

# Arabic synthetic compounds

Abdullah S. Al-Dobaian

King Saud University

This paper discusses some data of Arabic synthetic compounds in which regular plural inflection is included inside compounds. These data pose problems to Kiparsky's (1982) level-ordering lexical morphology model and Li's (1990) generalization on verb incorporation. I argue that such compounds are lexically formed based on some pieces of evidence. To support the analysis, I compare the compounds and the construct state constructions in Arabic and Hebrew. Then I show that the lexical analysis explains the morphological, syntactic properties, and the semantics of Arabic synthetic compounds. More specifically, I explain how the lexical analysis applies to theta-role assignment inside the compound and then discuss the number specification of the non-head in the compound of Arabic and English.

**Keywords:** Arabic synthetic compounds, lexical morphology, English compounds, theta-role, regular inflections in compounds

## 1. Introduction

In this study, I examine the morphological, semantic, and syntactic structure of Arabic synthetic compounds (ASC). Such compounds pose problems for level-ordering phonology as proposed in Kiparsky (1982) and Li's (1990) generalization on verb incorporation. More specifically, the challenge of Arabic compound structure stems from the morphological presence of regular plural inflection inside a compound, which poses a problem for the morphological ordering relation between compounding and regular plural inflection. I argue that the compounds are lexically formed and illustrate how some interesting properties of syntax, like merger and number checking, can be handled lexically. I propose that a lexical analysis provides a proper explanation for what is an otherwise puzzling set of data.

In the second section, I illustrate previous analyses relevant to the discussion of ASC, such as prosodic morphology, especially that of McCarthy & Prince (1990), Ratcliffe (1990; 1997; 2003). Then I discuss Distributed Morphology (Halle & Marantz 1993). Next I illustrate dual morphological systems (Pinker 1999), and

others. Moreover, Kiparsky's level-ordering model of lexical phonology is examined. Finally I review Li's (1990) analysis of compounds. The third section shows the data of Arabic synthetic compounds. The fourth section explains the problems of ASC: First I provide some pieces of evidence for the lexical nature of ASC; then I offer an analysis that can better explain the morphological and semantic properties of the data; while the final section summarizes the basic ideas of the study.

## 2. Previous studies

Here, I review the basic linguistic literature that achieves two important goals: first, it shows the basic theoretical framework of the study; and second, this study relates to this accumulative linguistic research and attempts therefore to fill a gap. The topic of interest is compounds involving plurality which intersects with important morphological issues at the heart of the lexicon-syntax interface. Among the issues, to name a few, are inflectional morphology and derivational morphology and the boundary between them. A basic question of interest in psycholinguistic and generative research is how words are formed and at what level.

The review of literature comprises five aspects of morphological research: prosodic morphology of broken plurals (BP) and sound plurals (SP) as discussed in McCarthy & Prince (1990) and Ratcliffe (1990; 1997; 2003); Distributed Morphology (Halle & Marantz 1993; Siddiqi 2006; Kelly 2013); dual morphological systems (Pinker 1999; Berent & Pinker 2008, et al.); lexical phonology and morphology (Kiparsky 1982); and compound and plurality.

### 2.1 The prosodic morphology of BP and SP

Since BP and SP may be part of compounds in Arabic, a literature review of them is in order. There are three generative approaches that attempt to examine the structure of BP and SP. I explain in this section McCarthy & Prince's (1990) proposal and McCarthy (1993). Then I illustrate Ratcliffe's analysis (1990; 1997; 2003). I address problems with these analyses that make them inadequate to account for the data.

#### 2.1.1 *McCarthy & Prince (1990)*

McCarthy & Prince (1990) propose a theory relating phonology to morphology in order to explain how a word formation process applies in Arabic BP. More specifically, they establish a close link between templates<sup>1</sup> and prosodic structure (cf. also

---

1. Templates are fixed shapes of morphological structures. For details, see McCarthy (1993).

McCarthy 1993). In other words, a complex morphological process like BP may be better explained, according to McCarthy & Prince, by mapping an iambic foot to the singular form base. McCarthy & Prince develop a theory of prosodic morphology which has three basic components (McCarthy & Prince 1990: 209).

- I. Prosodic morphology hypothesis. Templates are defined in terms of the authentic units of prosody: mora ( $\mu$ ), syllable ( $\sigma$ ), foot (F), prosodic word (W), and so on.
- II. Template satisfaction condition. Satisfaction of templatic constraints is obligatory and is determined by the principles of prosody, both universal and language-specific.
- III. Prosodic circumscription of domains. The domain to which morphological operations apply may be circumscribed by prosodic criteria as well as by the more familiar morphological ones. In particular, the **minimal word** within a domain may be selected as the locus of morphological transformation in lieu of the whole domain.

McCarthy & Prince (1990: 211, 212) differentiate between BP and SP in that the latter is more productive and involve suffixation and that it represents a marginal mode of pluralization compared to BP.

They analyse the different patterns of Arabic BP based on Wright (1971) (as cited in McCarthy & Prince 1990: 215). Wright divides the patterns of BP into four basic categories: iambic,<sup>2</sup> trochaic,<sup>3</sup> monosyllabic,<sup>4</sup> and others.<sup>5</sup> Of all these BP types, McCarthy & Prince only focus on the iambic type since they consider it the most productive type. Their analysis is to circumscribe the first two moras of singular noun as a result of the minimal word condition. Then to form the plural, the circumscribed part of the word is mapped to an iambic foot.

Let us consider the following Examples (McCarthy & Prince 1990: 221):

- |     |        |           |          |                                   |
|-----|--------|-----------|----------|-----------------------------------|
| (1) | [naf]  | [sultaan] | [jundub] |                                   |
|     | naf    | sul       | jun      | circumscribe a minimal word       |
|     | nafaa  | suluu     | junuu    | map to iambic template            |
|     | nufuus | salaatiin | janaadib | restore residue and change vowels |

---

2. An iambic pattern is a foot consisting of two syllables which can be: light-heavy, light-light, or heavy.

3. A trochaic pattern is a foot consisting of two syllables which can be: heavy or light-light.

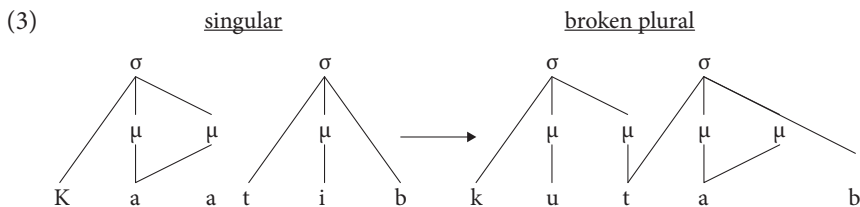
4. This type is restricted to a monosyllabic BP pattern *CuCC* involving adjectives of color and bodily defect as in *humr* 'red'. See McCarthy & Prince (1990: 214, 215).

5. This is represented by the BP type *CuCCaaC* as in *kuffaar* 'infidels'. It is derived from the active participle *kaafir*. It is discussed below.

As we observe in (1) McCarthy & Prince's analysis of BP accesses a bimoraic word and then maps it into an iambic foot, which is CvCv. Then the vowel melody is changed and the residue of the word is restored. McCarthy argues that Arabic has the following 'canonical noun patterns'<sup>6</sup> that are able to form the broken plural (McCarthy 1993, Example (20): 10):

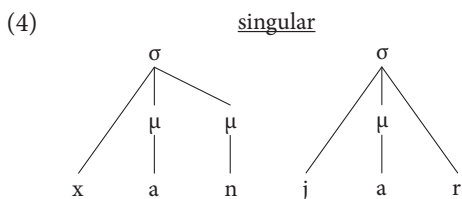
- |        |                |    |               |    |                |    |               |
|--------|----------------|----|---------------|----|----------------|----|---------------|
| (2) a. | H              | b. | LL            | c. | LH             | d. | HL            |
|        | CvCC           |    | CvCvC         |    | CvCvC          |    | CvCvC         |
|        | <i>baħr</i>    |    | <i>badal</i>  |    | <i>waziir</i>  |    | <i>kaatib</i> |
| e.     | HH             | f. | HL            | g. | HH             |    |               |
|        | CvCvCvC        |    | CvCCvC        |    | CvCCvC         |    |               |
|        | <i>jaamuus</i> |    | <i>xanjar</i> |    | <i>jumhuur</i> |    |               |

According to McCarthy, these patterns are canonically the basic noun patterns of Arabic being able to derive a broken plural with the exception of (2e) which he assumes to be rare and "probably an historical innovation" (1993, Example (20): 10). McCarthy regards the patterns (2a, b, c, and g) as the basic iambic types of Arabic. These templates are authentic prosodic units formed in the lexicon and used by the morphology. On the other hand, the patterns (2d and f) are "non-iambic" and are not basic templates themselves but they are integrated in a template through either a mora affixation or a templatic prosodic morphology (1993: 202). To illustrate, the pattern CvCvC derives a broken plural via the affixation of the mora as illustrated by (3).



In (3), the singular is formed by extending the mora of the vowel to form a long vowel. As a result a Heavy Light (HL) anti-iambic foot is derived. The broken plural is anti-iambic foot too and is formed by linking the second mora of the first syllable to *t* (McCarthy 1993, Footnote 8). This mora was occupied by the second member of the long vowel *aa* in the singular form. Moreover the vowel is lengthened in the broken plural by multiply linking *a* to two moras. As for CvCCvC, it is only restricted to nouns with four consonantal roots as illustrated by (4):

6. Final consonants in these patterns are extrametrical; thus they do not count in syllable weight.



According to McCarthy, this pattern is not an authentic template but an anti-iambic foot since it consists of four consonants with no “templatic constraint on the form” and is derived by “a-templatic” prosodic morphology (McCarthy 1993: 202).

McCarthy’s analysis of Arabic morphology is influential and effective. McCarthy’s basic assumption is that the plural is derived from the singular that combines the root and the template. However the circumscription of the broken plural to an iambic foot is nonetheless problematic. Given the unpredictability of the BP, the iambic analysis is incapable of accounting for the different types of Arabic BP. Such an analysis works for Arabic *ultimate plurals* like those in (1) in which an iambic CvCv can be circumscribed, but not for others like paucity, multiplicity, collective noun, and so on, as illustrated by (5):

- (5) himaar ‘donkey’ sariir ‘bed’
- |            |           |                             |
|------------|-----------|-----------------------------|
| him        | sar       | circumscribe a minimal word |
| hmaa       | saraa     | map to iambic template      |
| rahmaa     | -----     | ra- metathesis              |
| *rahmaarat | *saraaiir | restore residue             |

If we apply the iambic plural rule to the nouns in (5), we end up with ungrammatical broken plurals. The BP of these forms are *rahmirat* and *surur*. They are both trochaic broken plurals of the paucity *rafsilat* pattern and the multiplicity *fusul* pattern. McCarthy justifies that the trochaic patterns are of less generality and productivity (McCarthy & Prince 1990: 215, 278). Hence the iambic rule does not apply to them. However if we aim at providing an analysis to the BP phenomenon, it is essential that all types are taken into consideration. Another problem deals with the “anti-iambic” noun patterns that we discuss in the next section.

### 2.1.1.2 Ratcliffe (1990; 1997; 2003)

Even though Ratcliffe assumes that BP and SP are two different kinds of morphological processes involving two different functions, he, nevertheless, argues that both are formed in the lexicon. Ratcliffe adopts Siegel’s (1974) distinction between inflectional and derivational morphemes. Ratcliffe agrees with Siegel in associating inflectional and derivational morphemes with Class I and Class II respectively. Class II affixation involves inflections that are syntactically relevant while derivations belong to Class I which are irrelevant to syntax. Ratcliffe believes that

inflections and derivations employ the same formal rules of word formation and the same formal mechanisms. Thus he thinks it is redundant to assume that inflections and derivations are formed at the lexicon and the syntax. In order to eliminate such redundancy, he argues that *all* morphological processes are derived in the lexicon while the syntax has no role to play (Ratcliffe 1990: 96). This extreme position is reminiscent of the old strong lexicalist hypothesis that was proven to be missing an important generalization that the syntactic principles may not be completely eliminated from forming a morphologically complex word. Baker's influential work (1988) has robustly illustrated that incorporation in Mohawk is controlled by the syntactic principle Empty Category Principle (ECP). Therefore, Ratcliffe's claim cannot be maintained.<sup>7</sup>

Ratcliffe criticizes McCarthy's claim that only the canonical noun patterns in (2) form a broken plural since there are other patterns as well (Ratcliffe 2003: 239). For instance, a CvCCvCvC pattern form as *sankabuut* derives *sanaakib*. He also takes issue with McCarthy's non-iambic patterns. McCarthy proposes that pattern CvCvC (e.g. *kaatib*) and CvCCvC (e.g. *xanjar*) are not canonical noun templates and hence they have to produce a BP via mora suffixation and a-templatic prosodic morphology respectively. The fact that these patterns can develop a BP suggests that McCarthy's noun canonicity is empirically false (Ratcliffe 2003: 239). Ratcliffe observes that these patterns have lexicalized BP.<sup>8</sup>

In order to solve the problems of McCarthy's analysis, Ratcliffe (1990: 108) suggests a "Long Vowel Pluralization Rule" which derives a BP by inserting a -VV-segment after the second C in the prosodic template. This rule handles McCarthy's iambic and "anti-iambic" patterns:

---

7. Ratcliffe's analysis is problematic not because he uses a lexical analysis or he does not follow Baker's analysis as a reviewer suggests but because of two reasons. The first one is that his analysis denies syntax any role in morphological structure formation. Both the lexicon and syntax play important roles in deriving morphology. Baker (1988), among others, provides morphological data where syntactic principles determine the formation of words. However this does not entail that all morphology is syntactically formed, but there are irregular words that are not controlled by syntax and hence are lexically derived. Secondly, Ratcliffe's lexical rule has empirical problems. For example, Ratcliffe's "Long Vowel Pluralization Rule" does not apply to all BP patterns as explained below. Even though I argue for a lexical analysis of ASC due to specific pieces of evidence that I explain in details in § 3, I believe, nonetheless, that syntax has an important role in morphological structure (e.g. the construct state construction in Arabic and Hebrew that I also discuss in § 3).

8. Ratcliffe (2003: 239) points out that CvCvC pattern produces sound plural as *kaatibuun* and a broken plural *kuttab*. He does not however explain why this form redundantly takes two different types of plural.

(6)	Pattern	Singular Noun	BP
a.	CvCC	nafs	nufuus
b.	CvvCvC	kaatib	kuttaab
c.	CvCCvC	xanjar	xanaajir

As the examples in (6) show, the vowel is lengthened after the second consonant. However this rule partially applies to some broken plurals and is by no means comprehensive. To illustrate, the rule cannot handle the non-iambic patterns of BP, as for example trochaic patterns like *qiradat*. Another example of the trochaic pattern is *fusalaa?* where Ratcliffe's long vowel rule does not apply. He stipulates that when there are rational entities such as *wuzaraa?* 'ministers', the vowel after the third consonant is lengthened (1990: 109). However this condition is completely ad hoc and cannot be justified. Moreover there are non-rational entities resisting Ratcliffe's rule as the BP *xuṣabaa?* 'fertile', *ṛulamaa?* 'painful'. There is also BP *ṛan-sibaa?* 'shares'<sup>9</sup> which is problematic to the long vowel rule. These examples involve a trochaic pattern in which the long vowel is after the third unlike what Ratcliffe assumes. Finally the long vowel rule does not apply in monosyllabic *CuCC* pattern of the BP as in *ḥumr* 'red'. As a consequence of these counter-examples, Ratcliffe's rule cannot be maintained since it cannot account for the BP patterns as a whole.

## 2.2 Distributed Morphology (DM)

DM is a theory within the Minimalist program proposing that syntax is the only generative component of grammar. Therefore morphological structure building occurs in syntax (Halle & Marantz 1993). DM rejects the lexicon and distributes its constraints and operations between three lists: morphosyntactic features, vocabulary items, and the Encyclopedia (Kelly 2013).

According to DM, morphemes are associated with morphosyntactic features lacking any phonological content (Siddiqi 2006). There is no difference between words and phrases, they are all formed in syntax as a complex morphological structure. The structure of the word consists of a root (a lexical morpheme) and a functional category (usually v, n, adj). Syntactic operations like *merge* and *move* target abstract morpheme and build a hierarchical structure of terminal nodes (Siddiqi 2006).

The Vocabulary inserts Vocabulary Items (VIs) in the morphological structure which is at the spell out of syntax with phonology. At this level, the Vocabulary assigns phonological forms to the syntactic terminals (Siddiqi 2006; Kelly 2013). The morphosyntactic features are translated into a phonological content by means

9. The last three examples are taken from Alhalawaani (1987: 270, 271).

of the Subset principle stating that VI is inserted if it satisfies all or the subset of the features of the terminal node. The default VI satisfying the most features of the terminal node is inserted. However if there is a more narrowly specified VI, then the Elsewhere condition inserts it. After insertion takes place, readjustment rules apply (Siddiqi 2006). For example, a root merges with a verbal head in syntax. Then the verb merges with a tensed Inflection head (INFL) that is specified with [-ed] or [-ø]. At the morphological structure, the default [-ed] is inserted for a regular verb like *play*, whereas [-ø] morpheme is inserted for an irregular verb like *sing*. Then a readjustment rule changes [ɪ] to [æ] in *sang*.

Finally, the Encyclopedia at the syntax-semantics interface is the place for special meaning that is associated with idioms like *kick the bucket*. Moreover non-linguistic knowledge of the world is stored in the Encyclopedia.

DM has some problems. For instance, the distinction between special and regular meaning remains one of the debatable issues in DM. Siddiqi (2006) suggests that the Encyclopedia does not really distinguish between idioms and non-idioms since they all express idiosyncratic non-compositional meaning. To illustrate, *terrific* does not compose its meaning from *terror-ify-ic* but it has a meaning similar to *great*. In other words, *terrific* is like an idiom with non-compositional meaning. On the other hand, Kelly (2013) admits that DM does not explain how regular meaning is derived and where it can be added in the derivation. To solve this problem, he proposes that the Encyclopedia can handle special meaning as well as regular meaning and that these different meanings compete for insertion based on contextual factors. Wunderlich (2008) refers to another problem of DM with regard to the analysis of non-concatenative morphology. That is, a morphological word or root is not derived by syntax but by prosodic units like the mora, the foot, the syllable and so on. As illustrated by McCarthy & Prince (1990), discussed in § 1.1.1 above, Arabic broken plurals for example is formed by iambic foot as the examples in (1) show.

### 2.3 Dual morphological systems

Pinker argues for dual-route in inflectional morphology: full-form storage of words and rule-based. For example, the irregular past (i.e. *sank* of *sink*) is stored at the lexicon and therefore memorized while the regular past suffix is attached to a stem by a productive rule (i.e. *play+ed* → *played*). Therefore, the two systems differ in *compositionality*: the irregularly stored word in the lexicon is non-compositional, treated as a single lexical item. However the rule-based is composed from a stem and a suffix (Pinker 1999; Berent & Pinker 2008).

Pinker assumes that language is a combination of memorized lexical words and grammatical rules. The difference between the lexicon and grammar rules reduces

to morphological regularity. Even though the two mental systems seem to be distinct, they coherently work together and are constrained by a *blocking* principle that the child learns at an early age. That is, if the word’s past tense can be retrieved from memory (i.e. *sang*), the rule is blocked; otherwise, the productive rule takes place (i.e. *played*) (Pinker 1999; Berent & Pinker 2008).

The dual representation of morphology in the lexicon and syntax finds its roots in early generative grammar: Chomsky’s seminal work “Remarks on nominalization” (1970). Chomsky distinguished derived from gerundive nominal and proved that the former was derived lexically while the other syntactically using different pieces of evidence (Chomsky 1970).

The influence of Chomsky’s work was massive in reconsidering morphology as a basic component in generative linguistics. Before then, morphology was relegated to either syntax or phonology (Webelhuth 1995). Morphological research soon after “Remarks” went in some different directions within Chomsky’s generative tradition and led to the introduction of the *lexicalist hypothesis* like Halle (1973), Selkirk (1982) et al. Unlike Chomsky’s assumption in (1970) that derivational morphology is performed in the lexicon whereas inflectional morphology is in syntax (weak lexicalist), strong lexicalists argue that both types of morphology exist in the lexicon. Moreover, the insights of “Remarks on nominalization” founded other rivalling strong lexicalist generative theories such as Driven Phrase Structure Grammar, Generalized Phrase Structure Grammar, Lexical-Functional Grammar, and Kiparsky’s Lexical Phonology and Morphology (Webelhuth 1995).

### 2.4 Lexical phonology and morphology (Kiparsky 1982)

Kiparsky adopted a strong lexicalist approach to morphology in which derivations as well as inflections are all performed at the lexicon into three basic levels:

**Table 1.** Level-ordering morphology (adapted from Gordon 1985: 75)

	Examples	Properties
Level 1	+ion, +ous, +ity, +th, in+ mice, oxen, scissors	Derivational, irregular, semantically idiosyncratic, host deforming, stress shift, vowel reduction, unproductive
Level 2	#ness, #ism, #er, #ist, un# Compounding	Derivational, non-deforming, (more) semantically predictable, productive
Level 3	#s, #ed, #ing	Regular inflections, non-deforming semantically predictable

As Table 1 shows, Level 1 contains irregular inflections and primary derivational affixes that alter the bases they attach to. The semantics of such derivations is unpredictable. Level 2 includes secondary derivations and compounding. Regular inflections are added at Level 3. The ordering of these levels is strict. Therefore primary derivations apply first before secondary derivations and compounding is produced prior to attachment of regular inflections. As a result, compounds like *\*rats-infected* are ungrammatical because plurals may not appear inside compounds for compounding applies at Level 2 and as the word comes to Level 3 the plural may not be inflected inside the compound. However a compound like *mice-infested* is possible due to the fact the irregular plural is done at Level 1 and becomes therefore available for compounding rule at Level 2. Further details of compounds are discussed in the following section.

## 2.5 Compound and plurality

I review in this final subsection some studies discussing the constraints on number specification for the non-head of the compound and its impact on morphological level-ordering.

Berent & Pinker (2007) confirmed the fact that irregular plurals are allowed inside compounds unlike the case with regular plurals. In support of this view, they performed four experiments using compounds with irregular plurals that are phonologically unattested in English and compare them with attested regular plurals. They concluded that morphological distinction between regulars and irregulars determines the prevention of regular plurals inside compounds and not the phonological unfamiliarity of the word. In another related study, Pinker (1999) argued that the difference between irregular and regular plurals has to do with two mental systems that the human mind uses for processing. Namely, irregulars are stored in the memory whereas the system of symbolic computation generates regular plurals by means of rules joining the inflection to the base.

Gordon (1985) performed an experiment on English-speaking children from 3 to 5 exploring the relationship between compounding and inflection. He found that children, like adults, did not include regular inflections inside compounds but they were able to use irregulars as non-head position of compounds. Gordon suggested that children have an innate constraint preventing the occurrence of regular inflection inside a compound. In order to examine effect of this constraint, Alegre & Gordon (1996) showed 30 children a pair of pictures: one picture showed the eater was red while another picture illustrated that the rat was red. The researchers used syntactic recursion with the compound to contrast a plural compound like *red rats eater* and a singular like *red rat eater*. The results indicated that for the plural compound the children only chose the recursive interpretation or the noun pattern

(NP) interpretation: [[red rats] eater] in which the rats are red rather than the lexical compound or the non-recursive interpretation: [red [rats eater]] in which the eater is red. As for the singular compound, the children showed no preference between the recursive vs. non-recursive interpretation. This proved that children are innately aware of the morphological distinction or the level ordering between compounding and inflection at an early age and violations would be licensed through syntactic recursion.

To examine the genetic basis of the internal modularity or dual system of the language faculty, Clahsen & Almazan (2001) tested the plural formation inside a compound and noun plurals in Williams Syndrome (ws) subjects and Specific Language Impairment (SLI) children. They found that ws subjects correctly produced rule-based plurals and overgeneralized it to non-head of the compound while they failed to produce irregular plurals. On the other hand SLI subjects were able to use irregular plurals but were worse in using regular plurals. Thus ws and SLI differ in the lexical access of words and the rule-based component of language respectively. This investigation supports the distinction between the (irregular) lexical representation and the (regular) rule-based computational component of language.

The prevention of regular inflection between the two noun members of a compound was the basis upon which Li (1990) established his well-known generalization on verb incorporation (VI) stating that a verb necessarily takes a bare Verb Phrase (vp) as its complement.<sup>10</sup> According to Li, no inflection element should intervene between the two head verbs as the case in compounds *\*rats-infested* in which a (plural) inflection cannot intervene between the two compound noun parts. Li explained the ban of inflection as a result of violation condition compound (C) of the binding theory (1990).

Even though the studies in the literature agree for the most part that no plural inflections may appear inside a compound, Sneed (2002: 617) observed that this claim is not absolute. She noticed that there are many counter-examples like: admissions department, parks commissioner, assists kings. In fact, she believed there are many more examples which she was able to get based on searches of the General News category in the Lexis-Nexis Academic Universe for specific compound types. Other linguists reported other examples like: programs coordinator, buildings inspector (Clahsen et al. 1995: 117). These counter-examples as well as the Arabic examples in (7) pose problems for level-ordering morphology and for Li (1990). I provide a possible account of these data in § 3.

---

10. Li proposes the following generalization and argues that it is a necessary condition on VI:

A necessary condition on VI is that the matrix verb must be able to take a bare vp as complement.  
(Li 1990: 404, Number 12)

Li proves that this generalization is a consequence of Binding Condition C.

### 3. Data

In this section, I provide examples of Arabic compounds. Interestingly, the Arabic compounds differ from their English counterparts in that the non-head of the compound is plural in Arabic; it is singular in English:

(7) Arabic compound		English compound
raajimatu al-ṣwaariix	‘rockets launcher’	rocket launcher
haamilatu al-ṭaaʔiraat	‘aircraft carrier’	aircraft carrier
qanṣu al-baḡar	‘men hunt’	manhunt
naaqilatu al-jund	‘troops carrier’	troop carrier
(saaruux) ṣaabir	‘intercontinental	intercontinental ballistic
al-qaaraat	ballistic missiles’	missile
maaniṣatu al-sawaasiq	‘lightnings rod’	lightning rod
naaṭiḡhatu al-saḡaab	‘skies scraper’	skyscraper
ṭaakil al-ḡaḡaraat	‘insects eater’	insectivore
ṭaakil al-naḡl	‘ants eater’	anteater
mutasaddid al-maḡaam	‘multitasks’	multitask
mutasaddid al-ṭaḡzaab	‘multi-parties’	multi-party

As observed in (7), the non-head of the Arabic compounds are plural unlike the case in English. As a result, I marked the plurality in parentheses. I focus in this paper on this contrast and attempt to provide an account explaining the difference between Arabic and English compounds.

### 4. Lexical analysis of ASC

In this section, I provide some pieces of evidence for the lexical nature of ASC in Arabic and compare it to their counterpart compounds in Hebrew. Then I offer an analysis that can better explain the morphological and semantic properties of the data.

#### 4.1 ASC: lexical or syntactic?

I argue that ASC in Arabic is lexically formed based on different pieces of evidence. In order to fully understand the nature of compounds it might be fruitful to examine other phenomena such as construct state structures (cs) in Arabic and Hebrew and observe how these data relate to the discussion of compounds in Arabic. Below I provide five pieces of evidence supporting the lexical nature of compounds.

#### 4.1.1 The use of modifiers

To begin with, ASC do not permit the modification of part of the compound:

- (8) a. \**rasat haamilatu al-taaʔiraati al-nafaaḥati fii*  
 stopped-it carrier.SG-NOM the-aircraft-PL-GEN the-jet-GEN at  
*miinaaʔi Jeddah.*  
 port-GEN Jeddah  
 ‘The jet aircrafts carrier stopped at Jeddah’s port.’  
 b. \**ʔatlaqat raajimaatu al-swaariixi*  
 fired-it launcher.SG-NOM rocket.PL-GEN  
*al-kaḥiifati al-naara.*  
 the-intense-F-GEN the fire-ACC  
 ‘The intense rockets launcher fired its shots.’

The ungrammaticality of (8) is due to the modification of *alnafaaḥati* and *alkaḥiifati* to the non-head or complement element of the compounds: *al-taaʔiraati* and *al-swaariixi* respectively. These compounds are treated as one morphologically complex unit that can be modified as a whole and hence partial modification cannot work. Had the compound been formed syntactically, such partial modification of the compound would be possible. In fact, if we turn the above compounds into CS constructions the behaviour of the modifier may change:

- (9) a. *zinna ḥamla al-taaʔiraati al-nafaaḥati fii*  
 that carrying-ACC the-aircraft-PL-GEN the-jet-GEN in  
*safīnatin muṣaqqadun.*  
 a ship-GEN is.complicated-NOM  
 ‘Carrying jet aircrafts in a ship is complicated.’  
 b. *yaquumu al-israaʔiiliyyuun bi-rajmi*  
 start-SG-NOM the Israeali-NOM.PL PREP-launching-GEN  
*al-swaariixi al-kaḥiifati ʕala al-filastīniin.*  
 the-rockets.GEN.PL the intense-F-GEN on the Palestinian.GEN.PL  
 ‘The Israelis started launching intense rockets on the Palestinians.’

Unlike the case in compounds, adjectives modify *al-taaʔiraati* and *al-swaariixi* as the examples in (9) show. The difference between the compound and the construct state stems from their morphological representation in syntax. That is, a construct state in (9) is a phrase; hence the adjective can access the complement. However the compound functions syntactically as one morphological word preventing the adjective from modifying non-head part of the compound as indicated by the ungrammaticality of the examples in (8).

Interestingly, Hebrew shows exactly the same asymmetry between compounds in CS constructions as in Arabic. Borer (1988: 49, Example (9)) provides the following examples:

- (10) a. \**gan yeladim ktanim; shomer mitzvot yehudiyot*  
 garden children little; guard commandments Jewish  
 ‘\*a kindergarten for young children; \*practicing Jew’  
 b. *gan peyrot tropyim; shomer maxoniyot gnuvot*  
 garden fruit tropical; guard cars stolen  
 ‘a garden of tropical fruit; a guard of stolen cars’

In (10a), the adjectives *ktanim* and *yehudiyot* cannot modify the complement of the compounds: *yeladim* and *mitzvot* respectively. On the other hand, modification (*tropyim* and *gnuvoṯ*) is possible to the complements (*peyrot* and *maxoniyot*) of the construct states in (10b).<sup>11</sup> This is a strong reinforcement to the lexical nature of one complex morphological structure of compounds as opposed to the phrasal structure (i.e. two independent morphemes) of the construct states not just in Arabic but also in Hebrew.

#### 4.1.2 Referentiality

Reference to the complement of the compound is not possible unlike the case in cs construction. Consider the following examples:

- (11) a. \**sabarat naaqilatu al-ljunuudi<sub>i</sub> al-huduuda wa*  
 crossed-it carrier.SG-NOM the-troop.PL-GEN the border.PL-ACC and  
*kaanuu<sub>i</sub> mustasidiina li al-maṣṣrakati.*  
 were-they ready-ACC.PL for the battle-GEN  
 ‘\*The troop<sub>i</sub> carrier crossed the borders and they<sub>i</sub> were ready for the battle.’

11. A reviewer suggests that not only is the complement modification not possible in cs but also the head cannot be modified proving that the construct is a compound. However this claim is problematic. On one hand, the lack of head modification is observed in compounds as well as in phrasal constructs. To illustrate, Borer (2009) observes that the head in a phrasal construct cannot be modified \**beyt xadaš mora* ‘house new teacher’ as in compounds \**beyt xadaš xolim* ‘house new patients’. Thus the impossibility of head modification is not an evidence for making a construct a compound. On the other hand, the reviewer proposal will obliterate any possible distinction between compounds and phrasal constructs. More specifically, Borer (1988; 2009) distinguishes phrasal constructs from compounds by different semantic and syntactic and aspects. I only refer to some of them and the reader may check the sources directly. For instance, phrasal constructs allows referentiality of the complement of the construct (the topic of the following section) and the construct is associated with semantic transparency (I discuss this later) unlike the case in compounds. As a result, Doron & Meir (2013) conclude that compounds are less productive and thus generated in the lexicon whereas phrasal constructs are productive and produced in the syntax. Now the explanation of why the head in a compound may not be modified is beyond the scope of this paper, but I shall just refer to Sadock’s (2000) analysis that it stems from word ordering facts in Hebrew in which the modifier follows the noun and it is associated with it.

- b. *zinna naqla al-junuudi<sub>i</sub> alaḏiina sayuhaaribuun<sub>i</sub>*  
 that carrying-ACC the-troop.PL-GEN who FUT-fight-NOM.PL  
*al-aṣadduwa ṣabra al-ḥuduuda xatīirun.*  
 the enemy across the border is.dangerous-NOM  
 'Carrying troops who will fight the enemy across the border is dangerous.'

Reference of the pronoun in *kaanuu* (i.e. -*uu* waaw aljamaaṣah) to the complement of the compound (*alljunuudi*) is precluded in (11a) as indicated by the ungrammaticality. The index (<sub>i</sub>) represents referentiality between the pronoun and the noun. As for the construct state in (11b), the pronoun (waaw aljamaaṣah) in *sayuhaaribuuni* refers to *alljunuudi* and both have the same index (<sub>i</sub>). Again the contrast in reference is due to the lexical nature of compounds vs. the syntactic structure of the construct state. Hebrew also shows the same contrast (Borer 1988: 50, Example (11)) as the case in Arabic:

- (12) a. *hu bana lanu shney batey xolim ve-'exad le-zkenim.*  
 He built for-us two houses sick and-one for-old(s)  
 (cf. *beyt xolim; beyt zkenim*  
 house sick house olds  
 hospital retirement home)  
 b. *hu bana li shney batey sec ve-'exad mi-plastik.*  
 He built for-us two houses wood and-one from-plastic

*beyt xolim* and *beyt zkenim* are compounds and therefore the pronominal element *one* cannot be used to refer to part of the compound (house) because *house* works as a lexical integral part of the compounds associated with specific meaning that cannot be referred to by *one* in (12a). On the other hand the construct state in (12b), part of the construct state (house) can be referred to by *one*. So you have *batey sec* (house of wood) and *batey plastik* (house of plastic) and *one* can refer to the *house* part of the structure indicating that the construct state is fully decomposable into two morphemes and hence is subject to syntactic rules like reference unlike the situation in a compound in which the word is one morphological unit hence preventing reference rule from accessing the internal part of the structure.

#### 4.1.3 Conjunction of the complement

Conjunction is another testing ground for whether the compound is one complex word or not. Let us consider the following examples:

- (13) a. *\*rasat ḥaamilatu al-ṭaaʔiraati wa al-junuudi*  
 stopped-it carrier.SG-NOM the-aircraft-PL-GEN and the troop.PL-GEN  
*fī al-baḥr al-mutawassiṭ.*  
 at the Mediterranean Sea  
 'The aircraft carrier and the troops stopped at the Mediterranean Sea.'

- b. \**ʔaakil al-namli wa al-ḥaʕsaraati*  
 eater-NOM the ant-GEN and the insect-PL-GEN  
*ḥayawaanaan muxtalifaan.*  
 are.animal-DU-NOM different-DU-NOM  
 ‘\*Ants eater and insects are two different animals.’

The examples of (13) are ungrammatical because the conjunction of complement of compound is not allowed. The resulting meaning of (13a) is that there is a carrier for both aircrafts and troops; there be no way to have two compounds readings: aircraft carrier and troop carrier and that is why (13a) is bad.<sup>12</sup> The same reasoning applies to (13b) suggesting that syntactic operations like conjunction is not applicable to parts of the compound unlike the case in state nominals:

- (14) *zinna ḥamla al-ṭaaʕiraati wa al-ljunuudi*  
 that carrying-ACC the-aircraft-PL-GEN and the troop.PL-GEN  
*ṣamaliyyatun ṣaṣbatun.*  
 is.process-NOM is.difficult-NOM  
 ‘Carrying aircrafts and troops is a difficult process.’

Conjunction of the complement of the state nominal is possible unlike the case in compounds. This contrast between compounds and state nominals is also confirmed in Hebrew in which state nominals allow conjunction of their complement unlike compounds. Let us consider Hebrew data including conjunction from Borer (1988: 49–50) where (15a) shows compounds and (15b) construct state nominals:

- (15) a. \**gan yeladim ve-xayot*  
 garden children and animals  
 ‘a kindergarten and a zoo’  
 b. (cf. *gan yeladim*; *gan xayot*)  
 garden children; garden animals  
 kindergarten; zoo)  
 c. *shomer batim u-mexoniyot*  
 guard houses and cars  
 ‘a guard of houses and of cars’

Both compounds (*gan yeladim* and *gan xayot*) do not allow the conjunction of their possessor (*gan*) as opposed to construct states (*shomer batim*) that allows the conjunction of the complement (*batim*) to generate another construct state via conjunction (*shomer mexoniyot* ‘guard of cars’).

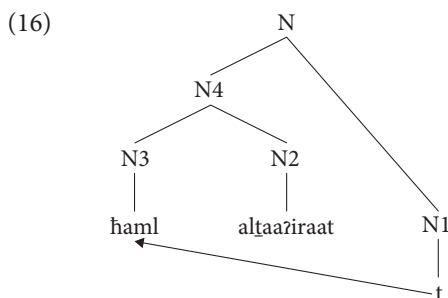
12. Another reading of (13a) in which ‘a carrier of aircrafts and troops’ is ungrammatical because the conjunction of the compound complement is not possible.

#### 4.1.4 *Semantic idiosyncrasy*

The compound is associated with specific meaning while the construct nominals involve predictable meaning. For example, *naaqilatu iljunuudi* ‘troops carrier’ is semantically associated with a special military vehicle that transports troops. However in construct state nominal like *naqalu iljuundi* ‘transporting troops’, the meaning is *compositional*, i.e. composed of the meaning of *naqal* and *iljuund*. As a result, the transporting of troops involves all possible means of transportation via plane, ship, car, etc. Another example is *raajimaatu asswaariixi* ‘rockets launcher’ which is semantically restricted to special weaponry tool launching rockets, whereas the semantics of the construct nominal *rajamu alsawaariixi* is general and hence suggests launching rockets through any possible means, be it by means of a highly-developed machine or a rudimentary tool. The opacity of semantics is one of the important evidence for the lexical nature of compounds vs. the semantic transparency of construct nominals which are non-lexical. Semantic opacity was one of the pieces of evidence that Chomsky (1970) presented in favour of the lexical formation of English gerundive nominals which paved the way for a new understanding of morphology and the role of the lexicon in morphology.

#### 4.1.5 *Theta-role satisfaction*

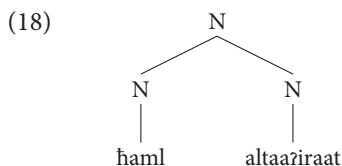
Based upon the evidence presented above, ASC is lexically formed as one complex structure while construct state nominals (cs) is syntactically formed as a phrase. One more argument in support of the distinction between ASC and cs is related to theta-role assignment. For the sake of the argument, let us assume that ASC is formed syntactically by means of movement:



The movement of *haml* from the lower N (N1) violates c-command condition as defined below (Chomsky 1986: 8):

- (17) A c-commands B if and only if A does not dominate B and every node that dominates A also dominates B.

*haml* starts in N1 and moves to N3 leaving a trace behind to adjoin to N2. According to (17), the first condition of c-command is met: N3, to which *haml* moves, does not dominate the trace because they are two independent nodes. However N4 dominates *haml* but not the trace. Thus the trace is not c-commanded; hence the movement of *haml* is not possible. This is a welcome result since we do not want a syntactic principle (e.g. movement) to access part of the word while ignoring the other part since the whole word functions syntactically as a morphologically complex word. Now it can be said that the compound may be syntactically derived but not through movement as in (16) but through merger as in (18):



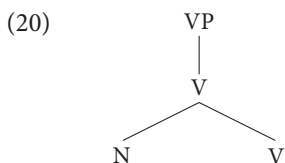
Let us assume that the compound is syntactically decomposed just as CS into two nouns as (18) illustrates. Even though the (four) pieces of evidence presented in this section preclude such an assumption, I shall add here another argument warning against a syntactic merger of ASC. To illustrate, let us consider the following examples (Sproat 1985: 222, Examples 112a–b):

- (19) a. \**John dog-kills*. (i.e. *kills dogs*)  
 b. \**Mary deer-hunts*. (i.e. *shot a deer*)

Sproat (1985) argues that the above examples are bad because of theta-role assignment. Namely, *dog-kill* and *deer-hunt* are not possible verbal or synthetic compounds; hence we cannot use them as verbal compounds in which *dog* and *kill* are grouped together deriving a verb. The thematic-role is assigned inside the synthetic compounds<sup>13</sup> from the head to the non-head. However because these are not possible synthetic compounds, if these are used as verbal compounds then we shall have a violation of theta-role assignment to the subject (i.e. *John* and *Mary*). The (external) theta-role is only assigned by the VP which is outside the compound:

---

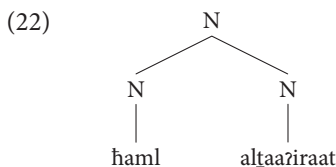
13. The types and mechanism of thematic-role assignment in synthetic compound is discussed in detail in the next section. For now, the discussion of theta-role is brief and general.



The verbs in (19) are non-synthetic compounds containing (external) subject argument which is assigned a theta-role by the *vp*. However the only theta-role possible, in case the verbs are treated as synthetic/verbal compounds, is assigned to *dog* and *deer* leaving the subject without a theta-role. Sproat's analysis can be extended to ASC. Let us consider the following example:

- (21) a. \**aljayf*      *ħaml-attaaʔiraat*.  
           The army carry-aircrafts.  
       b. \**aljayf*      *rajm-asswaariix*.  
           The army launch-rockets.

It is not permissible to use the verbs as synthetic compounds in which *ħaml* and *al-taaʔiraat* as one lexical verb via lexical merger. I shall extend Sproat's thematic-role analysis to rule out examples of (21). As a consequence, ASC may not be derived syntactically via syntactic merger due to thematic role violations: the verb does not have a theta-role to assign to their subjects. The only option for ASC is to be derived lexically as one morphological unit in which the theta-role is lexically satisfied inside the compound as will be explained in the next section. Finally, it is interesting to compare ASC to CS in terms of theta-role assignment. CS is represented syntactically as two parts:



*ħaml* assigns its theta-role to the non-head *attaaʔiraat*. Because CS involves two parts in syntax derived by syntactic merger, syntactic operations like modification, reference, conjunction can access any part of CS but not the internal structure of the compound as we explored in this section.

## 4.2 Lexical analysis of ASC

I argue above that ASC is lexically formed based on the use of modifiers, reference, conjunction, semantic opacity, and theta-role representation.<sup>14</sup> Therefore in this section, I provide a lexical analysis that explains the morphological, syntactic, and semantic properties of ASC. I first present the lexical theory and then explain how it applies to theta-role assignment inside the compound and afterwards discuss the number specification of the non-head in the compound.

### 4.2.1 *Lexical structure framework*

I adopt Lieber's Lexical Structure Framework (1983):<sup>15</sup>

---

14. A reviewer assumes that the analysis adopted here is a moribund Chomskyan theoretical model. As a matter of fact, the framework used in the analysis of the data follows the lexicalist approach by which compounds are lexically formed as proposed in other models that the reviewer suggests to be used like Construction grammar (Booij 2016) and Parallel architecture (Jackendoff 2002; 2007). The reviewer believes that such models are more successful than the Chomskyan model in dealing with compounds as they avoid the sharp division between the lexicon and grammar/syntax. Regardless of where and how complex morphological structures are formed, different models of language nonetheless observe a settled difference between a regular productive phrasal formation and irregular word formation. That is why in a non-Chomskyan theory as in psycholinguistic framework, Pinker (1999) argues that there are two mental subsystems of language: a memory system in which irregular words are listed in the lexicon and rule system by which regular words are combined via rules. Even in Construction grammar, Booij (2016) distinguishes between morphological construction (word formation) and phrasal construction (phrasal formation) even though he handles both words and phrases by lexical schematic rules. Moreover, Jackendoff (2009) realizes that compounds may be listed lexically while others are productive and hence do not need to be listed but instead are derived by a rule. As a consequence, Jackendoff, among others, suggest that the distinction between the lexicon and syntax should not be strict in order to allow for productivity and idiosyncrasy of a complex morphological structure as a compound. Main stream Grammar (Chomskyan theory) has realized this challenge and attempted to provide an account of it. For example, Li (2005) shows that this strict distinction between lexicon and syntax cannot explain empirical data that show complex morphological structure belonging to both levels of grammar. Therefore, he presents a theory which reconciles both lexical and syntactic aspects of word formation. For more specific details, see Li (2005). The analysis adopted in this study is influenced by Li's proposal and hence assumes a loose distinction between lexicon and syntax since both play a role in word formation.

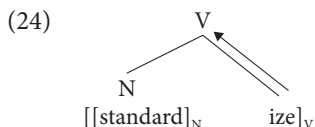
15. Lieber uses two different sets of principles: lexical category principles and argument-linking to account for the behavior of synthetic compounds. I combine these principles together in a Lexical Structure Framework.

## (23) Lexical structure framework:

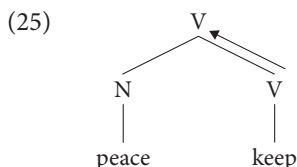
A compound is analysed by means of the following principles:

1. Principles of Feature Percolation (Lieber 1983: 252–254):<sup>16</sup>

- a. All features of an affix morpheme, including category features, percolate to the first branching node dominating that morpheme. For example:



- b. If two stems are sisters (i.e. they form a compound), features from the right-hand stem percolate up to the branching node dominating the stems. For example:



## 2. Argument-linking principle (Lieber 1983: 258):

- a. In the configuration  $[ ]_{\alpha}$  or  $[ ]_{\alpha} [ ]_{\left\{ \begin{smallmatrix} v \\ p \end{smallmatrix} \right\}}$  where  $\alpha$  ranges over all categories,  $\left\{ \begin{smallmatrix} v \\ p \end{smallmatrix} \right\}$  must be able to link all internal arguments.
- b. If a stem  $[ ]_{\alpha}$  is free in a compound which also contains an argument-taking stem, i.e. as a Locative, Manner, Agentive, Instrumental, or Benefactive argument.

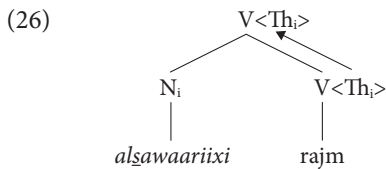
The Lexical Structure Framework consists of two sets of principles: principles (1) of (23) specifying how the class is determined within compounds and non-compounds and principles (2) determining how the thematic role is assigned from a head to non-head in a synthetic compound. To illustrate, Lieber (1983) assumes that a class or part of speech is a feature of some morphological head that percolates up to determine the category of the whole word as the diagram (24) illustrates. Hence he proposes some category or feature convention rules; I refer to two rules that are relevant to the main discussion of this study: one rule for the (non-compound) word with an affix in which the feature of that affix percolates up as in (24), another

<sup>16</sup> Lieber introduces four rules only two of which are relevant to the discussion of compounds in this study. Therefore I shall only include these two feature rules and ignore the other rules.

rule for the compound in which the right-most member is the head<sup>17</sup> and thus its feature percolates up as in (25). The argument-linking principles of (2) in (23) organise the thematic relationship between a head and non-head in a compound. Namely, if the head (v or p) has an internal argument,<sup>18</sup> then this argument is linked inside the compound. When the argument of the compound is not internal, it is interpreted as free and becomes argument which involves locative, agentive, instrumental, manner, or benefactive argument. E.g. in *handmade*, hand is not internal argument but a free instrumental argument. Another example is *panfry* where pan is a locative argument.

#### 4.2.2 *Theta-role assignment in ASC*

The synthetic compound like *raajimaatu alṣawaariixi* ‘rockets launcher’ involves a thematic relation between a head *raajimaatu* and the non-head *alṣawaariixi*. According to principle (2) of (23) the head derived from a verb assigns a theta-role to its internal argument *alṣawaariixi* lexically:



The thematic role is symbolized by *Th* in the angled bracket and has the index (<sub>*i*</sub>) which is similar to the index of N suggesting that *alṣawaariixi* is its internal argument. The thematic role percolates up to v and it is linked; hence the theme role is assigned. From the discussion of § 4.1.5, the theta-role may not be projected syntactically through movement of the non-head or merger of *alṣawaariixi* to the head because as we explained in case of movement, c-command will be violated. In case of merger, theta-role will be violated as shown in the examples in (8) above once we treat the compound *raajimaatu alṣawaariixi* as syntactically decomposed into two units representing the thematic role between the head and the non-head

17. Lieber builds on the right-hand rule (RHR) proposed by Williams (1981: 248). Williams argues that a head of a morphologically complex word is to the rightmost member of the word. I shall adopt this rule and it therefore will be part of the discussion of ASC. Of course, this rule is not without exception; however I shall not discuss the exception here.

18. Lieber (1983: 257) assumes internal arguments to be lexically specified obligatory arguments like object except the subject. The subject has a different nature and is never linked inside the compound. The theta-role (external argument) is satisfied by the predicate vp. For more details see Lieber (1983, Footnote 10).

of the compound. As a result, the theta-role in synthetic compound in (26) is only assigned lexically and then merged lexically as one complex morphological unit in syntax unlike the case in CS:

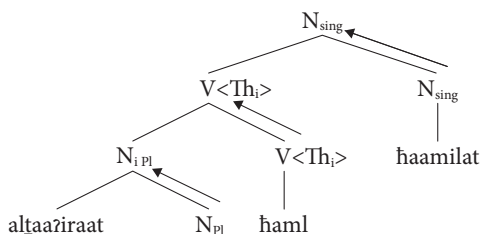
The theta-role of CS in (26) is represented in syntax as evidenced by the accessibility of syntactic operations like modification, reference, and conjunction to the parts of CS unlike the compound.

#### 4.2.3 The number specification in ASC

One of the most interesting aspects of ASC is the plural non-head of the compound. There is a general tendency at least in English for compounds to have a singular or irregular non-head member but never a regular non-head as reported by Gordon (1985), Alegre & Gordon (1996), Pinker (1999), Berent & Pinker (2007), et al. Interestingly, the Arabic data in § 2 includes compounds with both regular as well as irregular non-heads. Starting first with compounds involving irregular non-heads, *raajimaatu alṣawaariixi* is similar to the English compound *mice-infested* in which both have irregular non-head. Both examples are produced in accordance to level-ordering phonology. The irregular plural is done at Level 1 and becomes therefore available for compounding rule at Level 2. The problem stems however from other examples of ASC in which the non-head of the compound is a regular plural which directly violates the predictions against such types of compounds; thus it poses a problem for level-ordering phonology. Some of these examples are: *ʔaakil alḥaʔaraat* ‘insects eater’, (*ṣaruux*) *ṣaabir alqaaraat* ‘intercontinental ballistic missile(s)’, *mutaʔaddid aljinsiyyat* ‘multinational’.

Arabic is one of the languages that distinguishes between regular and irregular plurals just like English. In fact, Al-Dobaian (2014) argues that the (regular) sound plurals are derived at the syntax while the (irregular) broken plurals are formed in the lexicon. Evidence follows from the lexical access, morphological productivity, semantic distinctions of number, and the lexical representation that differentiate the two types of plurals in Arabic. Nonetheless and despite the apparent differences between the regular and irregular plurals in Arabic, the asymmetry between them is neutralized when they become part of the non-head element of the compound. That is, they are lexically derived. Let us consider the following diagram:

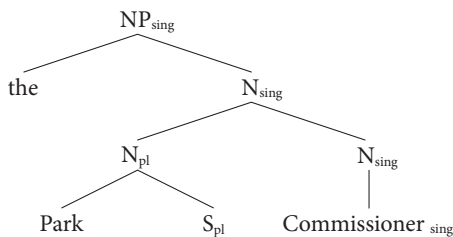
(27)



According to (23), Principle of Feature Percolation Convention (a) determines that the plural feature (pl) is the head and thus this feature percolates to the first branching node making the N plural and the affix *-aat* is attached forming *al-taaʔiraat*. After the noun gets the plural inflection, it is then merged lexically into another head *hamil* and thus forming the compound. According to (23), Principle of Feature Percolation Convention (b) states that the category of the head *haml* percolates up and the internal theta role of *al-taaʔiraat* is linked and satisfied inside the compound. The compound is then turned into a noun and the noun *haamilat* has singular feature as its head which percolates up in conjunction with principle feature percolation (a). Rules of syntax access only the features of the head of the compound: singular number and the noun category.<sup>19</sup> The information on how the compound gets the features and how the theta-role is assigned is not the business of syntax. Now suppose that the non-head compound member is irregular plural as in (26), the same plural mechanism applies lexically. So *al-saaruux* ‘rocket’ is headed by a plural feature percolating up to the branching N as in (27) to form *al-sawaariixi*. Thus the plural rule feeds the compound rule.

But how can we allow the presence of inflection inside a compound even though studies in the literature agree for the most part that plural inflections may not appear inside a compound? As we explained in § 1, Sneed (2002: 617) observed that this claim is not absolute. She noticed that there are many counter-examples like: admissions department, parks commissioner, assists kings. There were some proposals to offer solutions to such examples. For example, the word is pluralized and the regular inflection is added; then the word is returned to the lexicon for a

19. The number and theta-role checking of the non-head is performed lexically and are not accessed by rules of syntax while the features of the head of the compound are seen by the syntax. Di Sciullo & Williams (1987) argue that rules of morphology determine that the only features of the word that can be seen by syntax are the features of the head of the word and not non-head features. For example, *parks commissioner* has the following morphological diagram (Di Sciullo & Williams 1987: 49):



Di Sciullo & Williams argue that the plurality feature of *park* is the head of the noun *park* and it percolates up to the higher N and syntax cannot access this feature. However, the head singular number can be seen by syntax and therefore the compound is singular. I extend the same analysis to the compound in (27). For more details, see Di Sciullo & Williams (1987: 48–52).

compound rule (Sneed 2002: 622). But such a proposal weakens the level-ordering phonology because the model will not be able to distinguish a bad compound *\*rats-eater* from possible compound *parks commissioner*. Further evidence for the interaction of the regular plural and compounding comes from the study of Alegre & Gordon (1996). They illustrate that the interaction between the lexicon and syntax is licensed by syntactic recursion. They showed 30 three-year-old children pairs of pictures: one picture showed the eater was red while another picture illustrated that the rat was red. The researchers used syntactic recursion with the compound to contrast a plural compound like *red rats eater* and a singular like *red rat eater*. The results indicated that for the plural compound the children only chose the recursive interpretation or the NP interpretation: [[red rats] eater] in which the rats are red rather than the lexical compound or the non-recursive interpretation: [red [rats eater]] in which the eater is red. As for the singular compound, the children showed no preference between the recursive vs. non-recursive interpretation.

ASC combines both dual morphological processes: regular plural (the non-head of the compound) and the irregular (compound). Usually a word can be either irregular and be formed at the lexicon or be regular and be derived at the syntax. Therefore it is interesting to ask how a word can then be able to combine two morphological processes at the same time. Pinker (1999: 279) suggests that words may start as regular produced by productive rules, but through time these rules die and words become relics of history. This can be extended to ASC in which the plural of non-head members is used to be derived by rules and then it becomes frozen and lexicalised.<sup>20</sup>

Finally, it is worth mentioning that even though the non-head of Arabic and English synthetic compounds have different number specifications they are nonetheless controlled by the principles of feature percolation and argument-linking principles in (23). As a result, the non-head member of the English compound *rat-eater* checks the singular number feature which percolates up to branching

---

20. A reviewer assumes that if the non-head of the compound is derived by regular rules then it is not derived lexically and hence the compound does not violate Kiparsky's level-ordering morphology. However this assumption cannot be substantiated because compounds are not freely produced. For instance, *\*haamilat alkaasaat* 'cups carrier' is not an attested Arabic compound. If the compound non-head involves a regular inflection then this compound would be grammatical. However the ungrammaticality suggests that there is a restriction on forming compounds with regular inflections like *haamilat al-taaʔiraati*. But how can *haamilat al-taaʔiraati* be grammatical while *haamilat alkaasaat* cannot be? The answer is that the former is a lexicalized compound while the latter is not. Hence the compound being at Level 3 blocks the addition of regular inflection at Level 3. So one way to avoid this problem is to treat the regular inflection as frozen lexicalized element and therefore treat it as irregular affix at Level 1 and then add the compound rule at Level 2.

noun while the non-head of the Arabic compound *ʔaakil lhafaraat* ‘insect(s)ivore’ checks the plural number feature which percolates up to branching noun.<sup>21</sup> Finally, ASC is atomic (in the sense of Di Sciullo & Williams 1987) or a lexical unit and hence syntactic rules may not be able to access the internal structure of the word.

## 5. Conclusion

In this study, I provided data of ASC which are problematic to level-ordering morphology of Kiparsky (1982) as well as Li’s generalization on verb incorporation (1990). The challenge of the Arabic compound structure stems from the morphological presence of regular plural inflection inside a compound: a problem for the morphological ordering relation between compounding and the regular plural inflection and also a problem for Li’s generalization that no inflection can intervene between two head nouns. I briefly reviewed some relevant previous studies. I argued that ASC is lexically formed based on the use of modifiers, reference, conjunction, semantic opacity, and theta-role representation. To reinforce the analysis, I compared Arabic to Hebrew in terms of compounds and construct state constructions. I provided a lexical analysis that explained the morphological, syntactic, and semantic properties of ASC. More specifically, I illustrated how the lexical analysis applied to theta-role assignment inside the compound and then discussed the number specification of the non-head in the compound of Arabic and English.

## Acknowledgements

This research is supported by a grant from the Research Centre for the Humanities, Deanship of Scientific Research at King Saud University. I would like to thank King Saud University for this grant. I would like also to thank anonymous referees for useful comments on earlier versions of the paper.

---

21. The difference in number specification of the compound non-head in Arabic and English is a parametric variation resulting from language-particular rules. A reviewer mentions that compounds like *admissions department*, *parks commissioner*, *assists kings* are problematic to the proposal of the parametric number variation of the compound non-head in English and Arabic. Sneed (2002) mentions these as counter examples to the general observation, made by researchers like Gordon (1985), Alegre & Gordon (1996), Pinker (1999), and Berent & Pinker (2007), that there is a tendency for compounds non-heads in English to be regular but never regular. Even though Sneed observes that non-heads of a compound can have a regular inflection, she is quick to note that these regular non-heads do exist but not as common as singular non-heads (2002: 618).

## Abbreviations

ACC	accusative
ASC	Arabic synthetic compound
BP	broken plural
C	compound
CS	construct state structures
DM	Distributed Morphology
DU	dual
F	feminine
FUT	future
GEN	genitive
HL	Heavy Light
( <sub>i</sub> )	index
INFL	inflection
NOM	nominative
NP	noun pattern
Nsing	singular noun
PL	plural
PREP	preposition
SG	singular
SLI	Specific Language Impairment
SP	sound plural
V	vowel
VI(s)	vocabulary item(s)
VI	verb incorporation
VP	Verb Phrase
WS	Williams Syndrome

## References

- Al-Dobaian, Abdullah. 2014. Towards a morphological theory: The case of Arabic broken and sound plurals. *Journal of Arts* 26(3). 19–36.
- Alhalawaani, Mohammed. 1987. *AlWaaḍiḥ fiiʿilm alṢarf* [Explaining morphology clearly]. Damascus: Dar alMaʿmuun lilturaath [Almamuun Heritage Publishing Company].
- Alegre, Maria A. & Gordon, Peter. 1996. Red rats eater exposes recursion in children's word formation. *Cognition* 60. 65–82. doi:10.1016/0010-0277(95)00703-2
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Berent, Iris & Pinker, Steven. 2007. The dislike of regular plurals in compounds: Phonological familiarity or morphological constraint? *The Mental Lexicon* 2(2). 129–181. doi:10.1075/ml.2.2.03ber
- Berent, Iris & Pinker, Steven. 2008. Compound formation is constrained by morphology: A reply to Seidenberg, MacDonald & Haskell. *The Mental Lexicon* 3(2). 176–187. doi:10.1075/ml.3.2.02ber

- Booij, Geert. 2016. Construction morphology. In Hippiusley, Andrew & Stump, Gregory (eds.), *The Cambridge handbook of morphology*, 390–423. Cambridge: Cambridge University Press.
- Borer, Hagit. 1988. On the morphological parallelism between compounds and constructs. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of morphology*, 45–65. Dordrecht: Foris.
- Borer, Hagit. 2009. Afro-Asiatic, semitic: Hebrew. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of compounding*, 491–511. Oxford: Oxford University Press.
- Chomsky, Noam. 1970. Remarks on nominalization. In Jacobs, Roderick A. & Rosenbaum, Peter S. (eds.), *Readings in English transformational grammar*, 184–221. Boston: Ginn.
- Chomsky, Noam. 1986. *Barriers*. Cambridge: The MIT Press.
- Clahsen, Harald & Almazan, Mayella. 2001. Compounding and inflection in language impairment: Evidence from Williams Syndrome (and SLI). *Lingua* 111(10). 729–757.  
doi:10.1016/S0024-3841(00)00047-4
- Clahsen, Harald & Marcus, Gary & Bartke, Susanne & Wiese, Richard. 1995. Compounding and inflection in German child language. In Booij, Geert & van Marle, Jaap (eds.), *Yearbook of morphology*, 115–142. Dordrecht: Kluwer.
- Di Sciullo, Anna-Maria & Williams, Edwin. 1987. *On the definition of word*. Cambridge: The MIT Press.
- Doron, Edit & Meir, Irit. 2013. Construct state: Modern Hebrew. In Khan, Geoffrey (ed.), *Encyclopedia of Hebrew language and linguistics*, 581–589. Leiden: Brill.
- Gordon, Peter. 1985. Level-ordering in lexical development. *Cognition* 21(1985). 73–93.  
doi:10.1016/0010-0277(85)90046-0
- Halle, Morris. 1973. Prolegomena to a theory of word formation. *Linguistic Inquiry* 4(1). 3–16.
- Halle, Morris & Marantz, Alec. 1993. Distributed morphology and the pieces of inflection. In Hale, Ken & Keyser, Samuel Jay (eds.), *The view from building 20*, 111–176. Cambridge: The MIT Press.
- Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198270126.001.0001
- Jackendoff, Ray. 2007. *Language, consciousness, culture: Essays on mental structure*. Cambridge: The MIT Press.
- Jackendoff, Ray. 2009. Compounding in the parallel architecture and conceptual semantics. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford handbook of compounding*, 105–128. Oxford: Oxford University Press.
- Kelly, Justin Robert. 2013. *The syntax-semantics interface in distributed morphology*. Washington, DC: Georgetown University. (Doctoral dissertation.) Retrieved from [https://repository.library.georgetown.edu/bitstream/handle/10822/558391/Kelly\\_georgetown\\_0076D\\_12290.pdf?sequence=1&isAllowed=y](https://repository.library.georgetown.edu/bitstream/handle/10822/558391/Kelly_georgetown_0076D_12290.pdf?sequence=1&isAllowed=y)
- Kiparsky, Paul. 1982. From cyclic phonology to lexical phonology. In Hulst, Harry van der & Smith, Norval (eds.), *The structure of phonological representations* (Part 1), 131–175. Dordrecht: Foris.
- Li, Yafei. 1990. X<sup>0</sup>-binding and verb incorporation. *Linguistic Inquiry* 21. 399–426.
- Li, Yafei. 2005. *X<sup>0</sup>: A theory of the morphology-syntax interface*. Cambridge: The MIT Press.
- Lieber, Rochelle. 1983. Argument-linking and compounds in English. *Linguistics Inquiry* 14(2). 251–285.
- McCarthy, John J. 1993. Template form in prosodic morphology. In Staven, Laurel Smith (ed.), *Papers from the Third Annual Formal Linguistics Society of Midamerica Conference*, 187–218. Bloomington: IULC.

- McCarthy, John J. & Prince, Allan S. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8(2). 209–283. doi:10.1007/BF00208524
- Pinker, Steven. 1999. *Words and rules: the ingredients of language*. New York: Basic Books.
- Ratcliffe, Robert R. 1990. Arabic broken plurals: Arguments for a two-fold classification of morphology. In Eid, Mushira & McCarthy, John (eds.), *Perspectives on Arabic linguistics: Papers from the Annual Symposium on Arabic Linguistics. Volume II: Salt Lake City, Utah 1988*, 94–119. Amsterdam: John Benjamins. doi:10.1075/cilt.72.07rat
- Ratcliffe, Robert R. 1997. Prosodic templates in a word-based morphological analysis of Arabic. In Eid, Mushira & Ratcliffe, Robert R. (eds.), *Perspectives on Arabic linguistics: Papers from the Annual Symposium on Arabic Linguistics. Volume X: Salt Lake City, 1996*, 147–172. Amsterdam: John Benjamins. doi:10.1075/cilt.153.10rat
- Ratcliffe, Robert R. 2003. Toward a universal theory of shape-invariant (templatic) morphology: Classical Arabic reconsidered. In Singh, Rajendra & Starosta, Stanley (eds.), *Explorations in seamless morphology*, 212–269. London: SAGE.
- Sadock, Jerrold. 2000. *Decomposing the construct*. Chicago: University of Chicago. (Manuscript.)
- Selkirk, Elisabeth O. 1982. *The syntax of words*. Cambridge: The MIT Press.
- Siddiqi, Daniel. 2006. *Minimize exponence: economy effects on a model of the morphosyntactic component of the grammar*. Tucson: University of Arizona. (Doctoral dissertation.) Retrieved from <http://ling.auf.net/lingbuzz/000514/current.pdf>
- Siegel, Dorothy C. 1974. *Topics in English morphology*. Cambridge: MIT. (Doctoral dissertation.) Retrieved from <http://www.ai.mit.edu/projects/dm/theses/siegel74.pdf>
- Sneed, Elisa. 2002. The acceptability of regular plurals in compounds. In Andronis, Mary & Debenport, Erin & Pycha, Anne & Yoshimura, Keiko (eds.), *Chicago Linguistic Society 38: The main session*, 617–631. Chicago: The Chicago Linguistic Society.
- Sproat, Richard William. 1985. *On deriving the lexicon*. Cambridge: MIT. (Doctoral dissertation.)
- Webelhuth, Gert. 1995. *Government and binding theory and the minimalist program: Principles and parameters in syntactic theory*. Oxford: Blackwell.
- Williams, Edwin. 1981. On the notions “lexically related” and the “head of a word”. *Linguistic Inquiry* 12(2). 245–274.
- Wright, William. 1971. *A grammar of the Arabic language*. Cambridge: Cambridge University Press. (Translated with amendments and additions from the German of Caspari.)
- Wunderlich, Dieter. 2008. When morphology comes in. (Paper presented at Morphology Workshop, Grooßbothen, 21 June 2008.)

### *Author's address*

Abdullah S. Al-Dobaian  
 College of Art  
 Department of English and Literature  
 King Saud University  
 P. O. Box 2456, Riyadh 11451  
 Saudi Arabia  
[maliknabi@hotmail.com](mailto:maliknabi@hotmail.com)

### **Publication history**

Date received: 22 June 2015

Date accepted: 22 July 2016