-oriented approach by arguing that the reconstructed phrases may not suit the elliptic phrase, and cited the following example. If someone asks ‘Who loves Michael Jackson?’, the answer could be ‘Me’. The elliptic part of the phrase ‘Me’ does not suit the reconstructed phrase ‘I love Michael Jackson’, since the pronoun in the elliptic phrase is in the accusative case, whilst in the reconstructed phrase it is in the nominative case. Stainton gives another example in German. A German speaker would say ‘mein Vater’ [my [nom.] father], whilst pointing at someone that reminds him of his father; however the reconstructed phrase would be ‘Das erinnert mich an meinen Vater’ [that reminds me of my [acc.] father]. However, in the case of Stainton’s first example, answering the question posed with ‘I’ or ‘I do’ would be more grammatically correct English than answering with ‘Me’, even though the latter is generally accepted in colloquial usage; i.e., the elliptic phrase could originally have been composed with the pronoun in the nominative case. Alternatively, we can view the reconstructed phrase as being the (likewise grammatically correct) ‘It is me who likes Michael Jackson’. Similar arguments can be applied to Stainton’s German example. It should also be noted that ellipsis can be used to simplify an utterance, and that therefore, an elliptic phrase can be understood when the simplest case is used, even if it does not agree with the original/reconstructed phrase.

The example of marḥaban can be better explained via the concept of a zero element in a phraseme, given that a phraseme is a set phrase and a phrase by definition is more than one word (McGregor 2003: 77–78, 82). However, a one+

7 “When al-Khalīl (2003: 105) was asked about the accusative case of marḥaban he said “in it a hidden verb”; he meant: dwell or stay, so it became accusative by a hidden verb, then it became dead when its [the verb’s] meaning became well-known” (al-Farāhīdī 2005: 342).
8 McGregor (McGregor 2003: 77–119) provides a detailed account of the historical background of the concept of the zero-element. He also differentiates between ‘zero’ and ‘nothing’, for zero should fulfill two conditions provided by Haas (1962: 49): a) distinctive omission of overt forms,
zero element phraseme is not to be confused with a lexeme + pronoun phraseme, in which the second element is a suffix pronoun, e.g., ḥananayka [your [dual] mercies] = slowly. Though the zero element is the element that does not exist in some linguistic cases of a given language, its visible equivalent does exist in the majority of language cases (Haas 1962: 34). However, the zero element has an impact on its linguistic context (Haas 1962: 34). For example, the suffixes -ed and -t are morphemes that indicate the past tense in verbs in English, although such morphemes do not exist in verbs like cut and put (Haas 1962: 34). However, these verbs’ tenses can be understood from context; and the absence of an element signifying the tense is the zero-element. In Arabic, the zero-element applies to sukūn (Bishr 1998: 187; Firth 1957: 180–189):9 the case-ending used in the absence of any of the three case endings fatḥah \(\acute{a}\), ḍammah \(\dot{u}\), and kasrah \(i\) (Bishr 1998:187). Jazm is a syntactic case in which the case-ending is a zero-element (sukūn \(\emptyset\)) because it demonstrates an absence of the uttered morphemes. In the case of a one-uttered-word phraseme, the second non-pronounced element is a zero-element of the phraseme. For instance, the verb marḥb-ta, which functions as the action of the cognate object marḥaban, constitutes the zero-element in the phrase that forms the phraseme marḥaban.

In conclusion, applying the second criterion of phraseology to Arabic phrasemes creates a difficulty that needs to be overcome insofar as Arabic includes one-word phrasemes in which one element is uttered and the other is elliptic. This phenomenon is best classified as semantic ellipsis for two reasons. Firstly, the uttered element is a cognate object that needs an governor (ʿāmil) to justify its grammatical case. Therefore, a verb that coheres with it is reconstructed (taqdīr) – as arḥab-ka allāhu [[may] God [have] you in a spacious [place]], in the case of marḥaban. Secondly, a one word Arabic phraseme, e.g. marḥaban, does not constitute a syntactic ellipsis, since in such an ellipsis, the uttered element requires a reference to an uttered antecedent, but the action/verb has never actually been found in classical Arabic in the context of marḥaban. Finally, the application of the concept of the zero-element to one-word phrasemes allows them to meet the established definition of a phrase. The elliptic element of the phraseme marḥaban is a zero-element, as it exists only in parallel phenomena, and it has an effect – i.e., the formation of a phraseme – on the existing element. The one-

9 The concept of a zero-element in Arabic was first introduced by Firth (1957: 180–189) then further explored by Bishr (1998: 187).
word phraseme can thus be defined as a single word that is part of an elliptic phrase and therefore functions as a phraseme on its own.

Another issue that emerges when attempting to apply the second criterion to Arabic phraseology is the polylexical phenomenon of naḥt: two or more words that are merged into one, losing some of their letters in order to cohere with the structure of the quadrilateral root. For instance, ḥawqalah is derived from lā ḥawla wa lā quwwata illā bi-llāh [there is no might nor power except in God]: a sentence used in prayer or in response to an unpleasant situation. Such words function as phrasemes since they adhere to the other criteria; however, they require some further explanation. Semantic ellipses and zero-elements cannot be applied to the phenomenon of naḥt, since there is neither any ellipsis nor are there any non-pronounced elements. However, the original words are merged via contraction. Hence, manḥūt is a phraseme written as one word, but composed of fragments of other words that together formed a sentence-long phraseme (al-Khaṭīb 2003: 439).

2.3 The Number of Co-Occurrences Required before a Phrase can be Considered a Phraseme

Counting the instances of co-occurrence of a particular phraseme in classical or Modern Standard Arabic would normally require the existence of a corpus of relevant text. In its absence, classical collections of idioms and proverbs including Amṯāl al-ʿArab by al-Mufaḍḍal al-Ḍabbī (d. 784), al-Durrah al-Fākhirah fi al-Amṯāl al-Sāʾirah by Ḥamzah al-Aṣfahānī (d. 961), Majmaʿ al-Amṯāl by al-Maydānī (d. 1124), and collections of non-figurative set phrases like Thimār al-Qulūb fi al-Muḏāf wa al-Mansūb by al-Thaʿālibī (d. 1038) and Mā Yuʿawwal ʿalayh fi al-Muḏāf wa al-Muḏāf Ilayh by al-Muḥibbī (d. 1699), are key repositories of phrasemes for Classical Arabic and for MSA, which contains a large number of Classical Arabic phrasemes. Collections of Classical Arabic books such as Islamport.com or Shamila could be referred to, in order to measure the number of occurrences of the phraseme in Classical Arabic works. Also, the International Arabic Corpus is useful for MSA. Additionally, collections of eloquent phrases are an important source of phrasemes, reflecting prevalent metaphorical phrases. Two examples of this type of lexicon will be referred to: the Jawāhir al-Alfāẓ of Qudamah Ibn Jaʿfar (d. 949), and the al-Alfāẓ al-Kitabiyyah of al-Hamathānī (d.

---

10 The root in Arabic consists of either three letters 1-2-3 (a-k-l) or four letters 1-2-1-2 (w-s-w-s)/1-2-3-4 (ḥ-n-ẓ-l).
939). However, with Gries (2008: 5) stating only that this number should be “larger than [...] expected”, as previously noted in reference to Gries’s work, the lack of a well-established corpus of Classical Arabic does not support the corpus-based method of identifying a phraseme.

Accordingly, analyses of the idiomatic level and referring to the previous collections would be the potential methodology used to identify Arabic phrasemes. However, distinguishing idiomatic from literal meaning in Arabic can at times be problematic because of the lexemes that are affected by dead metaphors. Two conditions will merit the use of dictionaries for the purpose of tracking original meaning. First, the source should be written before the target era. Secondly, the original meaning, i.e. literal if it occurs in a secondary meaning, of the phraseme’s elements should be indicated. For example, for phrasemes that occur in late Andalusi works, e.g. the works of Ibn al-Khaṭīb, the dictionaries that can possibly be used for this purpose are the Al-ʿAyn of al-khalīl (d. 736), the Tahthib al-Lughah of al-Azharī (d. 981), the Maqāyīs al-Lugah of Ibn Faris (d. 1004), and the Tāj al-Lugah wa Ṣiḥāḥ al-ʿArabiyyah of al-Jawharī (d. 1003). Additionally, the original source domains of the phrasemes would be traced to their possible sources, with the aim of gaining a clear indication of their primary semantic level.

2.4 The Permissible Distance between the Elements of a Phraseme

A phraseme is a set phrase in which the elements cannot be substituted. These elements function as one semantic unit by being attached to each other (Gries 2008: 6). In a restricted-order language, the order in which an element occurs in a phrase is important to the reader’s understanding of that word’s grammatical class. Arabic is a free-order language, meaning that the grammatical class of a word is not affected by the order of the elements in the phrase in which it appears (al-Sīrāfī, 2008: 263).11 This raises an important question: What are the limits of order-change in an Arabic phraseme? To arrive at a definitive answer will require thorough analysis. It is reasonable to claim that a set phrase can be considered a

---

11 In some cases in Arabic, order is important for the identification of the grammatical class of a word: for instance, when a case-ending does not appear because it would render a long vowel at the end of a word un-pronounceable. One example of this is ḍaraba ʿĪsā Mūsā [Isa hit Musa]. Both Musa and Isa end with long vowels that cannot be pronounced alongside either the case-ending of the nominative case /u/, or the case-ending of accusative case /a/. Thus, only word-order reveals the meaning of the sentence, based on grammarians’ agreement that the subject comes before the object.
Phraseme as long as order-changes do not affect its metaphorical meaning. For instance, in the case of the phraseme *as-salāmu ʿalaykum wa raḥmatu l-lāhi wa barakātuhu* (may the peace, mercy, and blessings of God be with you), although the literal meaning of the phrase’s element does not express the meaning of greeting, the phrase is commonly used as a greeting. It can be found in various orders, such as *salāmu l-lāhi ʿalayka wa raḥmatuhū wa barakātu* and *ʿlayka wa raḥmatu l-lāhi s-salāmu*. In all three of these versions, despite changes to word-order, the phrase retains both the same metaphorical meaning and the same function as a greeting.

The second issue that must be addressed under this criterion is the size of any gap between the elements of a phraseme. As discussed earlier, in a broad sense and up to a certain point, a gap between the elements of a phraseme can be accepted. To identify the specifics of such limits in Arabic, a survey study would be required. However, as noted above the metaphorical meaning of a phrase is the main criterion for accepting a phrase, regardless of whether it is a phraseme or has lost its phraseological identity. For instance, *ʿalā qawmihā janat barāqish* [Barāqish has harmed her people] is a phraseme used to describe anyone who hurts their people unwillingly. In the context of an own-goal in a football match, for instance, the commentator might say *wa barāqishu hunā narāhā janat li-lʿas-aṭīʿ alā qawmihā* [and Barāqish, we can see her here, harmed, unfortunately, her people]. Although the phraseme has been changed syntactically – with Barāqish this time not a subject, but an object functioning as an antecedent of the omitted pronoun in the verb *janat* [harmed] – it still reflects its original metaphorical meaning. This, of course, works with sentence-long phrasemes but not with one-word phrasemes or with lexical idioms (e.g. compounds such as *manḥūt* or *tarkib mazjī*).

### 2.5 The Lexical and Syntactic Flexibility of Phraseme Elements’ Non-Substitutability

The concept of fixedness of an Arabic phraseme can be examined on two main linguistic levels: the syntactic and the lexical. Syntactically, phrasemes that “break the conventional grammatical rules” (Moon 1998: 21), known as ill-formed collocations, are completely fixed. Ill-formed collocations can be idioms, proverbs or even pragmatic phrasemes. A clear example of an Arabic pragmatic phraseme that is an ill-formed collocation is *murghamu-n akhāka lā baṭal* [your brother is forced (to do what he has done) not a hero]. Under the conventional grammatical rules of Arabic, *akhāka* should be written in the nominative case (*akhūka*) as the subject of a passive-voice sentence, or as it is known in Arabic,
nāʾīb fāʿil. In this pragmatic phraseme, the two conditions of free phrases are violated (Mel’čuk 1998: 30). Pronouns in Arabic are mostly morphemes, so in phrasemes that contain a pronoun, the pronoun changes with context. For instance, ḥanānay-ka [your (dual) mercies] is grammatically fixed in the accusative case, and its pronoun changes depending on the person(s) to whom it is addressed, as follows:

2MUS-ḥanānay-ka, 2FEM-ḥanānay-ki, 2DUL-ḥanānay-kuma, 2PLUR.FEM-ḥanānay-kum, 2PLUR-MAS ḥanānay-kunna.

Thus, the fixedness percentage of an Arabic phraseme can either be complete (in the case of ill-formed collocations/pragmatic phrasemes) or semi-flexible; and its status as completely fixed or semi-flexible affects whether its pronoun morpheme varies with context.

The lexical flexibility of an Arabic phraseme depends on the number of elements it has. Phrasemes with two elements, regardless of whether both are uttered or one is a zero-element, are completely fixed. Marḥaba-n (a lexeme + zero-element phraseme) and subḥāna l-lāh [exalted is God] (both elements of which are uttered) are both examples of two-element phrasemes that are completely lexically fixed. However, the lexical flexibility of phrasemes that are formed of more than two lexemes is merely restricted, due to the ability of the receiver/audience to comprehend the metaphorical meaning intended by the phraseme’s formation. Take the phraseme ċaraḥaba ʿuṣfūrayni bi ḥajari-n wāḥidi-n [(he) hit two birds with one stone]. If a speaker means to refer to finishing two or more tasks by performing just one action, he can either use the phraseme as it is, or change the word ʿuṣfūrayn [two birds] to ʿamalayn [two tasks], yielding ċaraḥaba ʿamalayni bi ḥajari-n wāḥidi-n. His audience will comprehend the reference to the metaphorical meaning because the semantic metaphorical meaning is still preserved in the remaining elements of the phraseme. Similarly, if the element ʿuṣfūrayn remains while ḥajari-n [a stone] is changed to another lexeme, such as tawqīʿ [signature] in the context of, say, paperwork, the phrase now being ċaraḥaba ʿuṣfūrayn bi tawqīʿi wāḥidi-n, the intended metaphorical meaning of the phraseme will still be obvious to the Arabic audience. In other words, the lexical flexibility of an Arabic phraseme is dependent on two conditions: 1) the phraseme must be formed of more than two elements, and 2) its metaphorical meaning must remain intact.

12 Unless the pronoun in the accusative case, it is either separated from the action or placed before the action. In these two situations, the pronoun is īyyā + (second-person or third-person pronoun). Pronouns in the nominative case are treated as separate lexemes.
2.6 The Semantic Unity and Unpredictability of a Phraseme

Fully fixed phrasemes are defined by the third and the fourth cases of the formula provided by Mel’čuk (1998: 30–31):

A [phraseme] AB of a language L is a semantic phraseme of L such that its signified ‘X’ is constructed out of the signified of one of its two constituent lexemes—say, of A— and a signified ‘C’ [‘X’ = ‘A[C]’] such that the lexeme B expresses ‘C’ only contingent on A.

(Mel’čuk 1998: 30)

The third case:

‘C’ = ‘B’, i.e. B has (in the dictionary) the corresponding signified; and ‘B’ cannot be expressed with A by an otherwise possible synonym of B.

(Mel’čuk 1998: 31)

As in the Arabic phraseme: Baytu l-Māl ‘the house of money’ (ministry of finance in the medieval era).

The fourth case:

‘C’ = ‘B’; ‘B’ includes (an important part of) the signified ‘A’, that is, it is utterly specific, and thus B is ‘bound’ by A.’

(Mel’čuk 1998: 31)

As in the Arabic phraseme: kharīru l-mā’i ‘the sound of falling water’.

Mel’čuk’s (1998: 30) formulae illustrate fully fixed phrasemes in which neither element can be substituted, at a semantic level. In the first formula, bayt as an individual lexeme means ‘house’, while al-māl means ‘money’. Yet the individual meanings of the lexemes do not add up to or predict the overall meaning of their phraseme: ‘ministry of finance’. Moreover, substituting a synonym for either of these elements will obscure the metaphorical meaning of the original phraseme. The same phenomenon can be observed with other figurative metaphors, and to a certain extent with non-figurative ones, e.g., kharīru l-mā’i [the voice of falling water] = a specific term for the sound of water like in a waterfall.

In the case of kharīru l-mā’i, the first element of the phraseme does not co-occur with any other lexeme, since the semantic field of the first element is included in the semantic field of the second element. This leads us to deem it a ‘cranberry collocation’: i.e., one of the elements – kharīr, in this instance – is unique to that collocation (Moon 1998: 21). Nevertheless, this unique element can
be replaced by another synonym that gives a broad sense of the target meaning. *Kharīr* is a special term to indicate the sound of falling water, but if a speaker uses *šawt* [sound] in the same context, it will be understood, provided that the hearer recollects the meaning of the original substituted element, *kharīr*.

In short, Arabic phrasemes occur as single semantic units, and their meanings cannot be predicted from the individual meanings of their elements. In non-figurative phrasemes, and in figurative ones (albeit with more difficulty), one of the elements can have a synonym substituted for it. However, when this happens, the resultant phrase 1) does not act as a phraseme, and 2) requires the audience to recall the original element of the phraseme, in order to understand the semantic unit that the collocation seeks to provide.

### 3 Conclusion

This paper has explored the challenges that emerge when the established criteria for phrasemes are applied to the Arabic language. The first criterion is affected by the fact that pronouns in Arabic are considered to be one-letter nouns; and the second, by the existence of numerous one-word Arabic phrasemes. The theory of the zero-element was found useful in overcoming the latter issue, insofar as a one-word phraseme can be construed as having two elements, one of which is a zero-element lexeme that was important in the formulation of the phraseme, but which no longer explicitly exists.

With regard to the third criterion, a lack of corpora prevents direct counting of the co-occurrence of the elements of a given phraseme in Arabic. We will therefore utilise metaphorical fixedness as a key parameter of the phrasemes sampled from that literature, supported by comparison with collections of fixed collocations in Arabic. In terms of the fourth criterion, the question of the distance between the elements of an Arabic phraseme will require further investigation. However, this chapter established that Arabic phrasemes exhibit a degree of flexibility based on the context, as long as the sixth criterion is fulfilled.

As to the fifth criterion, an Arabic phraseme can have some flexibility as regards accepting a substitute element, when the phraseme is formed of more than one uttered element and its semantic unity remains intact. Finally, Arabic phrasemes fit the sixth criterion in the sense that they occur as single semantic units. This criterion also supports the fifth one, by demonstrating the possibility of substituting one or more of the elements in a phraseme – but only if the audience recalls the original element(s).
References

al-Farāhīdī, al-Khalīl Ibn Aḥmad (2005): Kitāb al-ʿAyn. Beirut: Dār Iḥyāʿ al-Turāth al-ʿArabī.
Al-Khaṭīb, ʿAbd. Al-Laṭīf (2003): al-Mustaṣqā fi ʿIlm al-Taṣrīf. Kuwait: Dār al-ʿUrūbah.
al-Nīlī, Ibrāhīm (1999): Ṣafwat al-Ṣafiyyah Fi Sharḥ al-Durrah al-ʿAlfīyyah. Mecca: Jamiʿah Um al-Qurā.
al-Sīrāfī, Abū Saʿīd (2008): Sharḥ Kitāb Sībawayhi. Beirut: Dār al-Kutub al-ʿIlmiyyah.
Baccouche, Moufida Ghariani (2007): Arabic Phraseology. In Harald Burger, Dmitrij Dobrovolskij, Peter Kühn & Neal R. Norrick (eds.), Phraseology: An International Handbook of Contemporary Research, 752–758. Berlin, New York: De Gruyter.
Bishr, Kamaal (2003): Dirāsāt fī ʿIlm al-Lughah. Cairo: Dār al-Kutub al-ʿIlmiyyah.
Firth, John R. (1957): Papers in Linguistics. Oxford: Oxford University Press.
Gries, Stefan Th. (2008): Phraseology and Linguistic Theory: A Brief Survey. In Sylviane Granger & Fanny Meunier (eds.), Phraseology: An Interdisciplinary Perspective, 3–26. Amsterdam, Philadelphia: John Benjamins.
Haas, William (1962): Zero in Linguistic Description. In John R. Firth (ed.), Studies in Linguistics, 33–53. Oxford: Basil Blackwell.
Haddar, Kais & Abdmajid Ben Hamadou (1998): An Ellipsis Detection Method Based on a Clause Parser For Arabic Language. In Diane J. Cook (ed.), Proceedings of the Eleventh International FLAIRS Conference, 270–274. California: The AAAI Press.
Ibn ʿAqīl, Bahāʾ al-Dīn ʿAbdullāh (1980): Sharḥ Alfiyyat Ibn Mālik. Cairo: Dār al-Turāth.
Ibn Manẓūr, Muḥammad (2005): Lisān al-ʿArab. Bairut: al-Aʿlamī li-l-Maṭbūʿ¯.
Karabekyan, Samvel (2011): Ḥarf'. In Lutz Edzard & Rudolf de Jong (eds.), Encyclopedia of Arabic Language and Linguistics. Leiden: Brill. Retrieved from http://dx.doi.org/10.1163/1570-6699_eall_EALL_SIM_vol2_0015, accessed July 2, 2017.
Kees, Versteegh (2011): Taqdīr’. In Lutz Edzard & Rudolf de Jong (eds.), Encyclopedia of Arabic Language and Linguistics. Leiden: Brill. Retrieved from http://dx.doi.org/10.1163/1570-6699_eall_EALL_SIM_0134, accessed July 2, 2017.
Matthews, Peter (1997): Ellipsis. In Peter Matthews (ed.), The Concise Oxford Dictionary of Linguistics, 111. Oxford: Oxford University Press.
McGregor, William (2003): The Nothing That Is, The Zero That Isn’t. Studia Linguistica 57 (2), 77–87.
Mel'cuk, Igor (1998): Collocation and Lexical Functions. In Anthony Paul Cowie (ed.), Phraseology: Theory, Analysis, and Applications, 23–54. Oxford: Clarendon Press.
Moon, Rosamund (1998): Fixed Expressions and Idioms in English: A Corpus-Based Approach. Oxford: Clarendon Press.
Mughazy, Mustafa (2011): Ellipsis. In Lutz Edzard & Rudolf de Jong (eds.), Encyclopedia of Arabic Language and Linguistics. Leiden: Brill. Retrieved from http://dx.doi.org/10.1163/1570-6699_eall_EALL_COM_vol2_0003, accessed July 2, 2017.
Müller, Kathrin (1993): Und der Kalif lachte, bis er auf den Ruecken fiel. Ein Beitrag zur Phraseologie und Stilkunde des klassischen Arabisch. München: C. H. Beck.
Müller, Kathrin (2001): Da war ihm, als müsse er fliegen vor Freuden. „Tausendundeine Nacht“ als Fundus für arabische Phraseologie. München: C. H. Beck.
Stainton, Robert. J. (2005): In Defense of Non-Sentential Assertion. In Zoltan Gendler Szabo (ed.), Semantics Versus Pragmatics, 383–458. Oxford: Clarendon Press.
Taha, Zeinab Ahmed (2011): Mafʿūl. In Lutz Edzard & Rudolf de Jong (eds.), Encyclopedia of Arabic Language and Linguistics. Leiden: Brill. Retrieved from http://dx.doi.org/10.1163/1570-6699_eall_EALL_COM_vol3_0201, accessed July 2, 2017.