**The Role of Construct State in Multi-Word Lexical Units Formation in Arabic**

**Elma Dizdar**

*University of Sarajevo, Bosnia and Herzegovina*

**Abstract:**

As times change and new concepts become part of our reality, Arabic, just like any other language, is facing the challenge of building new lexical units into its vocabulary. However, its word formation system, relying heavily on derivation based on root and pattern system, characteristic of Semitic languages, defies the use of affixation, conversion and compounding as major word formation methods in English and other Indo-European languages. Although there are some examples of affixation and compounding in Modern Standard Arabic, their use remains largely limited, with blurred boundaries in linguistic description between the two and the Arabic equivalent of blending, i.e. *naḥt*.

Therefore concepts expressed by compounds in other languages are usually rendered by the construct state or annexation construction in Arabic. Still, such structures manifest a high degree of variation in terms of their lexicalization and semantic compositionality, thus ranging from free word combinations to multi-word lexical units. The paper focuses on description of lexicalized non-compositional instances of Arabic construct state, including examples from Modern Standard Arabic, as well as a phenomenon dating from Classical Arabic. In addition to analysis of syntactic and semantic structure of Arabic construct state multi-word lexical units, based on which they are divided into several classes, discussion centers on different stylistic mechanisms operative in formation of their meaning.

**Keywords**: Arabic, lexical, units, formation, construct, state.

**1. Introduction**

In the era of fast and rapidly progressing development of science and technology, languages of the world are faced with the constant onslaught of new concepts, as well as the constant struggle to find words to render them. One of the responses that languages resort to faced with this challenge is formation of multi-world lexical units (MLUs), also labeled multi-word expressions, extended lexical units, compound lexical items, fixed expressions.[[1]](#endnote-1)

Studies of MLUs mostly focus on their classification and properties (Hȕning & Schlȕcker, 2015; Sprenger, 2003; Zgusta, 1967), extraction and identification methods (Attia, Toral, Tounsi, Pecina, & Van Genabith, 2010; Boulaknadel, Daille, & Aboutajdine, 2008), as well as contrasting of MLUs from different languages (Alnaser, 2010; Štambuk, 1997). [[2]](#endnote-2) Apart from the practical issues of MLU extraction and contrasting, the issues remaining in the very center of the scholarly debate are the syntactic properties of words combined to form MLUs, as well as the degree of their semantic (non)compositionality.

This paper will focus on only one specific structure in Arabic, i.e. the construct state, exploring its potential in MLU formation in Arabic. The paper aims to show that, due to specific syntactic properties of the construct state, as well well as a degree of inherent non-compositionality of the structure itself, it provides abundant opportunities for MLU formation, and thus has a major role to play not only in development of Arabic lexicon but also in widening the range of stylistic choice.

**2. Arabic Construct State: From Phrase to Word and back**

The Arabic construct state is a construction typically comprised of two nouns, athough there is a possibility for the structure to include three or more nouns. As it will be discussed later, an adjective can also appear as the first member of the construction. However, the rules governing this structure show striking resemblence to those governing words, rather than phrases in Arabic.

 Thus the two nouns in the construct state have only one marker of (in)definiteness, carried by the last member of the construction,[[3]](#endnote-3) and only one marker of case, carried by its first member. In addition, as the main stress falls on the rightmost member of the construction, its phonological pattern is the one of a single word. Should we then conclude that, since it walks like a word and talks like a word, the construct state is a word, or to be more precise, a compounding mechanism in the language whose system defies the use of compounding? Far from that, the construct state, although syntactically functioning as a single word, is but a word combination as the two nouns comprising the structure may take adjectives as modifiers,[[4]](#endnote-4) and the demonstrative pronoun belonging to the second member of the construction may be inserted between its two members.

 Moreover, it is the semantic properties of the construct state, or the words comprising it that really have a decisive role in whether the construct state will be a free combination of words or an MLU, semantically functioning as a single word. As they move up the continuum of semantic non-compositionality, the construct state components tend to abandon the syntactic features defining the structure as word combination and keep only those identical to the features of words.[[5]](#endnote-5)

**2.1 Noun-Noun Combination in Construct State**

 Let us have a look at the following examples of the Arabic construct state:

|  |  |  |  |
| --- | --- | --- | --- |
| *waraqu al-ḥā̕ iṭi[[6]](#endnote-6)* | (paper for wall) | = wallpaper | (1) |
| *furšatu ̕ asnānin* | (brush for teeth) | = toothbrush | (2) |
| *safīnatu al-baḍā̕ i ̒ i* | (ship for goods) | = cargo ship | (3) |
| *qamaratu al-qiyādati* | (cabin of leadership) | = cockpit | (4) |
| *̒ arḍu al-̕ aḥlāmi* | (land of dreams) | = dreamland | (5) |

 With the exception of (3), examples (1) to (5) above have one-word equivalents in English, which is a positive score on one of the tests used for determining MLUs. Still, the word combinations in all of (1) to (5) show a high degree of compositionality, i.e. the meaning of the combination is easily derived from the meaning of its components, thus implying a negative score on another test proposed for determining MLUs, often perceived as the most important one. Consequently, they should be condsidered as being semantically closer to free combinations than MLUs. Their only non-compositional feature is the thematic relation stored in the modifier noun, i.e. the relation of *purpose* in (1) to (3) and *location* in (4) and (5). Nevertheless, even these combinations would not allow insertion of demonstrative pronoun between their components, although they might take adjectives as modifiers of only the first member of the construct state. In other words, they show syntactic and semantic characteristics of both, free combinations of words and MLUs.

 Unlike (1) to (5), the following examples clearly constitute MLUs:

|  |  |  |  |
| --- | --- | --- | --- |
| *ḥulmu al-yaqẓati* | (dream of wakefulness) | = daydream | (6) |
| *ribāṭatu al-ǧa ̕ ši* | (quality of beign tied, of a soul) | = composure; self-control | (7) |
| *̒ arūsu al-ši ̒ ri* | (bride of poetry) | = muse | (8) |
| *̒ arā ̕ isu al-nīli* | (brides of the Nile) | = lotus flowers | (9) |
| *̒ arūsu al-baḥri* | (bride of the sea) | = mermaid | (10) |
| *ma ̒ rifatu al-ǧamīli* | (knowledge of the beautiful) | = gratitude | (11) |
| *nukrānu al-ǧamīli* | (denial of the beautiful) | = ingratitude | (12) |

 In addition to having one-word equivalents in English, demonstrating a high degree of non-compositionality and meeting the requirements of other MLU tests, such as not allowing substitution of or adding components, the examples (6) to (12) above involve yet another very important feature of MLUs - stylistic mechanism of transfer of meaning. Thus in (7) to (10) there is obviously metaphorical transfer of meaning at work, while in (6) it is metonymical transfer that operates. As (11) and (12) generally appear to be a bit more compositional compared to other examples, metaphorical transfer, athough present, is less prominent in them. Interestingly enough, one would expect the antonym of the word *ma ̒ rifa* in (11), i.e. the word *ǧahāla* (ignorance) to be used in (12) to produce the antonymous pair, but that is not the case.

 In (10) above, the stylistic mechanism at work goes beyond the metaphore at hand as the construct state *̒ arūsu al-baḥri* is also used as an antonomastic substitute for few cities situated at the coast and apparently perceived as beautiful as brides of the sea, these being Venice, Alexandria, Aqaba and Jaffa. Such use of metaphorical or metonymical transfer in the construct state to produce antonomastic substitutes of toponyms or proper names is quite frequent in Arabic. Let us have a look at few other examples:

|  |  |  |  |
| --- | --- | --- | --- |
| *madīnatu al-ḍabābi* | (city of fog) | = London | (13) |
| *madīnatu al-nūri* | (city of light) | = Paris | (14) |
| *rahīnu al-maḥbisayni* | (hostage of two prisons) | = ̕ Abū al-̒ Alā ̕ al-Ma ̒ arrī[[7]](#endnote-7) | (15) |

**2.2 Fixed Noun-Noun Combination in Construct State**

In addition to random nouns combined in the construct state to form MLUs, there are two phenomena in Arabic that account for a large number of Arabic construct state MLUs. Although both phenomena involve the use of fixed noun or group of nouns in the position of the first member of the constuct state, their similarities end there. The first phenomenon involves the use of nouns which are not used at all or are hardly ever used outside of the construct state. As their meaning remains rather stable, they inevitably bring a degree of compositionality to their word combinations. However, from the syntactic point of view, they only allow those properties that the construct state shares with single words.[[8]](#endnote-8) Thus the only question related to their MLU status is whether they have more in common with affixation or compounding mechanisms than with MLU formation.

 Following are examples of the Arabic construct state involving the use of fixed nouns:

|  |  |  |  |
| --- | --- | --- | --- |
| *ḏawū al-mawaddati* | (owners of love) | = friends | (16) |
| *ḏawū al-ma ̒ rifati* | (owners of knowledge) | = acquaintances | (17) |
| *ḏātu al-bayni* | (essence of separation) | = disagreement | (18) |
| *ṣāḥibu al- ̕ amri* | (owner of command) | = ruler | (19) |
| *ṣāḥibu al-dayni* | (owner of debt) | = creditor | (20) |

 The other phenomenon referred to above is the one dating back from Classical Arabic. It is the use of *kunya*, originally a component of proper names, comprised of nouns denoting members of core family (father, mother, son and daughter),[[9]](#endnote-9) occupying the position of the construct state first member, and proper names of family members, occupying the position of its second member. However, *kunya* has over time evolved into a true MLU factory - a periphrastic mechanism used to express various concepts, from traits of humans and abstract nouns to names of animals and plants.

 Only few examples of *kunya* will be provided as an illustration:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *bintu al-baḥri* | (daughter of the sea) | = ship | = *al-safīna* | (21) |
| *bintu al-šafati* | (daughter of lips) | = word  | = *al-kalima* | (22) |
| *bintu al- ̒ ayni* | (daughter of eye) | = tear | = *al-dam ̒ a* | (23) |
| *bintu al-ǧabali* | (daughter of hill) | = echo | = *al-ṣadā* | (24) |
| *̕ aḫū al-firāši* | (brother of bed) | = patient | = *al-marīḍ* | (25) |
| *̕ aḫū al-nawmi* | (brother of sleep) | = death | = *al-mawt* | (26) |
| *̕ ummu al-nuǧūmi* | (mother of stars) | = sky | = *al-samā ̕* | (27) |
| *̕ ummu al-waḥši* | (mother of lonely/wild) | = desert | = *al-ṣaḥrā ̕* | (28) |
| *̕̕ ibnu al-layli* | (son of night) | = thief | = *al-liṣṣ* | (29) |
| *̕ abū al-baṣīri* | (father of the one endowed withn eyesight) | = blind one | = *al-kafīf* | (30) |

 Apart from meeting the requirements of all previously discussed MLU determination tests, the examples (21) to (30) also have one-word synonyms. Although the first component of MLUs in (21) to (30) is restricted to the use of but few nouns, this by no means influence the high degree of their non-compositionality. Moreover, the use of *kunya* instead of its one-word synonyms represents a stylistically much more potent way of rendering concepts. The transfer of meaning ranges from metonymical in (21) to (25), (27) and (29) to metaphorical in (26) and (28) and a specific form of euphemistic transfer in (30), where the concept is rendered by the use of its antonym.

**2.3 Adjective-Noun Combination in Construct State**

Although the Arabic construct state is generally a combination of two nouns, adjectives can also occur in the position of the construct state first member.[[10]](#endnote-10) These instances are rather frequent in Arabic and are often used in descriptions. However, even in this type of the construct state there are examples where the transfer of meaning, mostly metaphorical one, is at work, adding non-compositionality and producing MLUs. The following examples will illustrate how antonymy of Arabic adjectives *ṯaqīl* (heavy) and *ḫafīf* (light) is combined with metaphorically used nouns to produce MLU antonyms.

|  |  |  |  |
| --- | --- | --- | --- |
| *ṯaqīlu al-ẓilli* | (having a heavy shadow) | = insufferable | (31) |
| *ḫafīfu al-ẓilli* | (having a light shadow) | = likeable | (32) |
| *ṯaqīlu al-dami* | (having heavy blood) | = unpleasant | (33) |
| *ḫafīfu al-dami* | (having light blood) | = amiable | (34) |
| *ṯaqīlu al-rūḥi* | (having a heavy soul) | = dull, boring | (35) |
| *ḫafīfu al-rūḥi* | (having a light soul) | = charming | (36) |

**3. Conclusion**

Although there is nowadays a vast body of literature on MLUs, describing their properties, providing definitions and proposing tests for differentiating between them and other combinations of words, the boundaries of this group still seem to be fuzzy and fluid as there are many word combinations passing certain tests and yet failing the others. Consequently, they could be best described on a continuum, moving from those whose both syntactic and semantic features confirm their MLU status to those that either syntactically or semantically resemble free word combinations, retaining at the same time some key features of MLUs.

 The Arabic construct state has enormous potential in MLU formation, which frequently represents an invaluable substitute for compounding and affixation, inherently absent in Arabic word formation system. Due to its syntactic properties partly identical to those of single words, as well as a degree of inherent non-compositionality of its semantic structure, the construct state, as it has already proved, has all it takes to become and remain a significant factor in development of lexicon of Modern Standard Arabic. However, its importance in word formation does not end there. As shown by examples of *kunya*, the significance of the construct state is not only in its ability to produce MLUs, thus helping Arabic to face the challenges of modern times, but it is as much in widening the range of stylistic choice, in the chemistry it produces to render simple concepts in ways whose wealth of meaning and stylistic potential go far beyond the words denoting them.

1. As Hȕning & Schlȕcker (2015) point out, there is an abundance of terms used in contemporary studies to denote MLUs, i.e. multi-word expressions, which is the term used by the authors. The reasons for such an abundance could be attributed not only to difference in focus of various studies but also to different understanding of the notion of MLUs. Thus, although there are definitions and tests proposed to differentiate between MLUs and other combinations of words, such as those in Sprenger (2003) and Zgusta (1967), the boundaries as understood in literature still seem to be fluid and flexible. [↑](#endnote-ref-1)
2. Only several sporadic studies are mentioned serving as an illustration since going over the vast body of literature on MLUs would require much more room for elaboration than the strict boundaries of this paper could allow for. [↑](#endnote-ref-2)
3. This particular property of the construct state and its phonological pattern of a single word, are described as agreement asymmetries in Arabic DPs (Benmamoun, 2000). [↑](#endnote-ref-3)
4. Such modifiers, however, may only follow after the construct state, with modifiers of the first member of the construction following modifiers of its second member. [↑](#endnote-ref-4)
5. The enormous potential of the construct state in MLU formation could be related to inherent non-compositionality of the construction itself. According to relation-based approach to comprehension of conceptual combinations, developed in cognitive psychology, in addition to the two nouns expressed in the surface structure of the phrase such as the Arabic construct state, each noun-noun combination also contains a thematic relation linking the two nouns, stored in the noun functioning as the modifier. Moreover, such thematic relations compete with each others in comprehension of noun-noun combinations (Gagne, 2002; Ramey, 2005). [↑](#endnote-ref-5)
6. The DMG (Deutsche Morgenländische Gesellschaft) transcription system is used for transcription of Arabic names and words in the paper. [↑](#endnote-ref-6)
7. The two *prisons* of the famous writer being his blindness and the apartment to which he retreated from the outside world. [↑](#endnote-ref-7)
8. There might be some differences between the noun *ḏū/ḏāt* and other similar nouns, such as *ṣāḥib* with this respect, but the noun *ḏū/ḏāt* is most frequently used and is thus taken to be representative of the whole group. [↑](#endnote-ref-8)
9. Although nouns *̕ aḫ* (brother) and *̕ uḫt* (sister) were originally not part of *kunya*, as could be seen from descriptions in al- ̒ Arīḍī (1983) and Wright (1967), they are here included in the notion of *kunya* due to their semantic properties and syntactic behavior identical to the one of *kunya* component nouns in the construct state. This approach was also used by Muftić (1979) due to semantic reasons. [↑](#endnote-ref-9)
10. This has significant consequences in terms of syntactic features shared by the construct state and single words as earlier described, since it is in this type of the construct state possible for both the first and second member of the construction to carry the marker of definiteness.

**References:**

Alnaser, M. (2010). Multi-word Items in Dictionaries from a Translator’s Perspective. Ph.D. dissertation. Durham University: School of Modern Languages and Cultures. Retrieved from <http://etheses.dur.ac.uk/384/>.

al- ̒Arīḍī, M. M. (1983). *Mu ̒ ğam al-farā ̕ id al-maknūna fī al-̕ aṣwāt, wa al-kunya wa al-taġlīb, wa al-̕ aḍdād*. Bayrūt: Dār ab‘ād.

Attia, M., Toral, A., Tounsi, L, Pecina, P. & Van Genabith, J. (2010). Automatic Extraction of Arabic Multiword Expressions. In *Proceedings of the Workshop on Multiword Expressions: from Theory to Applications*, (pp. 18–26).

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford Studies in Comparative Syntax. New York: Oxford University Press.

Boulaknadel, S., Daille, B. & Aboutajdine, D. (2008). A multi-word term extraction program for Arabic language. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*. (pp. 1485-1488).

Gagné, C. L. (2002). The competition-among-relations-in-nominals theory of conceptual combination: Implications for stimulus class formation and class expansion. *Journal of the Experimental Analysis of Behavior 78*, 551-565.

Hȕning, M. & Schlȕcker, B. (2015). Multi-word expressions. In P. O. Mȕller, I. Ohnheiser, S. Olsen & F. Rainer (Eds.) *Word formation: An international handbook of the languages of Europe* (pp. 450-466). Berlin: De Gruyter.

Muftić, T. (1979). Osobenosti upotrebe 'kunye'. *Prilozi za orijentalnu filologiju* *27*. 133-164.

Ramey, C. H. (2005). Thematic typicality in modifier-noun conceptual combinations: Controlling for presentation-order effects. Ph.D. dissertation. Temple University. Ann Arbor: UMI. 3176844.

Sprenger, S. (2003). Fixed expressions and the production of idioms. Ph.D. dissertation. Nijmegen: Max Planck Instituut voor Psycholinguïstiek.

Štambuk, A. (1997). Multi-word lexical units in English and Croatian terminology of electronics. *Studia Romanica et Anglica Zagrabiensia, 42*. 373-388.

Wright, W. (1967). *A grammar of the Arabic language*. 3rd ed. vol. I-II. Cambridge: Cambridge University Press.

Zgusta, L. (1967). Multiword lexical units. *WORD, 23:1-3*, 578-587. doi: 10.1080/00437956.1967.11435507. [↑](#endnote-ref-10)