Identifying Arabic compounds other than the Synthetic Genitive Construction

Abdel Rahman Mitib Altakhaineh
Newcastle University
a.r.m.s.altakhaineh@ncl.ac.uk

Abstract: This study examines types of compounds other than the Synthetic Genitive Construction (SGC) in Modern Standard Arabic (MSA) and Jordanian Arabic (JA), discussing the word class of the parts of the compound and identifying the head. The analysis reveals that there are four types of compounds in MSA, and three in JA. The Prep + Prep combination is missing from JA. I also argue that the word class of the parts of the compound of Arabic in general, and of MSA in particular, is not diverse. Regarding the head, I suggest that N + N compounds other than the SGC, Adj + Adj compounds and reduplicated compounds can be either semantically double-headed or headless

Keywords: Arabic compounds; construct state; synthetic genitive construction; headedness; reduplication; numerals

1. Introduction

As noted by Altakhaineh (2016), most compounds in Arabic are examples of the Construct State/Synthetic Genitive Construction (henceforth, SGC) and the syntactic category of the internal parts of the compound is N + N or Adj + N.[1] However, there are certain N + N combinations that are not SGCs. Additionally, other closed sets of compounds may include

---

[1] The Construct State (referred to in Arabic as *Idâafah*) is defined as a construct that normally consists of two nouns or an adjective and a noun where the first element can be nominative, accusative or genitive based on the function of the whole construct in the sentence, whereas the second element is always genitive. Another important characteristic of the Construct State is that the first element is always indefinite, whereas the second can be definite or indefinite (see Fassi-Fehri 2012, 156). In this study, however, I opted for the term *Synthetic Genitive Construction* (SGC), which contrasts with 'analytic genitives', i.e., with the possessive markers, e.g., *li* 'for/of' in Modern Standard Arabic. In fact, the 'Construct State' refers to the morphological form of the possessum in a construct, e.g., the lack of nunation and in some Arabic dialects, e.g., Jordanian Arabic, the feminine suffix surfacing with a final /t/, etc. (see Altakhaineh 2016, 6–7).

adjectives, prepositions and particles. This study investigates these combinations. Firstly, it provides an analysis of N + N combinations other than SGCs, arguing that some of these combinations could be viewed as compounds. Secondly, this study identifies several further types of compounds on the basis of the syntactic category of their parts, e.g., Adj + Adj, N + Adj, etc. It also identifies the head of compounds other than the SGC. Subsequently, it discusses some combinations regarded as compounds by other researchers (e.g., Ryding 2005; Amer & Menacere 2013, 235, among others) who argue that these combinations are, in fact, not compounds. Finally, this study shows that reduplicated items and some types of numerals are best treated as compounds.

In order to achieve these five objectives, it is essential for the present study to provide a detailed description and analysis of the features of compounds other than the SGC in Arabic. Ultimately, the aim of this study is to shed light on analytical and theoretical questions in cross-linguistic morphology, especially concerning the process of compounding and its relationship with the formation of phrases and derived words. We begin with the discussion of the difference between SGC compounds and other potential types of N + N compounds.

2. Types of compounds other than SGCs

In this section, I examine types of Arabic compounds other than SGCs. According to Altakhaineh (2016, 134–135), an SGC compound is a complex word that consists of at least two adjacent words, where the second element is normally non-referential. He also notes that the second element of a compound within an SGC is not, in most cases, freely pluralised, as in:

(1) a. saaʕit l-yad   b. *saaʕit l- ʔayaadi
     clock the-hand   clock the-hand
     'the watch'   'the watches'

(2) a. ʕaruus l-baħr   b. *ʕaruus l-biħaar
     bride the-sea   bride the-seas
     'the mermaid' lit. 'the sea bride'   'the mermaids' lit. 'the seas bride'

(3) a. qawiyy l-qalb
     strong the-heart
     'a brave person' lit. 'the one with the strong heart'

   b. *qawiyy l-quluub
     strong   the-hearts
     'the brave people' lit. 'the ones with the strong heart'

Examination of Arabic compounds based on the most reliable criteria of compoundhood, i.e., adjacency, referentiality, and free pluralisation of the non-head shows that compounds other than SGC appear in Arabic (Altakhaineh 2016, 135).[2] Here, it is worth pointing out that adjacency and referentiality can be considered significant criteria when we are identifying compounding cross-linguistically (Altakhaineh 2016, 39–40). In English, Lieber & Štekauer (2009a, 11–12) suggest that the criterion of adjacency could be considered a reliable criterion for determining compound status. They show that while it is possible to insert a word such as *ugly* into the phrase *a black bird* (yielding *a black ugly bird*), it is not possible to insert such a word inside the compound *blackbird. Ugly* can only modify the compound as a whole (yielding *ugly blackbird*). Scrutinising the referentiality of the non-head in a compound, Bauer et al. (2013, 464) note that despite the fact that some complications pertaining to the referentiality of the non-head exist, e.g., when the non-head is a proper noun or has unique reference, it seems that the left element/the non-head of English compounds is normally non-referential (*idem.*). We begin the discussion of all this with an analysis of N + N combinations other than SGCs.

2.1. Noun + noun combinations

Various examples of N + N combinations other than SGCs can be found in Arabic. They are illustrated in (4)–(6).

(4)	sˤabaaħa masaaʔ	(5)	layla nahaar[3]	(6)	sˤayfa ʃitaaʔ
	morning evening		night daytime		summer winter
	'all day long'		'twenty-four seven'		'all year long'

In examples (4)–(6), the elements of the combinations, *sˤabaaħa* 'morning', *masaaʔ* 'evening', *layla* 'night', *nahaar* 'daytime', *sˤayfa* 'summer' and *ʃitaaʔ* 'winter', are all nouns. The syntactic category of the output

---

[2] Although the criterion of free pluralisation of the non-head is of great interest in SGC compounds, it is inapplicable to constructs other than SGCs due to the difficulty of pinpointing the head in these constructs (see section 3 for full discussion). Therefore, this criterion will not be pursued here any further.

[3] This is the form in MSA. In JA, it is phonologically realised as *leel nhaar* 'twenty-four seven'.

is therefore most plausibly a noun as well, although the function of these combinations is an adverbial of time, as in examples (7) and (8):

(7)   yadrus tˤ-tˤullaab   sˤabaaħa masaaʔ
      study   the-students morning evening
      'The students study all day long.'

(8)   yaʕmaal l-ʕummal   layla nahaar
      work   the-employees night daytime
      'The employees work twenty-four seven.'

The adverbial function of the combinations in examples (7) and (8) does not mean that they are adverbs; not all adverbials are adverbs and not all adverbs function as adverbials.

With regard to the compound or phrasal nature of these two combinations, it is important to note that the first and second N have to be adjacent, and neither the first nor the second element is referential. Inserting any element between them would result in ungrammaticality, as shown in (9)–(11):

(9)   sˤabaaħa (*wa) masaaʔ
      morning   and evening
      'all day long' lit. 'morning and evening'

(10) layla (*wa) nahaar
      night   and daytime
      'twenty four seven' lit. 'night and daytime'

(11) sˤayfa   (*wa) ʃitaaʔ
      summer   and winter
      'all year long' lit. 'summer and winter'

Therefore, the constructs in (4)–(6) are compounds.

2.2. Noun + adjective combinations

Ryding (2005, 59–60) suggests that N + Adj combinations in MSA, as in the following examples, are best regarded as syntactic phrases rather than compounds:

(12) a. l-walad   tˤ-tˤawiil   b. l-ʔawlaad   tˤ-tˤiwaal
          the-boy.MSG the-tall.MSG   the-boy.MPL the-tall.MPL
          'the tall boy'   'the tall boys'

      c. walad   tˤawiil   d. ʔawlaad tˤiwaal
          boy.MSG tall.MSG   boy.MPL tall.MPL
          'a tall boy'   'tall boys'

In all of (12a–d), the second element is an adjective that modifies the preceding noun *lwalad*. A well-known characteristic of such phrases in MSA and JA is that the adjective agrees with the noun in number, gender and definiteness (Ryding 2005, 59–60).

   However, while examples like (12a–d) are clearly phrasal, I argue that there is a closed set of N + Adj combinations that can be regarded as compounds, because they fulfil the adjacency criterion for compoundhood. Examples of this category are:

(13) l-baħr   l-ʔaħmar   (14) l-baħr   l-ʔabyadˤ   l-mutawassitˤ
      the-sea.MSG the-red.MSG   the-sea.MSG the-white.MSG the-middle.MSG
      'the Red Sea'   'the Mediterranean Sea', lit. 'the white middle sea'

(15) l-muħiitˤ   l-haadii   (16) l-qaamuus   l-muħiitˤ
      the-ocean.MSG the-calm.MSG   the-dictionary.MSG the-comprehensive.MSG
      'the Pacific Ocean'   'the Comprehensive Dictionary'

Similarly to example (12), the syntactic category of the combinations in examples (13)–(16) is the same as that of the first/left element. For instance, example (14) *lbaħr lʔabyadˤ lmutawassitˤ* 'the Mediterranean sea' is a noun phrase even though the elements *lʔabyadˤ* 'the white' and *lmutawassitˤ* 'the middle' are adjectives. However, examples (13)–(16) are different from example (12) in several other respects. Firstly, although the adjective in examples (13)–(16) agrees with the noun in number, gender and definiteness, this agreement does not change due to the fact that these examples have undergone lexicalisation. In other words, if the morphosyntactic features of the noun change, and in turn, those of the adjectives follow suit, the result will be unacceptable. This is shown in the following example:

(17) *l-qaawaamiis   l-muħiitˤa
      the-dictionary.MPL the-comprehensive.MPL
      'the Comprehensive Dictionaries', lit. 'the dictionaries the comprehensives'

Example (17) shows that if the adjective *lmuħiitˤa* 'the comprehensive' agrees with the noun *lqaawaamiis* 'the dictionaries' in number, the output is unacceptable. This is possibly because this is a name of a dictionary,

not a description of it. Hence, it cannot be pluralised. This phenomenon does not occur with phrases such as those in example (12). Note also that unlike example (12), examples (13)–(16) have proper names.

Secondly, unlike ordinary N + Adj sequences as in (18), the elements of examples (19)–(21) are inseparable in the sense that no element can intervene between them:

(18) l-walad   l-dʒamiil   tˤ-tˤawiil
      the-boy.MSG the-beautiful.MSG the-tall.MSG
      'the tall beautiful boy'

(19) l-baħr   (*l-waasiʕ)   l-ʔabyadˤ   l-mutawassitˤ
      the-sea.MSG   the-wide.MSG the-white.MSG the-middle.MSG
      'the wide Mediterranean Sea'

(20) l-baħr   l-ʔabyadˤ   (*l-waasiʕ)   l-mutawassitˤ
      the-sea.MSG the-white.MSG   the-wide.MSG the-middle.MSG
      'the wide Mediterranean Sea'

(21) l-qaamuus   (*l-dʒadiid)   l-muħiitˤ
      the-dictionary.MSG   the-new.MSG the-comprehensive.MSG
      'the new Comprehensive Dictionary'

In example (18), the adjective *ldʒamiil* 'the beautiful' can be inserted between the two elements of the phrase *lwalad tˤtˤawiil* 'the tall boy'. But no such insertion is allowed in examples (19)–(21). If the adjective *lwaasiʕ* 'the wide' is inserted either between *lbaħr* 'the sea' and *lʔabyadˤ* 'the white' or between *lʔabyadˤ* 'the white' and *lmutawassitˤ* 'the middle', the result is not acceptable. The same applies to example (21). Note that the adjectives found in examples (19) and (20) do not behave as normal adjectives in terms of agreement, i.e., number (see example 17). This could be due to that fact that examples (13)–(16) are lexicalised expressions, whose internal structure has been lost.

The previous argument is supported by the existence of similar examples in the Germanic languages. It has been argued that lexicalised Adj + N phrases may serve the same naming function as Adj + N compounds (e.g., Booij 2002; Jackendoff 2002). Giegerich (2005, 587) suggests that examples from English such as *dental care*, *solar system*, *postal service*, *polar bear* and *mental hospital* must be considered lexical even though they are phrasal in nature due to "the fore-stress pattern". Booij (2009, 214–215) points out that since adjectives in Dutch Adj + N combinations can be marked with the final inflectional ending *-e* (pronounced as schwa), it is ev-

ident that Dutch Adj + N combinations are phrases. However, some Adj + N combinations like this can not be modified by intensifiers such as *hele* 'very'. For example, it is unacceptable to say *hele zwarte doos* 'very black box' when using black box to refer to the registration device in airplanes; the intended meaning will be lost. Consequently, these combinations are to be considered lexical despite the fact that they are phrasal in nature (Booij *ibid.*).

For Italian, Gaeta and Ricca (2009, 43) argue that the difference between compounds and phrases is whether their elements are inseparable or not. Compounds should only consist of one uninterruptable phonological string, in which no intervening (non-inflectional) element can be inserted. Even though the impenetrability condition may be non-sufficient, since several lexicalised phrases are inseparable, it can still be maintained as a necessary condition for compoundhood.

In sum, I have shown that, in addition to the ordinary N + Adj phrases described by Ryding (2005), there are some N + Adj cases where lexicalisation has taken place, and those behave differently. These cases of N + Adj combinations can be regarded as compounds, since their behaviour is quite different from that of phrases. However, since they are syntactically phrase-like and semantically compound-like, these combinations could be identified as 'phrasal compounds'.

2.3. Adjective + adjective combinations

In MSA and JA, there exists a closed set of Adj + Adj coordinative compounds that have a compositional meaning. Examples of this type can be seen in (22) and (23):

(22) ħaamidˤ ħilw   (23) ħilw murr
       sour   sweet   sweet bitter
     'sweet-and-sour'   'bitter-sweet'

In examples (22) and (23), the syntactic category of the output is the same as that of the individual parts, i.e., adjective. The meaning of the whole compound in (22), 'sweet-and-sour', is derived from both elements *ħaamidˤ* 'sour' and *ħilw* 'sweet'. The two elements of the compound must be adjacent. Examples like (24a,b) are ungrammatical:

(24) a. ħilw (*wa) murr   b. ħilw (*ħaamidˤ) murr
            sweet   and bitter   sweet   sour   bitter
          'bitter and sweet'   'bitter, sour and sweet'

Examples (24a,b) show that any insertion between the two elements of the compound is unacceptable. Additionally, note that there seems to be an antonymic relationship between the two elements of the compounds in examples (22) and (23). Specifically, the meaning of *murr* 'bitter' is the opposite of *ḥilw* 'sweet'.

2.4. Particle + adjective combinations

Some researchers (e.g., Ryding 2005, 100; Amer & Menacere 2013, 235) discuss expressions in which the first element is *laa* 'no', as in (25) and (26), and they refer to these as compounds.

(25)	laa faqaari	(26)	laa markaziyyah
	no spine		no centralisation
	'invertebrate'		'decentralisation'

However, I argue that *laa* 'no' is a prefix, since it cannot stand on its own; it has to stand with adjectives, as in the following examples:

(27)	laa silki	(28)	laa ʔaxlaaqi
	no wired		no moral
	'wireless'		'immoral'

Examples (27) and (28) show that this prefix can be attached to several adjectives with a consistent meaning, yielding a potentially productive construction in which the first element is fixed, i.e., *laa* 'no', whereas the second element is changeable. Furthermore, *laa* 'no' cannot stand on its own, suggesting that it is a prefix similar to English *un-, il-, im-, in-,* etc.

Overall, contra Ryding (2005) and Amer & Menacere (2013, 235), I suggest that *laa* 'no' should be treated as a prefix. Therefore, examples (25)–(28) are instances of derived words, rather than compounds.

2.5. Particle + verb combinations

Some traditional grammarians treat a closed set of verbs in MSA as compounds consisting of the particle *maa* and a verb (e.g., Al-Rajihi 2000, 121–122). In these combinations, *maa* 'not/what' normally adds a sense of duration or negation to the second element, which is a verb (Ryding 2005, 638–640). However, I argue that these are not compounds on the grounds that the particle/element *maa* is a prefix that means 'not'. The relevant examples are the following:

(29) maa zaala   (30) maa bariħa
      not ceased   not left
      'did not cease/continue to be'   'did not leave'

(31) maa nfakka   (32) maa fatiʔa
      not be.disconnected   not ended
      'did not get disconnected'   'did not end'

In order for the element *maa* to add the sense of duration or negation, it has to appear with the verb as in examples (29)–(32). In other words, it cannot stand alone to convey that sense.

2.6. Numerals

Several linguists (e.g., Al-Rajihi 2000, Al-Humaydi 2005, 243; Booij 2010; Hurford 2011; Al-Hariiri 2013, 175, among others) have examined the structure of numerals in various languages, suggesting that a subset of numerals could be regarded as compounds. In MSA and JA, little attention has been given to either the structure or the content of numerals. Some numerals of MSA and their glosses in English are given in (33):

(33) a. xamsah   b. xamsat ʕaʃar[4]   c. xamsah wa xamsuun
          five   five   ten   five   and fifty
          'five'   'fifteen'   'fifty-five'

      d. maaʔah wa xamsah
          hundred and five
          'one hundred and five'

In MSA, all numerals above 10 are complex expressions. For example, unlike examples (33c,d), whose two elements are separated by a coordinating conjunction, the numeral *xamsat ʕaʃar* 'fifteen' in (33b) looks like a compound, consisting of two separate elements, *xamsah* 'five' and *ʕaʃar* 'ten'. Therefore, Al-Rajihi (2000, 75–76) considers cardinal numerals from eleven to nineteen to be compounds. Below is the full sequence from 11 to 19:

(34) ʔiħdaa ʕaʃar 'eleven', ʔiθnaa ʕaʃar 'twelve', θalaaθat ʕaʃar 'thirteen', ʔarbaʕat ʕaʃar 'fourteen', xamsat ʕaʃar 'fifteen', sittat ʕaʃar 'sixteen', sabʕat ʕaʃar 'seventeen', θamaaniyat ʕaʃar 'eighteen', tisʕat ʕaʃar 'nineteen'

---

[4] In JA, *tˤaʕiʃ* '-teen' is used instead of *ʕaʃar* 'ten' in numerals between eleven and nineteen. Note that *tˤaʕiʃ* '-teen' cannot stand on its own and *ʕaʃar* 'ten' is used in JA to mean ten, e.g., *ʕaʃar ʕyaal* 'ten boys'.

The Num + Num combinations in example (34) can be regarded as compounds, since they are inseparable; no element can be inserted between the internal elements of the compound. For example, the following are ungrammatical:

(35) xamsat (*wa) ʕaʃar
       five   and ten
       'fifteen'

The Arabic numerals in (35) are quite similar to Dutch, English and German numerals. Therefore, in the analysis of Arabic numerals I will pay special attention to Dutch numerals as analysed by Booij (2010).

As Booij (2010, 85) notes, "Most numerals of Dutch and English are complex linguistic expressions, formed by a recursive system of rules that enables the language user to form an in principle infinite set of numerals". In Dutch, English and German, all numerals above twelve are complex expressions. For instance, the numeral *vijf-tien* 'fifteen' in Dutch has the shape of a compound, because it consists of the lexemes *vijf* 'five' and *tien* 'teen'. It also has the stress pattern of Dutch compounds, with the main stress on the first element *op.cit.*, 88).

However, these numerals do not share other properties of regular compounds in Dutch, such as being right-headed. In particular, with regard to the word as a whole, the right-hand element of *vijftien*, i.e., *tien* 'ten', does not have the features characterising semantic heads (*ibid.*, 88). This is accounted for by viewing this special type of compound as being historically derived from (asyndetic) coordination (*idem.*). However, the exception to the position of the head cannot be used as a criterion to identify compounding in a language. First, Don (2009, 379) notes that there is a closed set of left-headed compounds in Dutch, in which new members cannot be added. These compounds comprise a verb stem and a body-part noun. The left member is a verb stem and the whole compound is a verb as follows (*idem.*):

(36) schuddebuik: lit. 'shake-belly' 'shake with laughter' (schud 'shake' + buik 'belly')

(37) reikhals: lit. 'reach-neck' 'reach anxiously' (reik 'reach' + hals 'neck')

(38) stampvoet: lit. 'stamp-feet' 'stamp with rage' (stamp 'stamp' + voet 'feet')

Here, it is worth noting that in Dutch left-headed compounds in (36)–(38), the non-head is always an argument of the verb. In addition, a well-known generalisation about compounding in English is the Right-Hand Head Rule

(RHHR), first suggested by Williams (1981, 248), who states that "in morphology, we define the head of a morphologically complex word to be the right hand member of that word." Nevertheless, the English compound *white collar* is an adjective like the first element, rather than a noun like the second. This does not mean that *white collar* is not a compound.[5]

Investigating other types of numerals, Booij (2010, 85) notes that examples (39) and (40) have the appearance of phrases due to being formed by means of coordination with the conjunction *en* 'and':

(39) een-en-vijftig '51' 'one-and-fifty'

(40) honderd (en) vijf '105' 'one hundred (and) five'

Examples (39) and (40) demonstrate that these numerals appear to be phrasal.[6] However, they can serve as bases of word-formation, especially for the formation of ordinal numerals by adding the suffixes *-ste* and *-de* (Booij 2010, 85), as in (41a,b):

(41) a. een-en-vijftig-ste 'one-and-fifty-th, fifty-first'
      b. honderd(-en)-vijfde 'hundred (and) fifth'

As a result, the numerals in examples (41a,b) are best treated as words, or it can be claimed that morphological operations may take phrases as their bases (*idem.*).

The same seems to apply partially to MSA numerals through adding the prefix *l-*, as in the following example:[7]

(42) a. l-xamsah   b. l-xamsat ʕaʃar   c. l-xamsah wa l-xamsuun
          the-five   the-five   ten   the-five   and the-fifty
          'fifth'   'fifteenth'   'fifty-fifth'

      d. l-maaʔah   wa xamsah
          the-hundred and five
          'one hundred and fifth'

---

[5] The issue of whether *white collar* is an adjective or a noun is still not settled in the literature.

[6] According to Booij (2010), it seems that there are two differences between examples (39) and (40). Firstly, it is possible to delete the conjunction in (40). Secondly, in (39), the conjunction *en* is pronounced as [ən], whereas in (40) it must be pronounced as [ɛn].

[7] The prefix *l-* normally functions as a definite article in Arabic. However, here it does not; it changes the number from cardinal to ordinal.

Examples (42a,b and d) show that by adding the prefix *l-* to the first element, cardinal numbers change into ordinal numbers. Example (42c) is different, since the prefix is added to both elements, i.e., *l-xamsah wa l-xamsuun* lit. 'the-five and the-fifty, fifty-five'. Therefore, Arabic ordinal numerals other than (42a, b) raise the question as to what extent their formation is morphological or syntactic. This issue needs further investigation.

2.7. Reduplication

Another category of compounding mentioned by traditional Arabic grammarians writing about MSA (e.g., Al-Rajihi 2000, 75) involves reduplicated words. The parts of such compounds are normally nouns, but certain prepositions can also be reduplicated. For example, in (43)–(47), the words *layl* 'night', *nahaar* 'daytime', *yawm* 'day' and *bayt* 'house' are all nouns, while *bayn* 'between' is a preposition.

(43) layla layl   (44) nahaara nahaar   (45) yawma yawm   (46) bayna   bayn
    night night   daytime daytime   day   day   between between
    'every night'   'all day long'   'daily'   'in-between'

(47) bayta bayt
    house house
    'close in distance'

Examples (43)–(47) show that the meaning of these reduplicated words can be compositional or non-compositional. In examples (43)–(46), the meaning of the whole compound could be predicted from the meanings of the individual parts, whereas example (47) has a meaning that is unpredictable from the meanings of the elements of the compound.

    Regarding the function of these compounds, they seem to function exclusively as time or place adverbials. That is, the compounds *yawma yawm* 'daily' or *bayta bayt* 'close in distance in reference to a building' have an adverbial function, as in examples (48) and (49):

(48) yaʕmal moħammad yawma-yawm
    work   Mohammad day-day
    'Mohammad works daily.'

(49) yaʕiiʃ moħammad wa ʕaliyy bayta-bayt
    live   Mohammad and Ali   house-house
    'Mohammad and Ali live close to each other.'

Reduplicated compounds can also be found in JA (see Musa & Altakhaineh 2015, 40). They are restricted to a few adjectives and a handful of set expressions, as in examples (50)–(52):

(50) kθiir kθiir   (51) ʃway ʃway   (52) basˤiitˤah basˤiitˤah
     much much   little little   simple   simple
     'very much'   'slowly'   (an expression denoting threat)

In examples (50)–(52) the elements of the replicated words *kθiir* 'much', *ʃway* 'little', and *basˤiitˤah* 'simple' are all adjectives. The whole reduplicated compound may have various functions in JA. Examples (50) and (51) can be used in an adverbial position, while example (52) functions as an interjection, as in (53) and (54) respectively:

(53) miʃaan   ʔallah suug   ʃway ʃway
     for.the.sake Allah drive.you little little
     'For Allah's sake, drive slowly.'

(54) basˤiitˤah basˤiitˤah raħ ʔaħki la ʔabuuy
     simple   simple   will tell.I to father.my
     'You will see, I will tell my father.'[8]

Note that these reduplicated items do not accept insertion of any other elements, as shown in (55) and (56):

(55) layla (*maʕ/fii) layl   (56) ʃway (*wu) ʃway
     night   with/in night   little and little
     'every night'   'slowly'

No element can intervene between the two reduplicated items. This indicates that they satisfy the adjacency criterion.

The idea of treating reduplication as compounding has been subject to some debate amongst linguists. For instance, Fabb (1998, 69) states that whole word reduplication can be considered to be a compounding process, since each part of the resulting word corresponds to an independently attested word. An example of whole word reduplication mentioned by Fabb (*ibid.*) is the Tamil compound *vantu-vantu* 'coming time and again', which is generated via reduplication of the word *vantu* 'coming'. Fabb's argument here looks plausible, since the internal elements are meaningful words that can stand alone. Henri (2012, 215) agrees with Fabb (1998) that reduplica-

---

[8] The meaning of *basˤiitˤah basˤiitˤah* is hard to convey, but this expression generally denotes threat. Additionally, this expression is usually accompanied by a hand gesture which is meant to intimidate the addressee.

tion is a type of compounding. However, Henri (ibid) claims that this type of compounding is peculiar in the sense that it deviates from the norm as far as compounding is concerned, specifically with regard to (1) the fact that reduplicated compounds are non-recursive; and (2) there is no change in category in such examples. Nonetheless, the force of these two arguments is not clear. Taking the first point into consideration, N + N compounding in Present-Day English, with examples like *library staff meeting room,* is undoubtedly recursive. Nevertheless, this option of multiple compounding is a rather recent one in the language, being attested only after c.1800, (e.g., Biber & Gray 2011, 237). With regard to the second point, many undoubted compounds have the same category as their components, as in *bookshop*, *bittersweet* and *stir-fry*. So the reduplicated compounds in Arabic in (43)–(52) are by no means exceptional in being non-recursive, and in having the same syntactic category as that of their internal elements.

More recently, Faraclas (2013, 244) argues that reduplicated items are to be treated as compounds. This is due to the fact that classical compounds and reduplicated items have the following characteristics in common:

1. Complexity: compounds consist of two or more lexical items which can appear as separate words in other contexts.

2. Attachment: the lexemes that make up a compound are inseparable so that no element can intervene between them, unless that intervening element is itself incorporated into the compound in order to form a more complex compound.

3. Phonological incorporation: phonologically, compounds behave as if they were simple lexical items.

Bauer et al. (2013, 463, 490) also classify certain reduplications in English as compounds. This concerns colloquial examples like *book book, friend friend, drink drink, home home, hot hot* and *green green*, which appear to be endocentric, with the compound as a whole being a hyponym of the head (*ibid.*). Based on the above discussion, it seems that several researchers acknowledge that full reduplicated items are compounds.

I argue that MSA and JA provide further evidence for treating some reduplicated items as compounds. In MSA and JA, it seems that semantic complexity (opacity) is another feature that classical compounds and reduplicated items share; examples are *bayta bayt* 'close in distance' and *ʃway ʃway* 'slowly', where the meaning of the combination cannot be straightforwardly deduced from that of the individual elements.

Note that similarly to classical compounds in English, reduplicated items in MSA and JA can be either compositional or non-compositional. For instance, the meaning of the compound *kθiir kθiir* 'very much' is com-

positional, since its meaning is derived from the meaning of the individual words. In contrast, the meaning of the whole compound ʃway ʃway 'slowly' is non-compositional, because its meaning ('slowly') is rather different from the meaning of its internal elements (with ʃway meaning 'little').

3. Headedness in compounds other than SGCs

3.1. Headedness criteria in the previous literature

Many criteria relevant to determining the head of a particular construction are discussed by Zwicky (1985). He proposes that the notion HEAD needs to capture the intuition that, in certain syntactic constructions, one element will dominate the rest (*op.cit.*, 2). He examines eight criteria in order to identify an element as a syntactic head. The publication of Zwicky's criteria sparked some debate about the correctness of some of these criteria and the possibility of adding further criteria (see in particular Hudson 1987). The consensus view that developed is summarised in Bauer (1990, 2–3), who also points out that "[...] although these criteria are neatly collected in the two articles mentioned, they do not originate there: the criteria have been widely discussed in earlier literature on the subject". Among the earlier scholars who addressed the notion of headedness are Bloomfield (1933); Marchand (1969, 214); Lyons (1977, 294); Williams (1981, 248), among others. Bauer's (1990, 2–3) useful summary of the previous research on headedness criteria is provided below:

1. A phrase is a hyponym of its head. Hudson (1987) calls this a "kind of" relation. This principle was originally proposed by Allen (1978, 11), who refers to it as the "IS A" condition. This condition suggests that the whole compound denotes a subclass of the concept that the head denotes.

2. The head is the subcategorizand; it is the item that selects its sisters.

3. The head is the governor.

4. The head is the distributional equivalent of the whole phrase.

5. The head is the obligatory element in the phrase.

6. The head is the "morphosyntactic locus".

7. The head is lexical (rather than phrasal).

The above-mentioned criteria can be used to identify the head in a phrase and have been adopted to identify the head in a compound (e.g., Arcodia 2012). However, some of Bauer's (1990) criteria may not be valid for identifying the head in a compound (Arcodia 2012, 368). In particular, criteria

4 (the head is the governor) and 7 (the head is lexical) are not applicable to English compounds (*idem.*). Along these lines, Arcodia (2012, 370) notes that "it should be evident that the characterization of heads is partly different for derivation and compounding".

However, the structure of Arabic compounds, especially those other than the SGC, is quite different from that of English compounds. This means that some criteria which are inapplicable to English compounds may, in fact, be applicable to Arabic ones. In the next section, I therefore employ all of the seven criteria compiled by Bauer (1990) to identify the head within a compound in Arabic. I group the seven criteria under three broad types, i.e., semantic (criterion 1), syntactic (criteria 2–5) and morphological (criteria 6–7). In the next section, I apply the above criteria to Arabic compounds other than SGCs.

3.2. Applying headedness criteria to compounds in MSA and JA other than SGCs

Compounds in MSA and JA which are not SGCs seem to behave differently from the SGC in terms of headedness. Here, I will start with Adj + Adj combinations and N + N combinations other than SGCs. Examples of these types are:[9]

(57) ħilw-un   murr-un
     sweet-NOM bitter-NOM
     'bitter-sweet'

(58) ħaamidˤ-un ħilw-un
     sour-NOM   sweet-NOM
     'sweet-and-sour'

(59) sˤabaaħ-a   masaaʔ-a
     morning-ACC evening-ACC
     'all day long'

(60) layla-a   nahaar-a
     night-ACC daytime-ACC
     'twenty-four seven'

Applying the semantic headedness criterion, Allen's (1978) condition is not applicable to examples (57) and (58), since adjectives do not form super/subsets. However, examples (57) and (58) clearly show that the meanings of the whole compounds are a mixture of both elements. With regard to (59) and (60), *sˤabaaħ masaaʔ* 'all day long' is neither *sˤabaaħ* 'morning' nor *masaaʔ* 'evening', indicating that both elements in this type of compound have semantically equal status. Here, it is worth pointing out that it is debated whether these compounds are semantically double-headed

---

[9] Note that the two elements of the compound in (57)–(60) are marked with the same case. The second element in SGC compounds, on the other hand, is always marked with the genitive case.

(Haspelmath 2002, 89) or semantically headless (Booij 2007, 80). Bauer et al. (2013, 443) note that the concept of headedness is problematic with regard to coordinative compounds. In the following paragraphs, I examine the concept of semantic headedness in coordinative compounds and its implications for the Arabic examples (57)–(60).

Booij (2007, 80–81) suggests that copulative/coordinative compounds (including dvandva and appositive compounds) do not have a semantic head, since the elements of these compounds have a semantically equal status. Examples from English are *blue-green* and *washer-dryer* (*idem.*). Conversely, Haspelmath (2002, 89) argues that English compounds, such as *bitter-sweet, deaf-mute* and *maidservant*, are semantically double-headed, since the two elements stand on an equal footing and they can be paraphrased with *and*. Haspelmath (2002, 89–90) also suggests that some compounds can be semantically headless (exocentric) based on his analysis of examples from Classical Tibetan, which he cites from Beyer (1992, 105). The head of the following examples is something like 'property', as in:

(61) mtho-dman 'height' (mtho 'high' + dman 'low')

(62) srab-mthug 'density' (srab 'thin' + mthug 'thick')

The semantic head of examples (61) and (62) is something like 'property', so *mtho-dman* literally means 'property in the dimension of high and low', i.e., height. Additionally, I note that in examples (61) and (62) the elements are both adjectives, yielding nouns, so I propose that neither adjective can be the head. Applying Haspelmath's (2002) analysis to Arabic compounds, it is clear that (57) and (58) are semantically double-headed, since the two elements are semantically equal and can be paraphrased with *and*. In addition, the two elements of examples (57) and (58) are adjectives, yielding an adjective.

In comparison with the Tibetan compounds, the Arabic examples (59) and (60) are syntactically different, since the latter consist of two nouns, yielding a noun. Additionally, the compounds in examples (59) and (60) can be used with an adverbial function and this is obviously also the case with the single nouns in examples (64) and (65) below. What is still special about (59) and (60) is that, while the word *sˤabaaħ* 'morning' or *masaaʔ* 'evening' can be used either as a noun or with an adverbial function, the compounds of the two nouns in (59) and (60) can only be used with an adverbial function. This makes them similar to English compounds like *mother–child*, which can only be used as a modifier of a noun, as in *mother–child relationship*. Furthermore, the meaning of the compounds in

(59) and (60) does not denote a combination of the two elements. For instance, if we coordinate the elements of the compound in (60), yielding a sentence like (63), the meaning and structure are different:

(63) ʔaʕmalu layl-an   wa nahaar-an
     work.I   night-INDF and daytime-INDF
     'I work at night and during daytime.'

The analysis of the Arabic data here is based on the ideas of Bloomfield (1933, 235), who notes that the copulative compound *bittersweet* 'bitter and sweet at the same time' is endocentric, since the compound, like its coordinated elements, *bitter* and *sweet*, functions as an adjective. However, the plant-name *bittersweet* is exocentric, since the grammatical function of the whole compound, as a noun, is different from that of the two adjective elements (*idem.*). That is, examples (57) and (58) are endocentric, whereas (59) and (60) are exocentric.

   To sum up, the situation with coordinative compounds is problematic, since the elements of a coordinative compound are always similar as far as their morphosyntactic and semantic properties are concerned (Bauer et al. 2013, 443). This indicates that either one of the elements can be viewed as the determinant of the compound's properties. In light of this situation, Bauer et al. state that "headedness seems not to be a useful concept in the analysis of coordinative compounds". Clearly, the concept of headedness in coordinative compounds cross-linguistically is worthy of further investigation.

   Returning to the cross-linguistic criteria of headedness, syntactically neither element selects the other in examples (57)–(60). As for which of the elements is the governor, it seems that both of the elements share the same case marking, i.e., nominative in (57) and (58), and accusative in (59) and (60). As far as the distributional equivalent is concerned, both elements share the same syntactic category, i.e., adjectives in (57) and (58), and nouns in (59) and (60). Regarding the final syntactic criterion, namely obligatoriness, both elements are obligatory; if one of them is deleted, the compound loses it meaning, as shown in (64) and (65):

(64) taftaħu l-maktabah sˤabaaħ-an   (65) taftaħu l-maktabah masaaʔ-an
     open   the-library morning-ACC   open   the-library evening-ACC
     'The library opens in the morning.'   'The library opens in the evening.'

Examples (64) and (65) can never denote 'all day long'.

   As for the first morphological criterion, both elements are marked equally for case, number and gender (see examples (57)–(60)). Thus, both

can be viewed as the morphosyntactic locus. Finally, both elements are lexical, rather than phrasal.

Concerning reduplication, Táíwò (2009, 44–45) suggests that reduplicated words in Yorùbá exhibit similar behaviour to coordinate compounds, since both the root/stem and the reduplicant have head-like features. Additionally, Táíwò (*idem.*) explains that the syntactic category of the reduplicated word can be the same as that of the root/stem, as in (66) and (67), or they can differ, as in (68) and (69):[10]

(66) ọmọ (N) 'child' → ọmọọmọ (N) 'grandchild(ren)'

(67) ńlá (Adj) 'big' → ńláńlá (Adj) 'very big'

(68) dára (V) 'be good' → dáradára (Adj) 'very good'

(69) fẹ́lẹ́ (V) 'be thin' → fẹ́lẹ́fẹ́lẹ́ (Adj) 'very thin'

Comparing data from Yorùbá to reduplicated compounds in Arabic, the syntactic category of the reduplicated words is the same as that of the stem, as in examples (43)–(52), some of which are repeated here for the readers' convenience:

(70) kθiir (Adj) 'much' → kθiir kθiir (Adj) 'very much'

(71) layla (N) 'night' → layla layl (N) 'every night'

(72) nahaara (N) 'daytime' → nahaara nahaar (N) 'all day long'

(73) basˁiitˁah (Adj)'simple' → basˁiitˁah basˁiitˁah (Adj) (an expression denoting threat)

Note that the reduplicated compounds in (70)–(73) have diverse functions. Specifically, example (70) consists of two adjectives and functions as an adjective. The reduplicated compounds in (71) and (72), which comprise two nouns, are always used with an adverbial function. The reduplicated compound in (73), which is formed from two adjectives, functions as an interjection. This suggests that example (70) is best treated as double-headed, whereas examples (71)–(73) are headless.

---

[10] It is worth pointing out that the fact that the syntactic category of the reduplicated word in examples (68) and (69) differs from that of the root/stem might be due to there being a null copula in these two examples, but when we reduplicate, only the adjective is reduplicated excluding the null copula. This issue requires further investigation.

Table 1: Types of compounds other than SGCs in MSA and JA

Language	First element	Second element	Examples	Gloss
MSA + JA	noun	noun	*sˤabaaħa masaaʔ* 'twenty four-seven'	*sˤabaaħa* 'morning' + *masaaʔ* 'evening'
MSA + JA	noun	adjective	*lbaħr lʔabyadˤ lmutawassitˤ* 'the Mediter-ranean sea'	*lbaħr* 'the sea' + *lʔabyadˤ* 'the white' + *lmutawassitˤ* 'the middle'
MSA + JA	adjective	adjective	*ħaamidˤ ħilw* 'sour-sweet'	*ħaamidˤ* 'sour' + *ħilw* 'sweet'
MSA	preposition	preposition	*bayna bayn* 'in-between'	*bayna* 'between' + *bayn* 'between'

All in all, Adj + Adj combinations, N + N combinations other than SGCs and reduplicated words can be double-headed or headless.

4. Conclusion

In sum, with regard to the classification of compounds other than SGCs, there are four types in MSA, and three in JA. The Prep + Prep combination is missing from JA. It seems that Arabic in general and MSA in particular do not have a wide diversity as regards the word class of the parts of compounds. Table 1 shows the components of possible syntactic categories of compounds in MSA and JA. MSA has a small number of compounds with word classes other than N + N and Adj + N combinations.

With respect to reduplications, in Arabic these items are compounds, since they are: (1) two separate lexemes; (2) inseparable; (3) simple lexical items; and (4) semantically non-transparent/non-compositional. Another fact about reduplicated compounds in MSA is that they can function as adverbials such as *bayna bayn* 'in-between' and *bayta bayt* 'close in distance', whereas in JA they can function as adverbials, e.g., *ʃway ʃway* 'slowly' or interjections, e.g., *basˤiitˤah basˤiitˤah* (an expression denoting threat). Arabic numerals from eleven to nineteen are compounds, whereas further investigation is necessary to explore the nature of other numerals. Finally, taking all the previous points about headedness into consideration, several gen-

eralisations can be made with regard to headedness in Arabic compounds. N + N compounds other than the SGC, Adj + Adj compounds and reduplicated compounds can be either endocentric (semantically double-headed) or exocentric (semantically headless).

Acknowledgements

References

Al-Hariiri, Mohammad. 2013. The structure of names in Arabic language. Al Hijaz International Refereed Journal for Islamic & Arabic Studies 2. 168–193.

Al-Humaydi, Abdelrahman. 2005. lasmaa lmurakkabah: anwaaha wa iraabha. diraasah nahwiyya [Compound nouns: Types and parsing. A syntactic study]. Ad Diriyah Journal 29. 243–264.

Al-Rajihi, Abduh. 2000. altatqbiiq nnaħwii [Applied syntax]. Alexandria: Dar Al-Maarifah Al-Jamiya.

Allen, Margaret R. 1978. Morphological investigations. Doctoral dissertation. University of Connecticut.

Altakhaineh, Abdel Rahman M. 2016. Compounding in Modern Standard Arabic, Jordanian Arabic and English. Doctoral dissertation. Newcastle University.

Amer, Walid M. and Karim Menacere. 2013. The challenges of translating English compounds into Arabic: For better or for worse. Babel 59. 224–243.

Arcodia, Giorgio Francesco. 2012. Constructions and headedness in derivation and compounding. Morphology 22. 365–397.

Bauer, Laurie. 1990. Be-heading the word. Journal of Linguistics 26. 1–31.

Bauer, Laurie, Lieber Rochelle and Ingo Plag. 2013. The Oxford reference guide to English morphology. Oxford: Oxford University Press.

Beyer, Stephan V. 1992. The classical tibetan language. New York: State University of New York.

Biber, Douglas and Bethany Gray. 2011. Grammatical change in the noun phrase: The influence of written language use. English Language and Linguistics 15. 223–250.

Bloomfield, Leonard. 1933. Language. New York: Holt, Rinehart and Winston.

Booij, Geert. 2002. Constructional idioms, morphology, and the dutch lexicon. Journal of Germanic Linguistics 14. 301–327.

Booij, Geert. 2007. Construction morphology and the lexicon. In F. Montermini, G. Boyé and N. Hathout (eds.) Selected proceedings of the 5th Décembrettes: Morphology in Toulouse. Somerville: Cascadilla Proceedings Project. 34–44.

Booij, Geert. 2009. Construction morphology and compounding. In Lieber & Štekauer (2009b, 201–216).

Booij, Geert. 2010. Constructions and lexical units: An analysis of Dutch numerals. In S. Olsen (ed.) New impulses in word formation. Hamburg: Helmut Buske Verlag. 81–100.

Don, Jan. 2009. IE, Germanic: Dutch. In Lieber & Štekauer (2009b, 370–385).

Fabb, Nigel. 1998. Compounding. In A. Spencer and A. M. Zwicky (eds.) The handbook of morphology. Oxford & Malden, MA: Blackwell. 66–84.

Faraclas, Nick. 2013. Nigerian pidgin. London: Routledge.

Fassi-Fehri, Abdelkader. 2012. Key features and parameters in Arabic grammar. Amsterdam & Philadelphia: John Benjamins.

Gaeta, Livio and Davide Ricca. 2009. Composita solvantur: Compounds as lexical units or morphological objects. Italian Journal of Linguistics 21. 35–70.

Giegerich, Heinz J. 2005. Associative adjectives in English and the lexicon–syntax interface. Journal of Linguistics 41. 571–591.

Haspelmath, Martin. 2002. Understanding morphology. London: Arnold.

Henri, Fabiola. 2012. 'attenuative verbal reduplication in Mauritian: A morpho-semantic approach. In E. O. Aboh, N. Smith and A. Zribi-Hertz (eds.) The morphosyntax of reiteration in Creole and Non-Creole languages. Amsterdam & Philadelphia: John Benjamins. 203–234.

Hudson, Richard A. 1987. Zwicky on heads. Journal of Linguistics 23. 109–132.

Hurford, James R. 2011. The linguistic theory of numerals. Cambridge: Cambridge University Press.

Jackendoff, Ray. 2002. Foundations of language: Brain, meaning, grammar, evolution. Oxford: Oxford University Press.

Lieber, Rochelle and Pavol Štekauer. 2009a. Introduction: Status and definition of compounding. In Lieber & Štekauer (2009b, 3–18).

Lieber, Rochelle and Pavol Štekauer (eds.). 2009b. The Oxford handbook of compounding. Oxford: Oxford University Press.

Lyons, John. 1977. Semantics. Cambridge: Cambridge University Press.

Marchand, Hans. 1969. The categories and types of present-day English word-formation: A synchronic–diachronic approach (Second, completely revised and enlarged edition). München: C. H. Beck.

Musa, Rebecca and Abdel Rahman M. Altakhaineh. 2015. An application of Optimality Theory (OT) on syllable structure within reduplication in Hausa spoken in Kano. International Journal of English Language and Linguistics Research 3. 37–49.

Ryding, Karin Christina. 2005. A reference grammar of Modern Standard Arabic. Cambridge: Cambridge University Press.

Táíwò, Oyè. 2009. Headedness and the structure of Yorùbá compound words. Taiwan Journal of Linguistics 7. 27–52.

Williams, Edwin. 1981. On the notions "lexically related" and "head of a word". Linguistic Inquiry 12. 245–274.

Zwicky, Arnold M. 1985. Heads. Journal of Linguistics 21. 1–29.