

## Lexical Variation in Arabic Phraseological Units Calqued from English: An Explanatory Approach

*Ahmed Seddik Al-Wahy*

Assistant Professor, Faculty of Al-Asun  
(Languages), Ain Shams University, Egypt.

### Abstract

Contrary to the long-standing view that regards fixedness as one of the defining features of phraseological units (PUs), investigation of authentic texts suggests that PUs display a higher degree of variability than was generally assumed. The present paper aims to classify and provide explanations for the types of lexical variation in Modern Arabic PUs that have been calqued from English. It shows that calqued PUs display types of lexical variation that are different from

those displayed by original PUs. The study also argues that explanations that are restricted to only one branch of linguistics cannot fully account for the types of variation found in Arabic calqued PUs and that an approach that combines both linguistic and extra-linguistic aspects can provide a more appropriate explanatory tool for such variation.

*Keywords:* phraseological units, calque, fixedness, lexical variation, Modern Standard Arabic

## Lexical Variation in Arabic Phraseological Units Calqued from English: An Explanatory Approach

*Ahmed Seddik Al-Wahy*

### 1. Introduction

The issue of variation in phraseological units (PUs) has recently received growing attention in phraseology and other branches of linguistics. This reflects the interdisciplinary nature of PUs, since, apart from their distinctive linguistic features, they combine cognitive, sociocultural, and ideological aspects. In addition, findings of studies in this area can have both theoretical and practical implications.

Interest in PUs in general and in their variability in particular is partly the result of the findings of corpus linguistics. Initial interest in phraseology in English was mainly triggered by the work of Sinclair (1991), who, depending on corpus evidence, suggested that much of the language people use consists of multi-word expressions rather than single words. More recently, corpus investigation has indicated that, in actual use, many such expressions are not as fixed as was traditionally assumed (e.g., Wray, 2002; Wulff, 2008; Naciscione, 2010). PUs and variation are also, incidentally, both aspects of linguistic irregularity, and, as Dobrovol'skij (2015, p. 275) observes, "present-day linguistics is devoting more and more attention to irregular phenomena in the structure of language." PUs often defy concepts such as grammaticality and compositionality as well as the sharp distinction between lexis and grammar,

which has been dominant in formal approaches to language, such as the generative approach. Variation is similarly problematic to such approaches. As Walker (2013, p. 455) observes, "in linguistic theory [by which he means the generative theory], variation has tended to be viewed as a problem to be avoided or solved." This is because the study of variation is concerned with performance, which is regarded as imperfect and unsystematic, rather than competence, the idealized knowledge of language that the generative theory focuses on.

The present study aims to identify the types of lexical variation in Arabic PUs that have been borrowed from English and to provide explanations, both linguistic and extra-linguistic, for such variation. Calqued PUs have been chosen in particular because they are different in many ways from original, non-calqued PUs, and thus are expected to display patterns of lexical variability that may not be found in original PUs. The differences relate to their origin (they appeared in a different culture and represent a different way of thinking), degree of institutionalisation (they are normally less conventionalized than original PUs), and context of use (they tend to be used in formal settings and are more characteristic of media discourse). Two main questions are addressed here:

- (1) What are the types of lexical variation in Arabic PUs that have been calqued from English?
- (2) What are the various factors beyond such variation?

The variety of Arabic investigated here is Modern Standard Arabic (MSA), which abides by the rules of Classical Arabic grammar, especially in terms of word inflection and case marking. It is a typically written variety, and it is sometimes referred to as “Modern Written Arabic” (e.g., Badawi, Carter, & Gully, 2004), which is the label used by Wehr (1974) in his Arabic-English dictionary. However, this variety is also encountered in speech, e.g., in news bulletins, media reports, formal speeches, pre-recorded telephone messages, or any kind of discourse where the aim is to impress the recipient or evoke a sense of seriousness. The characterisations of this variety given by Wehr (1974, p. VII) are still largely the same as those given by other linguists for MSA (e.g., Ryding, 2005, p. 4; Newman, 2007, p. 1; Badawi, Carter, & Gully, 2004, p. 37). MSA also displays the influence of Western languages, especially English and French. One aspect of such influence is the extensive process of borrowing, including the calque of PUs (see Al-Wahy, 2005, 2009 for further discussion). Unless otherwise indicated, ‘Arabic’ in the present paper refers to this modern standard variety.

The present corpus-based study adopts a taxonomic, explanatory approach to PU variation. First, it provides a classification of the types of calqued PUs based on formal criteria. Second, it seeks to explain variant phraseological forms with reference to any linguistic or extra-linguistic factors that may have resulted in

their co-presence in MSA. The examples and the frequencies of PU variants are drawn from the Arabic Language Corpus of King Abdul-Aziz City for Science and Technology (available at <http://corpus.kacst.edu.sa/index.jsp>), a 1,182,515,633-word corpus representing different varieties of written Arabic, which allows search according to region, historical period, type of publication, and topic. It allows search for combinations of words and collocations as well as single-word searches. The historical search available in this corpus has been used to ascertain that a given form is a calque rather than an original Arabic PU, while the geographical filter has been used to check whether there are regional preferences for certain PU variants.

In transliterating the Arabic examples, the LOC (Library of Congress) romanization system (see Appendix) has been used. It is beyond the scope of the present study to review all the 26 rules for the application of the system, and the reader can refer to the full version (available at <https://www.loc.gov/catdir/cpso/romanization/arabic.pdf>) for further details. However, it is important to bear in mind that it is a transliteration system; it is not intended to reproduce or guide the user to the natural pronunciation of Arabic. Therefore, features such as assimilation and elision, which are normally represented in phonetic, but not in phonemic, transcription, are not indicated in this system. For instance, the system does not represent the assimilation of the definite *al-* (the) when attached to words beginning with a “sun” letter, such as *sūq* (market), whose definite form is transliterated as *al-sūq* rather than *as-sūq*. For the sake of consistency, this rule is

observed here, even though it does not replicate the natural pronunciation of Arabic. The system does not represent the glottal stop (*hamzah*) in word-initial position, though it does represent it in middle and final positions. This rule is followed here, unless the *hamzah* is fully consonantal and is thus pronounced even if it occurs in middle position. Thus, the *hamzah* is represented in the words *'akhadha* (to take), *'ahmar* (red), and *'a'sāb* (nerves), but not in *ittakhadha* (to take) *ittiḥād* (union) or the definite *al-*, since in such cases the *hamzah* is lost in non-initial position. This modification is necessary, not because it is closer to natural pronunciation, but rather because it reflects the distinction between the consonantal *hamzah* and the non-consonantal glottal stop that occurs when uttering an initial vowel. Unlike phonetic transcription, the system allows the use of standard punctuation marks, such as the full stop and question mark, capital letters to indicate proper names, and hyphenation to show word boundaries in the case of the definite *al-* and inseparable prepositions and conjunctions, such as *bi-* (with) and *wa-* (and). While the system does not indicate case marks, except in some cases of nunation (*tanwīn*) involving nouns used adverbially, grammatical functions can be inferred from the glosses accompanying Arabic expressions. As for the translation of Arabic linguistic terms, the present study depends on Ryding (2005).

## 2. Theoretical Concepts

While it is generally accepted that phraseology is concerned with conventionalised multi-lexical expressions of different types, there is little agreement among phraseologists regarding the unit of study of the discipline and the terms

assigned to its various categories. Wray (2002, p. 9), for instance, lists more than fifty terms, each proposed as denoting the unit of study of phraseology, ranging from “amalgams,” “fixed expressions,” and “frozen phrases” to less rigid labels, such as “conventionalized forms,” “multiword items/units” and “recurring utterances.” The situation becomes more complicated when one considers the further distinctions and classifications under such general categories. As Wulff (2008, p. 11) observes, “the boundaries between idioms, collocations, and other multi-word units are fuzzy.” This confusing situation has not changed and is not expected to improve soon, given the accumulative and interdisciplinary nature of phraseology. Indeed, some phraseologists (e.g., Burger, Dobrovol'skij, Kühn, & Norrick, 2007, p. 18) suggest that unification of terms is “only possible and desirable to a certain degree.”

### 2.1. Phraseological unit

The term “phraseological unit” is used here as an umbrella term that covers various types of conventionalised, relatively fixed combinations of words. The term is not limited to combinations that are strictly “phrasal” in the grammatical sense, such as idioms and collocations, but is extended to multi-word expressions of higher ranks, such as proverbs and wise sayings. The term also covers pragmatic word combinations, such as salutations and greetings, which are transparent in meaning but are tied to special contexts of use. These have been referred to as “pragmatemes” (Mel'čuk, 2012, p. 41), “pragmatic formulae” (Pawley, 2007, p. 19), and “pragmatic idioms” (Bednarek & Bublitz, 2007, p. 129). The term “phraseological unit” also covers slogans, catchphrases, and famous

quotes, which may, but need not, be figurative. PUs from all the above types have been borrowed from English into MSA.

There are two types of narrow definition of PUs, one excluding non-phrasal expressions and the other excluding non-figurative expressions. Proverbs were traditionally excluded from the field of phraseology on the grounds that they are grammatically realized as full sentences and that they are best studied under paremiology (Jaki, 2104 :7), the discipline concerned with proverbs as units of folklore that express the values and cultures of different nations. While phraseology is also concerned with aspects of culture (e.g., Bragina, 2000; Skandera, 2007), paremiology focuses more on “the traditional nature of proverbs as items of folklore” (Norrick, 2014, p. 10). Most phraseologists, however, include proverbs and wise sayings as PUs, though the immediate purposes of individual researchers may impose some restrictions. For instance, Naciscione (2010, p. 8) defines PU as “a stable, cohesive combination of words with a fully or partially figurative meaning,” which includes proverbs under PUs. By contrast, she regards any set expression that is devoid of figuration, such as literal wise sayings and collocations, as a non-phraseological combination of words (Naciscione, 2010, p. 19).

Although proverbs and wise sayings are full sentences, they are parallel to idioms and collocations, respectively. Like idioms, proverbs are figurative, and like collocations, wise sayings are literal. Unlike idioms and collocations, however,

proverbs and wise sayings usually have a didactic function. While idioms typically “express a concept,” proverbs “typically state some commonly believed truth or advice” (Schmitt & Carter, 2004, p. 9). Proverbs such as *There is no use crying over spilt milk* and *People who live in glass houses shouldn't throw stones* have been calqued into Arabic. Examples of non-figurative wise sayings that have been borrowed into Arabic are *al-ittiḥād quwwah* (Union is strength) and *al-wiqāyah khayr min al-'ilāj* (Prevention is better than cure).

While both idioms and collocations are phrasal, they differ in that idioms are non-compositional, i.e., the meaning of the idiom as a whole is not composed of the individual meanings of its constituent words, which is not the case with collocations. In addition, idioms are figurative while collocations are literal, though the connection between the figurative image and the idiomatic meaning cannot be always established. For example, in an idiom like *to reinvent the wheel*, whose Arabic calque is *'a'āda ikhtirā' al-'ajalah*, the relationship between the figurative and idiomatic interpretation can be easily established. In an idiom like *a red line*, which has been calqued into Arabic as *khatt 'ahmar*, this connection is not equally clear. Phrasal PUs can be classified according to the type of phrase to which they belong. Table 1 provides examples of calqued PUs from each type together with their English origins.

**Table 1**

*Syntactic classes of calqued MSA PUs*

Type of phrase	Arabic PU	English origin
Nominal PU	<i>khatt 'aḥmar</i>	a red line
Verbal PU	<i>la 'iba dawrā</i>	to play a role
Adjectival PU	<i>khāli al-dasam</i>	fat-free
Adverbial PU	<i>yawman mā</i>	some day
Prepositional PU	<i>bi-dam bārid</i>	in cold blood

**2.2. Calque**

In general terms, calque is a kind of borrowing based on translation. It can apply at various levels of language, such as morphology, lexicology, or syntax, but the term is most commonly used with reference to the phraseological level. In this sense, calque is a direct translation procedure (as the term is used by Vinay and Darbelnet, 1995) in which the constituent words of a source language PU are translated by their default equivalents in the target language. In this sense, calque is one of the most productive translation procedures in the English-into-Arabic direction. It can also be regarded as a way of enriching Arabic phraseology by the introduction of new metaphors and, accordingly, new concepts that Arabic speakers can distinguish and encode

linguistically. As described by Capuz (1997, p. 88), calques are reproductions of foreign “lexical complexes” that add new lexical units to the borrowing language.

There are three main ways in which English PUs can enter MSA: full calque (giving each constituent in the phrase its default translation), partial calque (translating one or more constituents with the default equivalent but translating the other(s) more freely), and non-calque (translating the whole phrase freely, i.e., assigning Arabic words to the concept rather than its English wording). Examples of each type are shown in Table 2. It is observed that the Arabic PU for *skyscraper* does not include the word *as-samā'* (sky, heavens), probably because of its sacredness in the Arab culture.

**Table 2**

*Degrees of calque*

Degree of calque	Arabic PU	English origin
Full calque	<i>ka'b 'Akhīl</i>	Achilles' heel
Partial calque	<i>talawwuth sam 'ī</i> (lit. ‘hearing pollution’)	noise pollution
Non-calque	<i>nāṭihatu saḥāb</i> (lit. ‘cloud head-butter’)	skyscraper

### 2.3. Lexical variation

Variation, as a general linguistic term, refers to “differences in linguistic form with no apparent change in meaning or function” (Walker, 2013, p. 440). Such differences exist at various levels of language, from phonology to grammar. In phraseology, Naciscione (2010, p. 60) distinguishes three senses in which the term “variation” is employed: parallel variation in the base form of the PU (i.e., lexical variation), grammatical change, and creative variation for stylistic effect. In the case of calqued PUs, it is sometimes hard to decide what the base form is. Therefore, for the purposes of the present study, lexical variation means the presence of different lexemes in different calques of the same PU which are all used to express the same meaning or perform the same function.

A distinction is made here between inflectional and derivational modifications in the form of words in PUs. Inflectional modifications are not regarded as cases of lexical variation, since they are required for the appropriate grammatical usage in actual texts, which is typically different from the canonical, or dictionary form of the PU. This applies in particular to verbal and nominal PUs, such as idioms and collocations. In Arabic verbal PUs, the verb must be realized as either past or present if it is used in the indicative mood. The verb form also indicates the number, person, and gender of the subject. Such forms differ from the basic or dictionary form of the PU, in which the past tense, the masculine gender, and third person singular are normally used. Similarly, the forms of nominal and adjectival PUs can change in terms of gender, number, and case, while their canonical form is

normally masculine, singular, and nominative. For instance, changing *khaff 'ahmar* (a red line) into *khuṭūṭ hamrā'* (red lines) does not represent lexical variation, since the meaning of the PU is not tied to singularity. Different inflectional forms of verbs, nouns, and adjectives cannot be regarded as cases of PU variation (as assumed, for example, by Alqahtani, 2016, pp. 186, 188). Shifts related to grammatical categories, such as number, gender, person, tense, and definiteness are not regarded as cases of lexical variation unless the category in question is an integral part of the meaning of the PU.

Conversely, derivational changes, with the exception of nominalization, are regarded as cases of lexical variation, since they result in the creation of new words. For instance, *'akhadha khuṭwah* and *ittakhadha khuṭwah* (both meaning ‘to take a step’) are variants of the same PU, since *'akhadha* and *ittakhadha* are two different lexemes having the same root. These two variants are common in MSA calques of some other English collocations involving the verb *take*, such as *take a measure*, *take a decision*, or *take precautions*. This, however, is not possible with original, non-calqued Arabic PUs, such as *'akhadha ḥaqqah* (to receive one’s dues) or *ittakhadha ṣadīqā* (to have (someone) as a friend), in which the forms *'akhadha* and *ittakhadha* cannot be used interchangeably. This indicates that certain types of lexical variation are specific to calques and do not apply to original PUs.

Though a kind of derivation, nominalization is not regarded here as a type of lexical variation, since it is possible to derive a verbal noun (*maṣdar*) from any verbal PU in Arabic through morphological rules without creating a different semantic concept. For instance, a

verbal PU like *ḍaraba taḥta al-ḥizām* (to beat below the belt) can be nominalized as *al-ḍarb taḥta al-ḥizām* (beating below the belt), which is a change in the class of the PU as a whole rather than its lexical components. The relationship between the two forms is not paradigmatic since there is no context in which they can be used interchangeably.

### 3. Review of the Literature

Variation in PUs has been tackled from different perspectives in phraseology and other language-related disciplines, including generative grammar, semantics, cognitive linguistics, lexicography, computational linguistics, and machine translation. Each perspective has its own concerns and, if applicable, provides different explanations for variation. This section reviews the main approaches to PU variation adopted in these various disciplines as well as the objectives and/or explanations associated with each approach. This is not to imply that each discipline is totally separate from the others or that the present review covers all the ideas in each discipline. The aim here is to cast light on the dominant approach in each discipline and the explanations, if any, it suggests for variation.

#### 3.1. Semantics

Semantics is generally concerned with establishing correlations between form and meaning in language. In dealing with PUs, semanticists are mainly concerned with compositionality; they look into the structure of PUs and try to find in what way their constituents contribute to their overall meanings. According to Cruse (2011, p. 83), a phrase is compositional when “each constituent of the phrase carries an identifiable constituent of the meaning as a whole.” Central to this is the

idea of semantic constituency; for a word to be a semantic constituent, it must be substitutable with other words from the same class (to produce different meanings) and reproduceable in different texts with the same meaning. These two conditions are met in free phrases and in many PUs, though not in idioms, which are non-compositional, since their “grammatical constituents are not semantic constituents” (Cruse, 2011, p. 86). Any lexical variation in an idiom will thus result in the loss of its idiomatic meaning. Cruse (2011, p. 87–89) lists a number of linguistic properties of idioms, ranging from phonological to syntactic, which affirm their invariability. Relevant to the purposes of the present study are those related to substitution and modification. First, the constituents of an idiom cannot be substituted by synonyms. Second, an idiom may be modified as a whole, but no single constituent can be modified without damaging the idiomatic meaning.

Even within the confines of Cruse’s definition of idiom, the first condition does not apply to calqued idioms, which vary lexically and still keep their meaning (as shown in Section 4.1 below). In addition, it has been suggested that substitutability with a synonym is possible in English without damaging the idiomatic meaning. For instance, Zhu and Fellbaum (2015, p. 340) show that there can be variants of non-compositional idioms by synonym substitution, such as *vacant threat* as an alternative for *empty threat*. The second condition is also challenged by corpus evidence. For example, Wulff (2008, p. 37) observes that “(parts of) many idiomatic constructions can be modified by adjectives or relative clauses.” This means that idioms can be modified both externally, as whole units, and internally.

Moon (2015) also provides corpus examples of variation on the basic forms of many idioms, such as “hitting the political nail right on the head” (p. 325) as a variation on *to hit the nail on the head*. This involves modification of the noun *nail* and the prepositional phrase *on the head*, in addition to the use of the verbal noun, without changing the meaning of the idiom.

Many of the restrictions laid down by Cruse regarding idiom variability and shared by some generative linguists seem to focus on the idealised form of language rather than actual use and give precedence to logic over the realities of language (e.g., since the elements of an idiom lack semantic constituency, they cannot be substituted by a synonym or separately modified). In his explanation of compositionality, Cruse refers to the idea that the human mind cannot possibly store the infinite number of the grammatical structures that can be produced in language and that such structures are created by “rule-governed combinations of elements from a finite list according to generative rules” (2011, p. 84)—an account that clearly echoes a Chomskyan concept of language.

According to other scholars, variability cannot be explained in terms of compositionality. For instance, Zhu and Fellbaum (2015, p. 339) argue that “semantic compositionality and variation are in fact independent of each other,” preferring to attribute lexical variation in idioms to their “semantic transparency.” According to Langlotz (2006, p. 289), “systematic idiom-variation is a reflex of idiom-transparency,” which he understands as a cognitive concept composed of various factors and related to the way the speaker establishes a relation

between form and idiomatic meaning. The same idea is taken up by Abdou (2012), who applies Langlotz’s concept to Arabic idioms and who attributes many cases of idiom variation in Arabic to “the transparency of the underlying motivation of the idioms” (p. 108). In a similar vein, Heinonen (2014, p. 899) observes that “as long as the phraseological unit is recognized, its meaning can be modified according to the contribution the substitute part carries along.”

However, transparency is not in itself the reason why a particular variant appears in the language. It is less of an explanation than a factor that can facilitate the production, recognition, and currency of a given variant. If two or more variants of the same idiom are transparent, there is still need to explain why they are different even though they have the same idiomatic meaning and are used for similar situations, which is what the present paper attempts to do.

### 3.2. Generative grammar

For generative linguists, a number of problems are posed by idioms, which lie at the core of phraseology. The main problem has to do with their non-compositionality; applying the principle of compositionality to idioms would fail to account for their meaning, with the result that idioms have been generally neglected as peripheral to the linguistic system. This is one reason why “phraseology has generally played a rather limited role in the development of the various versions of generative grammar” (Gries, 2008, p. 10). Another problem has to do with the ungrammaticality of some idioms, such as *by and large*, which Cruse (2011, p. 90) calls “asyntactic idioms” or “grammatical idioms” (i.e., expressions that are peculiar

in terms of grammar), since the rules devised for the generation of ordinary grammatical structures do not apply to them. In this respect, it is sometimes argued that the vast majority of idioms follow the canonical grammatical rules of the language and that the cases that violate such rules are mere exceptions (e.g., Culicover, Jackendoff, & Audring, 2017, p. 554). However, the dominant view is that idioms are devoid of internal semantic structure and are stored in the mental lexicon as single units, in the same way as words. The variability of PUs poses a further problem to the generative model. The grammar needs to explain how a PU, which is treated as a single unit, can display lexical or grammatical variation. Variation also involves cases of manipulation and creative use, which have been dismissed by generative grammarians as instances of wordplay that need not be the concern of generative theory. As Langlotz (2006, p. 216) observes, “in generative frameworks, idiomatic wordplay has generally been excluded from theoretical description and disclaimed as a matter of performance.” The grammar needs to explain how a PU retains its meaning in spite of such variation.

Culicover et al. (2017, p. 553) admit that “the rich profusion of MWCs [multiword constructions] in English has been taken to provide evidence against derivational approaches to grammar,” and that many such constructions “cannot be accounted for by a straightforward compositional” approach. The examples they provide include idioms, whose interpretation is extraordinary and unpredictable, and collocations, which have predictable interpretation but display restricted and idiosyncratic arrangement of

words. Such constructions challenge the generative view that linguistic competence consists of knowledge of a set of syntactic rules and a mental lexicon in which words are stored, which, together, are responsible for the production of novel sentences. Given the large number of multiword constructions in language, Culicover et al. (2017) suggest an alternative approach in which entire multiword constructions are stored in the “extended lexicon” (p. 553) and are subject to the rules of grammar. Culicover et al. (2017, p. 565) suggest that PUs and morphologically complex words are probably acquired in the same way; just as complex words are first acquired as “unanalyzed units” and then their morphological structure is identified, PUs are acquired as single units and at a later stage their internal syntactic structure is recognized. This can help explain variations in the canonical forms of PUs at a later stage following their acquisition. The main problem with the generative model, however, lies in its account of the semantic idiosyncrasy of idioms rather than the generation of their syntactic structure.

### 3.3. Cognitive linguistics

In cognitive linguistics, the study of PU variation has generally focused on such areas as how PUs are mentally processed and whether variations in PUs can hinder their processing or prolong its time. Many cognitive accounts of variability suggest that PUs are not processed as single units, but rather allow for some internal change in form without affecting their meaning. An example is Naciscione (2010), who deals with unique stylistic uses of PUs from the speaker’s perspective, with reference to such mental processes as access, recall, and long- and short-term memory. She also deals with

PU's from the hearer's perspective, discussing how they are perceived, recognized, understood, and interpreted.

One of the experimental methods of cognitive linguistics is eye-tracking, in which special computer software is used to measure a participant's eye movement and fixation to recognize the time used for processing linguistic expressions. This method has been used by Geeraert, Baayen, and Newman (2017) to explore how different types of idiom variation, namely, lexical and syntactic modification as well as truncation of idioms, are processed and whether their processing is different from that of the canonical forms of idioms. In their experiment, Geeraert et al. (2017) do not use authentic variants of idioms, but rather manipulate idioms by replacing one of their constituents by a synonym, even though the resulting form may not be in actual use. The same applies to truncation, where one or more words are omitted from the idiom, which is also different from how PUs are truncated in actual use. Actual truncation of PUs is based on shared knowledge of the full PU and normally occurs with proverbs, where using only part can trigger recognition of the whole. The forms used by Geeraert et al. (2017) may be sufficient for their research purposes, e.g., to see the impact of such variations on the processing of PUs, but they remain unnatural or at least hypothetical forms that may or may not occur in actual language use.

According to Geeraert et al., the time of processing lexical variation is not significantly longer than that taken for processing the canonical form, which means that "altering a word within an idiom to a synonymous or a non-synonymous word does not result in a processing cost" (2017, p. 87). They also

show that truncated forms take shorter time than full forms, while the insertion of modifiers increases the processing time, both results being explained in terms of the relative length of the expression used. Such results question the view that idioms are stored as single units in the mental lexicon, since in that case, it would have to include not only the canonical form of the idiom but also all its possible variants.

### 3.4. Lexicography

Variability is also important for lexicography since the dictionary is generally viewed as a record of language and since the stance taken by most dictionaries is currently more descriptive than prescriptive. Lexicographers, therefore, seek to collect and record the various forms that a PU can take. As a result of the use of corpora in modern lexicography, a great deal of attention is currently paid to PU variants and the way they should be presented in general-purpose and phraseological dictionaries. For instance, in the Language Notes insert of the fourth edition of the *Longman Dictionary of Contemporary English* (LDOCE), which is based on the 300 million-word Longman Corpus Network, it is recognized that there are "variable idioms" that result from speakers' playing with idioms and creating new versions of them (LDOCE, p. 977). The example given in LDOCE is the idiom "to drive someone crazy," in which the word *crazy* can be replaced by *mad*, *nuts*, *bananas*, and *up the wall*. In LDOCE (976–977), the term "idiom" is used in the broadest sense, as it includes binomials (e.g., *high and dry*), similes (e.g., *as easy as a pie*), sayings, which can be reduced (e.g., *Too many cooks (spoil the broth)*), phrasal verbs (e.g., *chill out*), and metaphorical expressions representing feelings or types

of behaviour (e.g., *throw the baby out with the water*).

General-purpose dictionaries are not normally concerned with explaining PU variants, even though they may be listed in the dictionary. Lexicography, in the sense of dictionary-making, is more descriptive than explanatory, and, in the case of learner's dictionaries, it is pedagogical and even prescriptive. A lexicographer's job is mainly to define the meaning of the items of vocabulary of a language, not to explain why they take the forms they have. Some phraseological dictionaries (e.g., Ammer, 2013) mention the origin of PUs, especially idioms and proverbs, and list their variant forms, if any. Ideally, if there are established variants of a given PU, they should be listed in the dictionary. Dictionaries, however, cannot list variations resulting from word-play or creative use of PUs.

As far as Arabic lexicography is concerned, Jarad and Abu-Ssaydeh (2017) note that many aspects of Arabic-English dictionaries are in need of development, including their coverage of borrowed idioms and their treatment of idiom variation. They attribute this lack of coverage to failure to update these dictionaries, which were published before many borrowings had been established in Arabic. In their view, "variations that have acquired a canonical status must be recognized and recorded in the dictionary while incidental ones can be left out" (Jarad & Abu-Ssaydeh, 2017, p. 24). This implies that variations are not changes introduced into the canonical form, but are themselves canonical forms, judging by their frequency in the language. One problem with Jarad and Abu-Ssaydeh's study is the overlapping and inconsistency in the classification of idiom variation. In

the order listed in their study, idiom variation involves: the verb, the noun, synonyms, adjectives, addition, "non-synonymous words," "different word classes," "syntactic structure," and "borrowed idioms" (Jarad & Abu-Ssaydeh, 2017, p. 16). This involves a mixture of criteria, based on semantics, word class, syntactic structure, and origin of expression. Reference to the part of speech is also redundant, since the variant word is necessarily either synonymous or non-synonymous. In addition, it is not clear how borrowed idioms can fit in this classification, since they do not constitute a type of variation in themselves. Another problem with Jarad and Abu-Ssaydeh's study is that one of the dictionaries examined was compiled by one of the two researchers, which might impact the objectivity of data selection and procedures.

### 3.5. Computational linguistics and machine translation

From the point of view of computational linguistics and machine translation (MT), variability is a feature that adds to the complications that arise when dealing with PUs. One problem has to do with identifying PUs in natural language processing (NLP) systems, which is not an easy task, given the non-compositionality of certain PU types, such as idioms. One common method in computational linguistics is to regard such PU types as single units (e.g. Ramisch 2015). While this is helpful in many cases (e.g. *to give the green light* means 'to allow'), many PUs do not lend themselves easily to such analysis (e.g. *to cry over the spilt milk*, for which there is no single-unit substitute). In addition, any manipulation, insertion, omission, or replacement of PUs constituents can cause problems in the

automatic identification of the idiom. Ramisch (2015, p. 37) argues that “it is not possible to replace part of a MWE [multi-word expression] by a related (synonym/equivalent) word or construction.” This view is not supported by recent corpus-based investigations of PUs, as shown in Section 4.1 above.

PUs, even in their canonical form, have also constituted a problematic area in MT. Liu and Zhang (2015) note that word-based MT models, in which the word is regarded as the unit of translation, can fail to provide satisfactory translations of PUs since the alignment in such models occurs at the word level. An alternative is the phrase-based model, which aligns sequences of words rather than single words, and is thus more capable of dealing with idioms and other PUs. As Liu and Zhang (2015) observe, one main advantage of phrase-based MT models has to do with their ability to deal with non-compositional structures, by capturing their translations as whole units rather than “reconstructing them word by word awkwardly” (p. 207). The problem becomes worse when there is variation in the forms of PUs, especially in cases where such variation involves omission or insertion of words within the PU, thus separating its constituents from each other. Generally, however, MT studies are not concerned with providing explanations for variation in PUs, but rather with recognizing them in source texts and providing their phraseological equivalents in the target language.

In the English-Arabic language pair, problems of dealing with variation in calqued PUs in MT partly depend on the direction of translation. If the direction is from English into Arabic, the translation is more or less straightforward since it is

often the case that a single equivalent will be selected for the source expression. However, if the direction is from Arabic into English, there is need to recognize variants of calqued PUs, back-translate them into their source origin, and store them as synonymous expressions with a single target equivalent, provided that they are not stylistic or creative variations and that they do not reflect variation in their English original forms.

#### 4. A Taxonomy of Lexical Variation in Calqued PUs

This section provides a taxonomy of the types of lexical variation identified in the data collected. Investigation of the data indicates that there are two main types of lexical variation: paradigmatic, which is more dominant, and syntagmatic. In paradigmatic variation, two or more variants that differ with respect to one word or phrase are used in the same position of the PU. Syntagmatic variation refers to cases in which one word or phrase is added to the PU. Each major type includes subtypes, as shown in Figure 1.

##### 4.1. Paradigmatic variation

One common type of paradigmatic variation results from the use of synonyms in the variant forms. These can be independent Arabic words, as in (1), derivative forms of the same word, as in (2), or a borrowing and an Arabic translation in the variant PUs, as in (3). This type of variation may affect more than one word in the PU, resulting in what Heinonen (2014, p. 70) calls “covariance.” An example is (1f), in which both constituents of the PU can vary, yielding four possible calques of the English *national day*. In the examples below, the English word for which there are Arabic variants is underlined.

- (1) a. ka‘b/‘aqib ‘Akhīl (Achilles’ heel)
- b. ‘aghlaqa/’awṣada al-bāb fī wajhi *fulān/shay’* (to close the door in the face of somebody/something)
- c. sukhriyat al-qadar/mufāraqat al-qadar (irony of fate)
- d. ṭabaq/ṣaḥn ṭā’ir (flying saucer)
- e. lā fā’idah min al-bukā’ ‘alā al-laban/al-ḥalīb al-maskūb (It’s no use crying over spilt milk.)
- f. al-‘īd al-qawmī/al-‘īd al-waṭanī/al-yawm al-qawmī/al-yawm al-waṭanī (national day)

Another common type of paradigmatic variation involves the use of different derivational forms of the same word in the PU variants, which can be a noun (2a), a verb (2b), or an adjective (2c). In some cases, this kind of variation may involve a change of the part of speech, which may affect the morphology of the accompanying word; in examples (2d) and (2e), the variants are a noun and an adjective derived from that noun with relative adjective suffixation (*yā’ al-nasab*), which requires the use of a construct phrase (*‘idāfah*) in the first case and a noun plus adjective in the latter. This applies to both indefinite (2d) and definite (2e) PUs.

- (2) a. nazf/nazīf/istinzāf al-‘uqūl (brain drain)
- b. ḥaraqa/’aḥraqa qawāribah (to burn one’s boats)

- c. maṣādir mukhābarātiyyah/istikhbārātiyyah (intelligence sources)
- d. nashrat ‘akhbār/nashrah ‘ikhbāriyyah (news bulletin)
- e. ghurfat al-tijārah/al-ghurfah al-tijāriyyah (the chamber of commerce)

In the third type of paradigmatic variation, an Arabic word and a loanword are used in the variant forms, with two procedures, borrowing and calque, being used in the latter case. The words in the paradigmatic relation can be nouns (3a–c) or adjectives (3d and 3e). In (3), the Arabic word is mentioned first and the loanword second. Though the register in which such PUs are likely to occur is generally formal, the PU with the Arabic variant is regarded as purer than the one with the loanword.

3. a. ‘aqīdah/’aydiyūlūjiyyah siyāsiyyah (political ideology)
- b. mufāwaḍāt al-mā’idah/al-ṭāwilah al-mustadīrah (round-table negotiations)
- c. wathīqat/būlīṣat ta’min (insurance policy)
- d. istikhshāfāt niḩṩiyyah/bitrūliyyah (petroleum explorations)
- e. mukālamah hātifiyyah/tilifūniyyah (telephone call)

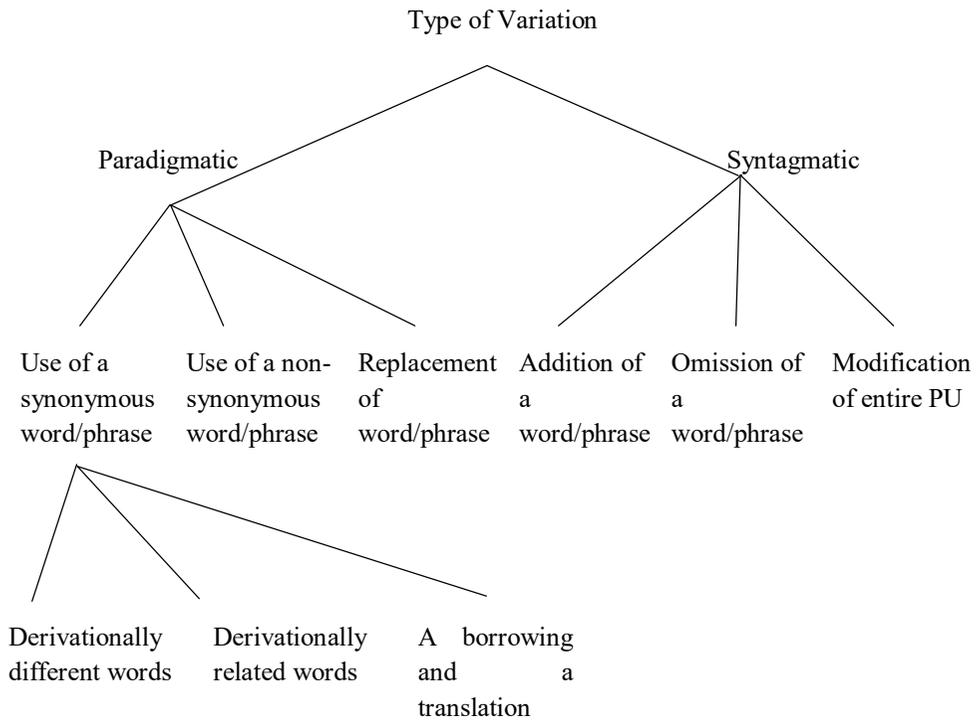


Figure 1. A Taxonomy of Lexical Variation in MSA PUs Calqued from English

In some cases, the synonymous forms are phrasal (in the grammatical sense), as shown in (4), where the variant forms are an adjective and a prepositional phrase. The phrasal form is closer to the English source in (4a), while the adjectival form is closer to the English in (4b). In (4c) and (4d), there are three Arabic variants of the English PU. The first variant uses the Arabic word *nafs* (same) as a head noun, though the second (and less common) form, in which the word *nafs* functions as appositive, is the correct one from a purist view. The third form in each case is a partial calque, equivalent to ‘in one boat’ and ‘to speak one language’, which in Arabic are synonymous with ‘in the same boat’ and ‘to speak the same language’.

- (4) a. 'a'ṣāb fūlādhiyyah/min fūlādh  
(nerves of steel)
- b. qabḍah ḥadīdiyyah/min ḥadīd (an iron grip)
- c. fī nafs al-qārib/fī al-qārib nafsih/fī qārib wāḥid (in the same boat)
- d. taḥaddathā nafs al-lughah/al-lughah nafsahā/lughah wāḥidah (to speak the same language)

The variant forms need not be synonymous, as shown in (5). This normally involves a full and a partial calque. In (5a), the first variant is a full calque of the English *password*, while the second, *kalimat al-sirr* (secret word), is a partial calque in which the word *pass* in the English compound is translated freely rather than literally. Similarly, in the first variant in (5b), the literal translation of the word *key* is used in the full calque, while a free translation *ra'ṭsiyyah* (major/main) is used in the partial calque. The same idea applies to the other examples, where the variant words in paradigmatic relation are not synonyms. Thus, we have *nizām/majmū'ah* (system/group) in (5c),

*naghamah/lahjah* (tune/dialect) in (5d), and *wathā'iqī/tasjīlī* (documentary/recording facts), in (5e).

- (5) a. kalimat al-murūr/al-sirr (password)
- b. kalimah miftāḥiyyah/kalimah ra'ṭsiyyah (keyword)
- c. al-nizām al-shamsī/al-majmū'ah al-shamsiyyah (the solar system)
- d. ghayyara naghamatah/lahjatah (to change one's tune)
- e. film wathā'iqī/tasjīlī (a documentary film)

Unlike the examples above, the examples of paradigmatic variation in (6) result from individual, creative uses of the PU and do not represent institutionalized usage. A word or phrase in the calque is replaced by another word or phrase to express a meaning based on the original calque but serving a new idea intended by the writer. In (6a), there is manipulation of the calque of the proverb *Don't put all your eggs in one basket* by replacing the phrase *all your eggs* with *Iraq and Afghanistan*. In (6b), there is play on the calque of *All is fair in love and war* to become *All is fair in the media*. In (6c), there is manipulation of the Arabic version of *All Quiet on the Western Front*, a novel title that has caught on among Arab writers and has appeared with many variants, replacing *the Western Front* with other phrases that serve their writers' purposes. In (6d), the word *al-marīḍ* (a sick man) is used in place of *al-gharīq* (a drowning man) in the Arabic calque of the English proverb *A drowning man will clutch at a straw*.

- (6) a. Irān waḍa'at al- 'Iraq wa 'Afgḥānistān fī sallah wāḥidah (Iran put *Iraq and Afghanistan* in one basket)

- b. kull shay' mubāḥ fi al-'i'lām (All is fair in *the media*)
- c. kull shay' hādi' fi Sūrya (All quiet in *Syria*)
- d. *al-marīd* yata'allaq bi-qashshah (A *sick man* will clutch at a straw)

- c. al-'insān laysa jazīrah/ al-'insān laysa jazīrah *mun 'azilah* (no man is an island/no man is an *isolated* island)

#### 4.2. Syntagmatic variation

As for syntagmatic variation, the first type involves the addition of a word or phrase to a calqued PU. In some cases, one variant is a full calque while the other has an additional word in the structure of the PU that helps clarify its meaning or make it sound more natural in Arabic. For instance, in (7a) the word *zimām* (halter) naturally collocates with the verb *'akhadha* (take) in Arabic and is therefore sometimes inserted in the PU to make it more acceptable. Variation here can be explained with reference to the tendency towards more naturalness of expression. In (7b), the insertion of the word *mustawā* (level), which collocates with the preposition *fawqa* (above), helps the creation of a semantically more acceptable version of the PU. The word *mun 'azilah* (isolated) in (7c) clarifies the meaning by providing the point of the island metaphor. This example also involves negation of opposites, a kind of modulation as the term is used by Vinay and Darbelnet (1995, p. 252); instead of *no man is*, the Arabic form in both variants is *al-'insān laysa* (*man is not*). The additional words in (7) are italicized in both English and Arabic.

- (7) a. 'akhadha al-mubādarah/'akhadha *bi-zimām* al-mubādarah (to take the initiative/to take *the halter* of the initiative)
- b. fawqa al-shubuhāt/fawqa *mustawā* al-shubuhāt (above suspicion/above *the level* of suspicion)

A different type of syntagmatic variation involves truncation of the PU. This usually occurs with proverbs and can be explained in terms of the assumption of common knowledge of the full form between the writer and the recipient. An example is (8), which is based on the proverb *Those who live in glass houses shouldn't throw stones*:

- (8) yuqaddimu lahum durūsā fi hīn 'anna baytahū min zujāj (He is giving them lessons while his house is of glass)

Another type of syntagmatic variation involves the modification of an entire PU by an additional word or phrase. In (9a), the adjective *muktamilah* (complete) is used to modify the whole calque of *a lame duck*. This is probably the only kind of modification recognized by generative linguists (e.g., Culicover et al., 2017) and semanticists who subscribe to generativist views (e.g., Cruse, 2011). In (9b), the word *nafaq* (tunnel) is used both literally and metaphorically, resulting in syntagmatic wordplay. The sentence is the title of an essay about an incident in which three girls were harassed by a group of youths while walking near the Nahdah Tunnel in Riyadh. The word *nahdah* (awakening) is used both literally and metaphorically in this context, and the whole expression indicates that awakening (represented here by the right of women to walk safely in public places) is a hard process that is faced with grave threats. The literalization of the word *nafaq* (tunnel) and the metaphorization of the word *nahdah* (awakening) are examples of the creative use of PUs for stylistic purposes.

- (9) a. qabla 'an yatahawwala 'ilā baṭṭah  
'arjā' *muktamilah* (before he turns into  
a *complete* lame duck)
- b. hal min ḍaw' fī nihāyat nafaq *al-Naḍah*? (Is there light at the end of the  
*Nahda* Tunnel?)

### 5. Explanation of Variation

The fact that calqued PUs are translations serves as a basis for explaining many types of variation. The act of translation in itself provides a fertile ground for variability, since there is no one single translation that everyone will produce or accept. If in original PUs it is not normally possible to replace a word with a synonym, in calqued PUs this is quite common. This explains the use of *ka'b* and *'aqib* (both meaning 'heel') in the Arabic calque of *Achilles' heel*, and *'aghlaqa* and *'awṣada* (both meaning 'to close') in the Arabic calque of *to close the door in the face of somebody/something*. Different translators can produce different renderings that can all be regarded as equivalent to the source PUs and that can all gain currency in the target language.

Variation caused by translation involves the use of synonyms or near-synonyms, including derivationally related words, such as *nazf/nazīf/istinzāf al-'uqūl* (brain drain), derivationally unrelated words, such as *ka'b/'aqib 'Akhīl* (*Achilles' heel*), and the use of a borrowing as opposed to the use of a pure Arabic equivalent, such as *'aqīdah/'aydiyūlūjīyah* *siyāsiyyah* (political ideology). All these are full calques and the variations displayed are due to different translation choices. Different translations also account for variation involving non-synonymous words in some calqued PUs, such as the two renderings of *the solar system*, namely, *al-niḡām al-shamsī* and *al-*

*majmū'ah al-shamsiyyah.*, the former being a full calque that is closer to the English original. Different translation choices can also account for syntagmatic variation in which a word is added in one of the PU variants to clarify its meaning or make the expression sound more natural in Arabic, as shown in (7) above.

Variation in calqued PUs can also be triggered by dialectal factors. Even within MSA, there are geographical variations which are manifested at the phraseological level. For instance, the variants in *ṭabaq ṭā'ir/ṣaḥn ṭā'ir* (flying saucer) are not only contrasted on the basis of synonymy, but also on the basis of geographical variation. The words *ṭabaq* and *ṣaḥn* (both more or less equivalent to *plate*) are not used equally frequently in the same geographical variety; the former is mostly used in Egyptian Arabic while the latter is more common in the Gulf and Levantine varieties. Words common among the speakers of a given dialect may find their way to their standard language usage. The same applies to *mā'idah* and *ṭāwilah* (both meaning 'table'), which appear in the Arabic calques of *round table meeting/discussion* and *negotiation table*. The word *mā'idah* is more common in Egypt while *ṭāwilah* is more common in the Gulf and the Levant. However, as is the case with many geographical variations, there are no sharp lines dividing the use of variant forms. While the corpus used can be searched by region, regions are determined by publications, not by individual writers. An essay by an Egyptian writer may appear in a Kuwaiti magazine and a statement by a Palestinian politician may be quoted in an Egyptian newspaper, which may affect the results of geographic search.

Variation in some PUs can be attributed to variation in the source PUs

from which they are borrowed. For instance, *haraqā qawāribah* and *haraqā jusūrah* are calqued from *to burn one's boats* and *to burn one's bridges*, respectively. Since they share the same idiomatic meaning and figurative idea, they are regarded as variants of the same PU. In some cases, the variant forms are calqued from PUs in different languages. The pragmatic PU *al-sayyidāt wa al-sādah* (Ladies and gentlemen), a salutation that is used in formal speeches, has two other variants in MSA, *sayyidātī, sādātī* (My ladies, my gentlemen) and *sayyidātī, 'ānisātī, sādātī* (My ladies, my misses, my gentlemen). The first variant is the closest to the English source, the addition of the definite *al-* being obligatory in this kind of vocative in which there is no particle like *yā*. Judging by their morphological structure, the second and third variants are probably borrowed from the French *Mesdames et Messieurs* and *Mesdames, Mesdemoiselles et Messieurs*, respectively. The third variant includes the word *'ānisātī* (my misses), which is used to refer to women who are not yet married. All three variants follow the English or French ordering, thus abiding by the Western rule of etiquette *Ladies first*, which has also been calqued into Arabic as *al-sayyidāt 'awwalā*.

Ideology also plays a role in variation in calqued PUs. Different political and religious ideologies in the Arab world can result in different forms of the same PU. One example is the political attitude towards *qawmiyyah* (nationalism) in different Arab countries, which is reflected in two different calques of expressions like *national heritage*, *national interest*, or *national day*. In one version, the adjective *national* in such expressions is consistently translated as *qawmī* (national), while in another version it is translated as *waṭanī* (related to homeland,

or *waṭan*), leading to two systematically opposing calques of the same expression, such as *al-turāth al-qawmī* and *al-turāth al-waṭanī* for the English *national heritage*. There are historical reasons for this political difference. After the independence of most Arab countries in the mid-twentieth century, some countries, notably Nasserist Egypt, adopted a political ideology calling for Arab nationalism. This ideology was rejected by some other countries, especially in the Gulf region, on the grounds that it is opposed to the tenets of Islam, which calls for unity on the basis of faith rather than land, race, or origin (see, e.g., Bayyumi, 2002). While in actual practice such countries follow the national state model, rejection of nationalism has continued at the level of language. Any word related to *qawmiyyah* (nationalism) is avoided in official statements and in the media, and the standard translation for *national* in such countries is *waṭanī* rather than *qawmī*. While the word *waṭanī* means 'related to homeland,' it can also mean 'patriotic,' and thus, unlike the controversial concept of *qawmiyyah*, *waṭaniyyah* evokes meanings of patriotism and love of one's country.

The Arabic variants of the PU 'national day' display the effect of religious as well as political ideologies. In addition to the *qawmī/waṭanī* contrast, the words *yawm* (day) and *'īd* (religious feast, or Eid) are used as translations of the word *day* in this collocation. Such covariance results in four different calques of the same PU: *al-yawm al-qawmī*, *al-'īd al-qawmī*, *al-yawm al-waṭanī*, and *al-'īd al-waṭanī*. The word *'īd* is sometimes rejected on the grounds that it is associated with religious feasts. According to a strict religious view, there are only two Eids in Islam and it is forbidden to call any other occasion Eid as this represents innovation in matters of

religion. Actual practice, however, is more moderate as it accepts extending the meaning of the word Eid to non-religious festivities that are celebrated every year, such as the national day, Mother's Day, and birthday. It is interesting that while the foreign concept is adopted in the Arab culture, it is given different labels that reflect different religious views, either by changing its wording or by justifying it as conforming to the principles of religion, though, in back translation, both variants will have a single form.

Perhaps more revealing is the calque of *irony of fate*, an expression used to refer to acts "of malice or mischief by fate," which is depicted as a person mocking people's normal expectations (Ruiz, 2009, p. 39). The idea of fate as a malicious, mischievous person mocking people is clearly at odds with Islamic principles, which advocate respect for and acceptance of fate and Divine Decree, whether good or bad. Moreover, this PU has its origin in Greek mythology, being related "to the idea that gods (or the Fates) are amusing themselves by playing with the minds of the mortals, with deliberate ironic intent" (Ruiz, 2009, p. 40)—an idea which is also clearly contrary to the tenets of monotheist religions. However, the expression *irony of fate* appears in Arabic in two versions: *sukhriyat al-qadar* (literally, 'irony of fate') and *mufāraqat al-qadar* (literally, 'incongruity of fate'), both used idiomatically in the same sense as the English PU. The former expression, which is closer to the English original, is sometimes rejected on the grounds that it constitutes disrespect for Divine Decree by depicting it as deriving amusement from mocking people. According to one *fatwa* (<https://www.binbaz.org.sa/fatawa/4218>), using this expression amounts to "disbelief, misguidance, and derision of God's decree" (my translation). Another

*fatwa* argues that this expression "indicates discontent with Divine Decree" and "implies attributing matters to [sic] irony to fate" (<http://www.islamweb.net/emainpage/index.php?page=showfatwa&Option=FatwaId&Id=26137>). The expression *mufāraqat al-qadar* (literally, 'incongruity of fate') is free from such implications and is therefore sometimes used to express the same meaning. However, corpus investigation shows that the more controversial PU is much more common than the supposedly more acceptable one: *sukhriyat al-qadar* occurred 273 times in the corpus while *mufāraqat al-qadar* occurred only 41 times.

Cases of manipulation and creative use of PUs are based on the principle of shared knowledge, as such cases would be pointless if the recipient did not know the original PU manipulated by the addresser. This applies to both paradigmatic manipulation, represented by the replacement of a PU constituent by another that suits the purposes of the addresser, and syntagmatic manipulation, including truncation of proverbs and wise sayings and modification of entire PUs. Creative variability represents a state of reconciliation between the idiom principle and the open-choice principle. It shows that, both at the level of decoding and the level of encoding, PUs are not stored in the mental lexicon as single units that do not accept alteration or modification.

## 6. Conclusion and Implications

A number of conclusions can be drawn from the above account of variability in MSA PUs calqued from English. First, it has been shown that calqued PUs display patterns of lexical variation that are different from those displayed by original, non-calqued PUs. In addition to synonyms, derivational forms of various

types are commonly used interchangeably in calqued PUs, which is not the case with original PUs. Second, the study indicates that individual approaches to PU variation cannot provide satisfactory explanations for the types of variation displayed by calqued PUs. Each approach has its own objectives and focuses on certain aspects of variation, such as the relationship between form and meaning, the processing of variants, or the way they can be explained within the framework of formal theory. Investigation of variation in MSA calques, however, indicates that there it is caused by different linguistic and extra-linguistic factors. The fact that such PUs are translations from another language explains the presence of certain types of paradigmatic variation, including the use of synonyms, which may be derivationally related or unrelated or may be loanwords and their translations. Given the interdisciplinarity of phraseology, any attempt at explanation cannot be limited to purely linguistic matters, but will necessarily refer to extra-linguistic factors, including ideology and culture. Variations can be attributed to different political and religious ideologies and world views, which sometimes apply a form of censorship on the wording of the PU to make it more acceptable to speakers sharing a given ideology or world view. Reliance on authentic data reveals that idioms and proverbs are subject to frequent creative manipulation and wordplay that put them to new uses to produce certain rhetorical effects.

The results of the study also have both practical and theoretical implications. At the practical level, the variant forms need to be taken into account when compiling dictionaries of Arabic idioms, both in the head-phrases themselves and in the

citations illustrating their actual use of language. In addition, even within MSA, there are still geographical variations, as manifested here at the phraseological level. Any accurate account of MSA needs to take such dialectal variations into consideration. Similarly, variants that carry the same meaning need to be included in Arabic-English MT systems as synonyms if they are not to be rendered literally and thus lose their idiomatic meaning. On the other hand, while the distinction between the idiom principle and the open-choice principle set the theoretical foundation that raised researchers' awareness of the prevalence of idiomaticity in language and inspired much work on phraseology, the variability of many PUs blurs this distinction. This particularly applies to cases of manipulation and creative use of PUs, where both principles are in operation. Elsewhere (Al-Wahy, 2009), it has been suggested that Arabic calques from English may diverge semantically from their English sources, leading to idiomatic false friends. Here, it is shown that the opposite can also be true; borrowing set phrases can result in phraseological synonymy in the borrowing language, where phrases with different forms have the same meaning.

It is hoped that the findings of the study will be of value to those interested in variation in PUs in general and in calques in particular, irrespective of the languages they deal with or the approach they adopt. It is also hoped that the present study will encourage further research in other language pairs to see whether calqued PUs display similar or different patterns of lexical variation and whether there are other explanations for variability at the phraseological level.

### References

- Abdou, A. (2012). *Arabic idioms: A corpus-based study*. London: Routledge.
- Alqahtani, H. (2016). Fixedness and variation in the structure of idiomatic expressions in Modern Standard Arabic: An analysis of the Saudi newspaper *Al-Riyadh*. *Humanities and Social Sciences Review*, 6(1), 183–194.
- Al-Wahy, A. S. (2005). The impact of English phraseology on Modern Standard Arabic: Idioms as a case in point. *English Language and Literature Studies*, 4(1), 145–189.
- Al-Wahy, A. S. (2009). Idiomatic false friends in English and Modern Standard Arabic. *Babel*, 55(2), 101–123.
- Ammer, C. (2013). *The American Heritage Dictionary of Idioms* (2nd ed.). Boston: Houghton Mifflin Harcourt.
- Badawi, E., Carter, M. G., & Gully, A. (2004). *Modern Written Arabic: A comprehensive grammar*. London: Routledge.
- Bayyumi, Z. S. (2002). *Al-'Arab bayn al-qawmiyyah wa al-Islām* [The Arabs between nationalism and Islam]. Cairo: Dar al-Qāhirah.
- Bednarek, M. & Bublitz, W. (2007). *Enjoy!:* The (phraseological) culture of having fun. In Skandera, P. (Ed.), *Phraseology and culture in English* (pp. 109–135). Berlin: Mouton de Gruyter.
- Bragina, N. (2000). Phraseology and the context of culture. *International Journal of Lexicography*, 13(1), 29–34.
- Burger, H., Dobrovolskij, D., Kühn, P., & Norrick, N. R. (Eds.). (2007). *Phraseologie/Phraseology: Ein internationales Handbuch der zeitgenössischen Forschung/ An international handbook of contemporary research*. Vol. I. Berlin: Walter de Gruyter.
- Capuz, J. G. (1997). Towards a typological classification of linguistic borrowing (Illustrated with Anglicisms in Romance Languages). *Revista Alicantina de Estudios Ingleses*, 10, 81–94.
- Cruse, A. (2011). *Meaning in language: An introduction to semantics and pragmatics* (3rd ed.). Oxford: Oxford University Press.
- Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. *Topics in Cognitive Science*, 9, 552–568. doi:10.1111/tops.12255
- Dobrovolskij, D. (2015). Introduction. *International Journal of Lexicography*, 28(3), 275–278. doi.org/10.1093/ijl/ecv022
- Geeraert, K., Baayen, R. H., & Newman, J. (2017). Understanding idiomatic variation. *Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017)*, Valencia, Spain, April 4 (pp. 80–90). Retrieved from: <http://www.aclweb.org/anthology/W17-17>

- Gries, S. Th. (2008). Phraseology and linguistic theory: A brief survey. In Granger, S. & Meunier, F. (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 3–26). Amsterdam: John Benjamins.
- Heinonen, T. R. (2014). Lexical variation within phraseological units. In Abel, A., Vettori, C., & Ralli, N. (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19 July 2014, Bolzano/Bozen (pp. 893–904).
- Jaki, S. (2014). *Phraseological substitutions in newspaper headlines: "More than meets the eye."* Amsterdam: John Benjamins.
- Jarad, N. I. & Abu-Ssaydeh, A. (2017). Idioms in the Arabic-English dictionary. *International Journal of Arabic-English Studies*, 17, 7–28.
- Langlotz, A. (2006). *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. Amsterdam: John Benjamins.
- Liu, Y. & Min, Z. (2015). Statistical machine translation. In Sin-wai, C. (Ed.), *The Routledge encyclopedia of translation technology* (pp. 201–212). London: Routledge.
- Longman dictionary of contemporary English* (4th ed.). (2003). Harlow: Pearson.
- Mel'čuk, I. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology*, (3)1, 31–56. doi:10.1515/phras-2012-0003
- Moon, R. (2015). Idioms: A view from the Web. *International Journal of Lexicography*, 28(3), 318–337. doi:10.1093/ijl/ecv023
- Naciscione, A. (2010). *Stylistic use of phraseological units in discourse*. Amsterdam: John Benjamins.
- Newman, D. (2007). *Arabic-English thematic lexicon*. London: Routledge.
- Norrick, R. N. (2014). Subject area, terminology, proverb definitions, proverb features. In Hrisztova-Gotthardt, H. & Varga, M. A. (Eds.), *Introduction to paremiology: A comprehensive guide to proverb studies* (pp. 7–27). Berlin: Mouton de Gruyter.
- Pawley, A. (2007). Developments in the study of formulaic language since 1970: A personal view. In Skandera, P. (Ed.), *Phraseology and culture in English* (pp. 3–45). Berlin: Mouton de Gruyter.
- Ramisch, C. (2015). *Multiword expressions acquisition: A generic and open framework*. New York: Springer.
- Ruiz, J. H. (2009). *Understanding tropes: At the crossroads between pragmatics and cognition*. Duisburg Papers on Research in Language and Culture, 75. Frankfurt am Main: Peter Lang.
- Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Schmitt, N. & Carter, R. (2004). Formulaic sequences in action: An introduction. In Schmitt, N. (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1–22). Amsterdam: John Benjamins.

- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Skandera, P. (Ed). (2007). *Phraseology and culture in English*. Berlin: Mouton de Gruyter.
- Vinay, J.-P. & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. C. Sager & M.-J. Hamel, Trans. and Eds.). Amsterdam: John Benjamins.
- Walker, J. A. (2013). Variation analysis. In Podesva, R. J. & Sharma, D. (Eds.), *Research methods in linguistics* (pp. 440–459). Cambridge: Cambridge University Press.
- Wehr, H. (1974). *A dictionary of Modern Written Arabic*. Ed. J. Milton Cowan. 3rd ed. Beirut: Librarie du Liban.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wulff, S. (2008). *Rethinking idiomaticity: A usage-based approach*. London: Continuum.
- Zhu, F. & Fellbaum, C. (2015). Quantifying fixedness and compositionality in Chinese idioms. *International Journal of Lexicography*, 28(3), 338–350. doi:10.1093/ijl/ecv018

## Appendix

### Symbols for the Transliteration of Arabic<sup>1</sup>

**Table A1.** Consonant Letters

Initial	Medial	Final	Alone	Romanization
أ، إ، آ	ء، آ، ؤ، ئ	ء	ء	'
ب	ب	ب	ب	b
ت	ت	ت	ت	t
ث	ث	ث	ث	th
ج	ج	ج	ج	j
ح	ح	ح	ح	h
خ	خ	خ	خ	kh
د	د	د	د	d
ذ	ذ	ذ	ذ	dh
ر	ر	ر	ر	r
ز	ز	ز	ز	z
س	س	س	س	s
ش	ش	ش	ش	sh
ص	ص	ص	ص	ṣ
ض	ض	ض	ض	ḍ
ط	ط	ط	ط	ṭ
ظ	ظ	ظ	ظ	ẓ
ع	ع	ع	ع	c
غ	غ	غ	غ	gh
ف	ف	ف	ف	f
ق	ق	ق	ق	q
ك	ك	ك	ك	k
ل	ل	ل	ل	l
م	م	م	م	m
ن	ن	ن	ن	n

هـ	هـ	هـ، هـ	هـ، هـ	h
و	و	و	و	w
يـ	يـ	يـ	يـ	y

**Table A2.** Vowels and Diphthongs

Diacritic/Vowel Letter	Symbol
اَ	a
اُ	u
اِ	i
اَ	ā
اُو	ū
اِي	ī
اَي	á <sup>2</sup>
اَو	aw
اَي	ay

<sup>1</sup> Based, with one modification shown in the Introduction, on the Library of Congress (LOC) romanization system, available at <https://www.loc.gov/catdir/cpsa/romanization/arabic.pdf>.

<sup>2</sup> This symbol, which has the same pronunciation as ā is used for the shortened 'alif (*al-alif al-maqṣūrah*) spelt as final *yā*' without dots, in words like *'alā* (on; over).