

**Newcastle University**

**School of English Literature, Language and Linguistics**



**Compounding in Modern Standard Arabic,  
Jordanian Arabic and English**

**Abdel Rahman Mitib Salim Altakhaineh**

**June 2016**

## Abstract

This study aims to identify types of compounds in Modern Standard Arabic (MSA) and Jordanian Arabic (JA) by applying the cross-linguistic criteria for compoundhood discussed in the relevant literature, with a special focus on English. These criteria -- orthographic, phonological, syntactic and semantic in nature -- have been proposed to make a distinction between compounds and phrases. The analysis reveals that the most reliable cross-linguistic criteria to distinguish between phrases and compounds in MSA, JA and English are adjacency and referentiality. With regard to the former criterion, no intervening elements can be inserted between the head and the non-head of compounds, whilst such insertion is allowed in phrases. With regard to the latter criterion, the non-head of a phrase is always referential, whereas the non-head of a compound is normally non-referential. Other criteria have been found to be partially applicable, e.g. compositionality, possibilities for modification and coordination, and free pluralisation of the non-head. In this study, I also suggest two reliable criteria that are exclusive to Arabic, or potentially Semitic languages in general. The first criterion is the appearance/absence of the possessive marker *li-/la* 'for/of' when the first element is definite. The second criterion deals with the appearance/absence of the possessive marker *li-/la* 'for/of' when the first element is preceded by a cardinal number.

In applying the various criteria, several properties of compounding in MSA and JA are examined in detail, such as stress assignment, the behaviour of serial verbs and V + V compounds, headedness, and types of compounds based on Scalise and Bisetto's (2009) classification. With respect to stress assignment, analysis shows that the default position of stress in both N + N compounds and phrases is on the first element. Concerning serial verbs and V + V compounds, the analysis shows that, although the distinction between them is not always clear-cut, V + V compounds are different from serial verbs with respect to the adjacency criterion. With regard to headedness, my study confirms that compounding in Arabic is predominantly left-headed. Regarding types of compounds, the Arabic data shows the usefulness of Scalise and Bisetto's (2009) classification, which originally was proposed on the basis of data from 23 languages, excluding Arabic. Finally, the study proposes a definition for compounds that may be applicable cross-linguistically and concludes with recommendations for further research.

## **Declaration and statement of copyright**

### **Declaration**

No part of the material within this thesis has previously been submitted for a degree at Newcastle University or any other university.

### **Statement of Copyright**

The copyright of this thesis rests with the author. No quotation should be published from it without his prior written consent and information derived from it should be acknowledged.

## **Dedication**

*To my mother, whom I love dearly, for everything*

*And to Maggie Tallerman and William van der Wurff for  
encouraging, motivating and supporting me*

## **Acknowledgments**

I would first like to thank my supervisors, Prof Maggie Tallerman and Dr William van der Wurff, for their patience, helpful advice, and being a source of inspiration. Thank you for all your enthusiasm and help in making what initially seemed terrifying into something that I enjoyed writing and found interesting. Without you, this piece of work would not have been possible.

I would also like to thank Dr SJ Hannahs, Prof Anders Holmberg, Dr Richard Waltereit and Dr Ghada Khattab for sharing their expertise with me. Their comments and suggestions helped to improve this thesis.

My special thanks go to my fellow colleagues here at Newcastle, Aseel Zibin, Marwan Jarrah, Bashayer Alotaibi, Maha Jasim, Murdhy Alshamari and Samer Hanafiyeh, for being such wonderful and supportive friends. Your discussions helped me broaden my perspective and clarify things.

And finally, I would like to thank all my family, friends and students in Jordan, the UK and the UAE, who have provided endless emotional support during my time here. I extend my undying gratitude to you all, thank you.

## List of abbreviations and symbols

(\*X) token not correct with material in brackets

\*(X) token not correct without material in brackets

1 first person

2 second person

3 third person

ACT active

ADJ adjective

ADV adverb

AP active participle

ART article

BP broken plural

DAT dative

DEF definite

DET determiner

F feminine

FUT future

GEN genitive

IMP imperative

IND indicative

INDF indefinite

INF infinitive

IPA International Phonetic Alphabet

JA	Jordanian Arabic
LE	linking element
M	masculine
MSA	Modern Standard Arabic
N	noun
NMLZ	nominalizer/nominalization
PASS	passive
PL	plural
POSS	possessive
PRS	present
PROR	progressive
PRT	particle
PST	past
PTCP	participle
SG	singular
SGC	Synthetic Genitive Construction
SP	sound plural
V	verb
*	ungrammatical sentence
'	preceding a stressed syllable
?	marked sentence/phrase

## Transcriptions and other conventions

Most of the symbols are drawn from the IPA but I use a few non-IPA symbols that are conventional in transcriptions of Arabic. For each symbol, the table specifies the exact phonetic value.

Arabic consonant/vowel	Symbol	Description
ء	ʔ	voiceless glottal stop
ب	b	voiced bilabial stop
ت	t	voiceless dento-alveolar stop
ث	θ	voiceless inter-dental fricative
ج	dʒ	voiced post-alveolar affricate
ح	ħ	voiceless pharyngeal fricative
خ	x	voiceless uvular fricative
د	d	voiced dento-alveolar stop
ذ	ð	voiced inter-dental fricative
ر	r	voiced alveo-palatal trill
ز	z	voiced alveolar fricative
س	s	voiceless alveolar fricative
ش	ʃ	voiceless alveo-palatal fricative
ص	s <sup>ʕ</sup>	voiceless alveolar emphatic fricative
ض	d <sup>ʕ</sup>	voiced alveolar emphatic stop
ط	t <sup>ʕ</sup>	voiceless dento-alveolar emphatic stop
ظ	ð <sup>ʕ</sup>	voiced alveolar emphatic fricative



ع	ʕ	voiced pharyngeal fricative
غ	ɣ	voiced uvular fricative
ف	f	voiceless labio-dental fricative
ق	q	voiceless uvular stop
ك	k	voiceless velar stop
ل	l	voiced alveolar lateral
م	m	voiced bilabial nasal
ن	n	voiced alveolar nasal
ه	h	voiceless glottal fricative
و	w	voiced labio-velar glide
ي	y	voiced palatal glide
ا	a	low short central unrounded
اُ	u	high short back rounded
اِ	i	high short front unrounded
آ	aa	low long central unrounded
وو	uu	high long back rounded
يي	ii	high long front unrounded
و	oo	mid long back rounded
او	aw	low short front unrounded + labio-velar glide
اي	ay	low short front unrounded + palatal glide
يي	ee	mid long front unrounded

# Table of Contents

Abstract.....	ii
Declaration and statement of copyright.....	iii
Dedication.....	iv
Acknowledgments.....	v
List of abbreviations and symbols.....	vi
Transcriptions and other conventions.....	viii
Chapter One: Introduction.....	1
1.1 Introduction.....	1
1.2 Background to the study.....	1
1.2.1 What is a compound?.....	1
1.2.2 Compounds and phrases.....	2
1.2.3 Types of compounds.....	4
1.3 Why compounding?.....	5
1.4 The significance of the study.....	6
1.5 MSA and JA: some background.....	8
1.6 Research questions and objectives.....	10
1.7 Methodology.....	10
1.8 The structure of the thesis.....	11
Chapter Two: Criteria for identifying compounding cross-linguistically.....	12
2.1 Introduction.....	12
2.2 The main distinguishing criteria between compounds and phrases.....	12
2.2.1 Orthography.....	12
2.2.2 Stress.....	13
2.2.3 Modification.....	16
2.2.4 Compositionality.....	19
2.2.5 Displacement.....	21

2.2.6 Insertion .....	22
2.2.7 Referentiality.....	23
2.2.8 Coordination .....	24
2.2.9 Replacement of the second element by a pro-form .....	26
2.2.10 Ellipsis.....	26
2.2.11 Inflection and linking elements.....	28
2.2.12 Tonal patterns.....	32
2.3 Language specific criteria.....	32
2.3.1 Postposed definite article in Danish.....	32
2.3.2 Word order .....	33
2.4 Compounding and derivation.....	35
2.5 Conclusion .....	39
Chapter Three: Identifying compounding in MSA and JA: The orthographic and phonological criteria .....	41
3.1 Introduction.....	41
3.2 The two possible N + N constructs in MSA and JA.....	41
3.3 Orthography .....	44
3.4 The sandhi phenomenon .....	45
3.4.1 Overview.....	45
3.4.2 Types of sandhi in MSA and JA .....	46
3.5 Stress assignment in N + N combinations .....	53
3.6 Stress assignment in N + N combinations in MSA and JA .....	53
3.6.1 Background on stress assignment in Arabic .....	56
3.6.2 Hypotheses.....	59
3.6.3 The experiment .....	61
3.6.3.1 <i>Pilot study</i> .....	61
3.6.3.2 <i>Sample</i> .....	62

3.6.3.3 <i>Tools and procedure</i> .....	62
3.6.3.4 <i>Data analysis</i> .....	65
3.6.4 Results and discussion .....	66
3.6.4.1 <i>N + N combinations marked with definiteness</i> .....	67
3.6.4.2 <i>N + N combinations not marked with definiteness</i> .....	77
3.7 Conclusion .....	84
Chapter Four: Identifying compounding in MSA and JA: the semantic and syntactic criteria .....	86
4.1 Introduction.....	86
4.2 The semantic criteria.....	86
4.2.1 Compositionality.....	86
4.2.1.1 <i>Compositionality of metaphorical and metonymical compounds</i> .....	92
4.2.2 Referentiality.....	98
4.2.3 Summary .....	105
4.3 The syntactic criteria.....	105
4.3.1 Modification.....	105
4.3.2 Adjacency .....	110
4.3.3 Coordination .....	114
4.3.4 Replacement of the second element by a pro-form .....	119
4.3.5 Ellipsis.....	119
4.3.6 Inflection and linking elements.....	121
4.3.7 Definiteness of the first element .....	125
4.3.8 Cardinal numbers before the first element.....	128
4.3.9 Summary .....	132
4.4 Conclusion .....	134
Chapter Five: Identifying compounding in Arabic: Adj + N combinations .....	136
5.1 Introduction.....	136

5.2 Adjective + Noun combinations: types and properties .....	136
5.2.1 Overview .....	136
5.2.3 Applying the cross-linguistic criteria of compoundhood to Adj + N combinations in MSA.....	141
5.2.3.1 <i>Orthography</i> .....	141
5.2.3.2 <i>Sandhi</i> .....	142
5.2.3.3 <i>Stress</i> .....	143
5.2.3.4 <i>Compositionality</i> .....	144
5.2.3.5 <i>Referentiality</i> .....	146
5.2.3.6 <i>Adjacency</i> .....	147
5.2.3.7 <i>Modification</i> .....	148
5.2.3.8 <i>Coordination</i> .....	151
5.2.3.9 <i>Inflection and linking elements</i> .....	152
5.2.3.10 <i>Adjacency and the order of elements</i> .....	156
5.2.4 Summary .....	160
5.3 Adjective-hood criteria .....	160
5.3.1 Overview.....	160
5.3.2 Cross-linguistic adjective-hood criteria .....	161
5.4 The syntactic category of Adj + N combinations in MSA.....	166
5.5 Other compound adjectival expressions within SGC .....	169
5.6 Conclusion .....	172
Chapter Six: Identifying compounds in Arabic: combinations other than SGC.....	173
6.1 Introduction.....	173
6.2 Noun + Noun combinations other than SGCs .....	173
6.3 Noun + Adjective combinations .....	175
6.4 Adjective + Adjective combinations.....	178
6.5 Particle + Adjective combinations .....	179

6.6 Particle + Verb combinations .....	180
6.7 Verb + Verb compounds or serial verbs .....	181
6.7.1 Serial verb criteria.....	182
6.7.1.1 <i>Overview</i> .....	182
6.7.1.2 <i>Applying verb serialisation criteria to V + V combinations in JA</i> .....	182
6.7.1.3 <i>Applying verb serialisation criteria to V + V compounds in English</i> .....	188
6.7.1.4 <i>Summary</i> .....	190
6.8 Reduplication .....	191
6.9 Numerals .....	195
6.10 Summary .....	199
Chapter Seven: Headedness and compound classification in MSA and JA .....	202
7.1 Introduction.....	202
7.2 Headedness .....	202
7.2.1 Overview.....	202
7.2.2 Headedness criteria in the previous literature .....	204
7.2.3 Applying headedness criteria to N + N compounds in MSA and JA within SGC205	
7.2.3.1 <i>The semantic criterion</i> .....	205
7.2.3.2 <i>The syntactic criteria</i> .....	207
7.2.3.3 <i>The morphological criteria</i> .....	214
7.2.4 Applying headedness criteria to Adj + N compounds in MSA within SGC .....	218
7.2.5 Applying headedness criteria to compounds in MSA and JA other than SGC ....	223
7.2.6 The implications of the headedness criteria.....	228
7.2.7 Summary .....	232
7.3 Arabic compounds within the cross-linguistic compound taxonomy of Scalise and Bisetto (2009).....	232
7.3.1 Overview.....	232
7.3.2 Bisetto & Scalise's (2005) classification .....	235

7.3.3 Scalise and Bisetto’s (2009) classification .....	237
7.3.3.1 <i>Subordinate compounds</i> .....	238
7.3.3.2 <i>Attributive compounds</i> .....	240
7.3.3.3 <i>Coordinate compounds</i> .....	241
7.3.4 Summary .....	243
7.4 Conclusion .....	244
Chapter Eight: Conclusion .....	245
8.1 Introduction .....	245
8.2 Main findings about compounding in MSA and JA .....	245
8.3 Recommendations for further research .....	247
References .....	249
Appendix: Information for participants and reading texts in MSA and JA .....	264

# Chapter One: Introduction

## 1.1 Introduction

It is well known that compounding is one of the most productive processes in the morphology of many languages. This has led to a large literature, exploring many issues in compounding. Nevertheless, as Lieber and Štekauer (2009: 4) point out, it is difficult to provide a universally applicable definition of a compound, firstly because compound elements in some languages are not free-standing words and secondly because there are often no clear boundaries between compounds and other linguistic entities such as phrases and derived words. There is also no comprehensive set of cross-linguistic criteria that can be used to distinguish compounds from phrases. Another – much debated – issue in compounding is the headedness of compounds. The criteria based on which the head of a certain construct can be pinpointed are still controversial.

This study aims to investigate some of these issues with regard to compounds in two languages that so far have not provided much input to the debate about compound structures and properties, i.e. Modern Standard Arabic (henceforth, MSA) and Jordanian Arabic (henceforth, JA). The differences between compounds and phrases in these languages will be investigated and possible types of compound in MSA and JA will be identified on the basis of the relevant literature on English compounds. In addition, this study explores the position of the head of compounds and the classification of compounds on the basis of Scalise and Bisetto's (2009) taxonomy. Finally, this study proposes a definition for compounds that may be applicable cross-linguistically.

The next section provides an overview of definitions and types of compound and some problematic issues that make the distinction between compounds and phrases a difficult task.

## 1.2 Background to the study

### 1.2.1 What is a compound?

Several scholars have provided definitions for compounds that are meant to be valid cross-linguistically. For instance, Marchand (1960: 11) indicates that compounds consist of two words or more which are combined to form a morphological unit. Katamba (1993: 54) proposes that compounds comprise two bases, at least, which could be words or root morphemes. According to Fabb (1998: 66), a compound can be defined as a word which itself consists of two or more words. Similarly, Olsen (2000: 280) states that compounding is a combination of



two free forms or stems, forming a new complex word. Carstairs-McCarthy (2002: 59) suggests that compounds are words which are coined by combining roots. Ralli (2013: 10) states that compounds consist of more than one lexeme which can be realised as words or stems based on the language under investigation. Note that all these definitions can be viewed as being too narrow, since they do not acknowledge the fact that phrases can be elements of compounds, at least in English, e.g. *jack-in-the-box*. In addition, these definitions do not provide help in distinguishing compounds from phrases.

Somewhat more precise definitions of compounding have been suggested by Bauer (2001: 695) and Plag (2003: 135). Bauer (2001: 695) posits that a “[c]ompound is a lexical unit made up of two or more elements, each of which can function as a lexeme independent of the other(s) in other contexts, and which shows some phonological and/or grammatical isolation from normal syntactic usage.” Finally, Plag (2003: 135) proposes that “a compound is a word that consists of two elements, the first of which is either a root, a word or a phrase, the second of which is either a root or a word.” I believe that the definition provided by Plag is more specific. Thus, his definition is my departure point to provide another definition that could be applicable cross-linguistically.

### **1.2.2 Compounds and phrases**

Several linguists (e.g. Katamba 1993; Bauer 2003; Booij 2007 among others) have attempted to differentiate between compounds and phrases in various languages. Katamba (1993: 332) defines a phrase as “a syntactic constituent whose head is a lexical category, i.e. a noun, adjective, verb, adverb or preposition”. A phrase may consist of one word, two words or more. Similarly, a compound consists of two words or more. This means that the number of words in a construct is not an indicator of whether this construct is a compound or phrase. Additionally, Bauer (2003: 135-136) shows that compounding is similar to phrase formation due to the fact that compounds are sequences of lexemes, unlike idioms, which are formed through rules of syntax. It is frequently the case that the meaning of a noun plus noun compound is indistinguishable from the meaning of an adjective plus noun. For example:

- |                          |                        |
|--------------------------|------------------------|
| (1) atom bomb            | atomic bomb            |
| (2) gold ring            | golden ring            |
| (3) verb paradigm        | verbal paradigm        |
| (4) language development | linguistic development |

These two combinations are equivalent alternatives despite the fact that N + N compounds are seen as products of morphology, while Adj + N compounds are products of syntax. Bauer (2003: 136) and Booij (2007: 82-83) explain that Adj + N compounds have an equivalent function to N + N compounds.

N + N	Adj + N
(5) city parks	urban parks
(6) ocean/sea life	marine life

The adjectives in (1-4) are derived from the nouns used in the competing construct, e.g. *verbal* from *verb* and *linguistic* from *language*. This is arguably also the case in (5) and (6), since *urban* is the only available relational adjective for expressing “related to cities” and *marine* is the only available adjective that expresses the meaning “related to seas”.<sup>1</sup>

A sequence of N + N in English can also be equivalent to possessive plus noun. The latter is usually seen as an example of syntax, whereas the former is viewed as a part of morphology. Relevant examples include the following (Bauer 2003: 136; Rosenbach 2007: 143):

	Compounds	Phrases
(7)	birdfoot	bird’s foot
(8)	dog house	dog’s house
(9)	summer day	summer’s day
(10)	lawyer fees	lawyer’s fees
(11)	Sunday lunch	Sunday’s lunch.

Thus, several linguists (e.g. Bauer 2003: 135-136; Katamba and Stonham 2006: 307; Lieber and Štekauer 2009: 11-12; Fàbregas and Scalise 2012: 120-122; Bauer *et al.* 2013: 431-436 among others) propose criteria to distinguish between compounds and phrases cross-linguistically. These criteria are explored in detail in chapter 2.

---

<sup>1</sup> Booij (2007: 33) discusses these adjectives, arguing that they should be viewed as cases of irregular derivation similar to the irregular inflection found in *bad-worse*. In this case, *urban* is derived irregularly from *city* and *marine* from *sea*. Booij (2007: 33-34) refers to these adjectives as suppletive adjectives.

Having shown that the boundaries between compounds and phrases are not crystal-clear, the main types of compounds in the literature are worthy of further investigation. Clarifying the general difference between the various types of compounds will facilitate the analysis of the compounds of Arabic in the following chapters.

### 1.2.3 Types of compounds

There are several types of compound discussed in the relevant literature (Fabb 1988: 66-67; Carstairs-McCarthy 2002: 64-65; Haspelmath, 2002: 89 and Booij 2007: 81). These types include, firstly, endocentric or semantically headed compounds, such as:

- (12) bookshop 'which is a type of shop'
- (13) board game 'which is a type of game'
- (14) high chair 'which is a type of chair'
- (15) graveyard 'which is a type of yard'

Secondly, there are exocentric compounds, which are not semantically headed, such as:

- (16) faintheart 'which is not a type of heart'
- (17) egghead 'which is not a type of head'
- (18) redskin 'which is not a type of skin'
- (19) scarecrow 'which is not a type of crow'

One type of exocentric compound, termed a *bahuvrihi*-compound, is defined as a compound which denotes the person in possession of the entity denoted by the compound (Booij 2007: 80). Booij notes that this type of compound exhibits a similar behaviour to adjectives although there is no adjectival head. An example is the Sanskrit word *bahuvrihi*, which consists of *bahu* 'much' and *vrihi* 'rice', i.e. 'having much rice' but denotes 'a rich man'. This word seems to have first been used as an adjective before becoming a noun. Examples from English are *blue-stocking*, *skinhead*, *baldhead* and *paleface* (Booij *ibid*). It appears that this type of compounding should fall under the definition of 'exocentric compound' as it has no semantic head. Bauer (2010: 169) points out that:

There is no surprise in having bahuvrihi compounds as one of the types of exocentric compound – or at least, if there is, it is because the Sanskrit label is sometimes appropriated for exocentrics as a group rather than for one type of exocentric.

Another type of compound mentioned in the relevant literature is copulative/coordinative compounds. In this type of compound, the relationship between the elements involves coordination, such as:

- (20) blue–green ‘is both blue and green’  
(21) washer–dryer ‘is both a washer and a dryer’  
(22) deaf–mute ‘is both deaf and mute’  
(23) bitter–sweet ‘is both bitter and sweet’

### 1.3 Why compounding?

Booij (2010a: 93) notes that in several languages, forming new compounds is the most productive type of word-formation. Furthermore, Joseph and Janda (1988: 204) observe that “we know of no language that lacks both affixation and compounding; Chinese, for example, certainly has compound morphology”. However, Joseph and Janda’s claim seems optimistic, since a few languages seem not to have compounding. In their corpus of 55 languages, Štekauer *et al.* (2008 cited in Scalise and Vogel 2010: 1) note that only 50 languages have compounds. Languages which they cite as lacking compounds include East Dangla, Karao, West Greenlandic, Diola Fogany and Kwak’wala (Štekauer *et al.* *ibid.*). Nonetheless, compounding is still a very productive word-formation process and examples of compounds from typologically different languages show the prominence of this process (Scalise and Vogel 2010: 1). In their corpus analysis of 23 languages, Scalise and Vogel (2010: 8-12) found the following results:

- There are 110 compound types in terms of how the internal word classes of a compound are combined.
- The most frequent patterns are: N+N, A+N, A+A, N+A, V+N, N+V, V+V, Adv+N, A+V, Adv+A.

- The order of the preference of the output category is exactly the same as that of the input categories:

N > A > V > Adv > P

These observations indicate that compounding is a common word-formation process worthy of further investigation, especially in languages in which compounding has not been examined in any detail yet.

#### **1.4 The significance of the study**

Despite the fact that there are many works discussing compounding cross-linguistically, e.g. Lieber and Štekauer (2009) and Scalise and Vogel (2010), these works do not include any discussion of Arabic. In addition, otherwise comprehensive handbooks on Arabic language and linguistics, such as Ryding (2005), Fassi-Fehri (2012) and Bauer, Lieber and Plag (2013), say little or nothing about compounds in Arabic either.

Some Arab researchers (e.g. Al-Humaydi 2005; Ryding 2005; Attia 2006; Al-Hariiri 2013; Amer and Menacere 2013) mention some examples of compounds in Arabic but do not specify the criteria based on which they consider a particular construct a compound. For instance, Attia (2006: 87) discusses multiword expressions in Arabic from a computational-linguistic perspective without making a distinction between compounds and phrases. According to Attia (2006: 92), "... a compound noun can be formed by a noun optionally followed by one or more nouns, optionally followed by one or more adjectives." This suggests that, for Attia (2006), the Arabic phrase *sayyarah mufaxxaxah* 'bombing car' is a compound. However, it is clear that *sayyarah mufaxxaxah* 'bombing car' is phrase, since the adjective *mufaxxaxah* 'booby-trapped, rigged (with explosives)' agrees with the preceding noun in number, gender and definiteness, which is a common characteristic of phrases in Arabic. More recently, in a study of English-Arabic translation, Amer and Menacere (2013) refer to some constructs as compounds. The two researchers consider examples of the Synthetic Genitive Constructions (henceforth, SGCs) as compounds without making a distinction between the subtypes that exist in SGCs. SGCs (referred to in Arabic as *Id'aaafah*) 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.<sup>2</sup> Another important characteristic of SGCs is that the first element obligatorily lacks a definite article, whereas the second can be definite or indefinite (see Fassi-Fehri 2012: 156), as in (24):

- (24) a. sayyaarat-u/a/i                      r-radʒul-i                      (MSA)  
           car-NOM/ACC/GEN                      the-man-GEN  
           ‘the man’s car’
- b. sayyaarat-u/a/i                      radʒul-i-n                      (MSA)  
           car-NOM/ACC/GEN                      man-GEN-INDF  
           ‘a man’s car’

Amer and Menacere (2013: 232) consider SGCs in which the syntactic relation between the two elements is possessive or genitive as compounds. As will become evident in later chapters, not all SGCs are compounds, an example of a genuine compound is the following:

- (25)    ʕaruus-u                      l-baħr-i  
           bride-NOM                      the-sea-GEN  
           ‘the mermaid’  
           lit. the sea bride

Amer and Menacere (2013: 235) also treat as compounds some combinations consisting of the negative particle *laa* ‘no’ used as a prefix followed by a N, such as *laasilkii* ‘wireless’ and *laaʔaxlaaqii* ‘impolitely’. Needless to say, it would be very problematic to recognise a category of compounds that is coined with affixes as proposed by Amer and Menacere (2013: 235). In chapter 6, I show that examples such as *laasilkii* ‘wireless’ are derivatives, rather than compounds.

On the basis of the above, it is clear that compounding in Arabic has received very little attention in the literature on word-formation. The present study aims to fill part of this gap by

---

<sup>2</sup> In Semitic languages, some scholars (e.g. Siloni 1997; Fassi-Fehri 2012) use the term Nominal State Construct or Construct State. However, I opted for Synthetic Genitive Constructions (SGCs) which contrast with ‘analytic genitives’, i.e. with the possessive markers, e.g. *li* ‘for/of’ in MSA. In fact, the ‘construct state’ refers to the morphological form of the possessum in a construct, e.g. lack of nunation and in some Arabic dialects, e.g. JA, the feminine suffix surfacing with a final /t/, etc. (see section 3.4).

providing detailed description and analysis of the features of compounds in Arabic, looking at all areas of full productivity as well as compound types which are less productive. Additionally, there will be full discussion of the distinction between compounds and phrases in Arabic, which may facilitate the comprehension of how these combinations operate, at least in Semitic languages. Ultimately, the goal 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.

### **1.5 MSA and JA: some background**

Arabic is the official language of 22 countries, stretching from the Arabian Peninsula up to Syria, and across the whole of North Africa. It is spoken by approximately 300 million people (Owens 2013: 2). In these countries, MSA is used in news broadcasts, official speeches, legal documents, books and newspapers. This variety of Arabic is also taught at universities around the world due to its standardisation and academic uses, especially in writing. MSA has no speech community; it is not natively acquired by Arab children. It is also not used in daily conversation, in which the spoken dialects of Arabic are used instead. Arab children acquire local and non-standard Arabic at home. By the time they go to school, they start to learn MSA. This phenomenon, where two different varieties of a language co-exist simultaneously in a community of speakers, each serving a distinct range of social functions, is referred to as diglossia (Saiegh-Haddad 2003: 432). The spoken dialect is the variety used to discuss relationships and to communicate with family and friends. These dialects vary within countries and across the region. For example, even though the countries of the Levant - Jordan, Palestine, Syria and Lebanon - seem to share a common dialect of Spoken Arabic, they still vary in certain linguistic aspects, i.e. phonological, morphological and lexical. However, outside of the Levant, the spoken dialects have less similarities with the spoken variety found in the Levant and some dialects of Arabic are mutually unintelligible. For instance, people in Jordan, Egypt or Oman may not be able to understand the Arabic variety spoken by someone from Morocco. The words and phrases that are used most often in conversation vary across countries. Based on my experience as a native speaker of JA, with a good working knowledge of MSA, I may not be able to fully understand a spoken variety of Arabic, e.g. Moroccan Arabic.

The focus in this thesis is on the varieties MSA and JA. These have some lexical,

phonological and morphosyntactic differences.<sup>3</sup> The most prominent morphosyntactic difference between them with respect to N + N combinations within SGCs is the presence (in MSA) versus absence (in JA) of the morphosyntactic feature of case, as in the following examples:

- |      |  |                              |       |
|------|--|------------------------------|-------|
| (26) | bayt-u/a/i<br>house-NOM/ACC/GEN<br>'the man's house'                             | r-radʒul-i<br>the-man-GEN    | (MSA) |
| (27) | raaʔid-u/a/i<br>pioneer-NOM/ACC/GEN<br>'the astronaut'<br>lit. the space pioneer | l-fadʕaaʔ-i<br>the-space-GEN | (MSA) |
| (28) | beet<br>house<br>'the man's house'   | z-zalameh<br>the-man         | (JA)  |
| (29) | raaʔid<br>pioneer<br>'the astronaut'<br>lit. the space pioneer                   | l-fadʕaaʔ<br>the-space       | (JA)  |

Note that in the MSA examples (26) and (27), the first element of SGCs can be nominative, accusative or genitive depending on the function of the whole N + N combination in the sentence, but the second element is always genitive. In contrast, JA examples (28) and (29) do not have case marking. In consecutive speech a default *-i* appears between the two elements for phonological purposes, i.e. breaking consonant clusters, but that is not a case marker. Furthermore, some lexical differences can be found between MSA and JA. For instance, the

---

<sup>3</sup> Within JA, at least two subvarieties can be distinguished, i.e. Urban Jordanian Arabic (UJA) and Bedouin Jordanian Arabic (BJA). There are some lexical-phonological differences between these two dialects, but they are identical in the nature and structure of their compounds.



word *man* is realised in MSA as *rradzuli* ‘the man’ in (26), whereas it is *zzalameh* ‘the man’ in JA (see example 28).

In general, the difference between MSA and JA is that the former is official, written, formal, learned and a “pan-language”, whereas JA is unofficial, spoken, informal, acquired and a local language (Owens 2013: 5-6).

## **1.6 Research questions and objectives**

This study aims to: (1) identify N + N compounds in MSA and JA, showing the differences between compounds and phrases by applying the cross-linguistic criteria for compoundhood; (2) discuss and investigate other possible types of compounding in MSA and JA, e.g. Adj + N, Adj + Adj and V + V compounds; (3) pinpoint the position and type of head in MSA and JA compounding; (4) examine to what extent the classification of compounds in MSA and JA fit into Scalise and Bisetto’s (2009) taxonomy; and finally (5) provide a universally applicable definition of compound. Thus, this study aims to provide a systematic and analytical description of compounding in MSA and JA, and check the validity of the cross-linguistic criteria used to identify compounding. The current study seeks answers to the following research questions:

1. What are the differences between N + N and Adj + N compounds and phrases within SGCs in MSA and JA based on the cross-linguistic criteria for compoundhood?
2. Are there any language-specific criteria that can be used to identify compounding in MSA and JA?
3. Are there types of compounding in MSA and JA other than SGCs?
4. Are there V + V compounds in MSA and JA?
5. What is the position of the head in a compound in MSA and JA?
6. What is the classification of compounds in MSA and JA on the basis of Scalise and Bisetto’s (2009) taxonomy?
7. What is the most widely applicable definition of a compound cross-linguistically?

## **1.7 Methodology**

The analysis in the current study is focused mainly on Arabic, especially MSA and JA, comparing them with English. When appropriate, examples are also considered from Hebrew, German, Dutch, Spanish, French, Italian and Danish. Since Arabic is my mother tongue, I use

my intuitions as a native speaker of Arabic together with grammatical descriptions of different constructs, i.e. Synthetic Genitive Constructions (SGCs) in the literature on Arabic. Grammaticality judgments of other native speakers of (Jordanian) Arabic were solicited when appropriate. For other languages, grammatical descriptions in the relevant literature are used. With regard to the experiment I conducted to determine the position of stress in N + N combinations (chapter 3), I used the latest version of Praat software (5.4.08), designed by Boersma and Weenink (2015), to pinpoint the position of the stress relying on the pitch and intensity contours.

## **1.8 The structure of the thesis**

This thesis is organised as follows: Chapter 2 identifies the cross-linguistic criteria for compoundhood. Chapter 3 applies the orthographic and phonological criteria discussed in Chapter 2 to identify compounding in MSA and JA. It also examines the assignment of stress in N + N combinations in MSA and JA. Chapter 4 applies the semantic and syntactic criteria to identify compounding in MSA and JA. It also suggests a scale of compositionality for compounding in MSA and JA and proposes two language-specific criteria to identify compounding in Arabic and potentially other Semitic languages. Chapter 5 applies the cross-linguistic criteria for compoundhood to identify Adj + N compounds in MSA. Chapter 6 investigates possible types of compounding apart from N + N and Adj + N combinations within SGCs. Chapter 7 identifies the position and type of headedness within compound words in MSA and JA. In addition, this chapter demonstrates the classification of compounding in MSA and JA on the basis of Scalise and Bisetto's (2009) taxonomy. Finally, Chapter 8 concludes and summarises the main points with some recommendations for further research.

## Chapter Two: Criteria for identifying compounding cross-linguistically

### 2.1 Introduction

There has been much discussion of what exactly a compound is and whether compounds can be distinguished from other word-formation processes such as derivation, on the one hand, and other syntactic constructs such as phrases, on the other. To answer the latter question, several criteria have been proposed (e.g. Bauer 1998a; Donalies 2004; Lieber and Štekauer 2009; Fàbregas and Scalise 2012; Bauer *et al.* 2013 among others), some of which deserve serious consideration, while others are less plausible. Hence, this chapter presents the criteria that have been proposed so far to draw borderlines between compounds, on the one hand, and phrases and derivation, on the other. In doing so, it aims to reveal universal criteria that can identify compounds cross-linguistically.

The chapter proceeds as follows: section 2.2 discusses the main general criteria that have been suggested in the literature to distinguish compounds from phrases. Section 2.3 provides an overview of some language-specific criteria said to apply in individual languages, in particular French and Danish. Section 2.4 discusses the boundary between compounding and derivation. Finally, section 2.5 summarises the main points and provides a working definition of compounding.

### 2.2 The main distinguishing criteria between compounds and phrases

#### 2.2.1 Orthography

Although spelling is usually regarded as a relatively superficial phenomenon, it has been considered a possible criterion for compoundhood in some languages. In Czech and Slovak, for example, orthography has been considered an important criterion, because all compounds are spelled as one word, whereas syntactic phrases are spelled as separate words (Lieber and Štekauer 2009: 7). Similarly, Szymanek (2009: 466) indicates that most Polish compounds are spelled as one word without a hyphen. However, he recognises the existence of some exceptions, especially with coordinate structures, such as *Bośnia-Hercegowina* ‘Bosnia-Herzegovina’ or *czarno-biały* ‘black and white’. In German too, compounds are usually spelled as a single word but coordinates like *rot-grün* ‘red and green’, *schwarz-rot-gold* ‘black and red and golden’ and *Dichter-Maler-Komponist* ‘poet and painter and composer’ are typically written with hyphens (Neef 2009: 396). The same applies to Dutch, where coordinates such as

*zwart-wit* ‘black and white’ and *directeur-grootaandeelhouder* ‘director and major shareholder’ are normally written with hyphens (see Booij 1992: 40-41).

In English, however, spelling offers no help in identifying compounds or distinguishing compounds from phrases. Some compounds are written as one word, with or without a hyphen, such as *horse-trade*, *ice-cream*, *overflow* and *egghead*. Many others are often written as two separate words, such as *body language* and *school bus*. It can be argued that orthography in English is unreliable, as there is no consistency in the orthographic representation of compounds. Examples given by Bauer (1998a: 69) include spellings such as *daisy wheel*, *daisy-wheel*, and *daisywheel*. Further examples of such inconsistency found in English dictionaries are *girlfriend* (*Hamlyn’s Encyclopaedic World Dictionary*), *girl-friend* (*Concise Oxford Dictionary* 7<sup>th</sup> Edition) and *girl friend* (*Webster’s Third New International Dictionary*). It may also be noted that some morphologists (e.g. Bauer 1983; Booij 2007: 91; Lieber 2009: 357) have different ways of writing the name of one of the topics that they study, with both of *word formation* and *word-formation* being found.

Thus, while spelling may offer help in identifying compounds in certain languages, it is by no means a universal or fail-proof criterion. More generally, Lieber and Štekauer (2009: 7-8) point out that it is usually assumed that the spoken language is primary, whereas writing uses an artificial system which only reflects aspects of the spoken language. This may suggest that spelling cannot be considered a criterion of compoundhood, since it only secondarily captures the words in the spoken language (Lieber and Štekauer *ibid*).

### 2.2.2 Stress

Stress has been the focus of a great deal of research in the last two decades, since it has been considered a useful criterion for distinguishing compounds from phrases in several languages (e.g. Bauer 2009a: 402 (Danish); Don 2009: 379-380 (Dutch); Kiefer 2009: 531 (Hungarian); Szymanek 2009: 472-73 (Polish); Zamponi 2009: 587, 592 (Maipure-Yavitero) among others). For instance, in Dutch, main stress tends to fall on the left-hand element of a compound, whereas most phrases have stress on the right-hand element (Don 2009: 379-380). Nonetheless, Don (2009: 380) states that:

There are some lexemes that, if they occur as left hand members of compounds, do not get the main stress. These lexemes include *stad* ‘city’, *staat* ‘state’, and *rijk* ‘national’. But other exceptions exist and no clear pattern seems to be present.

In English, stress can sometimes be used as a criterion to distinguish between compounds and phrases. For instance, 'blackboard is considered a compound, while *black board* a phrase (Booij 2012: 84). If the stress of *blackboard* falls on the initial word, as in /'blæk,bɔ:(r)d/, it denotes 'a large black or green surface which is fixed to a classroom wall for writing'. On the other hand, if the stress of *blackboard* falls on the second word, as in /,blæk'bɔ:(r)d/, it denotes 'a board which is painted black'. The idea that left-hand stress is often a mark of compounds, whereas right-hand stress is a sign of phrases was already discussed by Chomsky and Halle (1968: 17). They argue that the difference between compounds and phrases can be captured in a systematic way under the so-called nuclear stress rule (i.e. right-hand stress) and the so-called compound stress rule (left-hand stress).

Nevertheless, enough examples have been cited in the literature to show that stress as a criterion fails to distinguish reliably between phrases and compounds in English. For instance, Spencer (2003: 333) shows that stress can sometimes be used to convey different readings of the same combination of elements. For example, *apprentice instructor* is an instructor who is an apprentice, whereas *apprentice instructor* is one who instructs apprentices. The former reading is appositional, while the latter is associated with modification. Similarly, Giegerich (2004: 17) points out that *toy factory* is probably a factory where toys are made, but a *toy factory* is a factory which is also a toy. Examples have also been given in which there are compounds with right-hand stress and double stress. For example, Jones (1969: 259) states that when the second element of a compound seems to be especially important, the compound is double stressed, such as *'eye witness* and *'bow window*. Similar to Jones (1969), Bauer (2003: 134) provides the examples *'apple cake* which has single stress, and *'apple pie* which has two.

Unlike Bauer (2003), Lieber (2005: 376) notes that, while *'apple cake* is stressed on the left-hand stem, *apple 'pie* has stress on the right-hand stem. The difference in stress assigned to *apple cake* and *apple pie* could be ascribed to UK vs. US stress variation. Plag (2006: 144) posits that there is cross-varietal variation (e.g. British English vs. American English), which makes it difficult to examine the regularity of compounding stress patterns. Regional differences in terms of stressing certain forms or whole groups of forms can be found, such as *dry-'clean* in British English vs. *'dry-clean* in American English (Bauer *et al.* 2013: 445). Variation even within and across people who speak the same dialect can be also found in a given compound. According to Kunter (2011: 204), this kind of variation appears to be limited to particular compounds and is not present in others. Nevertheless, why certain

compounds exhibit variation, e.g. *ice-cream*, and others do not, e.g. *ice-cup* is still a mystery (Bauer *et al.* 2013: 445).

As Bauer (1983: 103) points out, a further factor influencing stress assignment is context. This type of variation can be attributed to many reasons. One of the most common ones is emphasis. For instance, *'undertaker* and *'underwriter* are usually pronounced with the stress on the first element. However, in the sentence *are we talking about undertakers or underwriters now?* the stress falls on *take* and *write*. Another example that shows the variation of stress assignment in context is: a person can say *would you like a 'milk 'shake?* using the same stress pattern as he/she would use in isolation. However, *an ice-cold 'milk shake is just what I need* has only one stress on *milk* (Bauer *ibid.*).

Other problems concerning the use of stress as a criterion are identified by Plag (2003: 138), who notes that there could be a systematic pattern in the following exceptions (cited in Bauer 1998a and Olsen 2000), where the stress falls on the right element of the compound:

Table 2.1. English compounds with right-hand stress

geologist-'astronomer	apple 'pie
scholar-'activist	apricot 'crumble
Michigan 'hospital	Madison 'Avenue
Boston 'marathon	Penny 'Lane
summer 'night	aluminum 'foil
may 'flowers	silk 'tie

Plag (2003: 139) suggests that there are some meaning relationships typically accompanied by right-hand stress, namely: (1) copulative compounds such as *geologist-'astronomer* and *scholar-'activist*, which are different from other compounds in that both elements refer to the same entity; (2) temporal or locative compounds such as *a summer 'night* and *the Boston 'marathon*; and (3) causative compounds, usually paraphrasable as ‘made of’, as in *aluminum 'foil* and *silk 'tie*, or ‘created by’, as in *a Shakespeare 'sonnet* and *a Mahler 'symphony*. However, Plag (*ibid.*) admits that it is not clear how many semantic classes should be set up to account for all the putative exceptions to the compound stress rule, which remains a problem for proponents of this hypothesis. Moreover, in some cases there does not appear to be a semantic basis for the exceptionality. For example, *'Madison Street* and *Madison 'Avenue* have the same structure (noun-noun), their respective elements have the same meaning relationships

and both are right-headed; nevertheless, they differ in their stress pattern. In a later experimental study, Plag (2006: 147-8) argues that stress is assigned in new compounds on the basis of analogy to existing N + N combinations. Specifically, the analogical pattern is determined by the head. In the case of *street* and *avenue* compounds, for example, the stress falls on the left element in *'Fifth Street*, whilst it falls on the right element in *Fifth 'Avenue* (as already pointed out by Bauer 1983). Nonetheless, Plag concludes with the following questions that require further investigation: what are the limits of this analogical approach? what are the factors that contribute to this kind of analogical behaviour?

Taking all the above arguments and open questions into consideration, the conclusion must be that stress, as a criterion for the differentiation between compound nouns and nominal phrases in English, is not foolproof. Therefore, further examination of other suggested criteria is needed.

### 2.2.3 Modification

Lieber and Štekauer (2009: 12) discuss another criterion to distinguish compounds from phrases, namely modification. It has often been said that the first element of a compound does not allow modification, whereas the first element of phrases can be modified. For instance, *very* can precede and modify an adjective that is part of a phrase, such as a *very black bird* said while pointing at a crow, but it is not possible to say *a very blackbird* if the reference is to the genus *Agelaius*. However, since some adjectives (i.e. relational ones) can never be modified by *very*, as in *\*a very mortal disease* (Lieber and Štekauer *ibid*), this criterion can only be applied to gradable adjectives, which means that it does not work across the board. In addition, the 'very' test can only be applied to compounds whose first element is an adjective. Therefore, the scope of this particular test is limited to Adj + N compounds.

Other researchers, such as Fàbregas and Scalise (2012: 120-121), argue that internal modification is found in English for both compounds and phrases, as in the following examples:

- (1) He sells [red balloons].
- (2) [Red balloons [sic]] seller.

This suggests that internal modification is a not reliable criterion in English, since both compounds and phrases can be internally modified. In Spanish, on the other hand, internal modification does not work in compounds, which means that it can be used as a criterion to

distinguish between compounds and phrases. For example, (3) shows that the compound *limpia ventanas* ‘window cleaner’ does not allow for modification of the element *ventanas* ‘windows’.

(3) \*un    limpia            ventanas    grandes  
      a    polish            windows    big

The intended meaning of example (3) is ‘a cleaner of big windows’. This may suggest that the impossibility of internal modification as a criterion to distinguish between compounds and phrases can be language-specific to Spanish.

Finally, in Germanic languages, compound structures are recursive; a compound can be an element in another compound, acting as a modifier (Bauer 2009: 350). Compounds like *Auckland architecture school library notice board* and *college teaching award committee member* are good examples of repeated modification that can be potentially unlimited. Furthermore, the Dutch compound *weersvoorspelling* ‘weather forecast’ can appear in another compound *weersvoorspellingsdeskundige* ‘weather forecast expert’, and the resulting compound can be used to form yet a further compound, *weersvoorspellingsdeskundigencongres* (Don 2009: 370-1):

(4) weersvoorspellingsdeskundigencongres  
      weers.voorspellings.deskundigen.congres  
      weather.forecast.experts.conference  
      ‘weather forecast experts conference’

Therefore, in languages like English, Dutch and German a compound can be built from another compound. Simply put, compounding can be recursive.

However, in practice, any string of more than five elements is very unusual in all three of these languages (Fleischer 1975: 82; Bauer 2009b) and several other languages, such as Slovak (Štekauer and Valera 2007) and Fongbe (Lefebvre and Brousseau 2002: 227), do not permit recursion at all. In some other languages, only some types of compounds can be recursive, whereas others cannot. For example, coordinative compounding (see section 1.2.3) is recursive in Romance languages; such a compound can be made longer by adding a third element, such as:



- (5) a) bar pizzería  
bar pizzeria
- b) bar pizzería discoteca  
bar pizzeria disco

On the other hand, in Romance languages, subordinative and attributive compounds (see section 1.2.3) are not recursive. For instance, in the Italian attributive compound *uomo lupo* ‘man-wolf, werewolf’, the addition of a third word that can be interpreted as an attribute is not possible (Fàbregas and Scalise 2012: 116), as in (6)<sup>4</sup>:

- (6) \*uomo lupo rana  
man wolf frog

Example (6) is intended to mean ‘a werewolf that has some properties of frogs’, but such a meaning cannot be conveyed through an attributive compound. Conversely, the subordinative compound in Germanic languages is well known for being highly recursive (Fàbregas and Scalise 2012: 116), as in (7):

- (7) garden decoration  
rose garden decoration  
tea rose garden decoration

Plag (2003: 134) points out that the longer a compound is, the more difficult it is for both the speaker and the hearer to produce it and understand it correctly. Thus, very long compounds are dispreferred for processing reasons (Plag *ibid*).

By comparison, it is well known that phrases are recursive. For instance, in English, phrases can be made longer and longer by putting a new phrase inside another one, as in possessives such as *John’s friend’s car’s motor* or prepositional phrases as in *in the kitchen in the cabinet in the corner...* and so on.

On the basis of the above, it is clear that the impossibility of modification cannot be

---

<sup>4</sup> One may argue that the Italian compound *uomo lupo* ‘man-wolf, werewolf’ can be interpreted as a coordinative compound. A detailed account of the interpretations of this type of compound and similar cases is explained in section 7.3.2.

used as a foolproof criterion for compoundhood. Some languages indeed disallow modification inside compounds but other languages do allow this, sometimes so productively that quite long compounds are routinely formed through a process of recursive modification. In other words, non-modifiability is not a universal property of compounds. Note, however, that the possibility of modification is not a sign of phrasehood.

#### 2.2.4 Compositionality

It has been said that “[a] complex linguistic expression is compositional if its meaning is determined by both the meanings of its parts and the way it is structured” (Neef 2009: 394). For instance, the English compound *bookshop* is compositional, because its meaning is derivable from its components, *book* and *shop* (Aronoff and Fudeman 2005: 104). The notion of compositionality pertains to the semantic head of the construct. With respect to semantic vs. syntactic headedness, a majority of compounds are interpreted in such a way that their grammatical and semantic heads coincide (Neef 2009: 395). The compositional meaning of a compound with the elements AB is ‘B that has something to do with A’. Essentially, every compositional compound which consists of two elements can be interpreted in a determinative way. The type of compound which shows this relationship most clearly is N + N compounds. For instance, a *Fisch•frau*, lit. fish•woman ‘is a woman that has something to do with fish’ (Neef *ibid*).

In English, three levels of semantic compositionality in compounds can be distinguished. The fact that there are levels within compositionality has been acknowledged by Fernando (1996: 36), who investigates idiom. In particular, Fernando argues that in addition to pure idiom which is completely non-compositional, there is another type of idiom which he refers to as a semi-idiom. The latter refers to a sequence that has one or more literal elements and at least one that has a non-literal sub sense. With regard to semantics, semi-idiom is not as complex as pure idiom, since its meaning is partially transparent. For instance, one can infer from the idiom “to promise someone the moon” that something is being promised. Similarly, Dirven and Verspoor (1998: 60) argue that compounds can be placed on a cline of transparency, which includes transparent compounds, partially transparent compounds, and non-transparent compounds. Examples of three classes are *apple tree*, *blackbird*, and *red tape*, respectively (for more detail see section 4.2.1). The three levels of compositionality are presented in Table 2.2.

Table 2.2. Levels of compositionality in English

	<b>Levels of compositionality</b>	<b>Examples</b>
1	Completely compositional	bookshop, houseboat, darkroom, physics teacher.
2	Semi-compositional	blackbird, green house, blackboard, cathouse, small talk.
3	Completely non-compositional	egghead, white-collar, redskin, faint-hearted, bluestocking

In the first level, the compounds are completely compositional in the sense that the meaning of the whole compound is the total sum of its parts. For instance, the compound *bookshop* is a shop that has something to do with books. In the second level, the meaning of the compound is not completely the total sum of its parts but the head makes a clear contribution to the meaning of the whole compound. For instance, the compound *blackbird* is a bird, even though it is not necessarily black, since the male is black, but the female is brown. Another example is *greenhouse*, which is a house, even though it is not green. The compounds in the third level are completely non-compositional in the sense that the meaning of the whole compound cannot be derived from the sum of its parts. For example, *egghead* refers to neither a head nor an egg. Its meaning, ‘intellectual’, is not related to both elements combined together.

Semantically speaking, by applying the ‘IS A’ condition, which was suggested by Allen (1978: 11), it seems that English compounds are usually semantically headed but there are some that are headless (see section 7.2 for more detail). This principle is normally used to differentiate between endocentric and exocentric compounds (see sections 1.2.3 and 7.3.1). Allen's ‘IS A’ condition is given in (8).

(8) In a compound [ [ ]<sub>A</sub> [ ]<sub>B</sub> ]<sub>C</sub> C ‘IS A’ B

This can be seen in the endocentric compounds in examples (9) and (10):

- (9) house boat IS A boat  
 (10) hand bag IS A bag

This principle can be used to show that *egghead* and *pickpocket* are exocentric compounds, as in (11) and (12):

- (11) egghead IS NOT A head  
 (12) pickpocket IS NOT A pocket

Bauer (1998a: 67) suggests that non-compositional compounds are listed in the dictionary, whereas syntactic constructs such as phrases are not, although he points out that this is more of a lexicographical criterion, rather than a linguistic one. In particular, Bauer (ibid) states that “many linguists seize one aspect of listedness - namely idiomaticity – and use that as a criterion for compound status”. Examples would be words like *blackboard* and *greenhouse*. Later, Kavka (2009: 33) argues that compositionality is the most important criterion that distinguishes compounds from free combinations, claiming that, like idiomatic expressions, compounds are non-compositional. Kavka (2009: 33) suggests that “their status will be understood more readily if they are viewed as parts of concrete, contextually defined utterances”.

On the other hand, Lieber (2005: 376) points out that compounding in many languages is highly productive and new compounds are very often compositional in meaning, especially when the context is taken into account. In other words, it is easy to dismiss this criterion for compoundhood at least in languages like English; the more productive the process of compounding in a language, the less chance that individual compounds will be lexicalized or listed (Lieber and Štekauer 2009: 7). Examples of compositional compounds are *houseboat*, *committee meeting* and *bookshop*, whilst *egghead*, *redskin* and *blue-stocking* are non-compositional. The same applies to phrases since *old hand* and *green fingers* are non-compositional, whereas *beautiful house*, *long journey* and *tall man* are compositional. Therefore, compositionality is not a reliable criterion to distinguish compounds from phrases in English. In other languages, things may of course be different. For instance, Borer (2009: 205) shows that compositionality in Hebrew is a reliable criterion to distinguish between compounds, on the one hand and various phrase types, on the other (see chapter 4).

### 2.2.5 Displacement

Fàbregas and Scalise (2012: 121) point out that in English it is possible to displace a constituent inside a phrase but not inside a compound, as in (13-14):

- (13) \*Truck is what he likes a [\_\_\_\_\_ driver].  
 (14) Trucks are what he [drives \_\_\_\_\_].

The gap shows the original position of the unit *truck* inside the structure. This criterion suggests

that compounds are not built by syntactic rules, as phrases are; compounds have no internal syntactic structure (Jackendoff 2009). Thus, this criterion can be viewed as reliable in English. Note that this criterion is closely related to the next one, ‘insertion’, which we now turn to.

### 2.2.6 Insertion

Insertion (also known as adjacency) is discussed by Lieber and Štekauer (2009: 11-12), who 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*). It has been noted that there is one potential exception to this general principle: the category of phrasal verbs (Lieber and Štekauer *ibid*). It has been suggested that these can be considered compounds, since they become inseparable when nominalised, as in *put-down*, *cop-out* and *carry-on*. Sentences like *he took his hat off* would then show that the criterion of non-insertion in compounds is not reliable). However, the idea that phrasal verbs are compounds is not accepted by all linguists. Jackendoff (2002: 90), for example, argues that phrasal verbs are not compounds, but rather *constructional idioms*, which he defines as syntactic schemas in which one position is lexically fixed while the remaining positions are variables that can be filled based on the particular rule (Jackendoff 2002: 188). This can be seen in the following example:

- (15) [AP V/N + -d [Prt out]]  
‘worn out from too much V-ing/ too much N’

The failure of phrasal verbs to observe the ‘insertion’ criterion (e.g. *pick it up*) can be regarded as an argument for following Jackendoff (2002) and considering phrasal verbs constructional idioms, rather than compounds. Removing phrasal verbs from the category of relevant data would mean that the criterion of non-insertion could be considered a reliable criterion for determining compound status.

I suggested above that the two separate criteria of insertion and displacement can be grouped under one single criterion, which can be called ‘adjacency’. The two criteria are closely related, since they both imply that the elements of a compound cannot be separated. That is, the ban against displacement posits that the elements of a compound should be impenetrable, while the ban against insertion posits that no intervening element can be inserted between the two components of a compound. Displacement and insertion can thus be seen as

two diagnostics to determine whether the string of words is separable or not. If the sequence of words is inseparable, we are dealing with a compound, rather than a phrase. Note, however, that the German so-called ‘separable verbs’, e.g. *aufhören* ‘to cease’ in *Hören Sie damit auf!* ‘Stop it!’ are separable, yet these can be thought of as compound verbs. This may suggest that adjacency, as a criterion for compoundhood, requires further investigation cross-linguistically.

### 2.2.7 Referentiality

Referentiality (also known as anaphoricity) can be defined as “the relationship by which language hooks onto the world” (Saeed 2003: 12). In particular, the underlined referring expression in *she is smart* picks out an entity or a specific person in the world. With regard to compounds, it has been observed that the first element of a compound is normally non-referential. For instance, the first element (the non-head) in *cat lover* does not refer to any specific cat (Lieber 2005: 376). In addition, any referential element used to modify a compound in English usually modifies the right element or the head as opposed to the first element or the non-head. For instance, in example (16), *these* modifies the second element, *accounts*:

(16)        these bank accounts

As a consequence of this lack of referentiality, Allen (1978: 113) claims that “individual elements of compounds...generally cannot function independently with respect to syntactic processes”. However, Bauer (1998a: 72) shows that a first element can occasionally serve as a discourse antecedent for pronouns, such as *so I hear you are a real cat-lover. How many do you have now?* More recent discussion of such examples is found in Bauer *et al.* (2013: 464), who argue that the context plays a pivotal role in making the first element of a compound referential. In particular, they point out that, in discussing the budget for the country’s army in a parliamentary debate, the word *army* in *army budget* has a specific reference, since it refers to the army of that particular country (Bauer *et al.* 2013: 464). This means that the interpretation of the first element of the compound is reliant on the context in which it occurs, especially in determining to which entity the first element can refer.

Scrutinising the referentiality of the non-head in a compound, Bauer *et al.* (2013: 464) note that although the non-head *truck* in *truck driver* is non-referential in nature, the non-referentiality of the non-head is limited to compounds in which the first element is a common noun. In contrast, they point out that there are compounds in which the first element is a proper

noun, e.g. *Beatles fans* or *Ahmadinejad supporter* (Bauer *et al.* *ibid*). Clearly, the first element of *Beatles fans* refers specifically to the band whose members are Lennon, McCartney, Harrison and Starr, while the first element of *Ahmadinejad supporter* refers to the former president of Iran. The same applies to compounded names of companies, businesses, countries or individuals (*Hewlett-Packard*, *Bosnia-Herzegovina*, etc.), where both elements of the compound refer to specific entities or individuals. Other examples in which the first element of compounds is referential are *earth science*, *sunrise*, *moonlight*, etc. In these examples, the first element has unique reference, i.e. *earth*, *sun* and *moon*. Despite the fact that some complications pertaining to the referentiality of the non-head exist, it seems that the left element/the non-head of English compounds is normally non-referential (Bauer *et al.* 2013: 464).

Finally, in languages such as Hebrew, referentiality can be used to differentiate between the three nominal constructs, i.e. R-constructs (i.e. possessive constructs) vs. M-constructs (i.e. modification constructs) and compounds (Borer 2009). The usefulness of this criterion will be further examined in chapter 4.

### 2.2.8 Coordination

It is well known that phrases can be coordinated using a conjunction such as the underlined phrase in *he wants to have biscuit and jelly*, whereas it is assumed that compounds are not normally coordinated using conjunction (Fàbregas and Scalise 2012). On the other hand, Fàbregas and Scalise (2012: 120) argue that coordination is possible in English for both compounds and phrases, for instance:

(17) He drinks tea and coffee. (phrase)

(18) He is a tea and coffee drinker. (compound)

This means that coordination cannot be relied on to differentiate compounds from phrases in English. It is worth pointing out that these combinations could have two interpretations. The first one is the case in which two compounds are coordinated and the head of the first compound is ellipted. An example of this case is *tea and coffee prices*, which is likely to mean ‘tea prices’ and ‘coffee prices’. The second interpretation is where there is coordination of two modifiers that are part of one single compound such as *a tea and a coffee break*, which is likely to mean ‘a break for tea/coffee’. However, in Spanish, coordination does not work with compounds:

- (19) \*un limpia [botas y ventanas]  
 a polish [boots and windows]  
 ‘a window and boot cleaner’  
 lit. a polish boots and windows

In (19), the coordination is not possible with one element inside the compound. It is grammatical to say *un limpia ventanas* ‘a window cleaner’, but when the word *botas* ‘boots’ is coordinated to the right element *ventanas*, the result is ungrammatical *\*un limpia botas y ventanas*, as can be noted in (19). In Hebrew, Borer (2009: 205) suggests that coordination is a reliable criterion to distinguish between compounds, on the one hand, and various syntactic constructs, on the other. Based on Borer’s (2009) analysis of Hebrew, it seems that coordination is reliable, because all compounds in Hebrew are non-compositional. Needless to say, neither elements of non-compositional compounds can be coordinated (see section 4.3.3).

Note that the English compounds [*tea and coffee*] *drinker* and [*wind and water*] *mills* can be classified as phrasal compounds, since the whole compound consists of two elements: the initial elements, *tea and coffee* and *wind and water*, are phrases, whereas the second, *drinker* and *mills*, are nouns (Lieber 2010: 152). Other instances which include syntactic phrases in the non-head position are [*floor-of-a-birdcage*] *taste*, [*slept-all-day*] *look*, [*pleasant-to-read*] *book* and [*connect-the-dots*] *puzzle* (Lieber 1992: 11). However, Jackendoff (2002: 90-93) remains sceptical about whether or not phrasal compounds are really compounds. It is clear that these compounds have function words inside them, for instance, the coordinate conjunction in the previous two examples. It is well known that phrases contain markers of grammatical functions, such as conjunctions or prepositions, and the meaning of a phrase which contains a conjunction is usually predictable. However, *rock ‘n’ roll* has a conjunction, but its meaning of ‘a type of music’ is not semantically predictable, in other words, it is non-compositional. Being semantically unpredictable and non-compositional is usually an indicator of compoundhood (cf. section 2.2.4). Furthermore, it is possible to replace any of the elements of a phrase with another word, whilst this is not possible in a compound. For instance, in *rock ‘n’ roll*, the second element cannot be replaced by another noun, e.g. *\*rock ‘n’ slide* and still have the meaning of ‘standard musical style’ (Fàbregas and Scalise 2012: 122). Thus, *rock ‘n’ roll* should be treated as a phrasal compound.



### 2.2.9 Replacement of the second element by a pro-form

Yet another possible criterion for compoundhood involves the use of pro-forms. Specifically, Bauer (1998a: 76-77) suggests that it is not possible to replace the second element of a compound with a pro-form. However, in a phrase, it is possible to replace the head noun with the pro-form *one*. For example, *a black one* can refer to our crow, but a *black one* cannot be the genus *Agelaius*. Nevertheless, Bauer shows that this criterion is not always valid. Despite being rare, examples such as *he wanted a riding horse, as neither of the carriage ones would suffice* are attested, where *riding horse* and *carriage horse* appear to be compounds (Bauer 1998a: 77). This means that the second element of a compound can be replaced as shown in the previous example. Hence, this criterion may not be viewed as reliable.

### 2.2.10 Ellipsis

Fàbregas and Scalise (2012: 120) argue that one of the elements of a phrase can undergo ellipsis as in (20), but not the internal elements of a compound:

(20) He drives a truck and he does it every day.

(Fàbregas and Scalise 2012: 120)

This kind of ellipsis utilises VP replacement, which in English requires the auxiliary *do*. Applying the same rule to the compound in example (21), Fàbregas and Scalise (2012) claim that ellipsis is not allowed:

(21) \*He is a truck driver and he does it every day.

(Fàbregas and Scalise 2012: 120)

Here, it is worth pointing out Fàbregas and Scalise (2012) seem to use the term ‘ellipsis’ inappropriately to refer to cases of verb replacement. Ellipsis refers to the deletion of one or more words from a clause that are nevertheless understood from the remaining context. For instance, in the sentence *He said that he would give me the money and he did (give me the money)*, the underlined part is deleted, since it can be understood from the context. In example (20), I argue that verb replacement takes place, rather than ellipsis through replacing *drives a truck* by *does it*. The same applies to example (21) too. It seems *is a truck driver* is being replaced by *does it*; the resulting sentence is ungrammatical, because the replacement is odd,

not because there is a problem with ellipsis which is not ellipsis all together in (21). The examples needed to illustrate this criterion would be of the following type:

(22) When he buys a car, he always buys the fastest \_\_\_\_.

Example (22) shows that an instance of ellipsis in which the head of the phrase, i.e. *car* is omitted.

(23) \*There was only one cup and that was a tea \_\_\_\_.

Example (23) shows a compound in which the head, i.e. *cup* is deleted. However, the sentence is ungrammatical, indicating that the head of a compound cannot be deleted. Nevertheless, ellipsis does work in other cases of compounds, as in (24):

(24) tea and coffee cups

The compound in (24) is likely to be interpreted as *tea (cups) and coffee cups*. It could be said that the first occurrence of *cups* is ellipsed (Fàbregas and Scalise 2012: 120). Similarly, the compound *truck and bus drivers* can undergo ellipsis in the same way as in example (24). The compound *truck and bus drivers* can be interpreted as *truck (drivers) and bus drivers*. It is worth noting that both ellipsis and coordination interact in both examples *tea and coffee cups* and *truck and bus drivers*. That is, whenever coordination applies, one element of the two compounds is not necessarily omitted (see 2.2.8 for the two possible interpretations of these constructs). Additionally, cases of ellipsis in English NPs are rather restricted, since normally the preform *one* has to be used, as in:

(25) \*When he buys a car, he always buys a fast\_\_\_\_.

Example (25) demonstrates that the sentence is ungrammatical, since the head of the compound cannot be omitted, unless it is replaced by the pro-form *one*. This takes us back to criterion 9, i.e. replacement of the second element by a pro-form (see section 2.2.9), in which the head can be replaced by a pro-form in both phrases and compounds.

All in all, whether we are dealing with ellipsis of the head of the first compound when two compounds are coordinated or coordination of two modifiers of a single compound, ellipsis cannot be used to distinguish between compounds and phrases in English.

### 2.2.11 Inflection and linking elements

The (im)possibility of inflecting words has also been advanced as a possible criterion to distinguish between compounds and phrases. In inflectional languages such as Czech, Slovak or Russian, the individual elements of syntactic phrases are inflected (Lieber and Štekauer 2009: 5). But compounds in these languages behave differently, since “[c]ompounds result from the combination not of words, but stems -- uninflected parts of independent words that do not themselves constitute independent words. It is the compound as a whole that is inflected” (Lieber and Štekauer 2009: 5).

However, some examples from Spanish which are considered compounds show that the first element can be inflected (Rainer and Varela 1992: 125):

- (26)        poet-isa-s        pintor-a-s  
              poet-F-PL        painter-F-PL  
              ‘women who are poets and painters’

In example (26), both elements of the compound have to exhibit feminine and plural inflection, so the compound has two instances of inflection.

In English, although the first element of compounds is in most cases inflectionless, as in *houseboat* and *spaceship*, there are counter-examples, referred to by Bauer *et al.* (2013: 436) as ‘descriptive genitives’, like *children’s hour* or *girls’ club* that carry inflection (Lieber 2005: 376). Other examples are *children’s home*, *arm’s-length*, *child’s play* and *no-man’s-land*. Selkirk (1982: 52) suggests that *arms race*, *sales slip*, *buildings inspector* and *weapons analysis* might be considered left-headed compounds, since the left elements are inflected for plurality. Selkirk (*ibid*) states that:

It would seem that the actual use of the plural marker ... might have the function (pragmatically speaking) of imposing the plural interpretation of the non-head, in the interest of avoiding ambiguity. This is probably the case with *programs coordinator* or *private schools catalogue*, for the corresponding *program coordinator* and *private school catalogue* are easily and perhaps preferentially understood as concerning only one program or private school.

Nevertheless, Katamba and Stonham (2006: 329-30) suggest that these compounds are pluralised by adding the plural suffix *-s* to the right element, yielding *arms races*, *sales slips*, *buildings inspectors* and *weapons analyses*. Semantically, *race*, *slip*, *inspector* and *analysis* are the heads. For instance, *buildings inspector* is a kind of *inspector*. Therefore, the *-s* in *arms race* is a plural marker of the non-head not of the whole compound.

In a recent study, Bauer *et al.* (2013: 436) examine examples of descriptive genitives such as *driver's licence*, *mother's milk*, *Broca's aphasia*, *men's room* and *smoker's cough*. Bauer *et al.* indicate that this type of compound could be potentially problematic. For instance, some of the examples of this type have competing forms, with and without the inflectional possessive *'s*. For example, based on the Corpus of Contemporary American English (COCA), Bauer *et al.* (2013) find that *lawyer's fees* and *people's power* can be both used without the genitive *'s*, i.e. *lawyer fees* and *people power*, whilst the deletion of the *'s* is not possible with *mother's milk*, i.e. \**mother milk*.

Bauer *et al.* (2013: 436-7) indicate that although on face value descriptive genitives appear to be phrases, such a classification is debatable. Specifically, descriptive genitives are different from other genitives in that their possessor is a noun, rather than a noun phrase. Therefore, descriptive genitives differ from determiner genitives in that in the latter, the possessor has a determiner function, expanding nominals into noun phrases (Huddleston and Pullum 2002: 354-5). Further, the possessor in determiner genitives causes the whole possessive construct to become definite even though the possessor itself is indefinite, e.g. *a smoker's car* which means 'the car of a smoker'. Here, one may notice that a car of a smoker does not have a corresponding *'s* possessive construct (Huddleston and Pullum *ibid.*).

In contrast, descriptive genitives are, according to Bauer *et al.* (2013: 437), similar to N + N compounds in many respects. Firstly, the first element of the former has word status, not phrasal status. Secondly, it has a classifying semantic function. Thirdly, it has the tendency to be non-referential. Finally, several descriptive genitives have left stress and lexicalised meaning. In fact, Rosenbach (2006: 83) indicates that the mixed behaviour of descriptive genitives makes their classification as compounds or phrases problematic. This confusion can be used as an argument to propose that descriptive genitives are gradient in nature, rather than categorical (Rosenbach 2006: 77).

To sum up, Bauer *et al.* (2013: 437) argue on the basis of the above discussion that there are a number of combinations which are "formally more or less syntactic and semantically more or less compound-like, with some gradience even within individual subtypes". They

conclude that descriptive genitives do have more in common with compounds compared to phrases. The appearance of inflection/linking elements in (potential) compounds has been attested in other languages and this will be discussed in the remainder of this section.

After analysing constructions from Germanic, Romance, Slavic, Finno-Ugric and Modern Greek, Donalies (2004: 76) suggests that one of the criteria which identify compounds is that they may contain a Linking Element (henceforth, LE), which is also known as ‘interfix’ (Dressler 1986). LE can be defined as a special kind of affix, which functions as an extension used to link two elements of a compound (Bauer 2003: 29). Booij (2012: 318) defines LE as a “meaningless element between two constituents of a complex word”. In Modern Greek, Ralli (2009: 454) argues that the first element of a compound is always followed by *-o*, which is semantically empty and is the historical remnant of a no-longer-existent theme vowel. Regardless of their etymological source, these LEs seem to be semantically empty. Examples of these elements in German, where they are common, are given in (27-30):

- (27) *Liebe-s-brief*            ‘love letter’    (*Liebe* ‘love’ + *s* ‘LE’ + *Brief* ‘letter’)  
 (28) *Arbeit-s-anzug*        ‘work suit’     (*Arbeit* ‘work’ + *s* ‘LE’ + *Anzug* ‘suit’)  
 (29) *Liebe-s-lied*            ‘love song’     (*Liebe* ‘love’ + *s* ‘LE’ + *Lied* ‘song’)  
 (30) *Familie.n.name*        ‘family name’ (*Familie* ‘family’ + *n* ‘LE’ + *Name* ‘name’)

In German, the most common LEs are *-s-*, *-es-*, *-(e)n-*, *-er-* and *-e*. In English, Allen (1978) argues that the *-s-* in *guard-s-man*, *craft-s-man*, *oar-s-man*, *trade-s-man*, *kin-s-man* and *deer-s-man* is LE, rather than a plural marker for two reasons: (1) the meaning of the first element in *guard-s-man* is singular; and (2) some elements, such as *deer* and *kin*, do not even inflect for plurality. In English, the *-o-* in *speed-o-meter* and *mile-o-meter* can also be regarded as a LE, since it neither has a meaning nor a specific function. Bauer (2003: 30) also argues that the *-o-* that occurs in the neo-classical compound *electrolyte* in English might also be seen as LE. In general, LEs in Germanic languages historically derive from plural and genitive markers (Bauer 2009: 346). In German, for example, the element *-s* can be found not only as LE but also as a suffix with genitive meaning, as in:

- (31)        das            Auto            mein-es        Bruder-s  
               the            car            my-GEN        brother-GEN  
               ‘the car of my brother’

Note, however, that LEs are not necessarily semantically empty, contrary to Booij (2012: 318). On the basis of a corpus study, Bauer and Renouf (2001: 116) note that the use of the plural is not only clarificatory, but sometimes necessary, as in *drugs-induced*, *forms-compatible*, *savings rate* and *singles-only*. The word *drugs* in *drugs-induced* is used to differentiate between legal drugs and illegal ones. The compound *drug-induced* in *drug-induced sleep* is something ordered by the doctor, whereas the compound *drugs-induced* in *drugs-induced teenage rampage* is something clearly related to drug abuse. Therefore, the plural in *drugs-induced* plays a crucial role in determining the meaning of the compound. This means that the plural marker *-s* is not semantically empty (cf. Selkirk 1982). This issue will not be discussed here any further.

Štekauer and Valera (2007) state that, in general, compounds of the stem + stem type, without any LE, are much more common than those with LE. But in case a language has both types, the LE type tends to be more productive (Štekauer and Valera 2007). Nevertheless, it can be argued that this criterion is specific to Germanic, Romance and Slavic languages (Di Sciullo 2009: 153), and even within these languages, compounds that do not include any LEs can be found. Hence, no generalisation can be made even within Germanic languages. An example from German that does not contain LE is:

(32) *Konzertreise* ‘concert tour’ (*Konzert* ‘concert’ + *Reise* ‘tour’)

Examples from Dutch are (Booij 1992: 37):

(33) *grootvader* ‘grandfather’ (*groot* ‘grand’ + *vader* ‘father’)

(34) *kookpot* ‘cooking pot’ (*kook* ‘cook’ + *pot* ‘pot’)

And finally, some English examples that do not contain LEs are *pickpocket*, *bookshop*, *schoolyard*, *bluestocking* and *truck driver*. Therefore, this criterion is typically found in Germanic languages, and even varies within German, English and Dutch, to be used as a criterion to identify compoundhood cross-linguistically. However, LEs can be used as evidence to show how compounds and phrases are related. If there is no inflection, the combination is a compound, whereas inflected N + N constructs could be compounds or phrases. Nevertheless, all in all, this criterion does not reliably identify compounds in English, German and Dutch.

### **2.2.12 Tonal patterns**

Bauer (2009b: 344) points out that, in some languages, there may be some specific phonological marking of compounding, such as tonal patterns. In Bambara, Creissels (2004: 30-31) argues that compounds are treated like derivatives in terms of their tonal patterns, showing only two patterns: either all syllables have a high tone or, if the first syllable has a low tone, all subsequent tones become high. This is independent of the lexical tone associated with later syllables. Thus, *sàga sògo* ‘sheep meat’ means ‘the meat of the sheep’ and it is a syntactic construct, while *sàgasogo* (with high tones on the last three syllables) is a compound and it means ‘sheep-meat, mutton’. In Hausa, Newman (2000: 116, 190) notes that a reliable criterion to identify compounds is their tonal pattern. Although this criterion is of great interest, it is obviously inapplicable to languages such as English, Dutch, German, Hebrew and Arabic, and will therefore not be pursued here any further, as it is beyond the scope of this study.

## **2.3 Language specific criteria**

In addition to the above-mentioned criteria, which – with due regard to the various problems and issues noted – can potentially be applied cross-linguistically, there are several language-specific criteria that can be applied to identify compounds in a particular language. For instance, all of the following have been proposed to differentiate between compounds and phrases in different languages: vowel reduction (Maipure), vowel deletion (Hebrew), vowel harmony (Chuckchee), voicing (Japanese), tonal sandhi phenomena (Fongbe), postposed definite article (Danish) and word order (French) (Lieber and Štekauer 2009: 12-14). However, in the majority of cases, the literature does not give us enough information on how these criteria distinguish compounding as a type of word-formation; therefore, further research is still needed (Lieber and Štekauer 2009: 14). In this section, I discuss two language-specific criteria in some more detail, i.e. the postposed definite article and word order in Danish and French, respectively. The next sub-sections explain these two criteria.

### **2.3.1 Postposed definite article in Danish**

Bauer (2009a: 404) mentions a language-specific syntactic criterion for distinguishing compounds from phrases in Danish. This criterion involves the use of a postposed definite article, which can be defined as the definite article suffixed to the end of the noun. In Danish, only a single N can take a postposed definite article. An example is given in (35):

- (35) hus-et  
house-the  
'the house'

However, the position of the definite article changes with the addition of another element, e.g. adjective. It has been observed that there is a consistent syntactic difference between an Adj + N compound and an Adj + N phrase. In the former, the definite article is attached to the end of the noun, whereas in the latter it precedes the noun; compare (36) and (37):

- (36) den hvid vin  
the white wine  
'the white wine'

- (37) hvid vin-en  
white wine-the  
'the white-wine'

The definite form of the phrase *hvidvin* 'white wine' is *den hvidvin* 'the white wine', as in example (36), but the definite form of the compound *hvidvin* is *hvidvinen* 'the white wine' as in example (37) (Bauer *ibid*). As expected, in the phrase *den hvid vin*, the adjective *hvid* 'white' is gradable and sub-modifiable, referring to the colour of wine. Conversely, the adjective *hvid* 'white' in the compound *hvidvinen* is not modifiable, referring to a kind of wine independent of its actual colour if contrasted with red wine and rosé. As a result, it is clear that *hvidvin* is a single complex word and not a noun with an independent premodifier.

To sum up, if a postposed definite article is possible, there is evidence that a sequence of two roots must be a compound in Danish (Bauer *ibid*).

### 2.3.2 Word order

Fradin (2009: 422-423) proposes that word order gives us evidence for compoundhood in French; if a sequence of lexemes displays an order that cannot be generated for syntactic phrases, we are likely dealing with a compound. For instance, Fradin (2009: 419) argues that the following examples must not be considered compounds, since they are instantiations of the



syntactic construct [N PP]<sub>NP</sub>, with a noun phrase consisting of a head followed by a PP complement:

- |      |                         |              |   |
|------|-------------------------|--------------|---|
| (38) | <i>avion à réaction</i> | ‘jet plane’  | ( <i>avion</i> ‘plane’ + <i>réaction</i> ‘jet’) |
| (39) | <i>chambre d’hôte</i>   | ‘guest room’ | ( <i>chambre</i> ‘room’ + <i>hôte</i> ‘guest’)  |
| (40) | <i>poêle à frire</i>    | ‘frying pan’ | ( <i>poêle</i> ‘stove’ + <i>frire</i> ‘fry’)    |

Fradin (2009: 419) also suggests that N + Adj and Adj + N syntactic substructures in French are not compounds, claiming that they occur in sentences where they are plain NPs and have no idiomatic meaning, as in the following examples:

- |      |                         |                  |  |
|------|-------------------------|------------------|--|
| (41) | <i>nature morte</i>     | ‘still life’     | ( <i>nature</i> ‘nature’ + <i>morte</i> ‘dead’)        |
| (42) | <i>beaux-arts</i>       | ‘fine arts’      | ( <i>beaux</i> ‘beautiful’ + <i>arts</i> ‘arts’)       |
| (43) | <i>premier ministre</i> | ‘prime minister’ | ( <i>premier</i> ‘prime’ + <i>ministre</i> ‘minister’) |

Such examples have been considered genuine compounds by Gross (1996) and Mathieu-Colas (1996). However, Fradin (2009: 419-20) argues that no sound argument has been provided by the two scholars, since they confuse compounding with idiomaticity. Fernando and Flavell (1981: 48) point out that the meaning of an idiom is not the total sum of the compositional function of its elements. For instance, the meaning of the idiom *kick the bucket* ‘die’ cannot be derived from the meaning of the individual elements, i.e. *kick* and *the bucket*. Similarly, Baker (1992: 63) rightly argues that idioms “are frozen patterns of language which allow little or no variation in form and [...] often carry meanings which cannot be deduced from their individual components” (see section 2.2.4). Therefore, the order of the words in idioms cannot normally be changed. The deletion or a replacement of a word is also not possible, neither is the change of the idiom’s grammatical structure, with the exception of intentional play on words (Baker, 1992). In this regard, Lattey (1986: 219) explains that idioms are a group of words whose meanings are not the same as the meaning of the internal elements. In sum, it seems that the non-compositionality aspect is still the most crucial property of idiomaticity. Due to the fact that idioms and compounds could be non-compositional, several researchers have confused between compounds and idioms, which is supported by Fradin’s (2009) argument regarding

idiomaticity not being used as a criterion to identify compounds.<sup>5</sup> Note that sheer idiomaticity is unlikely to be used as a criterion to differentiate between compounds and phrases, because compounds can also be compositional and productive (cf. Lieber 2005: 376; Rosenbach 2006: 83). All in all, Fradin (2009) concludes that a compound in French is a sequence of lexemes that cannot be generated by syntactic rules or principles, regardless of non-compositionality, e.g. *lave-vaisselle* ‘dishwasher’ which consists of *lave* ‘wash’ and *vaisselle* ‘dishes’.

To the best of my knowledge, the two criteria above have not been observed in other languages. Therefore, they are language-specific criteria, and will not be discussed here any further.

## 2.4 Compounding and derivation

It has been suggested that compounding and derivation may not be clearly distinct in some languages, including English. De Belder (2013: 40-41) suggests that compounds are prototypically constructed by free morphemes, and derivations by bound morphemes. One type of compound, namely, neoclassical compounds such as *biology*, *biography* and *anthropology* may be problematic under De Belder’s (2013) distinction, since it has been argued that neoclassical compounds are not composed of free morphemes. In addition, both combining forms and affixes can be added to lexemes, such as the combining form *-ology* in *music-ology* vs. the derivational suffix *-al* in *music-al*. A combining form can be defined as a “bound morpheme, more root-like than affix-like, usually of Greek or Latin origin, that occurs only in compounds, usually with other combining forms. Examples are *poly-* and *-gamy* in *polygamy*” (Carstairs-McCarthy 2002: 142). Booij (2007: 86) argues that neoclassical compounds occur when one of the elements is a root borrowed from Greek or Latin, which does not correspond to a lexeme. Booij (*ibid*) distinguishes three different cases:

- (44)      bio-logy, psycho-logy, socio-logy, geo-graphy, tomo-graphy  
(two combining forms)
- (45)      tele-camera, tele-phone, tele-vision, tele-gram, tele-kinesis  
(the final element is a lexeme)

---

<sup>5</sup> It seems that word order as a criterion was first suggested by Marchand (1969: 22) for English, especially for compounds with present or past participle as the second element such as *easy-going*, *high-born*, *man-made*. Even though these combinations have double stress, they are regarded as compounds by Marchand (*ibid*). In these cases, Marchand (*ibid*) points out that the first element cannot function syntactically as a modifier of the right-hand element, so that this criterion overrules the double-stress rule.

- (46)          magneto-hydro-dynamic, magnet-metry, bureau-crat  
(the first element is a lexeme)

Thus, the borderline between compounding and derivation is blurred at least in English. Bauer (1998b) argues that neoclassical compounds cannot be differentiated from prefixation. For example, in the word *geo-morphology*, the bound morpheme *geo* can be analysed either as a prefix attached to the lexeme *morphology*, or as a combining form attached to the lexeme *morphology* like the combining form *tele* in *tele-vision*.<sup>6</sup>

Furthermore, it is difficult to differentiate neoclassical compounding from blending and clipping, as in *Eurocrat* and *gastrodrama*. Neoclassical roots sometimes combine with affixes, such as *gynocidal*. Bauer (1998b) argues that if productivity is measured based on coining new forms unconsciously, we might hesitate to call neoclassical compounds productive. Nonetheless, some new neoclassical compounds have been formed in English (Bauer *ibid*). As a result, Booij (2009: 208) proposes the term ‘semi-affixes’ or ‘affixoids’ to refer to the constituents of neoclassical compounding, which are intermediate between affixes and lexemes. The terms ‘semi-affixes’ and ‘affixoids’ seem similar to the term ‘combining forms’, which is found in Carstairs-McCarthy (2002: 66).

The main characteristics of these combining forms that differentiate them from affixes are: (1) having positional freedom; (2) creating new words on their own; (3) containing linking elements; (4) having a higher degree of lexical density, i.e. ‘being semantically contentful’ (Bauer 1998b: 407); (5) tending to become free word; and finally (6) being the base of derivational suffixes (Bauer 1998b: 407; Carstairs-McCarthy 2002: 66; Fàbregas and Scalise 2012: 113; Ralli 2010: 59).

With respect to the first characteristic, Ralli (2010: 59) notes that affixes obey certain positional restrictions: prefixes precede the base, while suffixes follow, as in:

- (47)          a)     rewrite  
                b)     \*writere
- (48)          a)     happiness  
                b)     \*nesshappy

---

<sup>6</sup> Geo-morphology is the study of the evolution, features and configuration of the earth’s surface (from Greek *ge* ‘earth’; *morfé*, ‘form’ and *logos* ‘study’).

The prefix *re-* in *rewrite* and the suffix *-ness* in *happiness* cannot change their position, leading to unacceptable words, i.e. *\*writere* and *\*nesshappy*. However, in neoclassical compounds, some elements can appear before or after the base, like *phil* in *philharmonic* and *francophile*. Similarly, Fàbregas and Scalise (2012: 113) cite examples of neoclassical compounds that exhibit positional freedom, such as:

- (49)     a)     log-o-graph-y  
          b)     graph-o-log-y

The combining form *log-* can appear to the left of second element as in (49a) and to the right, as in (49b), exchanging its location with the combining form *graph-* (49a) vs. (49b). The constituents found in neoclassical compounds share properties of both lexemes and affixes (Ralli 2010: 59). Consequently, Fàbregas and Scalise (2012: 113) suggest that combining forms are like compounds, since those too sometimes have positional freedom, such as *apple* in *apple pie* ‘a type of pie made with apples’ and *pie apple* ‘type of apple specially used in pies’, and *white* in *white collar* and *milk white*.

Secondly, two combining forms may form a word such as *psych-o-logy*, *bi-o-logy*, *ge-o-graphy*, *electr-o-phile* and *tom-o-graphy* (with the *-o-* in each case being a linking element). In contrast, affixes cannot be used to create new words on their own, such as *\*re-ness*, *\*pre-ly* and *\*anti-tion*.

Thirdly, most neoclassical compounds behave like some other compounds, e.g. *guard-s-man*, *kin-s-man* and *speed-o-meter*, in terms of having LE. Examples of LEs, such as *-o-* and *-i-*, in some neoclassical compounds are *music-o-logy*, *anthr-o-pology* and *hom-i-cide* (Carstairs-McCarthy 2002: 66). On the other hand, LEs never appear with affixes. This means that the presence of LEs with bound morphemes is an indication that we are dealing with combining forms not affixes. Note that combining forms are not limited to English; some combining forms are productive across European languages, e.g. *afro-*, *compu-*, *crea-*, *cine-*, *cyber-*, *digi-*, *docu-*, *flexi-* and *euro-* (Booij 2007: 88).

Fourthly, Bauer (1998b: 407) differentiates between a combining form and an affix based on the kind of semantic information the morph conveys. It has been argued that the former has a higher degree of lexical content or density than the latter (Bauer 1983: 215). For example, the meaning of *neuro-* ‘related to the nervous system’ appears to be much more semantically contentful than the meaning of the prefix *re-* ‘again’ (Bauer *ibid*). In fact, Bauer

(1998b) suggests that there is a continuum from most semantically contentful to least semantically contentful bound morphemes. At the more contentful extreme, there are neoclassical combining forms, which are quite similar to independent lexical morphemes in meaning.

Fifthly, according to Fischer (1988: 57), if a combining form can be used as a free lexical element, preserving the same style and meaning, then at least synchronically, it should no longer be regarded as a combining form. For instance, since the 1980s, the combining form *electro* has been used to describe a type of electronic music. Due to a long period of use, *electro* has become a homophonous noun and adjective. As a result, synchronically, neither *electrobeat* nor *electo funk* are neoclassical combinations; rather they are compounds, consisting of two free morphemes. Similar cases that can be cited are *video*, *audio*, *hyper*, *poly*, *telly* and *porn*, which are not combining forms, rather free morphemes (Fischer *ibid*). If we take the combining form *hyper* as an example, it used to appear in technical and medical contexts, such as *hypertension*. At present, it can be used as a free lexical morpheme as in *he was very hyper yesterday*, which is an abbreviation of *hyperactive*. Similarly, instead of *polytechnic* and *television*, many speakers use *poly* and *tele* (usually spelled as *telly*), respectively. Affixes, by way of contrast, rarely change into lexical elements. Examples of affixes yielding lexical elements are *-ism* and *-ish*. Bauer (2005:101) notes that in English the derivational suffix *-ish* has developed into a separate word when it functions as a qualifier. Norde (2009: 223-25) mentions the examples below of *-ish* separated from the adjective it qualifies:

- (50) They have a pleasantly happy<sub>i</sub> ending (well, *t<sub>i</sub> ish*).
- (51) Is everyone excited<sub>i</sub>? I am— *t<sub>i</sub> ish*.
- (52) Can you swim well<sub>i</sub>?: *t<sub>i</sub> ish*.

Contrary to *-isms*, the development of *-ish* is not a case of lexicalisation of an affix for two main reasons. Firstly, it is known that lexicalised affixes become part of main word classes, i.e. nouns or verbs. However, *-ish* does not (which is best perceived as an adverbial ‘kind of’). Secondly, lexicalised suffixes are hypernyms of all the derived words with that suffix, i.e. *-isms* refer to all ideologies which end in *-ism*, such as capitalism and socialism. Conversely, *-ish* is not a hypernym of all adjectives ending in *-ish* (Norde 2009: 223-25). This issue, however, is beyond the scope of this study and thus is not pursued any further.

Finally, Carstairs-McCarthy (2002: 66) notes that combining forms can function as the base for derivational suffixes. Examples of such suffixes are *soci-* and *electr(o)-*, from which

*social* and *electric* can be formed. Affixes, on the other hand, are never used as bases for derivational suffixes, such as *tion* and *ic* \**tional* and \**ical*. In other words, affixes can be added to combining forms to form words, but affixes cannot be added to other affixes to create words.

All in all, it seems to me that the facts discussed here support the conclusion that the elements of neoclassical compounds are more root-like than affix-like. Bauer *et al.* (2013: 441-442) also suggest that the distinction between combining forms and derivational affixes is clear-cut, stating that “...neoclassical formations are best treated as compounds, and not as cases of affixation”.<sup>7</sup> As a result, I would argue that neoclassical formations are to be regarded as compounds.

## 2.5 Conclusion

In this chapter, several criteria used to distinguish between compounds on the one hand, and phrases and derivation on the other have been discussed. The majority of these criteria are potentially useful, even though not all of them can be straightforwardly applied to all languages. That is, some criteria are more reliable and widely applicable than others. For instance, stress can be applied to many languages (e.g. English, Dutch, Hungarian, Polish, German, Modern Greek, etc.), whereas some criteria are applicable to a certain language (e.g. postposed definite article in Danish). Furthermore, some criteria are partially useful to distinguish between compounds and phrases, i.e. compositionality. It has also become evident that drawing a boundary between compounding and phrases is not an easy task. It is for this reason that Bauer (1998a: 78) indicates that there is no criterion that gives a reliable distinction between the two types of construction, i.e. compounds and phrases, at least in English. In line with Bauer (1998a: 78), Plag (2006) is sceptical about what exactly a compound is, and the possibility of differentiating between N + N compounding and phrases. Nonetheless, assuming that phrasal verbs are not compounds but constructional idioms (Jackendoff 2002: 188), it can be concluded that in English, the most reliable criterion to differentiate between compounds and phrases is ‘adjacency’. This criterion can be applied to all the examples in Table 2.3 below.

---

<sup>7</sup> It is worth pointing out that there is an internal inconsistency in Bauer *et al.*'s (2013) book in which they suggest that the distinction between combining forms and derivational affixes is not clear-cut (Bauer *et al.* 2013: 486).

Table 2.3. Possible internal elements of compounds in English

<b>Compound</b>	<b>The internal elements of the compound</b>
windmill, egghead, truck driver, blackbird	two words
biology, sociology	two combining forms
television, telephone	combining form plus word
bureau-crat, magnetometry	word plus combining form
[water and wind] mill, [tea and coffee] cups, [pipe and slipper] husband, [floor-of-a birdcage] taste, [slept-all-day] look, [pleasant-to-read] book, [connect-the-dots] puzzle.	the first element is a phrase, the final element is a word
[jack-[in-the-box]], [mother-[in-law]], [bikini girls-[in-trouble]], [good-[for-nothing]]	the first element is a word, the final element is a phrase

Taking the examples in Table 2.3 into consideration, the following working definition of a compound, at least in English, can be suggested: a compound is a complex word that consists of at least two adjacent elements, in which each of these elements is either a word, combining form or a phrase, so that the whole compound is a combination of these elements.

And finally, although there are a few cases where referentiality, as a criterion, fails to distinguish between compounds and phrases, e.g. when the non-head is a proper noun or has unique reference, referentiality can be considered a significant criterion when we are identifying compounding cross-linguistically.

## Chapter Three: Identifying compounding in MSA and JA: The orthographic and phonological criteria

### 3.1 Introduction

After having identified the cross-linguistic criteria that can be used to distinguish between compounds on the one hand, and phrases on the other, in the previous chapter, the orthographic and phonological criteria are applied to MSA and JA in this chapter. Essentially, this chapter provisionally distinguishes between two types of N + N combinations in MSA and JA on the basis of some of the criteria discussed in chapter 2, e.g. adjacency. Later, this provisional division is tested, with the data to be examined in much more detail, against all the criteria, in the rest of this chapter and in the following chapter. In this chapter, the focus is mainly on the orthographic and phonological criteria, while the next chapter discusses the semantic and syntactic criteria.

This chapter proceeds as follows: section 3.2 discusses the basis upon which the provisional division between the two N + N combinations within SGCs in MSA and JA has been made through applying the most reliable criterion in chapter 2, i.e. adjacency. Section 3.3 discusses the orthographic criterion. Section 3.4 explains and discusses the phonological phenomenon of sandhi. Section 3.5 discusses stress assignment in N + N combinations in English. Section 3.6 examines to what extent stress can be used as a criterion to distinguish between the two N + N combinations within SGCs in MSA and JA, using Praat software. Finally, section 3.7 summarises the similarities and differences between the two nominal constructs.

### 3.2 The two possible N + N constructs in MSA and JA

In this section, I will be provisionally distinguishing two types of N + N combinations in MSA and JA on the basis of some of the criteria discussed in chapter 2. However, this will only be a first pass; a careful investigation is carried out with in much more detail, based on all the criteria, in the rest of this chapter and in the following chapter.

Examination of data in MSA and JA suggests that there are **two** possible nominal constructs that exhibit different behaviours.<sup>8</sup> One is the ‘the noun-noun genitive construct’

---

<sup>8</sup> These two nominal constructs are traditionally grouped under SGCs (*Id'aafa* in Arabic), which is generally defined as a possession relationship which links two nouns together, where the first element is the possessum and the second one is the possessor (Al-Rajihi 2000: 247).



(Ryding 2005: 205); it is equivalent to a possessive phrase containing *of* in English, such as *the king of Jordan*. This possessive construct is referred to as the P-construct henceforth. The other construct can be considered a type of compounding. Examples of the two types of construct are shown below:

- |      |  |                           |               |
|------|--|---------------------------|---------------|
| (30) | qamiis <sup>s</sup> -u<br>shirt-NOM<br>'the boy's shirt'     | l-walad-i<br>the-boy-GEN  | (P-construct) |
| (31) | bayt -u<br>house-NOM<br>'the man's house'                    | r-radʒul-i<br>the-man-GEN | (P-construct) |
| (32) | maaʔ-u<br>water-NOM<br>'the dignity'<br>lit. face water      | l-wadʒh-i<br>the-face-GEN | (compound)    |
| (33) | ʒaruus-u<br>bride-NOM<br>'the mermaid'<br>lit. the sea bride | l-baħr-i<br>the-sea-GEN   | (compound)    |

The relationship between the two elements of the constructs in (1-2) is a possession relationship, whilst this is not the case in examples (3-4). Regarding adjacency, which is discussed in detail in chapter 4, it seems that an element, e.g. a demonstrative, can be inserted between the two elements of the constructs in examples (1-2), whereas such insertion is impossible in examples (3-4). Note, however, that the two constructs are syntactically similar with regard to case marking. That is, the first noun, in either of these two constructs, can have any case, such as nominative, accusative or genitive on the basis of the function of the whole construct in the sentence. The second/right element is always in the genitive case, as illustrated in examples (1-4). Finally, based on compositionality, it appears that examples (1-2) are compositional, since the meaning of the whole construct is the total sum of the meanings of its two elements, whereas examples (3-4) are non-compositional. Here, note that in her analysis

of nominal constructs in Hebrew, Borer (2009) distinguishes between three nominal constructs, namely, R-constructs (i.e. possessive constructs), M-constructs (i.e. modification constructs) and compounds based on compositionality. Examples of these three constructs can be seen in examples (5-10):

- |      |                  |                               |                               |
|------|------------------|-------------------------------|-------------------------------|
| (34) | beyt<br>house    | (ha-)mora<br>(the-)teacher    | (R-construct)                 |
|      |                  |                               | ‘(the) teacher’s house’       |
| (35) | orex<br>editor   | (ha-)ma'amar<br>(the-)article | (R-construct)                 |
|      |                  |                               | ‘(the) editor of the article’ |
| (36) | kos<br>glass     | (ha-)mic<br>(the-)juice       | (M-construct)                 |
|      |                  |                               | ‘(the) juice glass’           |
| (37) | magevet<br>towel | (ha-)mitbax<br>(the-)kitchen  | (M-construct)                 |
|      |                  |                               | ‘(the) kitchen towel’         |
| (38) | beyt<br>house    | (ha-)sefer<br>(the-)book      | (compound)                    |
|      |                  |                               | ‘(the) school’                |
| (39) | orex<br>editor   | (ha-)din<br>(the-)law         | (compound)                    |
|      |                  |                               | ‘(the) lawyer’                |

According to Borer (2009), the clearest criterion to differentiate between compounds and R/M constructs in Hebrew is compositionality. Borer argues that in the case of productive syntactic constructs (i.e. R-constructs and M-constructs), the meaning is completely predictable from their parts, as shown in (5-8). In contrast, she argues that the meaning of compounds in Hebrew

is non-compositional, i.e. it is not predictable from the individual N components. The compounds in (9) and (10) are non-compositional, since the meaning of the whole compound is not the total sum of the meanings of its elements. For instance, the meaning ‘lawyer’ in (10), cannot be predicted from the meanings of its two elements, i.e. *orex* ‘editor’ and *hadin* ‘the law’.

However, this analysis of nominal constructs in Hebrew may not be applicable cross-linguistically. Compositionality fails to distinguish compounds from phrases in English and other languages, as explained in section 2.2.4. Therefore, one may suggest that the line between these nominal constructs cannot be drawn solely on the basis of compositionality as is the case with Hebrew.

Having established that there is some prima-facie plausibility to a division of N + N combinations in MSA and JA into phrases (P-constructs) and compounds, I will now test whether this division still holds up when investigated more carefully and against all the criteria in this chapter and the following one. The next section examines the orthography of these combinations and investigates whether it can be used to support the idea that they are indeed two different types of construct.

### 3.3 Orthography

In MSA and JA, there are numerous examples in which spelling as a criterion fails to differentiate between P-constructs and compounds. Both of these nominal constructs in MSA and JA usually contain two elements that are written as separate words. Thus, orthography does not distinguish between these two constructs. This can be seen in the following examples, where the word division given in the transliteration corresponds to what would be found in the Arabic-script versions of these combinations:

(40)	qamiis <sup>6</sup>	l-walad <sup>9</sup>	(P-construct)
	shirt	the-boy	
	‘the boy’s shirt’		

---

<sup>9</sup> The case markings in MSA, which are not realised in JA, will be neutralised throughout the thesis, and they will only be realised when necessary. Note, also, that the second element in both P-constructs and compounds can be definite, marked with the definite article *l-*, or the indefinite marker *-n*, i.e. nunation (see section 3.4.2). Even though the definite article is marked on the second element, the whole compound becomes definite, regardless of the indefiniteness of the first element.

- |      |   |                           |               |
|------|---|---------------------------|---------------|
| (41) | mazraʕat<br>farm<br>'the man's farm'                            | r-radʒul<br>the-man       | (P-construct) |
| (42) | duudat<br>worm<br>'the earth worm'                              | l-ʔardʕ<br>the-land/earth | (compound)    |
| (43) | raaʔid<br>pioneer<br>'the astronaut'<br>lit. the space pioneer. | l-fadʕaaʔ<br>the-space    | (compound)    |

All in all, orthography offers no help in differentiating MSA and JA compounds from phrases.

### 3.4 The sandhi phenomenon

I now turn to possible phonological criteria; one segmental (sandhi; discussed in this section) and the other prosodic (stress; discussed in 3.5).

#### 3.4.1 Overview

A prominent feature of N + N constructs in MSA and JA is sandhi. Andersen (1986: 1) defines sandhi as follows:

It refers to liminal phenomena: the junctures between segments, variation and alternations at the boundaries of constituents, or – from another point of view – the interfaces between phonetics and phonemics, and between phonology and morphology, including such truly liminal phenomena as allophones with apparently distinctive function, neutralizations with grammatical function, and so on.

More precisely, Jongen (1986: 119) defines sandhi as “all phonological modifications associated with a combination of signs and localised at their boundaries”. Another definition is suggested by Lipiński (2001: 202), who points out that sandhi refers to “the assimilative changes occurring in a word under the influence of neighbouring words uttered in consecutive speech”. He also indicates that this phenomenon is widely attested in Semitic languages in

general, and Arabic in particular. Simply put, sandhi refers to the processes which the form of a word or morphological formative undergoes as a result of the influence of an adjacent word or morphological formative. For example, the plural form *mes* of the French possessive determiner usually changes its pronunciation from /me/ to /mez/ before a word that begins with a vowel, as in, *mes amis* /mez ami/ ‘my friends’ in comparison with *mes raquettes* /me Raket/ ‘my rackets’ where there is no change. Sandhi can be either word-internal, at morpheme boundaries within words, as in the word *sympathy*, which consists etymologically of the two morphemes *syn-* + *pathy*; or word-external, at word boundaries, as in the pronunciation *i[m] Manchester* for *in Manchester* in some varieties of English.

With this by way of background, we now turn to the types of sandhi found in MSA and JA to determine whether it can be used as a criterion to distinguish between P-constructs and compounds, that is, to check whether sandhi can operate in certain types of N + N constructs but not others. As we shall see in section 3.5, this phenomenon also seems to play a role in stress assignment in MSA and JA.

### 3.4.2 Types of sandhi in MSA and JA

The most common type of sandhi in MSA occurs when feminine nouns ending in the bound-*h* morpheme ‘feminine form’ are pronounced with a final [t] instead of *-h* so that, for example, *zawdzah* ‘wife’ becomes *zawdzat rradzul* ‘the man’s wife’ as in the P-construct in (15), and *muʕallimah* ‘(female) teacher’ becomes *muʕallimat lfiizyaaʔ* ‘the (female) physics teacher’, as in the compound in (16):

- (44) a)      zawdz-ah  
              wife-FSG  
              ‘wife’
- b)      zawdz-at      r-radzul  
              wife-FSG      the-man  
              ‘the man’s wife’

- (45) a) muṣallim-ah  
 teacher-FSG  
 ‘(female) teacher’
- b) muṣallim-at l-fiizyaa?  
 teacher-FSG the-physics  
 ‘the (female) physics teacher’

Recall that the classification of N + N combinations into P-constructs and compounds in (15) and (16) is based on adjacency and the possessive relationship between the two elements (see section 3.2, examples 1-4). For instance, it is possible to insert *haaḏa* ‘this’ between the two elements of the P-construct in example (15b), as in *zawdzat haaḏa rradzul* ‘this man’s wife’. However, such insertion is impossible with the compound in example (16b), as in *\*muṣallimat haaḏihi lfiizyaa?* ‘teacher of this physics’.

With regard to the conditioning factor for the alternation between [h] and [t], Lipiński (2001: 205) posits that the deletion of the feminine ending *-t* at the end of a word is a pervasive phenomenon in Semitic languages. Across word boundaries, the *-t* ending reappears when preceded by a vowel in consecutive speech. At one point in the history of this language family, the deletion of the *-t* ending gave way to the appearance of *-a* at the end of the word. However, the residual final *-a* was then indicated in Arabic and Hebrew writing by the consonant *-h*.<sup>10</sup> Now, whether *-h* serves as a *mater lectionis* (i.e. the use of a certain consonant to indicate a vowel) is still unclear (Lipiński *ibid*). Borer (2009: 493), who discusses the sandhi phenomenon in Hebrew, provides data in which the *-t* ending reappears across word boundaries as well, as in example (17):

- (46) a) šmira  
 guarding.FEM
- b) šmirat sáf  
 guarding threshold  
 ‘gate keeping’

---

<sup>10</sup> In MSA and JA, the <h> sound is pronounced at the end of the word, e.g. *zawdzah* ‘wife’.

- c)      šmirat          yeladím  
           guarding      children  
           ‘guarding children’

On the basis of example (17), Borer (2009) argues that the feminine singular forms ending with *-a* have a distinct bound form when they appear as the head in both compounds and constructs, as can be seen by comparing *šmira* in (17a) with *šmirat* in (17b, c). In the latter case, the ending *-t* resurfaces across any word boundaries, especially when uttered in consecutive speech. This type of sandhi does not distinguish P-constructs from compounds.

Secondly, an examination of example (18b) shows that the first segment of the second/right element in P-constructs is sometimes attached to the final syllable of the first element of the construct when pronounced in consecutive speech:

- (47)      a)      bayt-u                      radzul-i-n                      (P-construct)  
                   house-NOM                  man-GEN-a  
                   ‘a man’s house’
- b)      bayt-r                      radzul-i                      (P-construct)  
                   house-NOM                  the man-GEN  
                   ‘the man’s house’

(Alexiadou, Haegeman and Stavrou 2007: 248)

In example (18b), the two words *bayt* ‘house’ and *rradzuli* ‘the man-GEN’, are connected together by sandhi. The whole construct in (18b) can be divided into five syllables, in which the second syllable has three sounds; one from the first noun, another from the nominative case marker *-u*, and the last one from the second noun, as in *bay.tur.ra.dzu.li* (the bold segments show where the connection occurs). Note that the definite article *l-*, which is attached to the second element and then undergoes assimilation to yield *rradzul* ‘the man’, (see section 3.6.3.3), connects with the final syllable of the first element, the whole process being an instance of sandhi. Ryding (2005: 40) describes the definite article *l-* (henceforth, the definiteness marker) as a prefix which is attached to nouns and adjectives, denoting definiteness. Looking more closely at example (18b) and the syllable outline of *bayt-r radzul* ‘the man’s house’, it should be noted that, when pronounced in consecutive speech, a vowel realising the case appears between the two consonants, i.e. *t* and *r* in *baytr* ‘house’, yielding

*bay.tur*. This vowel *u* is apparent in the syllable outline, but it is not represented in Alexiadou *et al.*'s example (18b). It is true, however, that the presence of the vowel *u* is also not always reflected in Arabic spelling; it is only realised as a diacritic in MSA. Example (19) gives the actual surface form of the combination:

- (48)      bayt-u-r                  radzul-i                  (P-construct)  
             house-NOM-the          man-GEN  
             'the man's house'

On the basis of example (19), it seems that sandhi appears in P-constructs, linking the two elements together phonologically. That is, the first segment of the second/right element, i.e. *rradzul* 'the man' is attached to the final syllable of the first/left element of the P-construct, i.e. *bayt* 'house'. However, it appears that the same applies to compounds, as illustrated with the following examples:

- (49)      faras-u-n<sup>11</sup>                  nahr                  (compound)  
             horse-NOM-the          river  
             'the hippopotamus'
- (50)      findzaan-u-f                  jaay                  (compound)  
             cup-NOM-the          tea  
             'the tea cup'

Examples (20) and (21) show that the second type of sandhi also applies to compounds; the first syllable of the second/right element in these compounds is attached to the final syllable of the first element of the construct, i.e. *fa.ra.sun.nahr* and *fin.dzaa.nuf.jaay* (the bold segments show where the connection occurs).

Thirdly, nunation and the final [n] of the masculine sound plural and dual suffixes are both omitted when the noun which contains either of them is in the construct head position.

---

<sup>11</sup> The assimilation of the geminate will be discussed in detail in section 3.6.



<sup>12,13</sup> In MSA, Al-Jahaawi (1982: 10-11) notes that Arab phoneticians define nunation (henceforth, indefiniteness marker) as “a short vowel plus the nasal consonant *n* ...”. Ryding (2005: 161) states that “Indefiniteness as a noun feature is usually marked by a suffixed /-n/ sound, which is written in a special way as a variation of the case-marking short vowel at the end of a word”. With regard to the masculine sound plural and the dual in MSA, they end with the segment *n*. The following examples demonstrate this phenomenon:

- (51) a) muṣallim-u-n  
 teacher-NOM-INDF  
 ‘teacher’
- b) muṣallim-u-l                      fiizyaa?  
 teacher-NOM.INDF-the            physics  
 ‘the physics teacher’
- (52) a) muṣallim-uun  
 teacher-NOM.MPL  
 ‘teachers’
- b) muṣallim-uu-l                      fiizyaa?  
 teacher-NOM.MPL-the            physics  
 ‘the physics teachers’
- (53) a) muṣallim-aan  
 teacher-NOM.DUAL  
 ‘two teachers’

---

<sup>12</sup> Fassi-Fehri (2012: 277) notes that the exact meaning of the [-n] ending of nunation is difficult to pin down, due to its debatable identity. He argues that it has been mistakenly identified as an indefinite marker (see Kouloughli 2007 for more detail). Instead, he views it as the head of a poss(essive) phrase, marking the absence of the possessor element, or absence of individuation. In fact, it does not appear in nouns heading a construct state (referred to in Arabic as *Id’aafah* and in this study as SGC), or individuated vocative nominals (see Fassi-Fehri 1993). I agree that nunation should not be treated as an indefinite maker. However it also needs to be recognised that nunation seems to be used for stylistic purposes, since it appears on proper nouns which are, by default, definite. The issue clearly needs further investigation.

<sup>13</sup> The masculine sound plural is formed by suffixation of *-uun* to a usually unchanged stem (Ryding 2005: 107).

- b)      muᶑallim-aa-l                      fiizyaaʔ  
             teacher-NOM.DUAL-the          physics  
             ‘the two physics teachers’

The segment *n* is deleted in the nunation case (22b), in the sound plural (23b) and in the dual (24b), when the element that contains these endings appears as the first element of N + N combinations. This is shown for P-constructs in (25) and (26), where in each case the a-example is the word in isolation and the b-example has it as the head of a P-construct:

- (54)      a)      bint-u-n  
                     daughter-NOM-INDF  
                     ‘a daughter’
- b)      bint-u                                      ḥamid-in  
                     daughter-NOM.INDF                  Hamid-GEN  
                     ‘Hamid’s daughter’
- (55)      a)      waalid-aan  
                     parents-NOM.DUAL  
                     ‘parents’
- b)      waalid-aa-l                                  ᶑariis-i  
                     parents-NOM.DUAL-the                  groom-GEN  
                     ‘the parents of the groom’

Examples (25) and (26) show that the final *n* in the nunation case and dual is deleted when they appear as the first element of P-constructs. The same is shown to apply to compounds in examples (27-29):

- (56)      a)      findʒaan-u-n  
                     cup-NOM-INDF  
                     ‘a cup’

- |      |    |   |                 |
|------|----|---|-----------------|
|      | b) | findzaan-u-l<br>cup-NOM.INDF-the<br>'the coffee cup'                                      | qahwa<br>coffee |
| (57) | a) | sikkat-aan<br>rail-NOM.DUAL<br>'two railways'   |                 |
|      | b) | sikkat-aa-l<br>rail-NOM.DUAL-the<br>'the two iron railways'                               | ħadiid<br>iron  |
| (58) | a) | muwað <sup>ç</sup> ð <sup>ç</sup> af-uun<br>employee-NOM.MPL<br>'employees'               |                 |
|      | b) | muwað <sup>ç</sup> ð <sup>ç</sup> af-uu-l<br>employee-NOM.MPL-the<br>'the bank employees' | bank<br>bank    |

Similar to P-constructs, examples (27-29) show that the final *n* in the sound plural and dual is deleted when they appear as the first element of compounds.

Taking all these observations into consideration, we can say the following: the first type of sandhi, which affects feminine nouns in which the bound-*h* morpheme 'feminine form' is pronounced as a final [t] instead of [h], can be found in both P-constructs and compounds. The second type of sandhi, connecting the definite article *l-* to the final syllable of the first element, also appears in both N + N types. Finally, the nunation and the final [n] of the sound plural and dual suffix are deleted when the noun which contains either of them is in the left element position in both P-constructs and compounds. Therefore, one may conclude that sandhi as a criterion fails to distinguish between P-constructs and compounds in MSA and JA. The next section sheds more light on one of the most useful phonological criteria to identify compounding cross-linguistically, namely, stress.

### 3.5 Stress assignment in N + N combinations

Although it is not without problems, stress has been considered a useful criterion to distinguish between compounds and phrases in many languages, as noted in section 2.2.2. In English for instance, *'blackboard* is considered a compound, while *black 'board* is a phrase (Booij 2012: 84). The argument is that left-hand stress is often a mark of compounds, whereas right-hand stress is a sign of phrases. However, there are some exceptions to this, which make stress an unreliable criterion in English. For example, *'apple cake* is stressed on the left-hand noun, but *apple 'pie* on the right-hand noun, despite the fact that both of them are compounds (Lieber 2005: 376). According to Bell and Plag (2012: 515-16), one relevant factor is a correlation between informativity and stress assignment in compounds; the more informative the second element is, the more likely it is to be assigned an accent, i.e. the more likely for the compound to be right-stressed. It is well known that, in Present-Day English, in some N + N combinations the main stress falls on the first element, e.g. *'table lamp*, while in others the stress falls on the second element, e.g. *silk 'shirt* (Bell and Plag 2013: 130). In fact, Bell and Plag (2013: 130) confirm the results of Plag *et al.* (2008) and Bell and Plag's (2012) studies. Specifically, they found that one third of the N + N compounds they studied are generally stressed on the second element, whereas two thirds are generally stressed on the first element. Whatever the nature of the factors at work, it means that there is no consistency in stress assignment in English compounds and hence it cannot be used to differentiate compounds from phrases (see section 2.2.2).

The stress criterion has never been applied to MSA and JA to check whether it can distinguish between compounds and phrases. Through applying this criterion, I will investigate stress assignment in N + N combinations in MSA and JA in the following section.

### 3.6 Stress assignment in N + N combinations in MSA and JA

With respect to stress assignment in N + N combinations in Semitic languages, several views have been proposed. For example, Siloni (1997: 21), writing about Hebrew, argues that in the construction known as the 'Nominal State Construct' (referred to as SGC in this study), the stress always falls on the second element, which is marked with the genitive case, whereas the first element remains unstressed. Lack of stress on the first element makes it eligible for the application of phonological rules which occur in words in unstressed environments, such as vowel reduction, compare *'bayit* in (30a) with *'beyt* in (30b) from Hebrew.

- (59) a) ha-bayit      šel      ha-'iš  
the-house      of      the-man  
'the man's house'
- b) beyt      ha-'iš  
house      the-man  
'the man's house'

Examples (30a, b) show that the stress is assigned to the second element, i.e. 'iš 'man'. Additionally, it demonstrates that the absence of the definite article *ha-* in the nominal state construct (30b) causes vowel reduction. Similarly, concerning stress assignment in MSA, Alexiadou *et al.* (2007: 248) claim that in examples like (31), the stress falls on the possessor *radzul* 'man', the genitive element.

- (60) a) bayt-u                      'radzul-i-n                      (P-construct)  
house-NOM                      man-GEN-a  
'a man's house'
- b) bayt-r                      'radzul-i                      (P-construct)  
house-NOM                      the man-GEN  
'the man's house'                      (Alexiadou *et al.* 2007: 248)

However, these two examples, which are the only ones discussed by Alexiadou *et al.* (*ibid.*), are not sufficient to conclude that stress in MSA always falls on the right element of N + N combinations. First, examples (31a, b) represent one specific construct, i.e. the P-construct. Compounds could have a different stress pattern compared to P-constructs. Second, these scholars have not addressed stress assignment in all phonological environments of N + N combinations in MSA, e.g. definite vs. indefinite and geminate vs. non-geminate (see section

3.6.3.3 for more detail). Third, there are combinations where stress would appear on the first element of P-constructs and compounds, as in the four examples below:<sup>14</sup>

- |      |   |                                   |               |
|------|---|-----------------------------------|---------------|
| (61) | <u>ḥaqībat-u</u><br>bag-NOM<br>'the girl's bag'                         | l-fataat-i<br>the-girl-GEN        | (P-construct) |
| (62) | <u>raaʔid-u</u><br>pioneer<br>'the astronaut'<br>lit. the space pioneer | l-fadʕaaʔ-i<br>the-space-GEN      | (compound)    |
| (63) | <u>ʕaruus-u</u><br>bride-NOM<br>'the mermaid'<br>lit. the sea bride     | l-baḥr-i<br>the-sea-GEN           | (compound)    |
| (64) | <u>ʔimtiḥaan-u</u><br>exam-NOM<br>'the chemistry exam'                  | l-kiimyaaʔ-i<br>the-chemistry-GEN | (compound)    |

Thus, it appears that some N + N constructs, like (31a, b), have stress on the second element, while others, like (32-35), do not. Clearly, these observations call for further investigation. In particular, experimental data are needed in order to determine the position of the stress in N + N combinations in MSA and JA. What follows in this section is driven by this argument.

Given the apparent significance of gemination (see the following section), I have included, in my test material, N + N combinations with and without gemination to examine the differences in stress assignment in both types. In the combinations that include a geminate, I have taken into consideration that different types of geminate exist, namely, gemination inside lexical items, e.g. *munassiq* 'coordinator' and assimilated gemination accompanied by sandhi, as in *f-fams* 'the sun' and in examples (20) and (21). The influence of gemination on stress

---

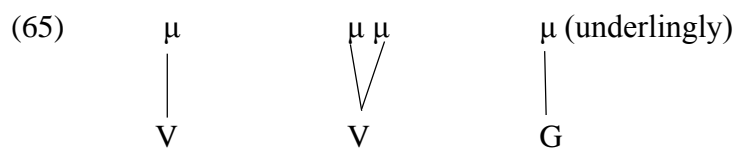
<sup>14</sup> To differentiate between word stress and prosodic stress, I used another convention to mark the latter, namely, underlying the word that bears it.

assignment is discussed in detail in the following subsection. The aim is to provide a full investigation of whether stress can help in distinguishing between P-constructs and compounds in MSA and JA by conducting an experiment, using Praat software.

### 3.6.1 Background on stress assignment in Arabic

Before outlining the experiment I have conducted, it is important to provide some background on stress assignment in Arabic. Arabic is a language with word stress, which means that one of the syllables in a content word is seen as prominent; thus, it receives primary stress (Watson 2002: 80). Four characteristics related to prominent or stressed syllables have been identified by Spencer (2002: 241). Firstly, a stressed syllable is in general louder than its neighbours. Secondly, a prominent syllable is longer in duration than its neighbouring unstressed syllables. Thirdly, the constituent sounds of a stressed syllable, particularly its onset consonants, are usually more clearly or more forcefully articulated than those in less stressed ones. Finally, a stressed syllable is the crux of pitch movement (accent), reflecting an intonation contour. This entails that a prominent syllable will usually be one uttered on a particularly high (or sometimes low) pitch (Spencer *ibid*).

Explaining stress assignment requires the use of the concept of mora, a phonological unit that determines syllable weight. Essentially, a short vowel is assigned one mora, long vowels two moras and geminate consonants are assigned one mora in the underlying representation (Hayes 1989). This is shown in (36):



A syllable is either heavy or light based on the number of moras it bears. A monomoraic syllable has one mora, a bimoraic syllable has two, and although rare, a trimoraic syllable has three (Hayes 1989). The type of syllable, whether heavy or light, influences the placement of stress (McCarthy 1979). In particular, Hayes (1989, 1995: 52) argues that in some languages stress falls on syllables that have a long vowel or diphthong (CVV) or those that have a geminate in the coda (CVG).<sup>15</sup> Syllables characterised by one of these three, i.e. long vowel,

---

<sup>15</sup> Although languages such as Leti, Malayalam and Ngalakgan, have syllables ending in a geminate, they are still considered light (see Hume *et al.* 1997; Baker 2008; Ringen and Vago 2011).

diphthong or a geminate, are often regarded as heavy or bimoraic. Conversely, syllables such as (CV) or (CVC) (if on the right edge of the word) are light or monomoraic (Hayes 1989, 1995; Watson 2002).<sup>16</sup> This suggests that CVC syllables that are not on the right edge of the word are considered heavy. Note that syllables in most Arabic dialects are maximally bimoraic and left-dominated. This means that they usually have trochaic feet (Watson 2002: 87). Trochaic feet are defined by Hayes (1995: 80) as ‘elements contrasting in intensity from groupings with initial prominence’. In other words, in trochaic feet, the first syllable of the foot is strong, whereas in iambic feet, the last syllable of the foot is strong. Hence, a trochaic foot is left-dominated, whereas an iambic foot is right-dominated. Iambs are asymmetrical binary feet with a weak followed by a strong element, whereas trochees consist of elements which differ in intensity that is, they are binary feet with a strong followed by a weak element (Hayes 1995).

Numerous studies have been conducted on stress assignment at the word level in Arabic, starting with Watson (2002: 81) who argues that a typical peninsular Arabic stress system in the present day follows the following rules:

- (66) Stress a final superheavy syllable (i.e. a syllable that has either one of these templates: CVCC or CVVC). An example of this is *da. 'rast* ‘I learned’. That is, when a word has only one heavy syllable, stress falls on it (cf. CA *da. 'rast* ‘I learned’).
- (67) Otherwise stress the rightmost non-final heavy syllable (up to the antepenultimate), e.g. *'mad.ra.sih* ‘school’.
- (68) Otherwise stress the leftmost light syllable, e.g. *'ka.tab* ‘he wrote’. Specifically, primary stress falls on the antepenultimate syllable when a word has only light syllables.

Stress assignment in JA follows similar rules to (37-39). In particular, Abu Abbas (2008: 15) describes stress in JA as follows:

---

<sup>16</sup> When a CVC syllable appears at the right edge of the word, it is regarded as extrametrical. Extrametricality was introduced by Hayes (1995), addressing the fact that syllables must include more segments to be heavy in word-final position in comparison with word-internal position. Thus, final consonants are extrametrical in some languages, including Arabic (Watson 2002: 90).



- (69) Stress the rightmost heavy syllable if it is not separated from the right edge of the word by more than two syllables. In other words, never stress pre-antepenultimate syllables. This rule is similar to Watson (2002) (see rule 38).
- (70) Stress the antepenultimate syllable in the absence of a heavy syllable. In other words, if the word has only light syllables, stress falls on the antepenultimate. Note that this rule is similar to Watson (2002) (see rule 39).
- (71) Never stress word-final CVC syllables. This rule is mentioned by several phonologists who have studied stress in Arabic (Hayes 1989, 1995; Watson 2002: 91): the peripheral (rightmost) foot is extrametrical and is thus invisible to the stress rules.

As a result, it is clear that syllable weight plays an important role in stress assignment in all dialects of Arabic, including JA (Abu Abbas 2008). Heavy syllables are more eligible to be assigned primary stress in Arabic compared to light syllables. With respect to geminates and stress assignment, Davis (2011: 845) points out that geminates, e.g. *dd*, are different from singleton consonants, e.g. *d*, in that word-final geminates, but not word-final singletons, attract stress to the word-final syllable. For instance, in Hadhrami Arabic spoken in Yemen, geminates attract stress onto the last syllable of the word as in [ʔa'xaff] 'lighter' vs. [ʔakbar] 'greater' (Bamakhramah 2009 cited in Davis 2011: 845). This kind of stress attraction onto the final syllable of the word is related to weight representation. Here, it is important to discuss whether geminates are capable of bearing weight or not. With regard to Lebanese Arabic (henceforth, LA), Khattab and Al-Tamimi (2014: 337) argue that, from the viewpoint of a weight-bearing account, syllables that end in a geminate are always heavy. This is because, like other varieties of Arabic, LA is a language which is characterised by being quantity-sensitive, i.e. syllable weight controls stress assignment. Additionally, Bamakhramah (2009 cited in Davis 2011: 845) argues that primary stress usually falls on the rightmost bimoraic syllable in Hadhrami dialect spoken in Yemen.

Taking the previous discussion into consideration, syllables ending in a geminate, which are considered bimoraic, will definitely be eligible to bear stress in Arabic. This is because geminates are underlyingly heavy or moraic as mentioned previously. The focus on stress assignment on syllables that have a geminate is particularly important in this experiment as is discussed later (see section 3.6.3.3).

Having considered the rules for stress assignment at the word level in Arabic, I move on to discuss stress assignment in N + N combinations in MSA and JA, for which I present data using Praat software. To the best of my knowledge, no prior study has been conducted on stress assignment in N + N combinations in MSA and JA. Therefore, this experiment will fill a gap, specifically in order to determine whether stress is a criterion that can distinguish between compounds and P-constructs (phrases) in MSA and JA. The following section provides an overview of the hypotheses and the procedures of the experiment.

### 3.6.2 Hypotheses

As far as N + N combinations are concerned, it is hypothesised that stress assignment depends on whether the second element is definite (i.e. marked with the definite article *ʔal*) or indefinite (i.e. marked with *tanwiin* ‘nunation’, the indefiniteness marker which is normally dropped when the speaker pauses at the end of the sentence). Note that nouns in MSA can be either marked with the definite article *ʔal-* or nunation *-n* (i.e. indefiniteness marker), but never both. In other words, these are in complementary distribution.<sup>17</sup> This can be illustrated with the following examples:

- |      |    |                              |  |               |
|------|----|------------------------------|--|---------------|
| (72) | a) | bayt-u<br>house-NOM          | r-raʔiis<br>the-president              | (P-construct) |
|      |    | ‘the president’s house’      |  |               |
|      | b) | *bayt-u<br>house-NOM         | r-raʔiis-i-n<br>the-president-GEN-INDF | (P-construct) |
|      |    | (*‘the/a president’s house’) |  |               |
| (73) | a) | ʕarus-u<br>bride-NOM         | l-baħr<br>the-sea                      | (compound)    |
|      |    | ‘the mermaid’                |  |               |
|      |    | lit. the sea bride           |  |               |

---

<sup>17</sup> Note that *ʔal-bayt* is pronounced *l-bayt* in consecutive speech.

b)	*ʕarus-u	l-baħr-i-n	(compound)
	bride-NOM	the-sea-GEN-INDF	
	(*‘the/a mermaid’)		
	lit. the/a sea bride		

On the basis of my intuition of stress patterns in several compounds and P-constructs, I formulated the following hypotheses about stress assignment in N + N combinations in both MSA and JA:

- 1) If the second/right element is definite, stress falls on the first element or on both elements.
- 2) If the second/right element is indefinite, stress falls on the first/left element.

The following are some N + N combinations which illustrate these hypotheses:

(74)	<u>qalam-u-t</u> <sup>18</sup>	<u>tʕaalib</u>	(P-construct/definite)
	pen-NOM-the	student	
	‘the student’s pen’		
(75)	<u>barnaamadʒ-u-t</u>	<u>tilfaaz</u>	(compound/definite)
	programme-NOM-the	television	
	‘the television programme’		
(76)	<u>qalam-u-l</u>	walad	(P-construct/ definite)
	pen-NOM-the	boy	
	‘the boy’s pen’		
(77)	<u>muʕallim-u-l</u>	fiizyaaʔ	(compound/ definite)
	teacher-NOM-the	physics	
	‘the physics teacher’		

---

<sup>18</sup> See the discussion of sandhi in section 3.4 for details on the assimilated geminate.

(78)	<u>bayt-u</u> house-NOM 'a man's house'	radzul-in man- GEN	(P-construct/indefinite)
(79)	<u>findzaan-u</u> cup-NOM 'a tea cup'	faay-in tea-GEN	(compound/indefinite)

The next section discusses the experiment I conducted in order to confirm or falsify these hypotheses.

### 3.6.3 The experiment

In order to determine whether stress can be used to distinguish compounds from P-constructs in MSA and JA, I tested several adults whose first language is JA. The participants were asked to read a number of compounds and phrases embedded in two paragraphs of running text (one written in MSA and another in JA). I recorded their voices and analysed them using Praat software. The procedure is fully explained in the following subsections.

#### 3.6.3.1 Pilot study

Before conducting the experiment, I tested the materials on two native speakers of Arabic to validate the text and methods used. The two participants were asked to read a short paragraph that contained five P-constructs and six compounds and I recorded their speech. The paragraph comprised four sentences. The results showed that one of the sentences was problematic. In particular, the stress on the two compounds in that sentence was not clear due to the presence of contrastive stress. This type of stress is defined as a stress which is assigned to a word or a syllable as opposed to its normal accentuation (Bolinger 1961: 83). This is done to contrast it with another word or syllable or to steer the attention towards it. For instance, in the phrase *parliament of the people, by the people, for the people*, the stress falls on the normally unstressed word *of* in order to focus on the contrast between *of*, *by*, and *for*. Note that, in the above example, two or more items are counterbalanced and a preference indicated for some members of the group (Bolinger *ibid*). Bauer *et al.* (2013: 445) note that contrastive stress may change the normal stress pattern assigned to a compound, e.g. *She meant Park 'Street, not Park*

'Road. Bauer *et al.* (ibid) suggest that this source of variation is to be ignored; the normal stress position of compounds is detected in non-contrastive environments.

The English translation of the Arabic sentence which caused problem in data analysis ran as follows:

At our school, the maths teacher, the physics teacher and the arts teacher drink a cup of tea every morning.

Here, the three compounds, *muṣallim rriyaad<sup>ṣ</sup>iyyaat* 'the maths teacher', *muṣallim lfiizyaa?* 'the physics teacher' and *muṣallim lfann* 'the arts teacher' are affected by contrastive stress. To avoid the effects of this type of stress, *muṣallim lfann* 'the arts teacher' was replaced by *mudaqqiq lkurraasah* 'the notes inspector' while *muṣallim lfiizyaa?* 'the physics teacher' was moved to the second paragraph, so that *muṣallim rriyaad<sup>ṣ</sup>iyyaat* 'the maths teacher' and *muṣallim lfiizyaa?* 'the physics teacher' are separated by six sentences. Additionally, the pilot study showed that the text contained two words which are not used in JA, namely, *ʔimtiḥaan θḥaanawiyya* 'the secondary school examination' and *s<sup>ṣ</sup>abaah masaa?* 'morning and evening'. These two words were replaced by their JA equivalents, i.e. *ʔimtiḥaan ttawdziihi* 'the secondary school examination' and *s<sup>ṣ</sup>ubḥ wu masaa* 'morning and evening'.

### **3.6.3.2 Sample**

Five adults (two female and three males) participated in the experiment, all native speakers of JA. The participants have a working knowledge of MSA, since they have studied it in detail for twelve years at school. They have also taken 2-3 modules of advanced MSA in their undergraduate degree in Jordan. Their mean age is 29 years old. Even though the participants are not native speakers of MSA (since MSA has no native speakers), the data collected from MSA will be of importance in this experiment, since it may provide a clear picture of stress assignment in N + N combinations in Arabic in general.

### **3.6.3.3 Tools and procedure**

In order to test whether stress can distinguish compounds from P-constructs in Arabic, the participants were asked to read two paragraphs (one written in MSA and another in JA). These two paragraphs contain a number of compounds (sixteen) and P-constructs (eight) in order to

examine whether the position of the stress differs in the two types of construct. With regard to the compounds, I included some examples such as *layla nahaar* ‘twenty four-seven’, *s‘abaaħ masaaʔ* ‘all day long’ and *ħulwun murrun* ‘bitter-sweet’ which are not representative of SGCs, or as referred to in Arabic, *Id‘aafah*. These examples are treated as coordinative compounds (see chapter 6). In addition, I included two instances of Adj + N combinations; one of them is a compound, i.e. *baʕiid nnað‘ar* ‘far sighted’, whereas the other example is a phrase, i.e. *dzadiid lkutub* ‘the new books’ (see chapter 5). These two instances are excluded from Table 3.1 below, which is based on data for N + N combinations only.

The compounds and P-constructs in the two paragraphs were chosen according to the hypotheses formulated in section 3.6.2. Note that although definiteness vs. indefiniteness is possibly important for stress assignment in N + N combinations in MSA and JA, as mentioned in section 3.6.2, the type of definiteness should be taken into consideration. In particular, if the second element is definite, then it is vital to examine which type of definite article is used in the P-constructs or compounds in question. It has been proposed that the definite article has an underlying phonological form /ʔal/ (Heselwood and Watson 2013: 34). This form surfaces as [ʔal] when the subsequent word starts with a non-coronal consonant. However, when the subsequent word starts with a coronal consonant, the /l/ completely assimilates to the following coronal, yielding a geminate coronal consonant (Heselwood and Watson *ibid*). For instance, if the definite article is added to *bint* ‘girl’, it surfaces as [ʔalbint] ‘the girl’, but if it is added to *t‘aalib* ‘student’, it surfaces as [ʔat‘aalib]. These aspects could play a role in stress assignment in N + N combination in MSA and JA. Hence, gemination has been taken into account in selecting the data investigated. Specifically, I have included in my test material N + N combinations with and without gemination in order to investigate the differences in stress assignment in both types. Furthermore, I have taken into account that there are two types of geminate, i.e. lexical geminate and assimilated geminate accompanied by sandhi. Table 3.1 shows the compounds and P-constructs selected for this experiment based on the above considerations. The full text of the two paragraphs is given in the Appendix.

Table 3.1. N + N compounds and P-constructs selected in the current experiment in MSA and JA

Definite 2 <sup>nd</sup> element		
The phonological environment of N+N combinations	Compound	P-construct
N1 non-geminate + N2 non-geminate	<ul style="list-style-type: none"> <li>• <i>ʔimtihaan lkiimyaah</i> ‘the chemistry exam’</li> <li>• <i>raʔiis lwuzaraah</i> ‘the prime minister’</li> <li>• <i>mudiiir lqaaʕah</i> ‘the head invigilator’</li> </ul>	• <i>maqsʕaf lmadrasah</i> ‘the school’s canteen’
N1 geminate + N2 non-geminate	• <i>muʕallim lfiizyaaʔ</i> ‘the physics teacher’	• <i>sayyarat lmudiiir</i> ‘the director’s car’
N1 non-geminate + N2 geminate	<ul style="list-style-type: none"> <li>• <i>ʕaamil nnaḏʕaafah</i> ‘the cleaner’</li> <li>• <i>ʔimtihaan ḏḏaanawiyya</i><sup>19</sup> ‘the secondary school examination’</li> <li>• <i>waziir ttarbiyah</i> ‘the Minister of Education’</li> </ul>	• <i>kitaab tʕaalib</i> ‘the student’s book’
N1 geminate + N2 geminate	<ul style="list-style-type: none"> <li>• <i>muʕallim rriyaadʕiyyaat</i> ‘the maths teacher’</li> <li>• <i>mudaqqiq lkurraasah</i> ‘the notes inspector’</li> </ul>	• <i>xutʕʕat lmuʕallim</i> ‘the teacher’s plan’
Indefinite 2 <sup>nd</sup> element		
The phonological environment of N+N combinations	Compound	P-construct
N1 non-geminate + N2 non-geminate	<ul style="list-style-type: none"> <li>• <i>layla nahaar</i> ‘twenty four-seven’,</li> <li>• <i>sʕabaaḥ masaaʔ</i> ‘twenty four-seven’<sup>20</sup></li> <li>• <i>kaʔs ʕasʕiir</i> ‘juice glass’</li> </ul>	• <i>kitaab fataah</i> ‘a girl’s book’
N1 geminate + N2 geminate	• <i>muʕallim sʕaff</i> ‘primary school teacher’	• <i>sidzill musʕahḥiḥ</i> ‘a marker’s record’
N1 non-geminate + N2 geminate	<ul style="list-style-type: none"> <li>• <i>ḥulwun murrun</i> ‘bitter-sweet’</li> <li>• <i>fatʕiirat tuffaaḥ</i> ‘apple pie’<sup>21</sup></li> </ul>	• <i>qalam muʕallim</i> ‘a teacher’s pen’
N1 geminate + N2 non-geminate	• <i>munassiq ʕuluuum</i> ‘a science coordinator’	• <i>qubbaʕat tʕaalib</i> ‘student’s cap’

<sup>19</sup> Note that the compound *ʔimtihaan ḏḏaanawiyya* (*ʔimtihaan* ‘exam’ + *ḏḏaanawiyya* ‘secondary school’) ‘secondary school examination’ in the MSA version is excluded from JA. The former is no longer used; it is replaced by *ʔimtihaan ttawdziihi* (*ʔimtihaan* ‘exam’ + *ttawdziihi* ‘secondary school’) ‘secondary school examination’.

<sup>20</sup> Note that the compound *sʕabaaḥ masaaʔ* (*sʕabaaḥ* ‘morning’ + *masaaʔ* ‘evening’) ‘twenty four-seven’ in the MSA version is excluded from JA. The compound is phonologically reduced and a conjunction is inserted between the two elements, i.e. *sʕubḥ wu masaa* (*sʕubḥ* ‘morning’ + *wu* ‘and’ + *masaa* ‘evening’) ‘twenty-four seven’.

<sup>21</sup> Note that the compound *fatʕiirat tuffaaḥ* ‘apple pie’ in the MSA version is changed in the JA version into *keekit tuffaaḥ* ‘apple pie’. This is due to the fact that *fatʕiirat* ‘pie’ is no longer used in JA.

Table 3.1 incorporates the two factors which may play a role in stress assignment in MSA and JA, i.e. (in)definiteness and gemination, as indicated in my hypotheses. Note that N1 refers to the first element of N + N combinations, whereas N2 refers to the second element. In MSA, nunation is the indefiniteness marker, whereas in JA, the absence of the definiteness marker is a sign of indefiniteness, since nunation is not realised in JA.

The selected compounds and P-constructs were embedded in a sequence of natural sentences forming a realistic piece of discourse as a way to reduce the effect of the observer's paradox. Labov (1972: 209) notes that the behaviour of a person being interviewed/tested is affected by the presence of the observer or interviewer. Thus, researchers may face problems in capturing naturally occurring data, since, in their presence, the respondents will be self-conscious and careful about what they say. However, the only way researchers can obtain natural data is via systematic observation; hence, the paradox. By putting the target items in a paragraph of natural text, the participants will not know exactly what the main focus of the experiment is. Campbell (2002) explains that if a researcher is interested in detecting certain phonetic features by asking respondents to produce certain expressions, the best method is to include these expressions in a longer piece of discourse, rather than asking the respondents to read them individually. Campbell (2002) also indicates that longer texts have the advantage of producing more natural speech as the respondents would not be aware of what the researcher's aims are, thus reducing the effect of the observer's paradox.

I recorded the participants' voices in a soundproof room to obtain high quality audio. For recording the paragraph in JA, the participants were given the text in a form where the spelling and diacritics were adapted to JA pronunciation. This was done to avoid the association of written stimuli with MSA. I made sure that the participants were comfortable and at ease. The participants were also assured that they could stop the recording at any time (see the Appendix).

#### **3.6.3.4 Data analysis**

The audio-recordings were analysed using the latest version of Praat software (5.4.08), which was designed by Boersma and Weenink (2015). The Praat speech analysis software was chosen because it is user-friendly, flexible, downloadable and free. It also provides a clear analysis of stress assignment, showing pitch, duration and intensity. Rutter (2008: 132) points out that the pitch contour, marked with a dashed blue line in the Praat diagrams, in addition to loudness (intensity, marked with yellow line) and duration are responsible for assigning stress. Spencer



(2002) notes that a prominent syllable will usually be one uttered on a particularly high (or sometimes low) pitch, which is one of the most important parameters responsible for stress assignment. Similarly, based on works like Hammond (1999) and Odden (2005), Plag, Kunter and Schramm (2011: 362) indicate that in English, the usual correlates of stress are pitch, intensity and duration, with stressed syllables having the tendency to be higher in pitch and longer in duration. The same applies to Arabic (see De Jong and Zawaydeh 1999).

With respect to the acoustic correlates of stress, note that the duration correlate are excluded in this study due to the presence of different segments in the examples of P-constructs and compounds, i.e. the target words in both types of construct are not the same. Duration can only be used as a criterion if both examples have the same segments, e.g. the English phrase *black 'bird* and the compound *'blackbird*. Therefore, the recordings in this experiment were fed into the software to examine both pitch (marked with a blue line in the spectrogram figures) and intensity (marked with a yellow line in the spectrogram figures) in order to determine how they can be used to pinpoint stress, since they are the most reliable acoustic correlates of stress in this case, as discussed by Rutter (2008) and Spencer (2002). Finally, I determined if there is a difference between the two types of N + N combination on the basis of visual inspection of the spectrogram figures together with auditory impression of the recordings.

### **3.6.4 Results and discussion**

This section presents a sample of the spectrogram figures used in order to test whether the hypotheses discussed in section 3.6.2 are supported or falsified. I start with the first hypothesis, which suggests that if the second/right element is definite, the stress falls on the first element or perhaps both. It will become clear in this section that this hypothesis is partially confirmed. The first element is assigned stress in all cases. If the second element begins with a non-coronal sound (i.e. there is no assimilated geminate), stress clearly falls on the first/left element only. If the second element starts with a coronal sound, causing assimilated gemination accompanied by sandhi, the N+ N combination is still assigned stress on the first element, but with a subtle fall of the intensity contour on the second element. With respect to the second hypothesis, if the second/right element is indefinite, stress always falls on the first/left element. These two hypotheses are investigated in detail in the following section. Note that even though subtle differences between MSA and JA were detected at the word level, the spectrogram figures show no differences between MSA and JA in terms of stress assignment in N + N combinations. In other words, although the pitch and intensity values of some words are slightly higher in

MSA than those of JA, both of these values were higher on the first element, rather than on the second. Hence, the analysis below applies to both varieties of Arabic. In addition, based on two acoustic contours, i.e. pitch and intensity, the data analysis shows that stress assignment in the target N + N combinations is the same for all participants in virtually all cases. However, I have selected the clearest tokens out of 210 tokens (105 for JA and 105 for MSA) to represent the stress patterns.

#### **3.6.4.1 N + N combinations marked with definiteness**

Since gemination plays a significant role in stress assignment, this section is divided into four sub-sections on the basis of the presence vs. absence of gemination (see Table 3.1).

##### **➤ 1<sup>st</sup> non-geminate + 2<sup>nd</sup> non-geminate**

For this case, three compounds, e.g. *raʔiis lwuzaraaʔ* ‘the prime minister’, *ʔimtiḥaan lkiimyaaʔ* ‘the chemistry exam’ and *mudiir lqaaʕah* ‘the head invigilator’, and one P-construct, e.g. *maqsʕaf lmadrasah* ‘the school’s canteen’ were tested. The following spectrogram figures represent two of these combinations.

Figure 1. Spectrogram for *raʔiis lwuzaraaʔ* [ra.ʔii.sil.wu.za.raaʔ]<sup>22</sup> ‘the prime minister’ (compound)<sup>23</sup>, MSA

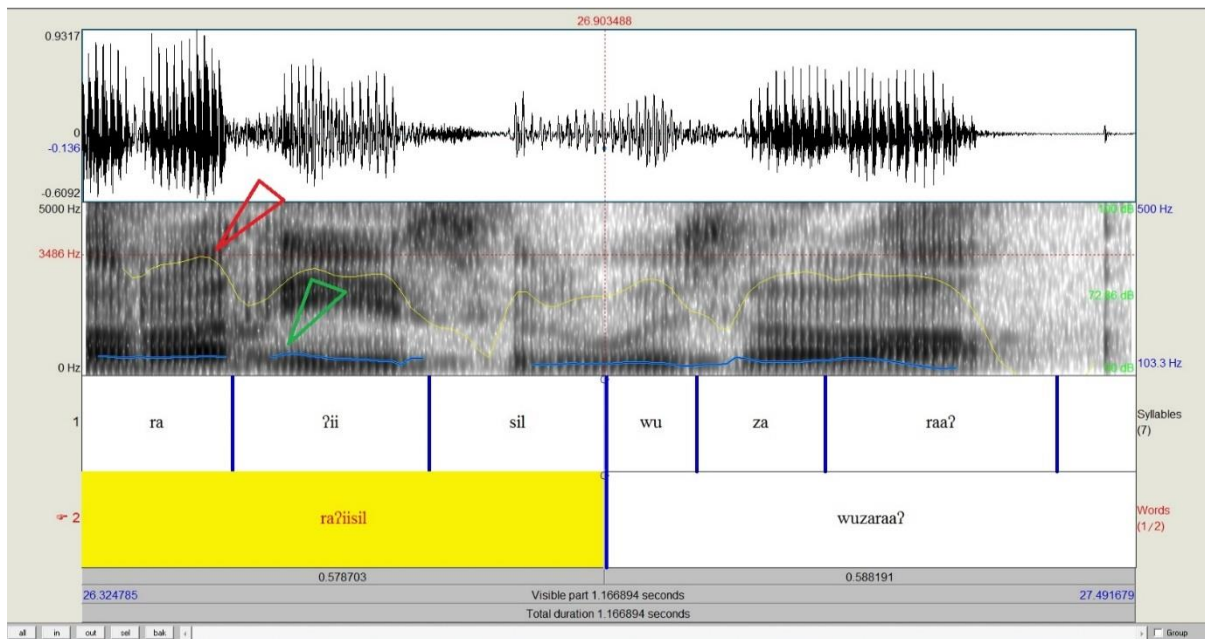
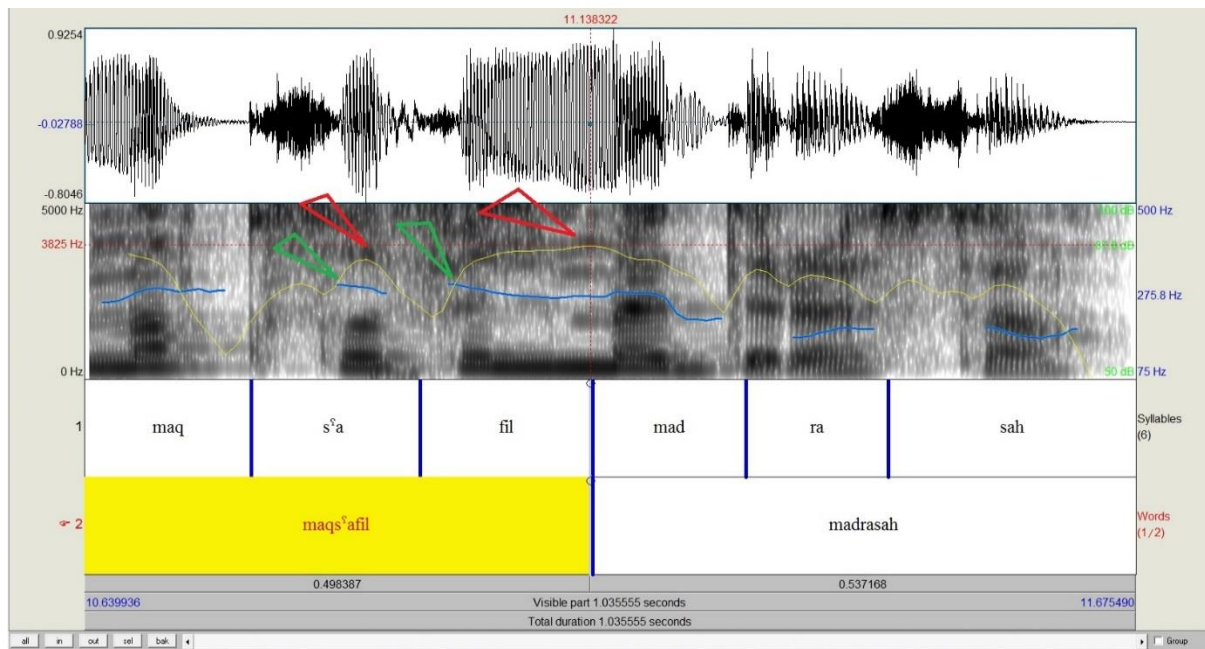


Figure 2. Spectrogram for *maqṣʻaf lmadrasah* ‘the school’s canteen’ [maq.sʻa.fil.mad.ra.sah] (P-construct), MSA



<sup>22</sup> The syllabification given in all spectrograms is performed based on the actual pronunciation of the constructs in the text. The case marked on the N + N combination in MSA is based on its function in the context. In particular, the first element can be nominative *-u*, accusative *-a* or genitive *-i*, whereas the second element is always genitive *-i*. Remember, however, that the genitive case on the second element is not realised in JA.

<sup>23</sup> The triangles shown in the spectrograms are used to indicate the high values of pitch and intensity, i.e. the red triangles pinpoint intensity, whereas the green ones indicate the pitch.

Looking at the pitch and intensity correlates, the spectrograms in Figures 1-2 show that in the absence of gemination, the stress falls on the first element. This is clear from the fact that in Figures 1 and 2, the values of both the blue (i.e. pitch) and yellow (i.e. intensity) lines are slightly higher on the first element compared to the second element. According to Spencer (2002: 241), a stressed syllable can be more stressed than its neighbours. As a result, we do not need to consider all syllables as either stressed or unstressed. More often than not, in addition to the main stress we will find other stressed syllables in a word or phrase, but these are not stressed to the same degree as the main stressed syllable (Spencer *ibid*). This is known as secondary stress (as opposed to that assigned on the most stressed element, primary or main stress). Data in Figures 1 and 2 may indicate that the second elements *lwuzaraa?* ‘the minister’ and *lmadrrasah* ‘the school’ are not completely unstressed. In effect, they could bear secondary stress.

This confirms my hypothesis that in definite N + N combinations, the default position of the stress is on the first element or both, but not on the second. Note that the position of the stress so far does not provide a clear distinction between compounds and P-constructs in MSA and JA. In both cases, the stress falls on the first element.

➤ **1<sup>st</sup> geminate + 2<sup>nd</sup> non-geminate**

For the cases where the first word has a geminate, one compound, e.g. *muʕallim lfiizyaa?* ‘the physics teacher’ and one P-construct, e.g. *sayyarat lmudair* ‘the director’s car’ were included in the experiment. Despite the slight difference in intensity in Figure 3, the pitch is clearly higher on the first element. In general, the spectrogram Figures 3-4 demonstrate that these two combinations seem to be assigned stress on the first element as shown below.

Figure 3. Spectrogram for *muṣallim lfiizyaaʔ* [mu.ṣal.li.mil.fiiz.yaa.ʔi] ‘the physics teacher’ (compound), MSA

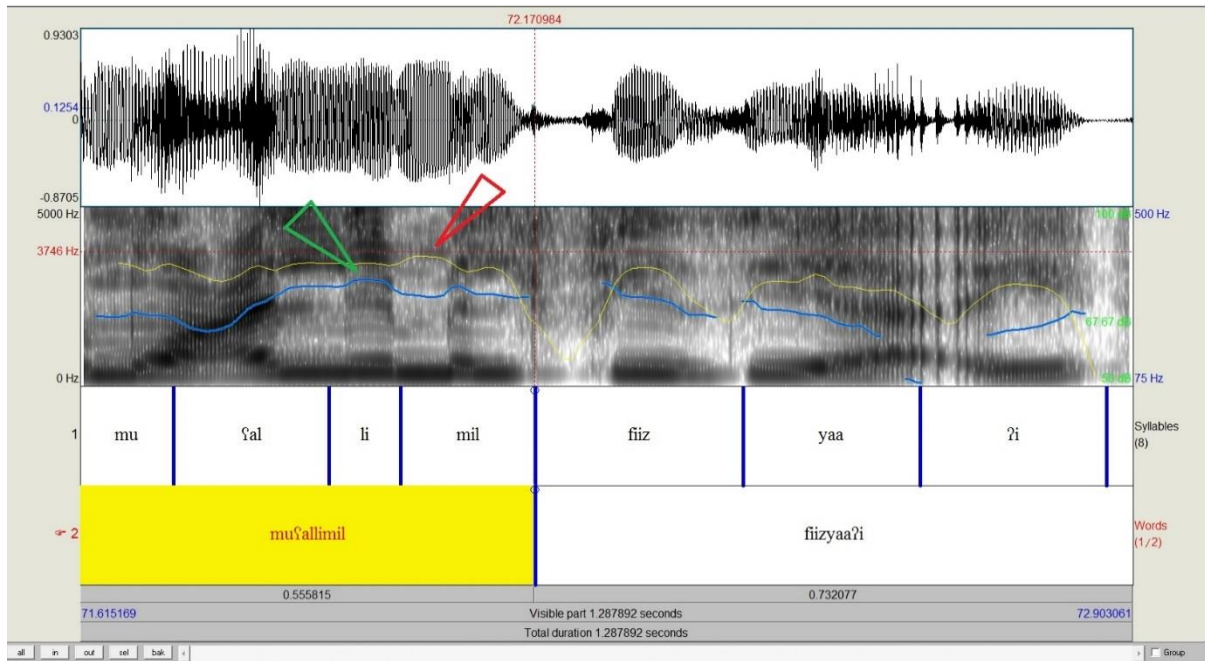
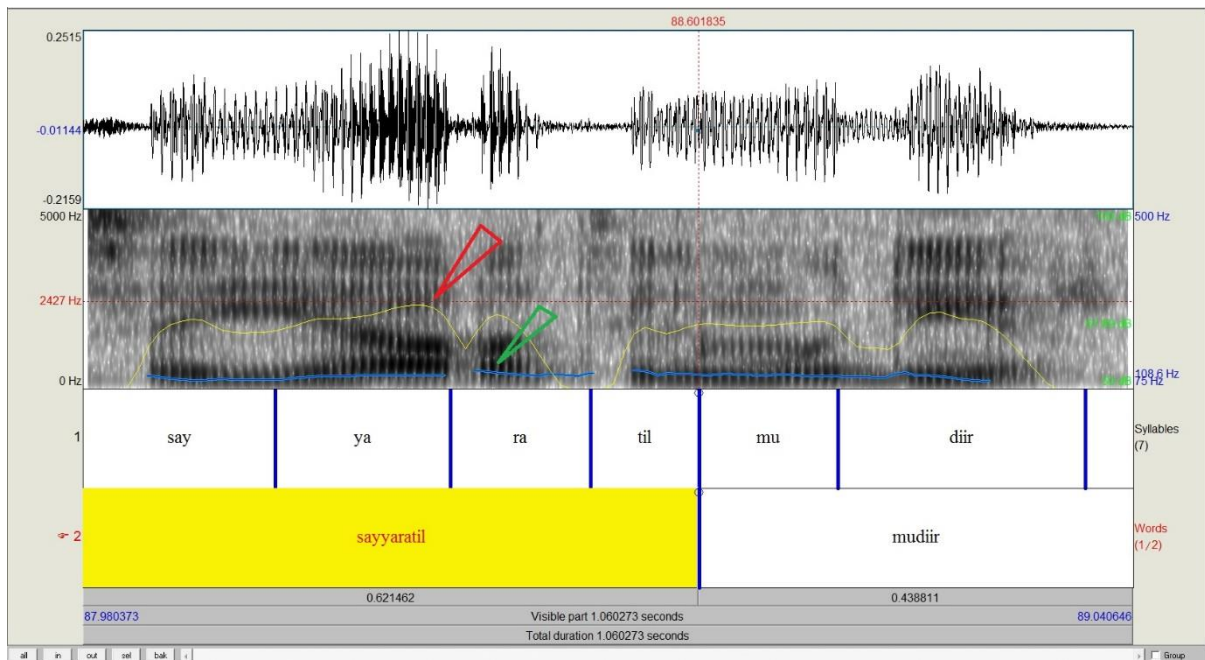


Figure 4. Spectrogram for *sayyarat lmudir* [say.ya.ra.til.mu.diiir] ‘the director’s car’ (P-construct)



Note that in both the compound and the P-construct in Figures 3-4, the first element has a geminate: *mu.ṣal.lim* ‘teacher’ and *say.ya.rat* ‘car’. However, the presence of the geminate here

does not affect stress assignment. Stress appears to fall on the first element also in environments in which no gemination occurs on the first element, as mentioned in the previous section. Even though sandhi also operates in both the compound *mu.ʕal.li.mil.fiz.yaa.ʔi* ‘the physics teacher’ and the P-construct *say.ya.ra.til.mu.diir* ‘the director’s car’, it does not have any impact on stress across word boundary. For this to happen, there has to be an assimilated geminate accompanied by sandhi as shown in detail in Figures 5-7.

➤ **1<sup>st</sup> non-geminate + 2<sup>nd</sup> geminate**

For the case in which the first element does not contain a geminate and the second does, I included three compounds, namely, *ʕaamil nnaðʕaafah* ‘the cleaner’, *ʔimtihaan θhaanawiyya* ‘the secondary school examination’, and *waziir ttarbiyah* ‘the minister of education’, and one P-construct, e.g. *kitaab tʕʕaalib* ‘the student’s book’ in the experiment. Figures 5-7 below show one compound, i.e. *ʕaamil nnaðʕaafah* ‘the cleaner’, in both MSA and JA, and one P-construct, i.e. *kitaab tʕʕaalib* ‘the student’s book’.

Figure 5. Spectrogram for *ʕaamil nnaðʕaafah* [ʕaa.mi.lin.na.ðʕaa.fa.ti] ‘the cleaner’ (compound), MSA

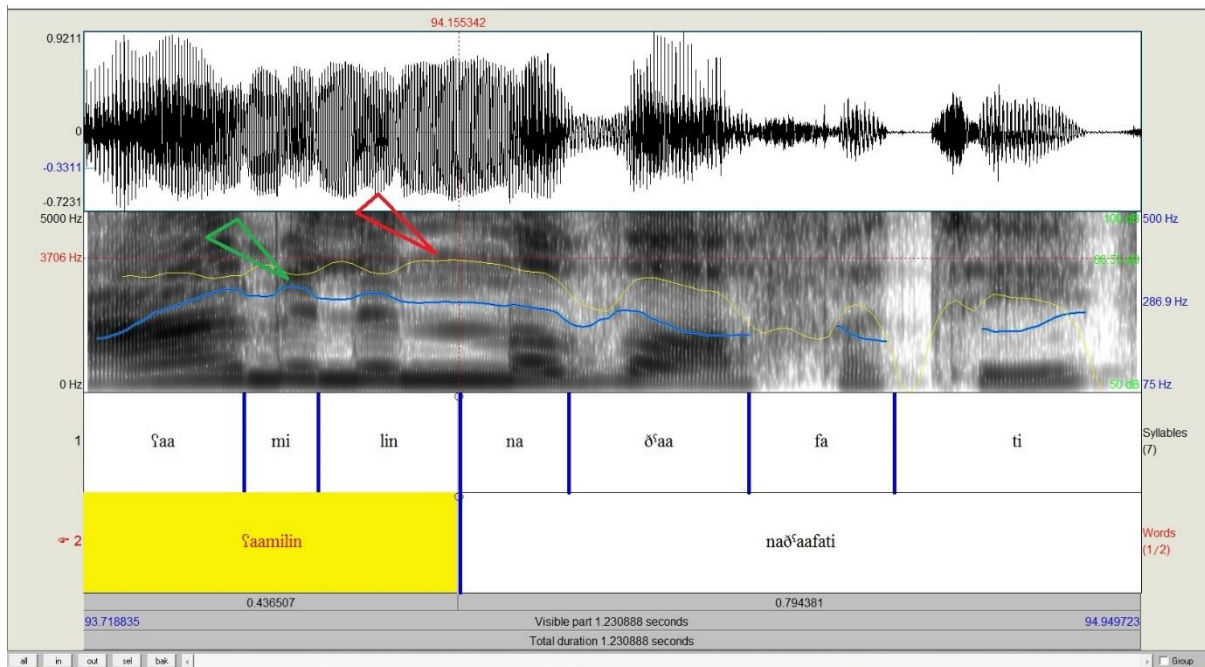


Figure 6. Spectrogram for *ʕaamil naðʕaafah* [ʕaa.mi.lin.na.ðʕaa.fah] ‘the cleaner’ (compound), JA

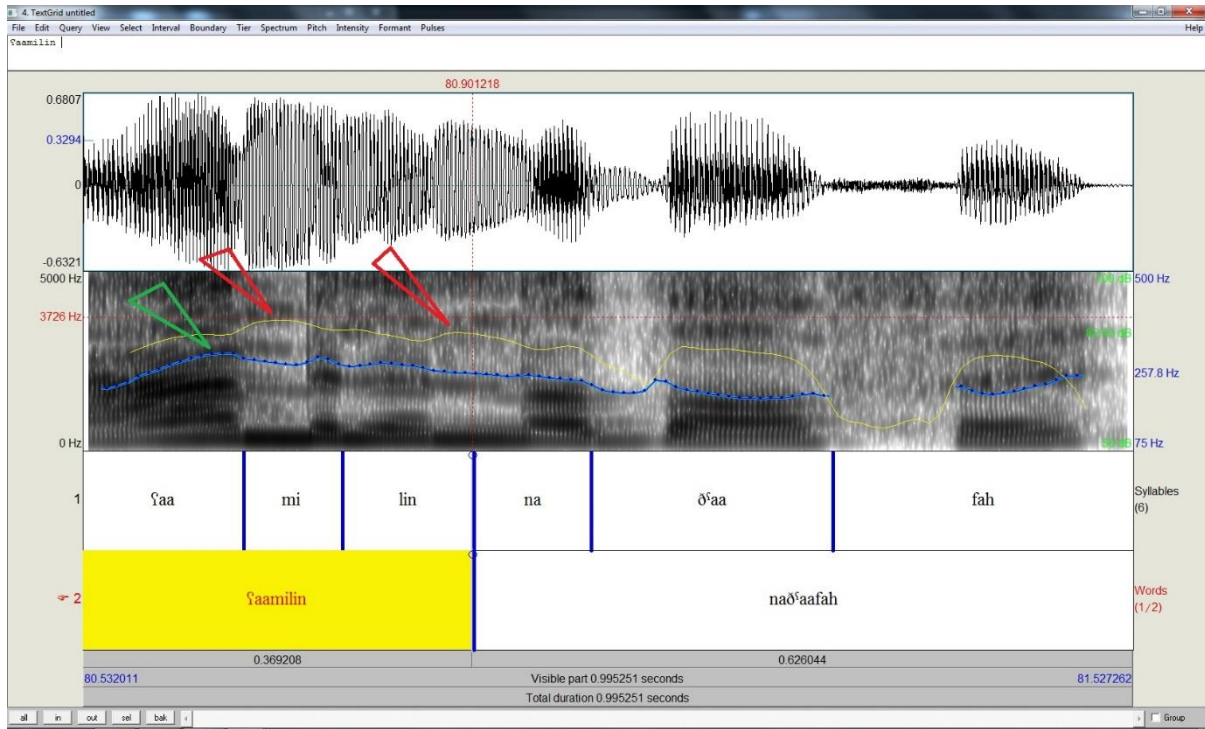
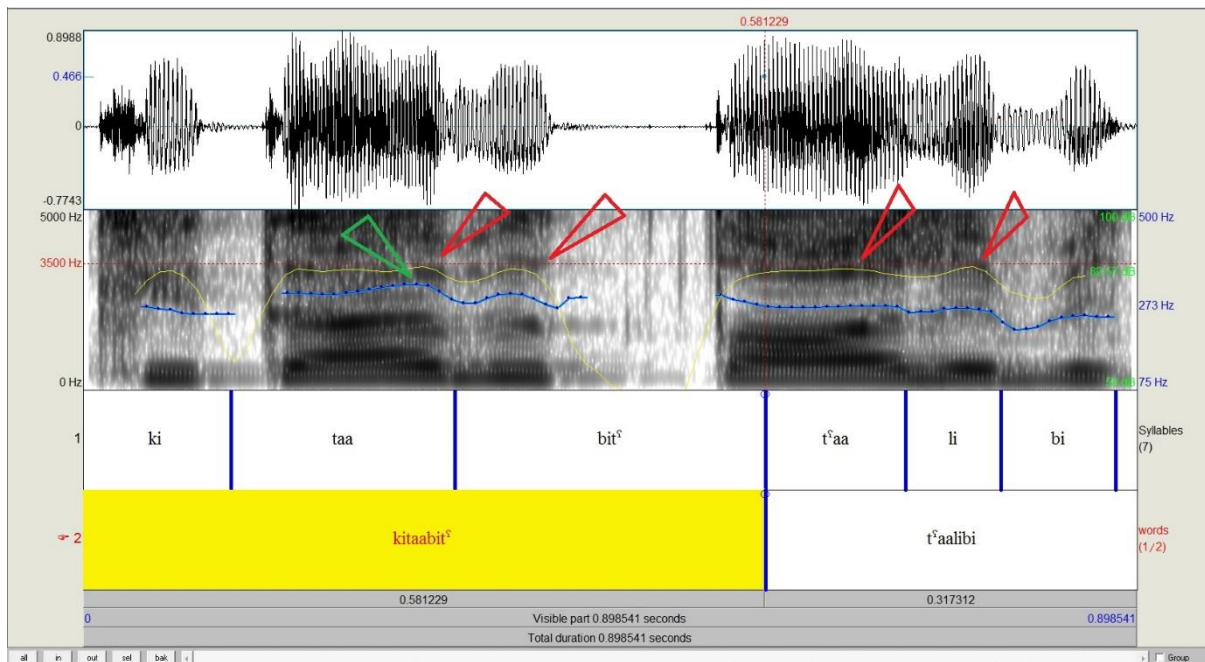
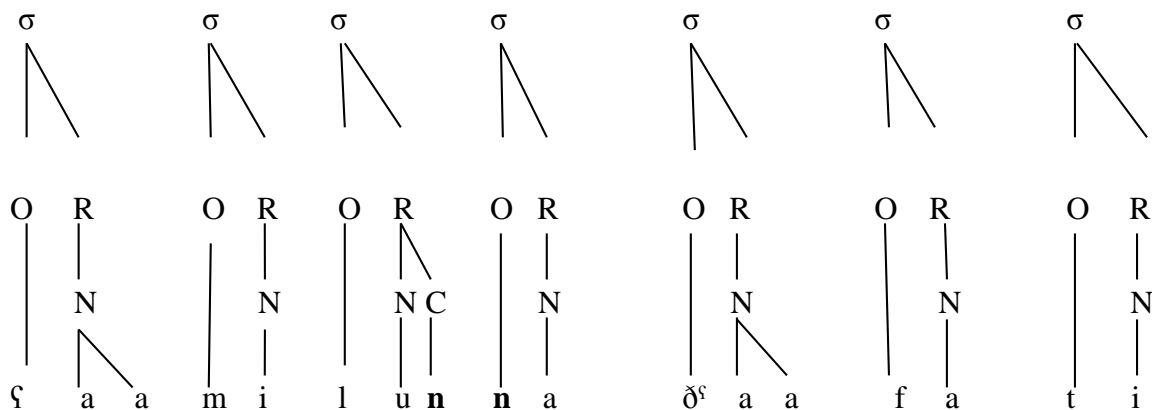


Figure 7. Spectrogram for *kitaab tʕaalib* [ki.taa.bitʕ.tʕaa.li.bi] ‘the student’s book’ (P-construct), MSA



Figures 5-7 show that in both *ʕaamil nnaðʕaafah* ‘the cleaner’ and *kitaab tʕaalib* ‘the student’s book’, stress seems to fall on the first element, characterised by a peak, i.e. one prominent syllable. It is clear that, as shown by the red and green triangles in the figures above, the first elements *ʕaamil* ‘worker’ and *kitaab* ‘book’ have higher pitch (the blue line) and intensity (the yellow line) than the second elements. Note, however, that there is a tiny decrease in the intensity contour across word boundaries, making it appear as though the second element is also stressed. The common factor between the elements of both N + N combinations is that the second element contains a geminate and that the two elements are connected together by sandhi as mentioned in section 3.4. Specifically, the first consonant of the geminate in *ʕaa.mi.lun.na.ðʕaa.fa.ti* ‘the cleaner’ forms the coda of the last syllable of the first element, i.e. *lun*, whereas the second consonant of the geminate syllabifies as the onset of the first syllable of the second element, i.e. *na*. The syllabification process of *ʕaa.mi.lun.na.ðʕaa.fa.ti* ‘the cleaner’ is demonstrated in Figure 8 below, in which O stands for onset, R for rhyme, N for nucleus and C for coda:

Figure 8. Syllabification of *ʕaa.mi.lun.na.ðʕaa.fa.ti* ‘the cleaner’



When they are connected by sandhi, i.e. *ʕaa.mi.lun.na.ðʕaa.fa.ti* ‘the cleaner’ and *ki.taa.butʕ.tʕaa.li.bi* ‘the student’s book’, the intensity seems to be affected by the presence of the assimilated geminate, causing the fall at the onset of the first syllable of the second element to be only very slight, as shown in Figures 5-7. Simply put, the geminate does not only affect the coda of *lun* and *butʕ*, but also the onset of the second syllables *na* and *tʕaa*, respectively. The ability of the geminate to affect both the onset and the coda is not surprising because it is the same segment which appears in the onset and in the coda; the segment has the same acoustic



properties. Consequently, I would suggest that in N + N combinations with assimilated geminates on the word boundary, a secondary stress or perhaps double stress is assigned.

The presence of a geminate in their examples may explain why Alexiadou *et al.* (2007: 248) claim that the stress falls on the second element in N + N combinations such as *baytr 'radzul* 'the man's house'. The assimilation of *l-* to *r* creates a geminate in the coda. As explained previously, the geminate is bimoraic; thus, it attracts stress. Therefore, according to Alexiadou *et al.* (2007: 248), the second element is assigned the primary stress. However, the experiment I conducted provides somewhat a more detailed picture of stress assignment in N + N combinations. The gemination accompanied by sandhi found in such examples affects stress, but it does not shift it completely from the first element. This argument is supported by the fact that if the second definite element starts with a non-coronal consonant so that there is no assimilation, and in turn no gemination, the stress clearly falls on the first element only (see Figures 3-4). The spectrogram figures also show that stress is assigned to the first element in both compounds and P-constructs, which means that it does not help in differentiating between them.

➤ **1<sup>st</sup> geminate + 2<sup>nd</sup> geminate**

For this case, one P-construct and two compounds are included in the paragraph which the five participants read, namely, *muṣallim rriyaad'iyyaat* 'the maths teacher' and *muṣallim ttaariix* 'the history teacher' (compounds), and *xut't'at lmuṣallim* 'the teacher's plan' (P-construct). Figures 9-10 show the spectrograms for the two items of N + N combinations.

Figure 9. Spectrogram for *muṣallim rriyaad<sup>ṣ</sup>iyyaat* [mu.ṣal.li.mur.ri.yaa. d<sup>ṣ</sup>iy.yaa.ti] ‘the maths teacher’ (compound), MSA

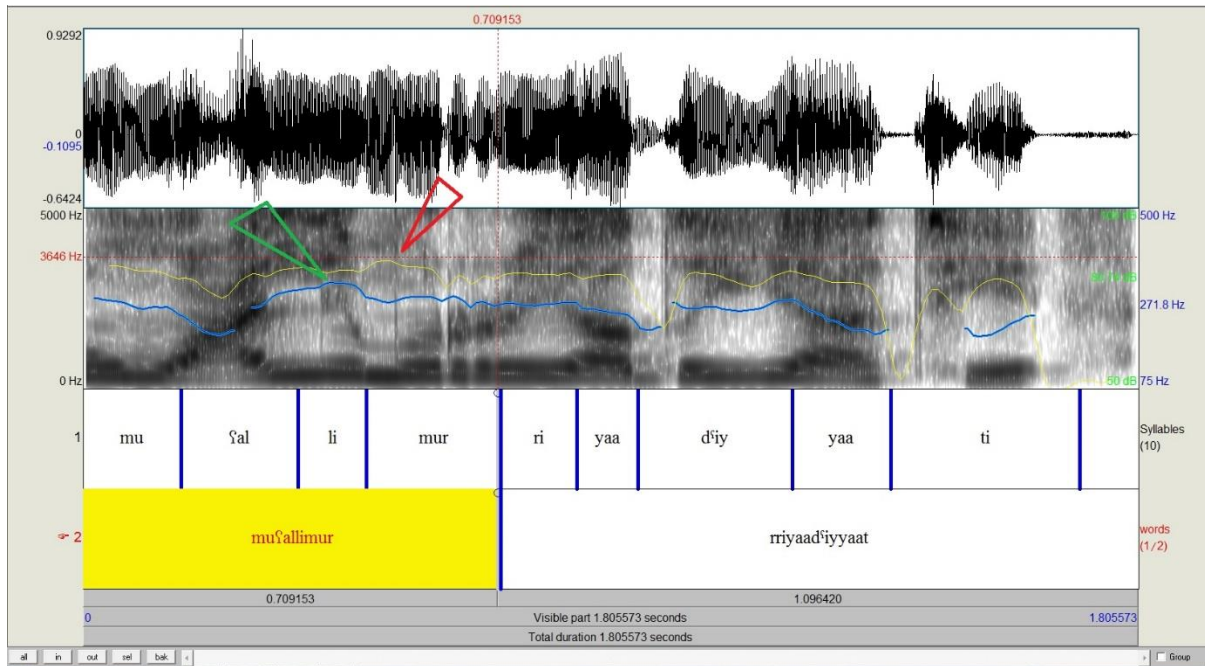


Figure 10. Spectrogram for *xuṭ<sup>ṣ</sup>t<sup>ṣ</sup>at lmuṣallim* [xuṭ<sup>ṣ</sup>.t<sup>ṣ</sup>a.til.mu.ṣal.li.mi] ‘the teacher’s plan’ (P-construct), MSA

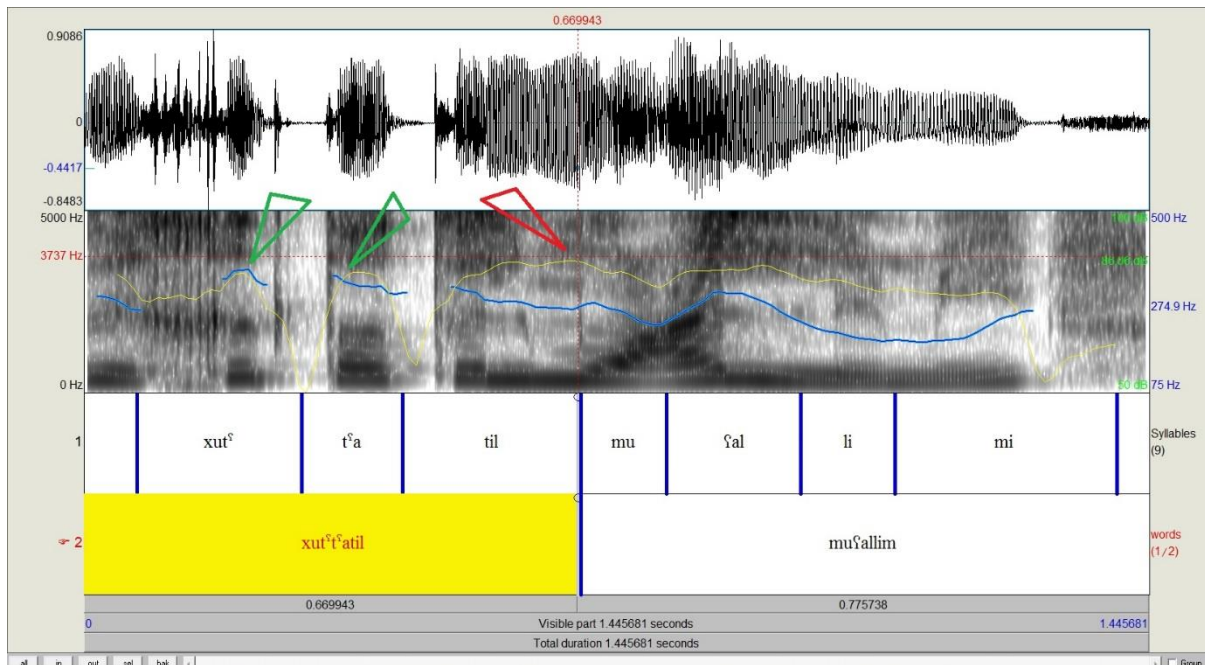


Figure 9 shows that the compound, i.e. *muṣallim rriyaad<sup>ṣ</sup>iyyaat* ‘the maths teacher’ is assigned stress on the first element. However, the presence of an assimilated geminate together with

sandhi means that the intensity contour falls only very slightly across the word boundary. Figure 10 shows that the P-construct, i.e. *xut<sup>ʕ</sup>t<sup>ʕ</sup>at lmuʕallim* ‘the teacher’s plan’, in which the second element does not start with a coronal sound is assigned stress on the first element with a fall of both intensity and pitch contours. Although sandhi also operates in the P-construct, i.e. *xut<sup>ʕ</sup>.t<sup>ʕ</sup>a.til.mu.ʕal.lim* ‘the teacher’s plan’, it does not have an effect on stress across word boundaries. The decisive factor seems to be that there has to be an assimilatory gemination for stress to be influenced across word boundaries. Hence, stress is assigned to the default position, i.e. the first element. Similar to the first element *xut<sup>ʕ</sup>t<sup>ʕ</sup>at* ‘plan’ in which a geminate is present, the second element also has a lexical geminate, i.e. *lmuʕallim* ‘teacher’. Yet, the stress falls on the first element not on the second.

Even though the position of the stress in the compound in Figure 9 and the P-construct in Figure 10 is on the first element, there is a subtle difference in the decline of the pitch and intensity contours across word boundary. This difference is purely a phonetic one caused by assimilatory geminate. The examples *ʕaamil nnað<sup>ʕ</sup>aafah* ‘the cleaner’, and *kitaab t<sup>ʕ</sup>t<sup>ʕ</sup>aalib* ‘the student’s book’ represented in Figures 5-7 respectively have a similar stress pattern to *muʕallim rriyaad<sup>ʕ</sup>iyyaat* ‘the maths teacher’ in Figure 9 due to the assimilatory geminate together with sandhi (the intensity contour falls slightly on the second element affecting stress across word boundary in both cases). Note that all other cases show that stress does not differentiate between compounds and P-constructs in MSA and JA.

The paragraph in the test included another compound that has a geminate in the second element but not an assimilated one, which can be used to check whether *muʕallim rriyaad<sup>ʕ</sup>iyyaat* ‘the maths teacher’ is indeed a special case. Figure 11 below shows this example.

Figure 11. Spectrogram for *mudaqqiq lkurraasah* ‘the notes inspector’ [mu.daq.qi.qil.kur.raa.sah] (compound), JA

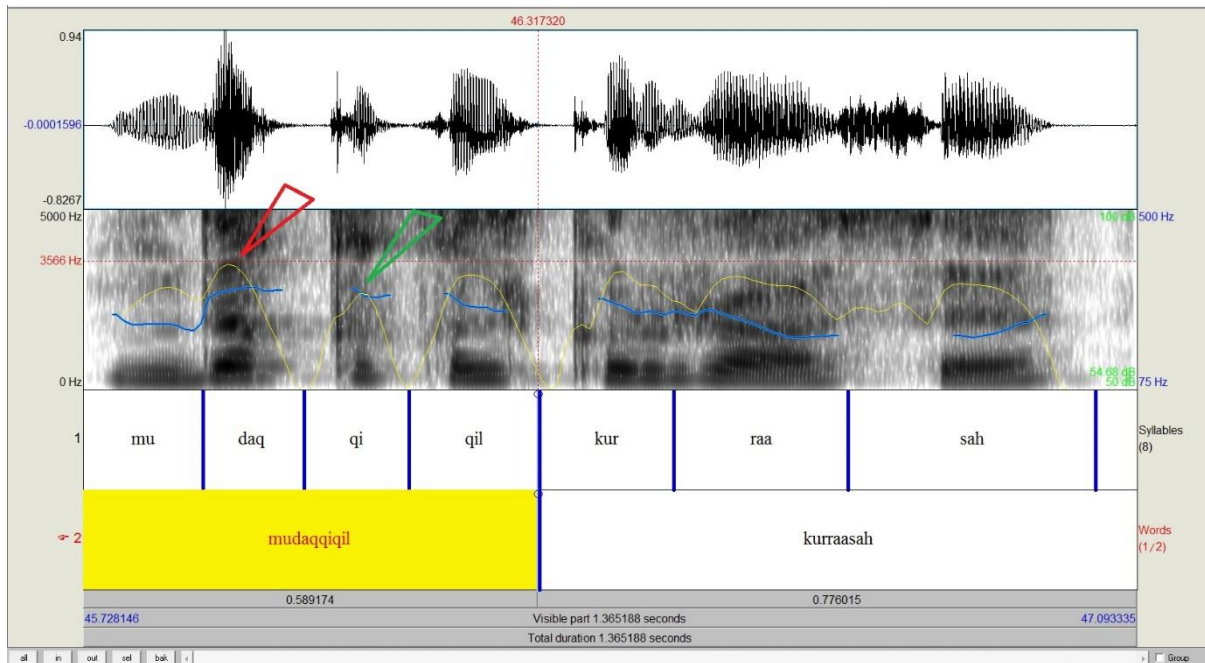


Figure 11 shows that in the absence of an assimilated geminate, the stress seems to fall on the first element even though sandhi operates in this example: *mu.daq.qi.qil.kur.raa.sah* ‘the notes inspector’ (with the position of sandhi shown in bold). Note that both the P-construct in Figure 9 and the compound in Figure 11 are assigned stress on the first element, which means that stress cannot be used to make a distinction between the two types of N + N combination.

This section has shown that the first hypothesis about stress assignment in N + N combinations (i.e. if the second/right element is definite, stress falls on the first element or on both elements) is confirmed. The next section provides an examination of stress assignment in indefinite compounds and P-constructs.

### 3.6.4.2 N + N combinations not marked with definiteness

This section is divided into four sub-sections on the basis of the presence vs. absence of gemination (see Table 3.1).

#### ➤ 1<sup>st</sup> non-geminate + 2<sup>nd</sup> non-geminate

For this case, three compounds were included in the experiment, i.e. *layla nahaar* ‘twenty four-seven’, *s’abaaḥ masaa?* ‘twenty four-seven’, and *ka?s fa?siiir* ‘juice glass’ and one P-construct, i.e. *kitaab fataah* ‘a girl’s book’. The analysis shows that the stress tends to fall on the first

element in all of them. Figures 12-14 below represent this case.

Figure 12. Spectrogram for *kaʔs ʕasʕiir* (*kaasit ʕasʕiir*, in JA) [kaa.sit ʕa.sʕiir] ‘juice glass’ (compound), JA

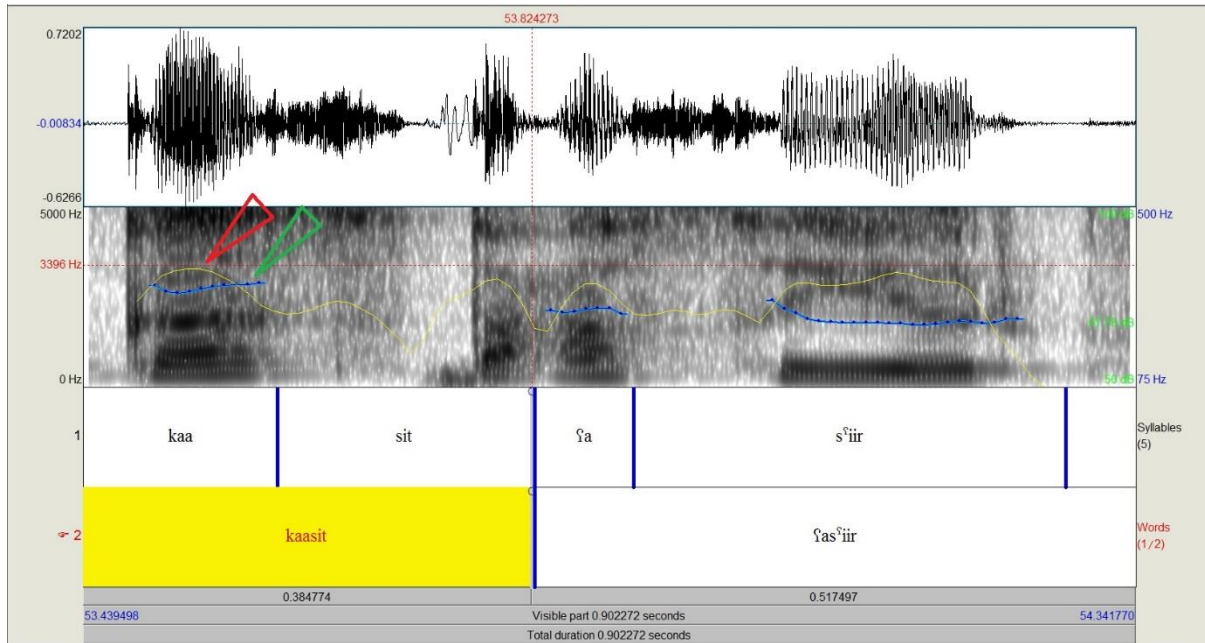


Figure 13. Spectrogram for *sʕabaah masaaʔ* [sʕ.a.baa.ha ma.saa.ʔin] ‘twenty four-seven’ (compound), MSA

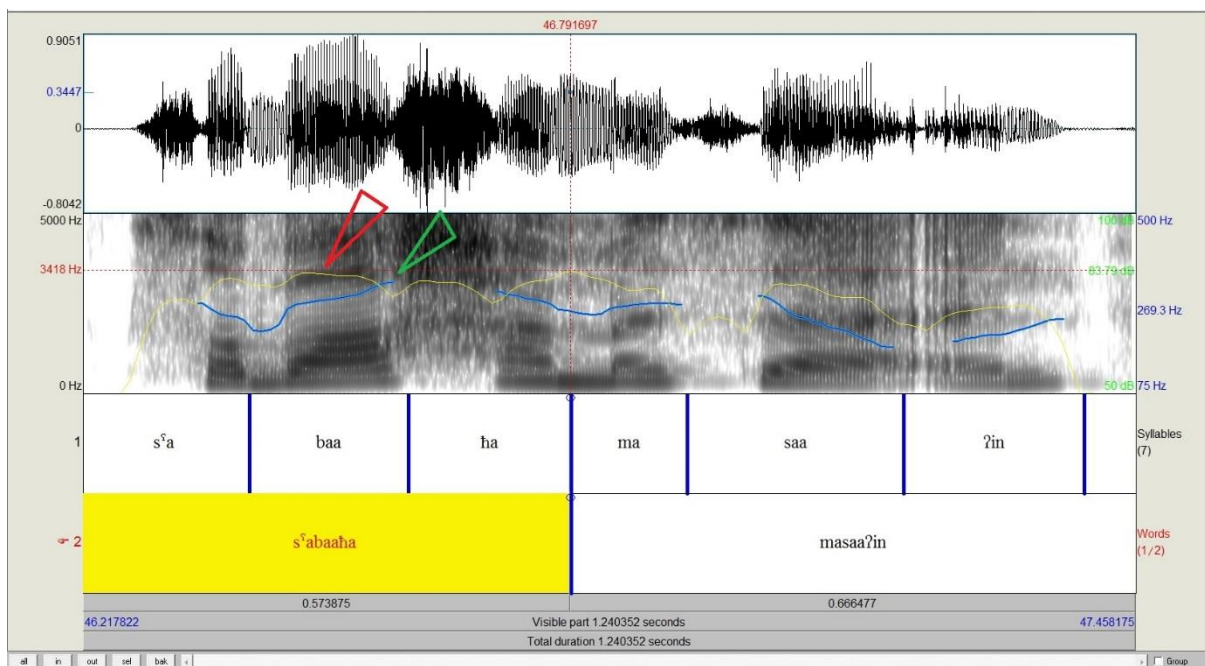
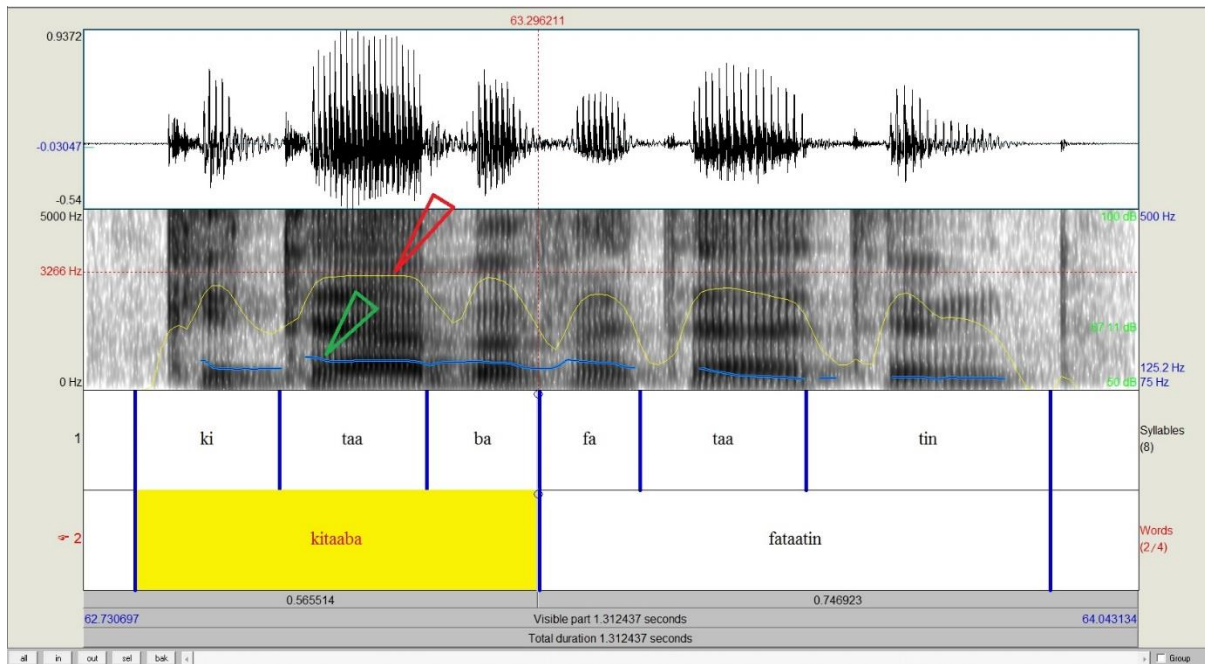


Figure 14. Spectrogram for *kitaab fataah* [ki.taa.ba fa.taa.tin] ‘a girl’s book’ (P-construct), MSA



The three figures above show that the first elements of both the compounds (see Figure 12 and 13) and the P-construct (see Figure 14) are pronounced more forcefully and clearly. The pitch and intensity contours have higher values on the first element, which means that the first element is more stressed than the second (see the red and green triangles above). It is worth pointing out that the intensity and pitch values are slightly high at the end in Figure 12 because of the presence of a superheavy syllable, i.e. *sʕiir*, which has the template CVVC (cf. Watson 2002)

➤ **1<sup>st</sup> geminate + 2<sup>nd</sup> non-geminate**

For the purpose of the experiment, one compound, e.g. *munassiq ʕuluum* ‘a science coordinator’ and one P-construct, e.g. *qubbaʕat tʕaalib* ‘a student’s cap’ were tested. Figures 15-16 below show that stress is assigned on the first element on both the compound and the P-construct.

Figure 15. Spectrogram for *munassiq* *ʕulum* [mu.nas.siq ʕu.luum] ‘a science coordinator’ (compound), JA

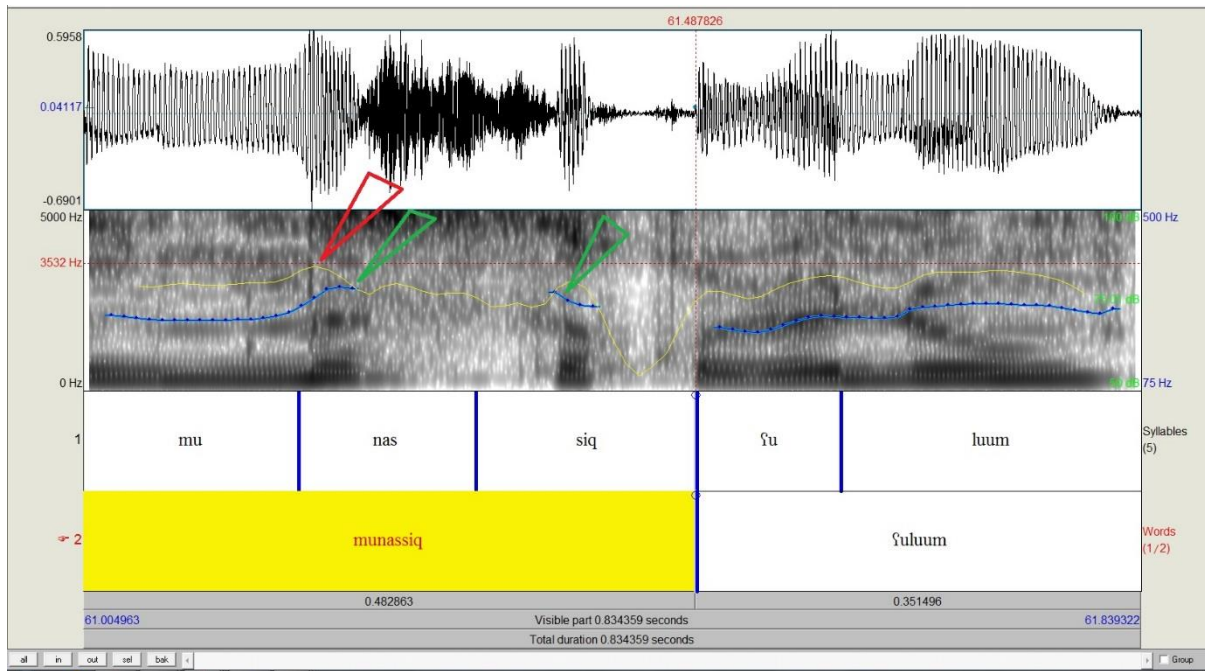
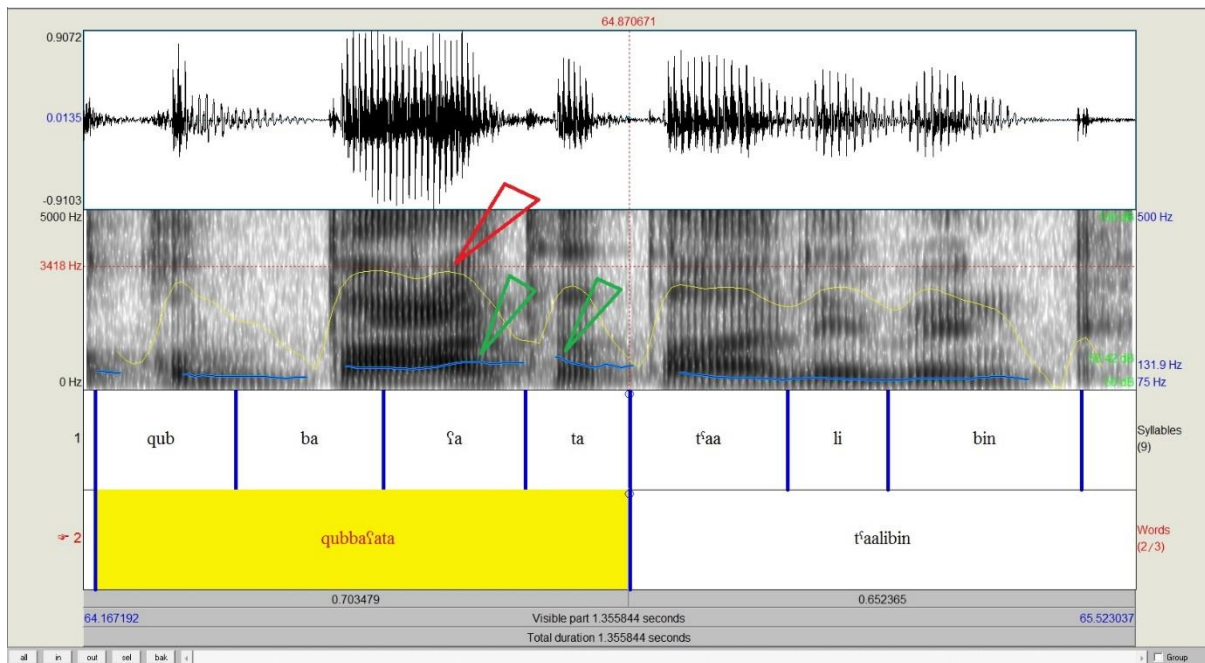


Figure 16. Spectrogram for *qubbaʕat tʕaalib* [qub.ba.ʕa.ta tʕaa.li.bin] ‘a student’s cap’ (P-construct), MSA



Even though the first element in both the compound and the P-construct has a geminate, i.e. *mu.nas.siq* ‘a coordinator’ and *qub.ba.ʕat* ‘a cap’, this does not have an effect on stress. In environments in which the first element does not contain a geminate (see the previous section),

stress tends to fall on the first element, which means that stress is assigned by default on the first element. Here too, it is worth pointing out that the intensity and pitch values are slightly high at the end in Figure 15 (the same was observed in Figure 12) because of the presence of a superheavy syllable, i.e. *luum*.

➤ **1<sup>st</sup> non-geminate + 2<sup>nd</sup> geminate**

For this case, two compounds, e.g. *fat<sup>ʿ</sup>iirat tuffaaḥ* ‘apple pie’ and *ḥulwun murrun* ‘bitter-sweet’, and one P-construct *qalam muṣallim* ‘a teacher’s pen’ were included in the experiment. Spectrograms for the two types of construct are presented in Figures 17-19.

Figure 17. Spectrogram for *fat<sup>ʿ</sup>iirat tuffaaḥ* [fa.t<sup>ʿ</sup>ii.ra.ti tuf.faa.ḥin] ‘apple pie’ (compound), MSA

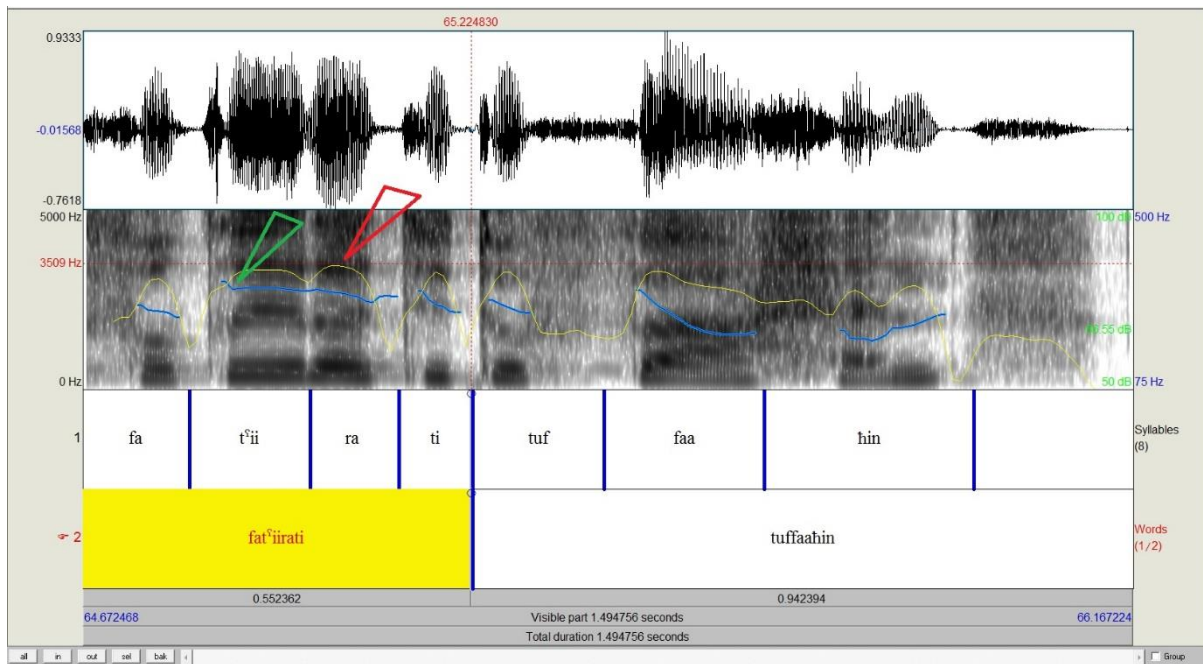




Figure 18. Spectrogram for *ḥulwun murrun* [ḥul.wun mur.run] ‘bitter-sweet’ (compound), MSA

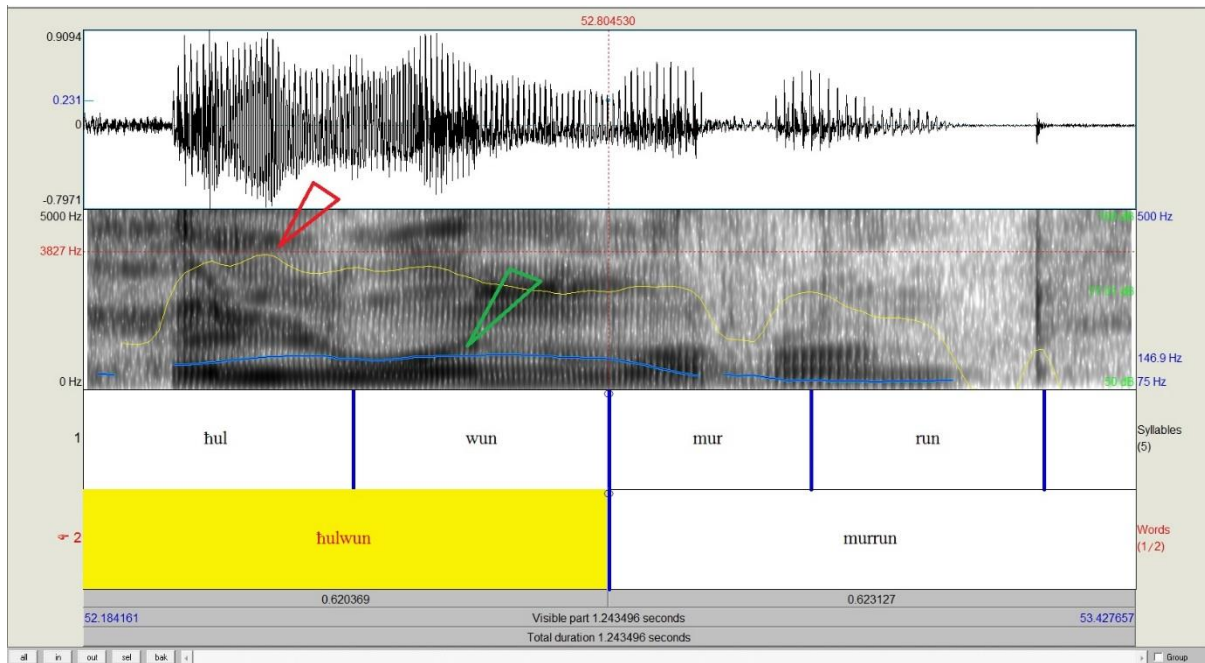
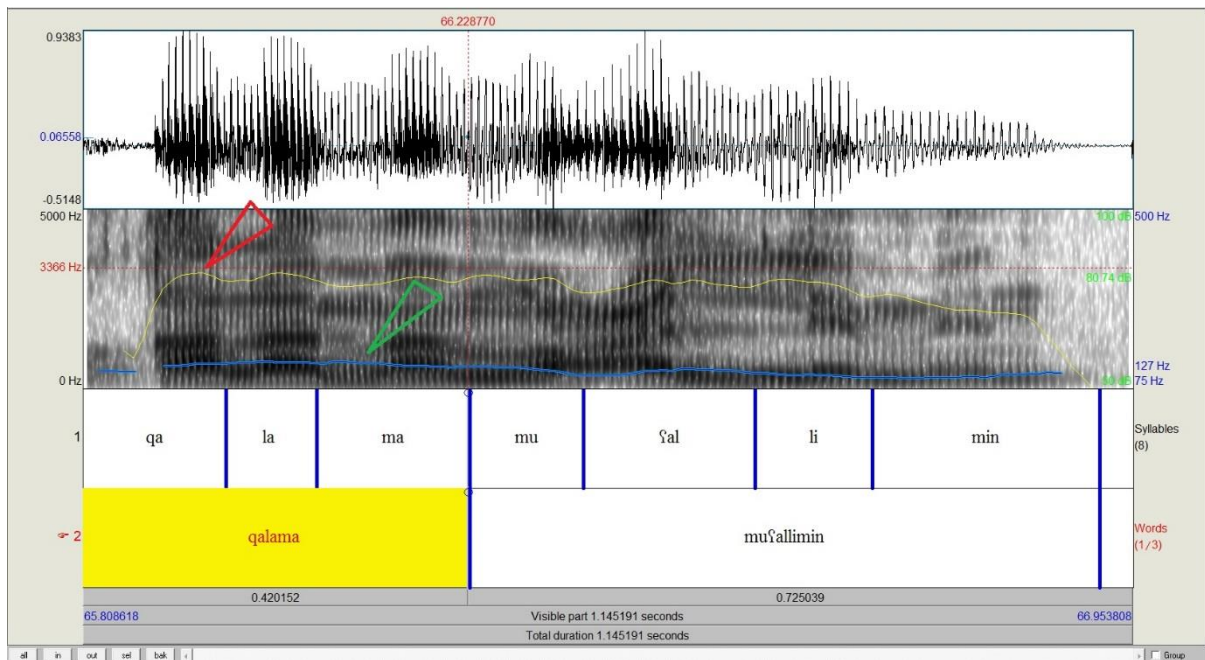


Figure 19. Spectrogram for *qalam muʿallim* [qa.la.ma mu.ʕal.li.min] ‘a teacher’s pen’ (P-construct), MSA



The three figures above show that stress tends to appear on the first element in both the compounds (see Figures 17-18) and the P-construct (see Figure 19). Despite the fact that the second elements of both constructs contain a geminate, i.e. *tuf.ʕaah* ‘apple’, *mur.run* ‘bitter’

and *muṣallim* ‘teacher’, the main stress falls on the first element, rather than on the second. Note, however, that even if the main stress falls by default on the first element, that does not mean that the second element is not stressed at all (cf. Spencer 2002). While the main stress clearly falls on the first element (i.e. *fat‘iirat* ‘pie’, *ḥulwun* ‘sweet’ and *qalam* ‘pen’ in *fat‘iirat tuffaah* ‘apple pie’, *ḥulwun murrun* ‘bitter-sweet’ and *qalam muṣallim* ‘a teacher’s pen’), the second element (i.e. *tuffaah* ‘apple’, *murrun* ‘bitter’ and *muṣallim* ‘teacher’) is not unstressed; it is assigned secondary stress.

Note that the stress placement in *ḥulwun murrun* ‘bitter-sweet’, which is an Adj + Adj combination, is the same as that in the N + N combination *fat‘iirat tuffaah* ‘apple pie’. It thus appears that the stress tends to fall on the first element of the compound regardless of the word class of the elements.

Again, with regard to this case, it is evident that stress fails to differentiate between compounds and P-constructs in MSA and JA.

➤ **1<sup>st</sup> geminate + 2<sup>nd</sup> geminate**

In this case, one compound, i.e. *muṣallim s‘aff* ‘a primary school teacher’ and one P-construct, e.g. *sidzill mus‘ahhiḥ* ‘a marker’s record’ were tested. Figures 20-21 represent the two types of construct.

Figure 20. Spectrogram for *muṣallim s‘aff* [mu.ṣal.li.ma.s‘af.fin] ‘a primary school teacher’ (compound), MSA

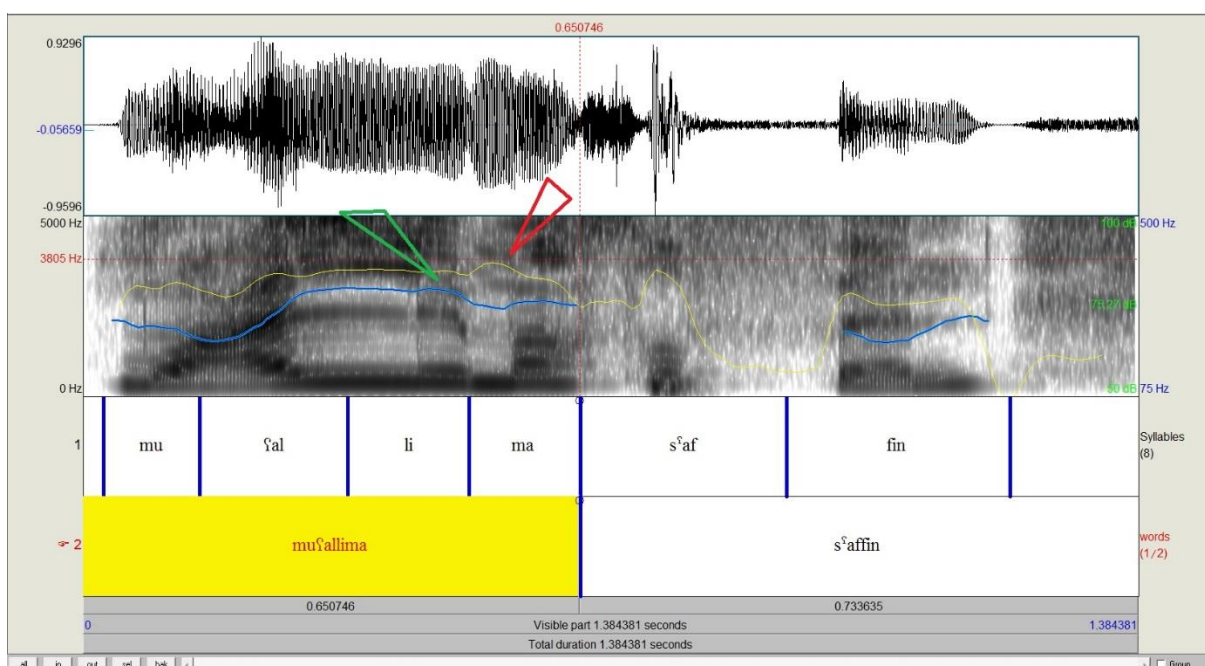
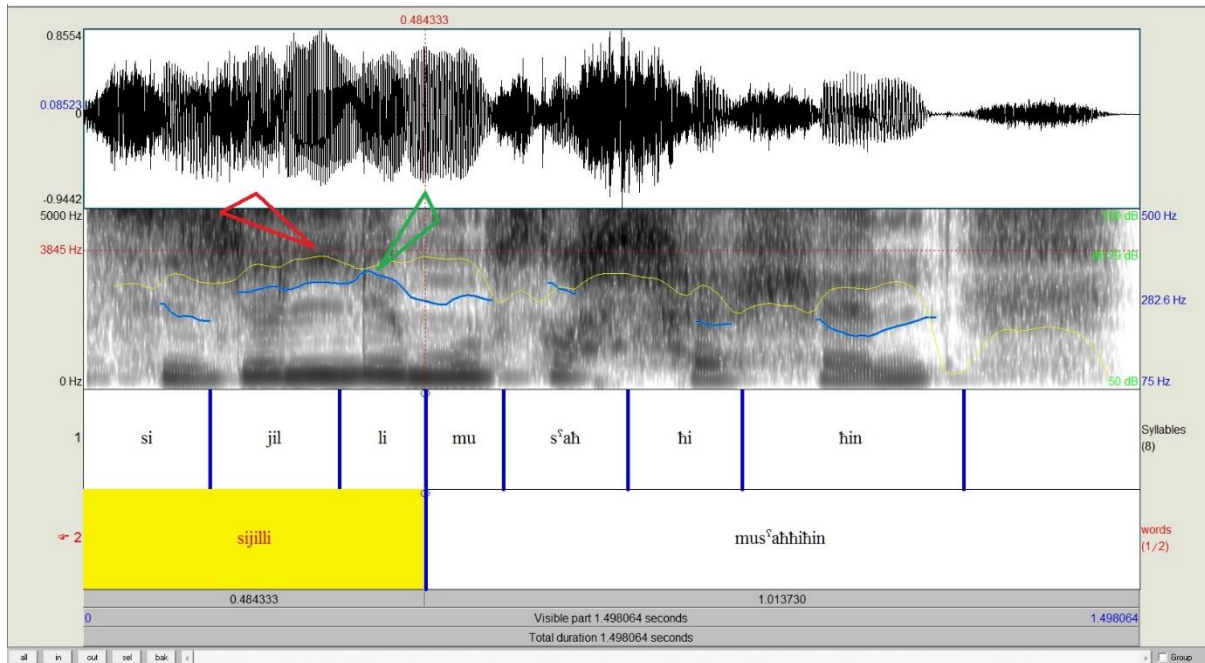


Figure 21. Spectrogram for *sidzill musʿahhih* [si.dzil.li.mu.sʿah.hi.hin] ‘a marker’s record’ (P-construct), MSA



The two figures show that both pitch and intensity seem to have higher values on the first element. Therefore, the first element seems to be more stressed, in both the compound (see Figure 20) and the P-construct (see Figure 21). Hence, in this case too, stress offers no help in differentiating between compounds and P-constructs in MSA and JA.

### 3.7 Conclusion

In sum, it has become apparent in this chapter that orthography makes no distinction between compounds and P-constructs in MSA and JA; both types of constructs are always written as two separate words. In addition, sandhi operates in both types of construct; thus, it fails to distinguish between them. Examination of spectrogram data shows that stress plays no role in distinguishing between various N + N combinations (i.e. compounds and P-constructs) in MSA and JA. Both types of construct show similar stress assignment in all cases. The analysis has shown that the default position of the stress in N + N combinations seems to be on the first element. However, the presence of an assimilatory geminate on the boundary between the two words has been shown to cause a very slight fall in intensity contour, making it appear as though

a secondary stress or perhaps double stress is assigned. This indicates that the first hypothesis in which I proposed that primary stress falls on the first element or both elements if the second element is definite is confirmed.

Regarding the second hypothesis, data analysis demonstrates that if the second element is indefinite, primary stress falls on the first element. In other words, the second hypothesis is confirmed. Since this pattern is found in both the compounds and P-constructs, the conclusion is that phonological criteria fail to differentiate between these two types of construct. The analysis has shown that stress assignment in N + N combinations follows the same patterns in both MSA and JA.

Finally, although the conclusion to this chapter looks negative (compounds and P-constructs cannot be distinguished orthographically or phonologically), we are now in a position to state this conclusion with confidence, rather than with the uncertainty that prevailed before. The investigation of these aspects of N + N combinations has also led to some new results which are valuable independently of the phrase-compound distinction, such as patterns in the location of stress in N + N combinations, especially the effect of assimilatory gemination on stress assignment.

## **Chapter Four: Identifying compounding in MSA and JA: the semantic and syntactic criteria**

### **4.1 Introduction**

As discussed in the previous chapter, orthographic and phonological criteria fail to distinguish between the two possible types of construct, i.e. compounds and P-constructs within N + N combinations in MSA and JA. Therefore, other criteria need to be examined in order to differentiate between these two possible nominal constructs which were identified in the previous chapter. This chapter discusses the differences between these constructs by applying the semantic and syntactic criteria delineated in chapter 2. The chapter is organised as follows: section 4.2 discusses the differences between the two possible N + N constructs in MSA and JA by applying the compositionality and referentiality criteria. Section 4.3 applies the syntactic criteria discussed in chapter 2 to the two types of construct in MSA and JA. These syntactic criteria include modification, adjacency, coordination, replacement of the second element by a pro-form, ellipsis, and inflection and linking elements. In addition to the general criteria discussed in chapter 2, I propose two language-specific criteria that are exclusive to Arabic. The first criterion deals with whether the definiteness of the first element triggers the appearance of the possessive marker *li-* ‘of/for’ on the second element, while the second criterion is concerned with the appearance of the possessive marker *li-* ‘of/for’ when the first element is preceded by a cardinal number. Finally, section 4.4 summarises the similarities and differences between the two suggested N + N constructs in terms of the semantic and syntactic criteria.

### **4.2 The semantic criteria**

In this section, I apply to MSA and JA the semantic criteria of compositionality and referentiality, used in the relevant literature to distinguish between compounds and phrases cross-linguistically.

#### **4.2.1 Compositionality**

As we saw in section 2.2.4, it has been suggested that compositionality can be used to distinguish between different types of N + N combinations (cf. Borer 2009). In MSA and JA, P-constructs are always compositional in the sense that their meanings are the total sum of their parts, as in (1) and (2):

- (1)        *matʕaar*        *ʕammaan*        (P-construct)  
              airport        Amman  
              ‘Amman’s airport’
- (2)        *θawb*        *ʔumm-ii*        (P-construct)  
              gown        mother-my  
              ‘my mother’s gown’

In example (1), *matʕaar ʕammaan* ‘Amman’s airport’ IS AN airport, using Allen’s (1978) ‘IS A’ condition. This means that it has a semantic head, which is *matʕaar* ‘airport’. Semantically, it is an airport which has something to do with Amman, particularly, it is located in Amman. The same applies to example (2), in which *θawb ʔummii* ‘my mother’s gown’ IS A gown; with regard to its meaning, it is a gown that belongs to my mother. On the other hand, compounds can be either semantically compositional or non-compositional. In the former case, the meaning of the compound is the total sum of its parts, whereas in the latter case, the meaning of the compound is not the total sum of its parts. In fact, we shall see that in some cases, the meaning of the whole compound can be completely unrelated to the internal elements.

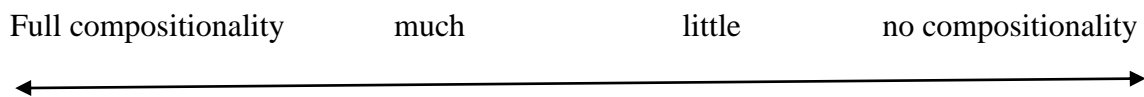
In relation to degrees of semantic compositionality, Bauer (1983: 56) argues that compounds that exhibit different degrees of compositionality are attested in the literature.<sup>24</sup> For example, unlike *understand* which has been completely lexicalised, i.e. the two components *under* and *stand* have lost their original meanings (Lipka 1977: 160, cited in Bauer 1983), compounds like *playboy* and *bedstead* still have some relation with the meaning of their parts to some extent; hence, they are only partially compositional (Bauer 1983: 56-57). According to Marelli and Luzzatti (2012: 653) and Ji *et al.* (2011), semantic transparency or degrees of semantic compositionality is all about how well the combination of the two elements of the compound, rather than the two individual words, determines the meaning of the whole compound. Along these lines, semantic compositionality can be said to lie on a synchronic and

---

<sup>24</sup> There is general consensus concerning the concept of semantic transparency in that it is usually considered to mean semantic compositionality. In fact, in a number of cases, compositionality, in relation to compounds, has a similar definition to semantic transparency. Roelofs and Baayen (2002: 132) indicate that “A morphologically complex word is semantically transparent if its meaning is compositional”. According to Girju *et al.* (2005: 488), “the meaning of compositional compounds can be successfully derived from the meaning of the noun constituents”. The distinction between semantic compositionality vs. semantic transparency will not be discussed here any further.

diachronic continuum with full compositionality at one end and no compositionality at the other end, with various degrees of compositionality in between (Waugh 1994: 64):

Figure 1. Degrees of semantic compositionality, based on Waugh (1994)



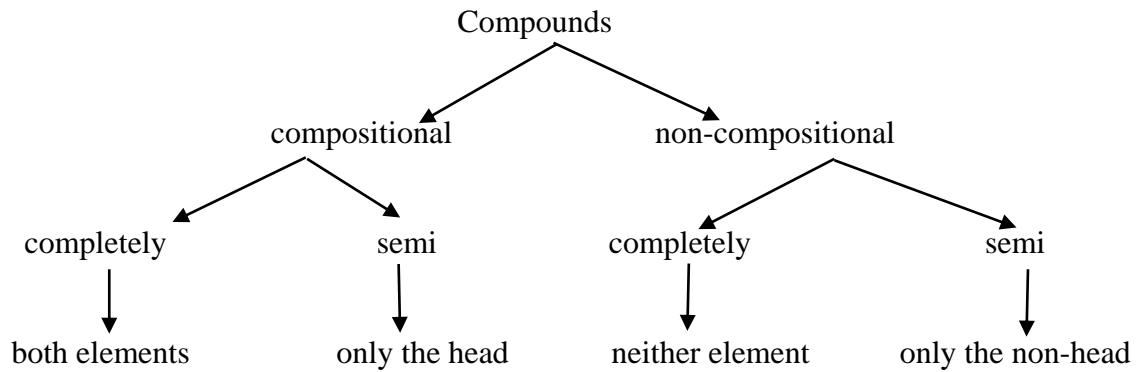
Similarly, Kavka (2009) argues that compositionality as an interplay of variability and literalness should be regarded as a scalar phenomenon. This is because lexical and grammatical features are not merely present or absent; they construe a continuum containing three stages, i.e. full variability, partial variability and zero variability or invariability. Likewise, literalness exhibits a texture of interrelations, which is mainly gradient (Kavka *ibid*). Thus, this makes compositionality a scalar phenomenon, in which multiword expressions are viewed as fully compositional, e.g. *shoot a bird* and *red ink*; semi-compositional, e.g. *shoot a film* and *red carpet*; or non-compositional, e.g. *shoot the breeze* and *blue blood*. With respect to the difference between fully compositional and semi-compositional, Kavka (2009) states that *playground* is fully compositional, since it literally means a ground to play on, while *life boat* does not mean ‘a living boat’, rather ‘a boat used for saving lives’. Similar examples are *bulldog*, *horse-fly*, *stone-fish*, etc. He also indicates that the majority of non-compositional compounds are invariable that is, the sequence of elements is fixed, e.g. *lazybones* not *\*boneslazy*. With regard to their lexical flexibility, modifications are always external, that is, they modify the whole compound, rather than one or the other element, e.g. in *an intolerable lazybones*, the word *intolerable* describes the whole compound.

According to Kavka (2009), the meaning of a multiword expression may become fixed through time, which reduces the degree of variability, resulting in it having a figurative meaning. This may suggest that the expression moves along the scale of compositionality. Historically, compositionality can be regarded as a cline on which multiword expressions are situated. Expressions which are referred to as idiomatic are on the non-compositional side of the cline.

Waugh’s (1994: 64) classification and Kavka’s (2009) proposal, which are quite similar, can also be applied to compounds in MSA and JA; the degrees of compositionality being based on the semantic contribution of the head and the non-head to the meaning of the whole compound. To provide a clear picture of the degrees of compositionality in MSA and

JA, Waugh's (1994) classifications will be slightly modified. If only the head of the compound contributes to the meaning of the whole compound, then it is semi-compositional, whereas if only the non-head of the compound contributes to the meaning of the whole compound, then it is semi non-compositional. This is illustrated in Figure 2 below.

Figure 2. Levels of compositionality vs. non-compositionality in MSA and JA compounds



Examples of each type of compound described in Figure 2 are provided below:

- (3)      *yazl/faṣar*<sup>25</sup>    *l-banaat*      (completely non-compositional compound)  
          spinning/hair    the-girls  
          ‘the candyfloss’  
          lit. the girls’ hair
- (4)      *Ṣarus*            *l-baḥr*            (semi-non-compositional compound)  
          bride            the-sea  
          ‘the mermaid’  
          lit. the sea bride
- (5)      *Ṣas‘iir*            *t-tuffaaḥ*      (completely compositional compound)  
          juice            the-apple  
          ‘the apple juice’

<sup>25</sup> The word *yazl* ‘spinning’ is the one used in MSA, while the word *faṣar* ‘hair’ is used in JA.



- (6)           burdʒ           l-ʕarab  
                  tower           the-Arab  
                  ‘the Arab tower’                   (semi-compositional compound)

Examples (3) and (4) that illustrate that two levels of non-compositionality can be distinguished in MSA and JA compounds. In example (3), *faʕar lbanaat* ‘candyfloss’ IS NOT *faʕar* ‘hair’. Semantically, it is neither related to *faʕar* ‘hair’ nor to *lbanaat* ‘the girls’. The meaning of the whole compound (i.e. candyfloss) is also not related to the meaning of both elements combined. Hence, it is completely non-compositional. Conversely, *ʕaruus lbaħr* ‘mermaid’, in example (4), IS NOT A *ʕaruus* ‘bride’. However, *ʕaruus lbaħr* ‘mermaid’ is a mythical creature that lives in *lbaħr* ‘the sea’. Hence, it is semi-non-compositional. Note that both completely non-compositional and semi-non-compositional compounds are exocentric. In particular, both types are semantically headless, since they do not denote hyponyms of their heads, i.e. the left elements (for more on headedness, see chapter 7).

Similarly, two levels of compositionality can be detected in compounds in MSA and JA. For example, *ʕasʕiir ttuffaaħ* ‘the apple juice’, in example (5), IS A *ʕasʕiir* ‘juice’. Semantically, *ʕasʕiir ttuffaaħ* is juice made from apples. Hence, this compound is completely compositional. In contrast, *burdʒ lʕarab* ‘the Arab tower’, in example (6), IS A *burdʒ* ‘tower’, located in Dubai, but it is mostly occupied by foreigners. In other words, it is not related to the second element *lʕarab* ‘the Arab’. Thus, this compound is semi-compositional, because only the head contributes to the meaning of the whole compound (see Figure 2).

The discussion above shows that the compositionality criterion can only distinguish between P-constructs and non-compositional compounds such as (7-10), shown below:

- (7)           Abuu           l-hawl           (completely non-compositional compound)  
                  father           the-terror  
                  ‘the sphinx’  
                  lit. the father of terror
- (8)           maaʔ           l-wadʒh           (completely non-compositional compound)  
                  water           the-face  
                  ‘the dignity’  
                  lit. the face water

- (9) faras n-nahr (semi-non-compositional compound)  
horse the-river  
‘the hippopotamus’  
lit. the river horse
- (10) raʔs l-maal (semi-non-compositional compound)  
head the-money  
‘the start-up capital’  
lit. the money head

However, compositionality cannot distinguish between P-constructs, on the one hand, and compositional compounds, on the other. This can be seen in the following examples:

- (11) matʕaar ʕammaan IS A matʕaar (P-construct)  
airport Amman airport  
‘Amman’s airport’
- (12) mazraʕat r-radʒul IS A mazraʕah (P-construct)  
farm the-man farm  
‘the man’s farm’
- (13) haassat ʃ-ʃam IS A haassa (compositional compound)  
sense the-smell sense  
‘the sense of smell’
- (14) dʕiffat n-nahr IS A dʕiffat (compositional compound)  
bank the-river bank  
‘the river bank’

It seems that both P-constructs and compositional compounds are compositional. From a semantic viewpoint, the meaning of the whole construct is the total sum of the meaning of the two elements. Furthermore, examples (11-14) are endocentric, since they have a semantic head, i.e. the left element. This means that items that are non-compositional are compounds.

Note that many examples of non-compositional compounds in MSA and JA may have

been borrowed from other languages, such as English, and integrated into Arabic at some point in history. MSA and JA do not have one-to-one equivalents for words like ‘mermaid’, i.e. *ġarus lbaħr* ‘sea bride’ in MSA, *ħuuriyyat lbaħr* ‘beautiful woman in the sea’ in JA.<sup>26</sup> Therefore, MSA and JA use two words to describe one lexical item. This can explain why N + N constructs are very productive. Note, also, that the outcome of this integration is, in many cases, non-compositional (see examples 7-10).

#### ***4.2.1.1 Compositionality of metaphorical and metonymical compounds***

In relation to compositionality, it is also important to consider the metaphorical and metonymical compounds that exist in English. According to Bauer *et al.* (2013: 465, also in Bauer 2010), *bahuvrihi* compounds (i.e. person or thing that has X, in which X is the property described by the compound), are often meant metaphorically or metonymically. Metonymy is defined as “a figure in which one word is substituted for another on the basis of some material, causal, or conceptual relation” (Preminger and Brogan 1993, cited in Papafragou 1996: 169). An example of metonymy is: *have you read Jane Austen?* In this example, the speaker means: *have you read a novel by Jane Austen?* Here the name of the author is used to refer to her books. Bauer *et al.* (2013: 465) discuss two examples from English to explain metaphorical and metonymical meanings, respectively. Firstly, in the compound a *blockhead*, a person whose head is like a block is perceived as a stupid person. In other words, there is a similarity between a block and the inability to understand; hence, the metaphor. Secondly, in the compound *redshank*, the *shank* ‘the lower part of an animal’s leg, or meat from an animal’s leg’, which is a part of the bird, is used to refer to the bird itself. Thus, a *redshank* is a bird with red shanks.

Based on the above examples, several scholars (e.g. Marchand 1960; Bauer 1978; Bauer 2008; Scalise and Fàbregas 2010; Bauer *et al.* 2013 among others) have observed that *bahuvrihis* are essentially formed on the basis of PART FOR WHOLE metonymy. However, the endocentricity vs. exocentricity of *bahuvrihis* has been subject to some debate in the relevant literature due to their figurative nature. Due to the importance of *bahuvrihis* in this study, I discuss their possible analysis in the two following paragraphs.

Bauer (2008: 65) suggests that *bahuvrihis* possibly need to be regarded as endocentric compounds which are used metaphorically, not exocentric compounds even though they “form a recognisable group, and are included [in discussion of exocentric compounds] for

---

<sup>26</sup> *ħuuriyyah* is the most beautiful young woman with a fair skin found in heaven.

completeness”. In other words, bahuvrihis may not be genuinely exocentric; they can be interpreted as endocentric compounds with a metaphoric reading. Similarly, Scalise and Fàbregas (2010: 121) suggest that bahuvrihi-compounds can be endocentric, because they can be interpreted through metonymy. For example, due to the fact that all humans have a face, the bahuvrihi-compound *paleface* could denote a human being characterised by a notable property of his/her face.

On the other hand, some researchers (e.g. Booij 2002: 143; Benczes 2006) propose that bahuvrihis are exocentric, since there is an implicit element in the head position, i.e. a person. Bauer himself seems not completely decided on this issue. Bauer (2010: 169) defines bahuvrihis as exocentric compounds (see section 1.2.3). More recently, Bauer *et al.* (2013: 478) suggest that bahuvrihis are exocentric, but leave open the possibility of analysing them as endocentric because of their figurative reading. They indicate that bahuvrihis of this type are productive in English, e.g. *air head*, *bone head*, *acid head*, *cheesehead*, *butterhead*, *bottle head* and *redhead*, which are all metonymic in nature. Bauer *et al.* (2013: 478) conclude that the main difference between bahuvrihis and endocentric attributive compounds is that the referents of bahuvrihi compounds are characterised figuratively, normally being metaphorical or metonymic. Therefore, analysing them as regular endocentric compounds with a metonymic or metaphorical interpretation of the head noun is possible (Bauer *et al.* 2013: 478-9). In this study, bahuvrihis are treated as exocentric compounds, since the semantic head, i.e. person, is absent. In addition, Bauer (2008, 2010) and Bauer *et al.* (2013) still use the term exocentric throughout, even though they acknowledge another possible analysis.

Resuming our discussion of the existence of metaphor and metonymy in compounding, their existence is hardly surprising if metaphorical and metonymical thinking is regarded as a normal, everyday ability of humans (see Lakoff and Johnson 2003). As argued by Langacker (1987), Talmy (1988), and Croft and Cruse (2004) among others, metaphor and metonymy can be viewed as a type of construal operation, which may suggest that they facilitate the process of interpreting/conceptualising the world around us. Therefore, it is generally accepted that the use of metaphors and metonymies in word-formation should be regarded as a natural process, rather than a rhetorical one.

Similarly, some compounds in MSA have metaphorical and metonymical meaning.<sup>27</sup> These are illustrated below<sup>28</sup>:

---

<sup>27</sup> Metaphors are identified using Group’s (2007) Metaphor Identification Procedure.

<sup>28</sup> Examples (15-17) are not used in JA. *ʃʃams* ‘the sun’, *ldʒamal* ‘the camel’ and *lmiʕdeh* ‘the stomach’ are used instead of *ʕayn ssamaaʔ* ‘eye of the sky’, *safiinat sʕsʕahraaʔ* ‘the ship of the desert’ and *bayt ddaaʔ* ‘house of

- (15)      *ʕayn*                  *s-samaaʔ*  
             eye                  the-sky  
             ‘the sun’  
             lit. the eye of the sky
- (16)      *safiin-at*                  *s<sup>ʕ</sup>-s<sup>ʕ</sup>aḥraaʔ*  
             ship-FSG                  the-desert  
             ‘the camel’  
             lit. the ship of the desert
- (17)      *bayt*                  *d-daaʔ*  
             house                  the-illness  
             ‘the stomach’  
             lit. the house of the illness

In example (15), a metaphor can be detected. The eye of the sky refers metaphorically to the sun. Specifically, the part that mainly captures people’s attention when they look at someone’s face is their eyes. Similarly, the thing that mainly captures people’s attention when they look at the sky is the sun. Hence, in example (15) *ssamaaʔ* ‘the sky’ is portrayed as a face and the sun is portrayed as *ʕayn* ‘eye’.

In example (16), the metaphor is manifest in the word *safiinat* ‘ship’. The camel is described as a ship in terms of its purpose. Specifically, just like a ship is the means of transportation across water, the camel used to be the only means of transportation across the desert in the past. There is another metaphor, in which the desert is perceived as the sea in which the camel and the ship travel, respectively.

In the last example (17), the compound *bayt ddaaʔ* ‘stomach’ is by far the most complex compound I came across. In particular, *bayt ddaaʔ* ‘stomach’ is not a type of *bayt* ‘house’; it is an organ inside the body. The stomach is the source of many types of diseases that affect humans and animals. The meaning of the metaphor could be explained in terms of containers or ontological metaphors (see Lakoff and Johnson 2003). That is, the use of *bayt* ‘house’ is

---

illness’, respectively. Note that *ʕams* ‘the sun’, *ldzamal* ‘the camel’ and *lmaʕidah* ‘the stomach’ are also used in MSA.

meant to denote a container in which things can be contained. Thus, the stomach is compared to a house, since both of them can contain things within them, e.g. in *bayt* ‘house’ *furniture, clothes, tables*, etc. are found, whereas in the stomach, *water, food*, etc. which can be a cause for illness are found. The illness is contained within the stomach. In other words, this metaphor means that the stomach is the house of illness. The metonymy, on the other hand, is the reason for the link which is established between illness and stomach, not any other part of the body. That is, there is an inherent link between illness and the stomach, since it is the place where one feels pain. So, the metonymy in this example is that the house of illness stands for stomach irrespective of whether someone is ill or not. This is a case of metaphonymy, i.e. metaphor with a built-in-metonymy discussed by Goossens (1990).

Examples of metaphorical and metonymical compounds can also be found in JA, as in:

- (18)        *masakit*        *xet<sup>ʕ</sup>*        *l-muʃkileh*  
               held.I        thread        the-problem  
               ‘I found the first clue of the problem.’  
               lit. I held the thread of the problem

- (19)        *kull*    *n-naas*        *btdawwir*        *ʕala*    *lugmit* *l-ʕeef*  
               all    the-people    searching    for    bite    the-bread  
               ‘All people are working to obtain the basic means of survival.’  
               lit. all the people are searching for a bite of bread

The compound *xet<sup>ʕ</sup> lmuʃkileh* ‘the thread of the problem’ in example (18) is a metaphorical expression, in which *lmuʃkileh* ‘the problem’ is portrayed as a garment. The thread is what holds the garment together and the key of the problem is what leads to its solution. Holding a thread of that garment is used to describe finding the first clue that may lead to identifying the problem and thus finding a solution for it. A metonymy (synecdoche) can also be found in example (18) where the thread is the key part of the garment and the clue is the key to finding the solution. The first element of the compound *xet<sup>ʕ</sup>* ‘thread’ provides the connection to the cloth metaphor.

The compound in example (19) is a metonymic expression, where *lugmit lʕeef* ‘bite of bread’ is used to refer to the basic food. Another metonymy can be found in example (19),

since there is an inherent link between basic food, i.e. bread, and the ability to survive.<sup>29</sup> Note, here, that the two elements of the compound yield the metonymic expression.

There has been a wide debate on the issue of compositionality in metaphorical and metonymical compounds (Langacker 1987; Dirven and Verspoor 1998; Benczes 2006 among others). Researchers such as Langacker (1987), and Dirven and Verspoor (1998) argue that compounds can be placed on a cline of transparency based on whether their meanings are transparent. At one end of the continuum, fully productive and transparent compounds can be found, and at the other end non-transparent compounds or darker compounds, which refer to metaphorical and metonymical compounds, can be found. Between the two extreme ends of the continuum, partially transparent compounds can be found. With regard to transparent compounds, Dirven and Verspoor (1998: 60) argue that both parts of the compound and the semantic link between them “are unequivocally analysable and hence immediately transparent”. An example of this case is *apple tree* ‘a tree which bears apples’. In partially transparent compounds, the elements of the compound are still analysable. However, the semantic link is less transparent and not enough to determine to which subcategory the meaning of the compound belongs, e.g. *blackbird* does not denote a black bird, rather a bird species. Finally, Dirven and Verspoor (1998) argue that non-transparent compounds are cases in which metaphorical and metonymical processes are involved. These are difficult to interpret, since the semantic link between the elements of the compound is not transparent. An example of this case is *red tape*, which does not describe a type of tape; rather it refers to a very long and irritating bureaucratic procedure (cf. Heyvaert 2009).

Benczes (2006: 75) argues against Dirven and Verspoor (1998) and identifies two main problems with their account. Firstly, their categorisation of the various degrees of transparency is opaque. Based on their statement “unequivocally analysable”, Benczes (2006) questions whether some transparent compounds can be more transparent than others. She also questions when compounds can be viewed as partially transparent; their definition is vague. Secondly, Dirven and Verspoor’s (1998) classification of metaphorical and metonymical compounds as non-transparent is flawed (Benczes 2006: 76). In fact, Dirven and Verspoor (1998) cite an example that contradicts their claim about metaphorical and metonymical compounds being unanalysable, i.e. *information highway*, which refers metaphorically to the internet. The metaphorical meaning of *highway* can be interpreted on the basis of the metaphor ‘the internet

---

<sup>29</sup> Note that bread is the basic food in Arab countries in general and in Jordan in particular. In other countries, it could be other types of food, e.g. rice.

is a highway' (Rohrer 1997). Rohrer explains that in cyberspace, the Internet facilitates our movement in space virtually to another destination. Since the Internet is perceived as a highway on which humans can move and accumulate information, the meaning of the metaphorical compound *information highway* is easily predicted. According to Dirven and Verspoor (1998: 60-61), *information highway* is "easily analysable", despite the fact that the two researchers claim that metaphorical and metonymical compounds are non-transparent and unanalysable (Benczes 2006: 76).

The reasons for these problems, according to Benczes (2006), is that their classification of metaphorical and metonymical compounds as non-transparent is inaccurate from the beginning. If metaphor and metonymy are processes that are used in everyday conversations and are an integral part of our lives (Lakoff and Johnson 2003), then metaphorical and metonymical compounds are similar to everyday expressions just like non-metaphorical or non-metonymical ones. This may indicate that metaphorical and metonymical compounds are transparent. Then one may ask: what is the difference between non-metaphorical compounds such as *apple tree* and metaphorical ones such as *information highway*. The answer to this question, according to Benczes (2006: 77), is linguistic creativity. In the case of metaphorical and metonymical compounds, the words are combined together creatively, rather than opaquely (see Heyvaert 2009: 246). This may suggest that semantic transparency is not regarded as a property of the entire multiword expression, rather it is a property of individual constituents. Since lexical words can be used metaphorically and metonymically, this means that compounds with metaphorical and metonymic meanings can be predicted on the basis of the lexical meaning of their individual elements. It is well known that every word has a literal meaning and one that can be used non-literally or figuratively. For instance, the word *general* 'someone who is highly ranked in the army' can be used metaphorically to refer to a boss in a company or organisation. The comparison is usually made on the basis of resemblance or correlation. The metaphorical use of the word *general* being used to refer to a boss is predictable and transparent. Thus, one may argue that cases of metaphors and metonymies in compounding are compositional.

Finally, I suggest that the compositionality of metaphorical and metonymical compounds is gradable, since not all native speakers of the language would be able to interpret them in the same way. For instance, in example (15), two native speakers of JA out of three were able to guess the meaning of *ʕayn ssamaa?* lit. the eye of the sky 'the sun' easily. However, none of them was able to interpret the metaphor in example (17) *bayt ddaa?* 'the stomach'. At the beginning, they thought that *bayt ddaa?* 'the stomach' meant 'the hospital'.

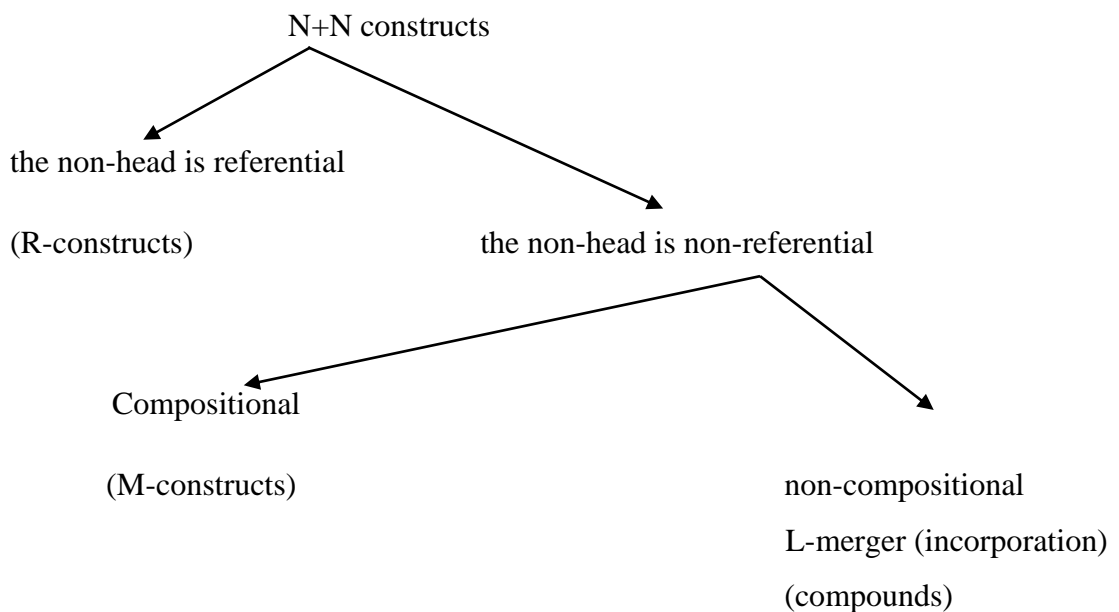


Speakers' inability to guess the meaning of metaphors such as *bayt ddaa?* 'the stomach' may imply that some metaphors are more linguistically creative than others. In MSA for example, poets and writers used to employ metaphors abundantly to make money through composing poems for royalty. However, these observations require rigorous empirical investigation, and is thus beyond the scope of the present study.

#### 4.2.2 Referentiality

Referentiality is one of the most important criteria that can differentiate between P-constructs on the one hand, and both compositional and non-compositional compounds, on the other (Bauer *et al.* 2013: 464). In Hebrew, Borer (2009) suggests that referentiality of the non-head plays a pivotal role in distinguishing between R-constructs, which refer to phrases where the relation between the two elements is a possessive one, on the one hand, and M-constructs and compounds, on the other. The following typology summarises Borer's (2009: 511) criteria of differentiating between compounds and phrases:

(20)



The typology in (20) shows that N + N combinations in Hebrew are mainly divided on the basis of referentiality of the non-head. If the non-head of the N + N combination is referential, then

it is an R-construct, as in (21):

- |      |  |                            |               |
|------|--|----------------------------|---------------|
| (21) | beyt<br>house<br>'(the) house of (the) minister' | (ha-)sar<br>(the-)minister | (R-construct) |
|------|--|----------------------------|---------------|
- (Borer 2009: 491)

Example (21) is an instance in which the non-head *hasar* 'the minister' is referential. On the other hand, if the non-head of the N + N combination is non-referential, then we are dealing with either an M-construct, i.e. a modification construct, or a compound. Examples of M-constructs and compounds in Hebrew are provided below:

- |      |  |                           |               |
|------|--|---------------------------|---------------|
| (22) | melaxex<br>chewer<br>'(the) (one who) chews grass' | (ha-)'esev<br>(the-)grass | (M-construct) |
|------|--|---------------------------|---------------|
- 
- |      |                                  |                       |            |
|------|----------------------------------|-----------------------|------------|
| (23) | orex<br>editor<br>'(the) lawyer' | (ha-)din<br>(the-)law | (compound) |
|------|----------------------------------|-----------------------|------------|
- (Borer 2009: 491-2)

In examples (22) and (23), the non-heads *ha'esev* 'the grass' and *hadin* are non-referential. The main difference between examples (22) and (23) is that the former is compositional, whereas the latter is non-compositional. In this regard, Borer (2009: 509) states that "Why, the reader may now wonder, is incorporation necessary for the formation of compounds? Such a question, however, appears to be ill-phrased. Incorporation is not necessary for compounds. Rather, compounds, by definition, are constructs that have undergone incorporation". This argument, however, is purely theoretical and is thus beyond the scope of this study. Additionally, in relation to Borer's argument on compositionality, several authorities (Bauer 1983, 2003, 2009b; Booij 2009, 2010, 2012; Lieber 2005, 2009, 2010; Plag 2003, 2006 among others) have not suggested that compounds should be non-compositional.

Concerning the limitations of referentiality, Borer (2009) does not discuss whether there are any exceptions in which the non-head of M-constructs and compounds in Hebrew is referential. In other words, the referentiality criterion in Hebrew according to Borer (2009) is

foolproof. However, there are examples in which the non-heads of Hebrew compounds have unique references, indicating that they are referential:

(24) mishkafej (ha-)shemesh  
glasses (the-)sun  
'(the) sunglasses'

(25) or (ha-)jare'ax  
light (the-)moon  
'(the) moonlight'

(26) tapuach (ha-)adama  
apple (the-)earth  
'(the) potato'

Examples (24-26) show that the non-heads of the compounds are referential, since they refer to one specific entity, i.e. *hashemesh* 'the sun', *hajare'ax* 'the moon' and *haadama* 'the earth'. This suggests that the non-head of Hebrew compounds are usually non-referential, but not always. There are a few instances in English where proper nouns can appear in compounds, such as *Beatles fans* or *Ahmadinejad supporter* (Bauer *et al.* 2013: 464). Other examples of coordinating compounds, either in company names, e.g. *Sony Ericson* or in geographical names, e.g. *Alsace-Lorraine* are cited by Borgwaldt and Benczes (2011: 231).

Regarding Arabic P-constructs and compounds, the first element (the head) of P-constructs and compounds is always referential:

(27) bayt l-marʔah haaða (P-construct)  
house.M the-woman.F this.M  
'this woman's house'

(28) muʕallim l-fiizyaaʔ haaða (compound)  
teacher.M the-physics.F this.M  
'this physics teacher'

- (29)        ǧaruus         l-baħr.M        haađih-i        (compound)  
               bride.F         the-sea.M       this.F  
               ‘this mermaid’  
               lit. this bride sea

Examples (27-29) show that the demonstrative *haađa* ‘this’ refers to the entity denoted by the head (compare gender markings on the demonstrative and the head) in both P-constructs and compounds, indicating that the head is always referential. Note that the demonstrative *haađa* ‘this’, as shown in example (27-29), follows SGCs rather than precedes them (Fassi-Fehri 2012: 214-15). The normal position of demonstratives in Arabic is before the noun they modify in cases other than SGCs, e.g. *haađa lbayt* ‘this house’.

With respect to the second element (the non-head) of P-constructs and compounds, it seems that it is always referential in P-constructs:

- (30)        baytu             r-radzul                             (P-construct)  
               house             the-man  
               ‘the man’s house’

- (31)        ħaqqibat         l-bint                               (P-construct)  
               purse             the-girl  
               ‘the girl’s purse’

In examples (30) and (31), the second elements *rradzul* ‘the man’ and *lbint* ‘the girl’ refer to specific entities in the outside world. That is, the second elements of P-constructs are referential. On the other hand, the non-heads of compounds, whether compositional or non-compositional, are non-referential:

- (32)        minfaar         l-xařab                             (compositional compound)  
               saw             the-wood  
               ‘the wood saw’

- (33)       $\zeta$ aruus          l-baħr                                (non-compositional compound)  
               bride                the-sea  
               ‘the mermaid’  
               lit. the sea bride

In examples (32) and (33), the non-heads *lxafab* ‘the wood’ and *lbaħr* ‘the sea’ do not refer to specific entities in the outside world. For instance, in *\zetaaruus lbaħr* ‘mermaid’, *lbaħr* ‘the sea’ does not refer to a specific sea; *\zetaaruus lbaħr* ‘mermaid’ refers to the mythical sea creature that lives in salty water. The same applies to *minfaar lxafab* ‘wood saw’; *lxafab* ‘the wood’ does not refer to a specific type of wood.

There is additional evidence that the non-head of P-constructs is referential, whereas the non-head of compounds is non-referential. If a demonstrative, e.g. *haađa* ‘this’ is placed before the second element in (34) and (35), *minfaar lxafab* ‘the wood saw’ and *\zetaaruus lbaħr* ‘mermaid’, the compounds become ill-formed. Demonstratives are usually used when the speaker has a referent to that demonstrative in mind. In other words, the referent is “perceptually anchored” (Sigel 2002: 1). However, the non-heads in examples (34) and (35) are non-referential; therefore, the demonstrative cannot refer to them. This can be seen in the following examples:

- (34)      minfaar          (\*haađa)          l-xafab          (compositional compound)  
               saw                    (\*this-M)        the-wood.M  
               ‘the saw of (\*this) wood’

- (35)       $\zeta$ aruus          (\*haađa)          l-baħr          (non-compositional compound)  
               bride                (\*this.M)        the-sea.M  
               ‘the bride of (\*this) sea’

By contrast, the demonstrative *haađa* can be placed before the second element of P-constructs, since it is always referential:

- (36)      baytu                haađa                r-radzul          (P-construct)  
               house                this.M                the-man.M  
               ‘this man’s house’

- (37) qamiis<sup>ʕ</sup>      haaðih-i      l-fataah      (P-construct)  
 shirt      this-F      the-girl.F  
 ‘this girl’s shirt’

The reason why the demonstrative *haaða* ‘this’ does not occur with compounds is because the second element of compounds, whether compositional or non-compositional, is non-referential.

Nonetheless, there are certain cases in which the non-heads of compounds can be referential, i.e. the second elements have a unique reference (see examples 24-26) or they are proper nouns. Examples (38-40) represent the former case:

- (38) d<sup>ʕ</sup>awʔ      l-qamar  
 light      the-moon  
 ‘the moonlight’

- (39) ʃuʃaaʕ      ʃ-ʃams  
 ray      the-sun  
 ‘the sun’s ray’

- (40) duudat      l-ʔard<sup>ʕ</sup>  
 worm      the-earth  
 ‘the earth worm’

The non-heads in examples (38-40) are referential, since they refer to a unique entity, i.e. *lqamar* ‘the moon’, *ʃʃams* ‘the sun’, and *lʔard<sup>ʕ</sup>* ‘the earth’. Examples of the latter case where the non-heads are proper nouns are provided below:

- (41) raʔs      kulayb  
 head      Kulayb  
 ‘very valuable object’

- (42)        mawaafiid    ʕurquub  
               appointments Urquub  
               ‘someone who is not punctual’
- (43)        ʕasʕaa        muusaa  
               staff         Moses  
               ‘a divine object’
- (44)        mismaar        dzuḥaa  
               screw         Juhaa  
               ‘excuse’
- (45)        yazw         l-ʕiraaq  
               invasion      the-Iraq  
               ‘the invasion of Iraq’

In examples (41-45), the meanings of the compounds are not transparent. In particular, in example (41), *Kulayb* refers to a king in the Arabian Peninsula, who was killed by his cousin with a spear thrown at his back. The reason for the value of his head is that *Kulayb* was a very strong and invincible man. Thus, the meaning of the compound indicates that getting his head is very difficult and thus valuable to his enemy. In example (42), *ʕurquub* is a person who used to come late or not at all. Therefore, if someone comes late, people would describe his appointments as *mawaafiid ʕurquub* ‘Urquub’s appointments’. It is well known, in example (43), that *ʕasʕaa muusaa* ‘Moses’ staff’ was used to perform miracles; thus, it is a divine object. In example (44), *Juḥaa* is a person who sold his house unwillingly to another man, but kept a screw on the wall as an excuse to keep coming back to the house whenever he wanted. The buyer could not do anything about it, since the fact that the screw belonged to *Juḥaa* was included in the contract. Because of *Juḥaa*’s frequent visits to the house to see his screw, the buyer was forced to leave the house to *Juḥaa*. Therefore, *mismaar dzuḥaa* ‘*Juḥaa*’s screw’ has become as a symbol for excuses. On the other hand, the meaning of example (45) is transparent; it is the total sum of the meanings of both words *yazw* ‘invasion’ and *ʕiraaq* ‘the Iraq’.

Based on the above discussion, it can be suggested that referentiality of the non-head in MSA and JA can be used to distinguish between P-constructs and compounds, since the non-head of the former is always referential, whereas that of the latter is usually non-referential.

Therefore, a distinguishing criterion between P-constructs and compounds has now been discovered.

### 4.2.3 Summary

In sum, there are some similarities and some differences between P-construct and compounds in MSA and JA on the basis of the semantic criteria discussed above. These similarities and differences are summarised in Table 4.1.

Table 4.1. The similarities and differences between P-constructs and compounds in MSA and JA

<b>The semantic criteria of compound-hood in MSA and JA</b>	<b>Compounds</b>	<b>P-constructs (possessive phrases)</b>
Compositionality	Compositional or non-compositional	Always compositional
Referentiality of the head	Always	Always
Referentiality of the non-head	Rarely (only with unique referents and names)	Always

## 4.3 The syntactic criteria

In this section, I discuss the syntactic criteria that can be used to distinguish between P-constructs and compounds cross-linguistically, including modification, adjacency, coordination, replacement of the second element by a pro-form, ellipsis, and inflection and linking elements. In addition, two language-specific criteria that are exclusive to Arabic, and potentially Semitic languages more widely, are investigated.

### 4.3.1 Modification

The modification criterion can be applied to N +N combinations in MSA and JA. To begin with, the heads of P-constructs and compounds (both compositional and non-compositional) can be modified by an adjective, as in (46-48):



- (46) dzalast-u fi bayt-i l-marʔat-i l-qadiim-i  
 sat-I in house.M-GEN the-woman.F-GEN the-ancient.M-GEN  
 ‘I sat in the woman’s ancient house’ (P-construct)
- (47) ʃaribt-u bi findzaan-i l-qahwat-i l-dzadiid-i  
 drank-I by cup.M-GEN the-coffee.F-GEN the-new.M-GEN  
 ‘I drank in the new coffee cup’ (compound)
- (48) marart-u bi ʃaruus-i l-baħr-i l-dzamiilat-i  
 passed-I by bride.F-GEN the-sea.M-GEN the-beautiful.F-GEN  
 ‘I passed by the beautiful mermaid’  
 lit. I passed by the beautiful sea bride (compound)

In examples (46-48), the heads *bayti* ‘house’, *findzaani* ‘cup’ and *ʃaruusi* ‘bride’ can be modified by the adjectives *lqadiimi* ‘the ancient’, *ldzadiidi* ‘the new’ and *ldzamiilati* ‘the beautiful’, respectively. This modification is indicated by gender agreement between the heads of the N + N constructs and the adjectives that follow the N + N combinations. The non-heads in examples (46-48) agree with neither the heads nor the modifying adjectives. Here, it is worth pointing out that when a P-construct is modified by an adjective, ambiguities can arise if the construct head and inner-NP are of the same gender, number, and case, as in (49):

- (49) dzalastu fi bayt-i r-radʒul-i l-dzamiil-i  
 sat-I in house.M-GEN the-man.M-GEN the-beautiful.M-GEN  
 ‘I sat in the man’s beautiful house’ or ‘I sat in the beautiful man’s house’  
 (P-construct)

Example (49) has two different readings, since the adjective *ldzamiili* ‘the beautiful’ agrees with both the head and the non-head in gender, number, and case. In such cases, the nature of the adjectives sometimes determines which element, is being modified, as in the following examples of P-constructs:

- (50) fi bayt-i r-radʒul-i l-qadiim-i  
 in house.M-GEN the-man.M-GEN the-ancient.M-GEN  
 ‘in the man’s ancient house’

- (51) fi bayt-i r-radʒul-i ð-ðakiyy-i  
in house.M-GEN the-man.M-GEN the-clever.M-GEN  
‘in the clever man’s house’

The same applies to compositional compounds; ambiguities can arise if the construct head and inner-NP are of the same gender, number, and case, as in the following example:

- (52) maʕ ʕasʕiir-i t-tuffaaħ-i l-laðiið-i  
with juice.M-GEN the-apple.M-GEN the-delicious.M-GEN  
‘with the delicious apple juice’ or ‘with the juice of the delicious apple’  
(compositional compound)

Two different readings can be detected in example (52), i.e. ‘the delicious apple juice’ or ‘the juice of the delicious apple’. However, the nature of the adjective usually determines which element is being modified, as in:

- (53) maʕ ʕasʕiir-i t-tuffaaħ-i n-naadʕidʒ-i  
with juice.M-GEN the-apple.M-GEN the-ripe.M-GEN  
‘with the juice of the ripe apple’ (compositional compound)
- (54) maʕ ʕasʕiir-i t-tuffaaħ-i l-baarid-i  
with juice.M-GEN the-apple.M-GEN the-cold.M-GEN  
‘with the cold apple juice’ (compositional compound)

In example (53), the adjective *nnaadʕidʒi* ‘ripe’ cannot be used to modify the head *ʕasʕiir* ‘juice’, since the reading \*‘the ripe juice of the apple’ is not possible. Similarly, it is not possible to use the adjective *lbaaridi* ‘the cold’ to describe the non-head *ttuffaaħ* ‘the apple’; hence, the reading ‘the juice of cold apple’ is not acceptable.

The head of non-compositional compounds can be externally modified by an adjective without any ambiguities. If the whole compound is modified by an adjective, it is always the head which is modified (see example 55), rather than the non-head. A reading with modification of the non-head, as in (56), is impossible:

- (55)      ʕaruus                      l-baħr                      l-dzamiil-at-i  
             bride.F                      the-sea.M                      the-beautiful-F-GEN  
             ‘the beautiful mermaid’  
             lit. the beautiful sea bride
- (56)      ʕaruus              l-baħr              (\*l-dzamiil-i)  
             bride.F              the-sea.M              (\*the-beautiful.M-GEN)  
             ‘the beautiful mermaid’  
             lit. the bride of the beautiful sea’

Examples (55) and (56) show that the left element is the only element that can be modified in non-compositional compounds. The fact that the compound *ʕaruus lbaħr* ‘mermaid’ is non-compositional makes the modification of the non-head completely impossible. The non-head of non-compositional compounds can be viewed as the concept of something, rather than the entity itself, i.e. we are talking about the concept of SEA, but not referring to an actual sea. Thus, it can never be modified. This means that *lbaħr* ‘the sea’ is used in the compound *ʕaruus lbaħr* ‘mermaid’ to denote salty water in which mermaids presumably live, not an actual sea.

Finally, as far as recursion is concerned, P-constructs are recursive in the sense that P-constructs can be used inside other P-constructs, as in examples (57) and (58):

- (57)      ʔaθaaθ              bayt              r-radʒul  
             furniture              house              the-man  
             ‘the furniture of the man’s house’
- (58)      ʔazraar              qamiis<sup>6</sup>              l-walad  
             buttons              shirt              the-boy  
             ‘the buttons of the boy’s shirt’

Examples (57) and (58) show that P-constructs in MSA and JA are recursive. However, their recursiveness seems to be limited to three words. This state of affairs is not unheard of; it applies to English as well. For example, English four-word phrases such as *the boy’s shirt’s buttons’ colour* are rarely used. In fact, both English and Arabic are similar in this regard.

Similarly, compositional compounds are recursive. This means that a compound can be

used inside another compound, as in (59-62):

- (59)      xuuḏat          raaʔid          l-fadʕaaʔ  
helmet          pioneer          the-space  
‘the space helmet’  
lit. the space pioneer helmet
- (60)      raʔiis          taḥriir          l-madʒallah  
chief          editing          the-magazine  
‘the editor-in-chief of the magazine’
- (61)      wuzaraaʔ      duwal          madʒlis          t-taʕaawun  
ministers      countries      council          the-cooperation  
‘the ministers of the countries of the (Gulf) Cooperation Council’
- (62)      ʔiḥtifaal      ziraaʕat      ʕadʒarat      (z)-zaytuun  
celebration    planting      tree            (the)-olive  
‘the celebration of planting (the) olive tree’

In examples (59-62), the whole constructs are definite because of the non-head (the last element). Simply put, the definite article *l-* ‘the’ cannot be marked on the remaining elements. In this respect, Arabic compounds are similar to English compounds. According to Biber and Gray (2011: 237), NNN sequences started to appear in English in the late nineteenth and early twentieth centuries. However, they are still relatively rare. The two researchers cited a number of examples they found in the ARCHER corpus, e.g. *army reorganisation scheme*, *home rule bill*, *river colony politics*, etc. In late twentieth century, four-noun sequences started to appear in the corpus, e.g. *life table survival curves*, *peak mean plasma concentration*, *plasma concentration time curve*, etc. (Biber and Gray 2011: 238).

Finally, compounds in MSA and JA can appear inside P-constructs, as follows:

- (63)      [qalam          [muʕallim      l-fiizyaaʔ]]  
pen          teacher          the-physics  
‘the physics teacher’s pen’

- (64) [qiiθaarat      [ʒaruus      l-baħr]]  
 guitar/lyres    bride            the-sea  
 ‘the mermaid’s lyres’

Examples (63) and (64) show that the compounds *muʒallim lfiizyaaʔ* ‘the physics teacher’ and *ʒaruus lbaħr* ‘the mermaid’ appear inside P-constructs, where the possessums *qalam* ‘pen’ and *qiiθaarat* ‘lyres’ belong to the possessors *muʒallim lfiizyaaʔ* ‘the physics teacher’ and *ʒaruus lbaħr* ‘the mermaid’, respectively.

In sum, modification as a criterion can be partially used to distinguish non-compositional compounds from P-constructs, since the non-head of the former cannot be modified by an adjective, whereas that of the latter can be. On the other hand, modification cannot make a distinction between P-constructs, on the one hand, and compositional compounds, on the other. Finally, recursion can only be used to distinguish between P-constructs and non-compositional compounds, since the latter cannot be recursive.

#### 4.3.2 Adjacency

The criterion of adjacency concerns the question of whether the two elements of the N + N constructs allow or resist the insertion of any intervening elements. In MSA and JA, it is unusual to insert any word, such as an adjective, between the first and second element in both P-constructs and compounds (compositional or non-compositional):

- (65) a.      ʒaamil            l-masʕnaʕ      l-kabiir            (P-construct)  
           worker.M        the-factory.M    the-big.M  
           ‘the big factory worker’  
           lit. the worker of the big factory
- b.      ʒaamil            (\*l-kabiir)      l-masʕnaʕ  
           worker.M        (\*the-big.M)    the-factory.M  
           ‘the big factory worker’  
           lit. the worker of the big factory

- (66) a.  $\zeta$ as'iir          t-tuffaaḥ          l-laḏiiḏ (compositional compound)  
 juice.M          the-apple.M          the-delicious.M  
 ‘the delicious apple juice’  
 lit. the juice of the delicious apple
- b.  $\zeta$ as'iir          (\*l-laḏiiḏ)          t-tuffaaḥ  
 juice.M          (\*the-delicious.M)          the-apple.M  
 ‘the delicious apple juice’  
 lit. the juice of the delicious apple
- (67) a. bayt          d-dawaaʔ          l-mufiid (non-compositional compound)  
 house          the-medicine          the-useful  
 ‘the useful pharmacy’  
 lit. the house of the useful medicine
- b. bayt          (\*l-mufiid)          d-dawaaʔ  
 house          (\*the-useful)          the-medicine  
 ‘the useful pharmacy’  
 lit. the house of the useful medicine

The above examples show that the two elements of both P-constructs and compounds are inseparable. On face value, adjacency as a criterion to distinguish P-constructs from compounds may therefore appear not to work, since neither construct allows any intervening elements to be inserted between the head and the non-head. However, in section 4.2.2, on referentiality, we saw that a demonstrative, e.g. *haaḏa* ‘this’, which refers to the entity denoted by the second element, can be inserted between the head and the non-head of P-constructs, as in:

- (68) saaḥ-at          haaḏa          r-radʒul          (P-construct)  
 watch-F          this.M          the-man.M  
 ‘this man’s watch’

- (69)        kitaab                    haaðih-i            t<sup>ʕ</sup>-t<sup>ʕ</sup>aalib-ah            (P-construct)  
               book.MSG                this-F                the-student-F  
               ‘the book of this student ’

These examples make clear that the elements of P-constructs are in fact separable. On the other hand, the demonstrative *haaða* cannot be inserted between the head and the non-head of compounds whether compositional or non-compositional. Possibly, this is because the second element of both compositional and non-compositional compounds is non-referential (section 4.2.2). The five examples below support this claim:

- (70)        muʕallim        (\*haaðih-i)        l-fiizyaaʔ        (compositional compound)  
               teacher        (\*this-F)        the-physics.F  
               lit. the teacher of (this) physics
- (71)        qaaʔid        (\*haaða)        l-dʒayf        (compositional compound)  
               leader        (\*this.M)        the-army.M  
               lit. the leader of (\*this) army
- (72)        waziir        (\*haaðih-i)        t-tanmiyy-ah        (compositional compound)  
               minister        (\*this-F)        the-development-F  
               lit. the minister of (\*this) development
- (73)        ʕaruus        (\*haaða)        l-baħr        (non-compositional compound)  
               bride        (\*this.M)        the-sea.M  
               ‘(\*this) mermaid’  
               lit. the bride of (\*this) sea
- (74)        qaat<sup>ʕ</sup>iʕ        (\*haaða)        t<sup>ʕ</sup>-t<sup>ʕ</sup>ariiq        (non-compositional compound)  
               crosser        (\*this.M)        the-road.M  
               ‘(\*this) bandit’  
               lit. the crosser of (\*this) road

Examples (70-74) show that the elements of both compositional and non-compositional compounds are indeed inseparable. Therefore, one may argue that adjacency can be used as

criterion to make a distinction between P-constructs, which accept the insertion of the demonstrative *haaḏa*, and compounds, which do not accept the insertion of the demonstrative *haaḏa*, except in the case of nouns which have a unique reference, as in (75) and (76):

(75)      ?afīfī-at      (?haaḏih-i)      ʃ-ʃams  
          rays-F            this-F            the-sun  
          ‘the rays of this sun’

(76)      dʕuuʔ            (?haaḏa)            l-qamar  
          light            this            the-moon  
          ‘the light of this moon’

Even though the demonstrative *haaḏa* ‘this’ can be inserted between the two elements of the compounds in (75) and (76), it is marked since the non-head has a unique reference, i.e. *ʃams* ‘the sun’ and *lqamar* ‘the moon’.<sup>30</sup> In other words, the addressee would be able to tell what referent the speaker has in mind, because there is only one sun and one moon, at least in our solar system. This means that adjacency is still valid.

Not only the demonstrative *haaḏa* ‘this’ can be used to show that two-element of P-constructs are penetrable, but also the quantifiers, i.e. *baʕdʕ* ‘some’ and *kull* ‘all’ can be inserted between the two elements, as in (77) and (78):

(77)      haqaaʔib      baʕdʕ            tʕ-tʕalabah            (P-construct)  
          bags            some            the-students  
          ‘the bags of some students’

(78)      malaabis      kull            tʕ-tʕalabah            (P-construct)  
          clothes            all            the-students  
          ‘the clothes of all students’

Conversely, the two elements of non-compositional compounds are impenetrable, whereas those of compositional compounds can be separable on the basis of the discourse context, as

---

<sup>30</sup> As a native speaker of Arabic, I have never used examples (75) and (76). Nonetheless, these examples do not sound wrong when I hear them. People rarely use the demonstrative *haaḏa* ‘this’ with unique references.



illustrated in (79-81):

- (79)       $\text{\textcircled{S}aruus}$       ( $*\text{ba}\text{\textcircled{S}}\text{d}^{\text{\textcircled{S}}}/\text{kull}$ )    l-bihaar      (non-compositional compound)  
 bride            ( $*\text{some/all}$ )      the-seas  
 lit. the bride of ( $*\text{some/all}$ ) sea( $*\text{s}$ )
- (80)       $\text{\textcircled{h}aafilat}$       ( $?\text{ba}\text{\textcircled{S}}\text{d}^{\text{\textcircled{S}}}/\text{kull}$ )    l-madaaris      (compositional compound)  
 bus              some/all            the-schools  
 ‘the bus of some/all schools’
- (81)       $\text{\textcircled{q}aa}\text{\textcircled{?}id}$       ( $?\text{ba}\text{\textcircled{S}}\text{d}^{\text{\textcircled{S}}}/\text{kull}$ )    l-dzuyuuḥ      (compositional compound)  
 leader            some/all            the-armies  
 ‘the commander-in-chief of some/all armies’

Example (79) shows that the two elements of non-compositional compounds are inseparable: the quantifiers *baṣd* ‘some’ and *kull* ‘all’ cannot intervene between them. In contrast, the two elements of compositional compounds in examples (80) and (81) can be separated by quantifiers in specific contexts. In other words, the compositional compound *qaa?id kull ldzuyuuḥ* ‘the commander-in-chief of all armies’ is more acceptable if followed by the adjective *lʿarabiyyah* ‘the Arab’, resulting in *qaa?id kull ldzuyuuḥ lʿarabiyyah* ‘the commander-in-chief of all Arab armies’, as in (82):

- (82)       $\text{\textcircled{q}aa}\text{\textcircled{?}id}$        $\text{\textcircled{k}ull}$     l-dzuyuuḥ      l-ʿarabiyyah  
 leader            all            the-armies      the-Arab  
 ‘the commander-in-chief of all Arab armies’

This is due to the fact that the non-head is made referential by the discourse context. It is unacceptable to refer to someone as the commander-in-chief of all armies without specifying which armies, as pointed out by Bauer *et al.* (2013: 464) see section 2.2.7.

### 4.3.3 Coordination

Coordination can be used as a criterion to differentiate between compounds and phrases in some languages, such as Hebrew. In MSA and JA, it can be argued that coordination cannot be

used to differentiate P-constructs from compositional compounds. In particular, the head/left element can be coordinated with another noun in both P-constructs and compositional compounds, as exemplified by (83-86)<sup>31</sup>:

- (83)      bayt              wa      sayyaarat      r-radʒul  
house              and      car              the-man  
‘the man’s house and car’              (P-construct)
- (84)      qalam              wa      mastʿarat      l-walad  
pen              and      ruler              the-boy  
‘the boy’s pen and ruler’              (P-construct)
- (85)      munassiq      wa      muʕallim      l-fiizyaaʔ  
coordinator      and      teacher              the-physics  
‘the physics coordinator and teacher’      (compositional compound)
- (86)      findʒaan      wa      ʔibriiq      ʃ-ʃaay  
cup              and      pot              the-tea  
‘the tea cup and pot’              (compositional compound)

Examples (83-86) show that the head of P-constructs, i.e. *bayt* ‘house’ in (83) and *qalam* ‘pen’ in (84), and the head of compositional compounds, i.e. *munassiq* ‘coordinator’ in (85) and *findʒaan* ‘cup’ in (86) can be coordinated using the conjunction *wa* ‘and’ with other nouns, without any syntactic or semantic problems. On the other hand, the head of non-compositional compounds cannot be coordinated with other nouns. This can be seen in the following examples:

- (87)      ʕarus              (\*wa      ʕariis)              l-baħr  
bride              (\*and      bridegroom)      the-sea  
‘the female (\*and male) merpeople’  
lit. the/a bride (\*and bridegroom) of the sea (non-compositional compound)

<sup>31</sup> Constructions of the kind in (83-86) are very common in MSA, but not found in Classical Arabic, and not considered correct by prescriptivists.

- (88) *faʕar* (\*wa dʒadaaʔil) l-banaat  
 hair (\*and braids) the-girls  
 ‘the candyfloss (\*and braids)’  
 lit. the hair (\*and braids) of the girls  
 (non-compositional compound)

Examples (87) and (88) show that the heads of non-compositional compounds, i.e. *ʕaruus* ‘bride’ in (87) and *faʕar* ‘hair’ in (88) cannot be coordinated with other nouns, without resulting in an impossible reading. Thus, examples (83-88) demonstrate that the possibility/impossibility of head coordination in N + N constructs can be used to distinguish between P-constructs and non-compositional compounds. However, coordination cannot differentiate between P-constructs and compositional compounds.

The non-head of both P-constructs and compositional compounds can also be coordinated with other nouns, as in (89-92):

- (89) bayt r-radʒul wa l-marʔa  
 house the-man and the-woman  
 ‘the house of the man and woman’ (P-construct)
- (90) qalam l-walad wa l-bint  
 pen the-boy and the-girl  
 ‘the boy’s and the girl’s pen’ (P-construct)
- (91) munassiq l-fiizyaaʔ wa l-kiimyaaʔ  
 coordinator the physics and the-chemistry  
 ‘the physics and chemistry coordinator’ (compositional compound)
- (92) ʕasʕiir t-tuffaaħ wa l-burtuqaal  
 juice the-apple and the-orange  
 ‘the apple and orange juice’ (compositional compound)

Examples (89-92) show that it is possible to coordinate the non-head of both P-constructs and compositional compounds with other nouns. However, such coordination is impossible with

non-compositional compounds:

- (93)       $\zeta$ aruus          l-baħr          (\*wa    n-nahr)  
               bride            the-sea        (\*and   the-river)  
               ‘the mermaid of the sea (\*and the river)’  
               lit ‘the bride of the sea (\*and river)’            (non-compositional compound)

- (94)       $\zeta$ ar            l-banaat        (\*wa    l- $\zeta$ awlaad)  
               hair            the-girls        (\*and   the-boys)  
               ‘the candyfloss (\*and the boys)’  
               lit. the hair of the girls’ (\*and the boys)        (non-compositional compound)

Examples (93) and (94) demonstrate that if the non-head of non-compositional compounds, i.e. *lbaħr* ‘the sea’ in (93) and *lbanaat* ‘the girls’ in (94), is coordinated, the compound loses its non-compositional reading. Hence, the possibility/impossibility of non-head coordination plays a role in distinguishing between P-constructs and non-compositional compounds. However, it makes no distinction between P-constructs and compositional compounds.

Additional evidence comes from Hebrew. Specifically, Borer (2009: 496) states that “while at least prescriptively the head of the construct may not be coordinated directly, the entire construct may be coordinated with the identical non-head realised as a pronoun on the second conjunct” as shown below:

- (95)      a.      beyt            ha-mora<sub>2</sub>        ve-xacer-a<sub>2</sub>        (P-construct)  
                   house            the-teacher<sub>2</sub>    and-yard-her<sub>2</sub>  
                   ‘the teacher’s house and her yard’
- b.      beyt            mora<sub>2</sub>            ve-xacer-a<sub>2</sub>        (P-construct)  
                   house            teacher<sub>2</sub>        and-yard-her<sub>2</sub>  
                   ‘a teacher’s house and her yard’
- (Borer 2009: 496)

The same fact is found in MSA and JA, as follows:

(96) bayt r-radʒul<sub>i</sub> wa sayyaratu-h<sub>i</sub> (P-construct)  
house the-man and car-his  
‘the house of the man and his car’

(97) qalam l-walad<sub>i</sub> wa mastʿarati-h<sub>i</sub> (P-construct)  
pen the-boy and ruler-his  
‘the boy’s pen and his ruler’

In examples (96) and (97), the coordinated nouns, i.e. *sayyaratuh* ‘his car’ in (96) and *mastʿaratih* ‘his ruler’ in (97) are marked with a resumptive pronoun referring back to the non-head, i.e. *rradʒul* ‘the man’ in (96) and *lwalad* ‘the boy’ in (97). In contrast, such coordination is impossible with both compositional and non-compositional compounds:

(98) munassiḡ l-fiizyaaʔ<sub>j</sub> (\*wa muʕallimu-ha<sub>j</sub>)  
coordinator.M the physics.F (\*and teacher-her)  
lit. the physics coordinator (\*and its teacher) (compositional compound)

(99) \*ʕasʕiir t-tuffaaḥ<sub>k</sub> (\*wa ʕadʒaratu-h<sub>k</sub>)  
juice.M the-apple.F (\*and tree-its)  
lit. the apple juice (\*and its tree) (compositional compound)

(100) \*ʕaruus l-baḥr<sub>i</sub> (\*wa nadʒmatu-h<sub>i</sub>)  
bride.F the-sea.M (\*and star-his)  
‘the mermaid of the sea (\*and its star)’  
lit. the bride of the sea (\*and its star) (non-compositional compound)

(101) \*ʕaʕar l-banaat<sub>e</sub> (\*wa dʒadaaʔili-hum<sub>e</sub>)  
hair the-girls (\*and braids-their)  
‘the candyfloss (\*and braids)’  
lit. the hair of the girls (\*and their braids) (non-compositional compound)

Examples (98-101) show that the non-head of both compositional and non-compositional compounds cannot be coordinated with another noun marked with a resumptive pronoun referring back to the non-head (cf. Borer 2009: 496-97). This is due to the fact that the non-

head of compounds is non-referential. Therefore, this type of coordination, which I will refer to as ‘resumptive coordination’, can be used to distinguish between P-constructs, on the one hand, and compositional and non-compositional compounds, on the other. Conversely, the other types of coordination can only differentiate between P-constructs and non-compositional compounds.

#### 4.3.4 Replacement of the second element by a pro-form

According to Bauer (1998a: 76-77), it is unusual to replace the head of a compound with a pro-form, whereas this works in phrases. This criterion has been suggested to distinguish compounds from phrases. For instance, in English, a *green one* can refer to my grandfather’s house which is painted green, whereas a *green one* cannot refer to a building made of glass used for growing plants. However, as mentioned in chapter two (section 2.2.9), this criterion is invalid, at least in English. In MSA and JA, this criterion cannot be applied, because there are no pro-forms that can replace nominal elements.

#### 4.3.5 Ellipsis

In English, one of the elements of a phrase can undergo verb replacement (cf. Fàbregas and Scalise 2012: 120), but not any of the internal elements of the compound. An example from English is:

(102) He drives a taxi and he does it every day.

This construction is possible, but barely acceptable in both MSA and JA, as in (103) and (104):

(103) ?huwa yuʕallim l-fiizyaa? wa yafʕal ʕaalika kulla yawm  
 he teaches the-physics and he.does that every day  
 ‘He teaches physics and he does it every day.’

(104) ?huwa yudiir ʃ-ʕarikah wa yafʕal ʕaalika kulla yawm  
 he manages the-company and he.does that every day  
 ‘He manages the company and he does it every day.’

Comparing examples (103) and (104) with the compositional compound and P-construct in

examples (105) throughout (106), it seems that just like in English, e.g. *\*he is a taxi driver and he does it every day*, VP replacement is not allowed in either type of construct:

(105) huwa muƣallim l-fiizyaa? (\*wa yafƣal ɗaalika kulla yawm)  
 he teacher the-physics (\*and he.does that every day)  
 lit. he is a physics teacher (\*and he does it every day).

(106) huwa mudiir ƣ-farikah (\*wa yafƣal ɗaalika kulla yawm)  
 he manager the-company (\*and he.does that every day)  
 lit. he is the company manager (\*and he does it every day).

This may indicate that VP replacement fails to make a distinction between P-constructs and compounds, since it is not allowed in either type of construct. On the other hand, verb replacement works in other cases of compositional compounds, in which two compounds are coordinated and the head of the first compound is ellipped, especially if the non-head is coordinated with another noun (cf. phrasal compounds) such as:

(107) muƣallim-uu l-fiizyaa? wa l-kiimyaa? (compositional compound)  
 teacher-PL the physics and the-chemistry  
 ‘the physics and chemistry teachers’

The compositional compound in example (107) is likely to be interpreted as *physics (teachers) and chemistry teachers*, where the word *muƣallimuu* ‘teachers’ is ellipped. This compound is equivalent to the English compound *physics and chemistry teachers*. In addition, as mentioned before in section 2.2.8, the non-head of P-constructs can also be coordinated:

(108) mudiir-u ƣ-farikah wa l-muʔassasah (P-construct)  
 manager-PL the-company and the-institution  
 ‘the company and institution managers’

Example (108) is likely to be interpreted as *the company (managers) and institution managers*. Again, one may suggest that this type of ellipsis makes no distinction between P-constructs and compounds, as it works in both types. All in all, ellipsis as a criterion fails to differentiate between P-constructs and compounds.

### 4.3.6 Inflection and linking elements

Inflection can be used to distinguish compounds from phrases in some languages that exhibit inflectional markings. In MSA and JA, the head can be freely pluralised in both P-constructs and compounds, whether compositional or non-compositional, as in:

- |       |   |                           |                              |
|-------|---|---------------------------|------------------------------|
| (109) | buyuut<br>house.PL<br>'the man's houses'            | r-radʒul<br>the-man       | (P-construct)                |
| (110) | qalaaʔid<br>necklace.PL<br>'the woman's necklaces'  | l-marʔah<br>the-woman     | (P-construct)                |
| (111) | muʕallim-uu<br>teacher-PL<br>'the physics teachers' | l-fiizyaaʔ<br>the-physics | (compositional compound)     |
| (112) | bana-at<br>daughter-PL<br>'the disasters'           | d-dahr<br>the-time        | (non-compositional compound) |

While the non-head of P-constructs can be freely pluralised, such pluralisation is impossible in compounds, as in:

- |       |    |                                    |                     |               |
|-------|----|------------------------------------|---------------------|---------------|
| (113) | a) | bayt<br>house<br>'the man's house' | r-radʒul<br>the-man | (P-construct) |
|-------|----|------------------------------------|---------------------|---------------|



- b)      bayt              r-ridzaal  
             house            the-man.PL  
             ‘the men’s house’
- (114)    a)      dzawaaz              s-safar              (compositional compound)  
             permission/passport    the-travel  
             ‘the passport’
- b)      \*dzawaaz              l-ʔasfaar  
             permission/passport    the-travel.PL  
             ‘the passports’  
             lit. the permission of travels
- (115)    a)      ḥajar              z-zaawiyah      (semi-non-compositional compound)  
             stone              the-corner  
             ‘the foundation’
- b)      \*ḥajar              z-zawaayaa  
             stone              the-corner.PL  
             ‘the foundations’  
             lit. the stone of the corners

Examples (113-115) demonstrate that the possibility/impossibility of free pluralisation of the non-head can be used as a criterion to differentiate between P-constructs, in which such pluralisation is possible, and compounds, in which such pluralisation is impossible. Examples (114b) and (115b) are incorrect, because the second element is non-referential and the pluralisation does not contribute to the meaning of the whole compounds (compare with the English compound *sale(s) slip*). Therefore, it cannot be pluralised.

Note, however, that there are some compounds in which the plural marker appears on the non-head. This is because such compounds were originally formed with the non-head carrying the plural marker. In fact, if the non-head in such a combination is made singular, the compound becomes ungrammatical:

- (116)
- |    |                                     |                  |
|----|-------------------------------------|------------------|
| a) | *raʔiis                             | l-waziir         |
|    | president.MSG                       | the-minister.MSG |
|    | ‘the prime minister’                |                  |
|    | lit. the president of the minister  |                  |
| b) | raʔiis                              | l-wuzaraaʔ       |
|    | president.SG                        | the-minister.MPL |
|    | ‘the prime minister’                |                  |
| c) | ruʔasaaʔ                            | l-wuzaraaʔ       |
|    | president.MPL                       | the-minister.MPL |
|    | ‘the prime ministers’               |                  |
| d) | *ruʔasaaʔ                           | l-waziir         |
|    | president.MPL                       | the-minister.MSG |
|    | ‘the prime ministers’               |                  |
|    | lit. the presidents of the minister |                  |

Example (116) demonstrates that the non-head of the compound *raʔiis lwuzaraaʔ* ‘prime minister’ must be plural; otherwise, the compound will be ungrammatical. It should be pointed out that some non-compositional compounds can be found in different forms, where both the head and the non-head are also marked with the plural marker, as in example (117):

- (117)
- |    |                                 |             |                              |
|----|---------------------------------|-------------|------------------------------|
| a) | qaatʕiʕ                         | tʕ-tʕariiq  | (non-compositional compound) |
|    | crosser.SG                      | the-road.SG |                              |
|    | ‘the bandit’                    |             |                              |
|    | lit. the crosser of the road    |             |                              |
| b) | qaatʕiʕ                         | tʕ-tʕuruq   |                              |
|    | crosser.SG                      | the-road.PL |                              |
|    | ‘the bandit’                    |             |                              |
|    | lit. ‘the crosser of the roads’ |             |                              |

- c)      qut<sup>ʕ</sup>t<sup>ʕ</sup>aaʕ                      t<sup>ʕ</sup>-t<sup>ʕ</sup>ariiq  
             crosser.PL                      the-road.SG  
             ‘the bandits’  
             lit. ‘the crossers of the road’
- d)      qut<sup>ʕ</sup>t<sup>ʕ</sup>aaʕ                      t<sup>ʕ</sup>-t<sup>ʕ</sup>uruq  
             crosser.PL                      the-road.PL  
             ‘the bandits’  
             lit. ‘the crossers of the roads’

When the non-head of the non-compositional compound *qaat<sup>ʕ</sup>iʕ t<sup>ʕ</sup>t<sup>ʕ</sup>ariiq* ‘the bandit’ in (117) is pluralised, the meaning of the whole compound is not affected. In other words, the pluralisation of the non-head in (117) is only an empty morphological marking that makes no difference to the meaning. The meaning of the compound will still denote *bandit* or *bandits* on the basis of the head, regardless of the pluralisation of the non-head. Example (117) supports Katamba’s (1993: 317) argument, which suggests that compounds are pluralised by adding the plural suffix *-s* to the right element/the head, thus, yielding *arms races*, *sales slips*, *buildings inspectors* and *weapons analyses*. Semantically, *race*, *slip*, *inspector* and *analysis* are the heads. In (117), the plurality of the whole compound *qaat<sup>ʕ</sup>iʕ t<sup>ʕ</sup>t<sup>ʕ</sup>uruq* ‘the bandit’ is determined based on the plurality of the left element/the head *qut<sup>ʕ</sup>t<sup>ʕ</sup>aaʕ* ‘crossers’. Thus, the compound will be plural only if the head is pluralised, rather than the non-head. However, one case in which the pluralisation of the non-head influences the meaning of the compound is cited below:

- (118)      a)      ʔibrat                      l-muxaddir  
                   needle                      the-drug  
                   ‘the anaesthesia needle’
- b)      ʔibrat                      l-muxaddiraat  
                   needle                      the-drugs  
                   ‘the drugs needle’

Example (118) demonstrates that the pluralisation of the non-head has an impact on the meaning of the compound. In example (118a), when the non-head, i.e. *lmuxaddir* ‘the drug’ is singular, it refers to ‘anaesthesia’. However, when the non-head *lmuxaddir* ‘drug’ is pluralised,

i.e. *Imuxaddiraat* ‘the drugs’ in (118b), the meaning of the word changes into ‘illegal substances’. Nevertheless, such cases where the pluralisation of the non-head changes the meaning of the compound are rare.

After applying the general criteria used to distinguish compounds from P-constructs cross-linguistically, in the next two sections, I discuss two language-specific criteria that can be used to distinguish P-constructs from compounds in MSA and JA.

#### 4.3.7 Definiteness of the first element

The first criterion is related to definiteness in N + N combinations in MSA and JA. Specifically, when the first element of a P-construct is marked with the definite article *l-*, a possessive maker *li-* ‘of/for’, which is equivalent to the English possessive maker *-’s*, has to be attached to the second element of the P-construct in MSA. This is illustrated below:

(119)	l-bayt-u the-house-NOM ‘the woman’s house’	li-l-marʔa of-the-woman	(MSA P-construct)
-------	--	----------------------------	-------------------

(120)	l-qalam-u the-pen-NOM ‘the teacher’s pen’	li-l-muʕallim of-the-teacher	(MSA P-construct)
-------	---	---------------------------------	-------------------

Examples (119) and (120) show that when the first element of the P-construct in MSA is marked with the definite article *l-*, the possessive marker *li-* ‘of/for’ appears on the second element, i.e. *lmarʔa* ‘the woman’ and *lmuʕallim* ‘the teacher’. In contrast, even when the first element is marked with the definite article *l-* ‘the’ in either compositional or non-compositional compounds, the possessive marker *li-* cannot appear:

(121)	*l-raaʔid the-pioneer ‘the astronaut’ lit. the pioneer for space	li-l-fadʕaaʔ-i for-the-space-GEN	(MSA non-compositional compound)
-------	---	-------------------------------------	----------------------------------

- (122) \*l-ṣaruus                    li-l-baħr  
 bride.FSG                    for-the-sea.MSG  
 ‘mermaid’  
 lit. bride for the sea                    (MSA non-compositional compound)
- (123) \*r-raʔiis                    li-l-wuzaraaʔ-i  
 the-president                    for-the-minister-GEN  
 ‘prime minister’                    (MSA compositional compound)
- (124) \*l-muṣallim                    li-l-fiizyaaʔ-i  
 the-teacher                    for-the-physics-GEN  
 ‘physics teacher’                    (MSA compositional compound)

Examples (121-124) show that when the first element of a non-compositional compound is marked with the definite article *l-*, the possessive marker *li-* ‘of/for’ cannot be attached to the second element, i.e. *lfaḍʿaaʔ* ‘the space’ and *lbaħr* ‘the sea’. The same holds true for compositional compounds. That is, the possessive marker *li-* is not allowed to be attached to the second element when the first element of the compound is marked with the definite article. This means that the definiteness of the first element, accompanied by the attachment of the possessive marker to the second element as a criterion can distinguish between P-constructs and compounds, whether compositional or non-compositional, in MSA.

In JA, the situation is quite different. The possessive marker used in JA is not the same as in MSA. In particular, instead of *li-*, the word *tabaṣ*<sup>32</sup> ‘of/for’ acts as a possessive marker in JA. Regarding P-constructs, the possessive marker *tabaṣ* ‘of/for’ can be inserted between the two elements of P-construct in JA, as shown in the following examples:

---

<sup>32</sup> Note that the possessive marker *tabaṣ* ‘for’ has undergone semantic bleaching at one point in time. This process has caused it to lose its semantic content, i.e. ‘follower/property of’ and act as a functional or grammatical word instead, i.e. a possessive marker.

- (125)      s-siyyaarah                  tabʕat                  z-zalameh      (P-construct)  
                  the-car.FSG                  for.FSG                  the-man.MSG  
                  ‘the man’s car’
- (126)      l-ʔalʕaab                  tabʕaat                  l-walad              (P-construct)  
                  the-toys.FPL                  for.FPL                  the-boy.MSG  
                  ‘the boy’s toys’

Examples (125) and (126) demonstrate that the possessive marker *tabaʕ* ‘for’ can be inserted between the two elements of a P-construct in JA. As shown in example (125), *tabʕat* ‘for’ agrees with the first element *ssiyyaarah* ‘car’ in gender and number. The second element *z-zalameh* ‘the man’ is masculine, whereas the first element *ssiyyaarah* ‘car’ is feminine, resulting in the feminine *tabʕat* not the masculine form *tabaʕ*. The same applies to example (126), in which the possessive marker agrees with the first element in gender and number. This is different from the possessive marker *li-* in MSA, which has a default form, meaning that it does not agree with other elements in the construct. Another difference between *li-* and *tabaʕ* is that the former is bound, while the latter can stand on its own (compare 121-124 and 125-126).

Insertion of the possessive marker *tabaʕ* ‘for’ is possible in compositional compounds, but it cannot appear in non-compositional compounds in JA. The following examples illustrate this point:

- (127)      l-findzaan                  tabaʕ                  l-gahwih  
                  the-cup                          for                          the-coffee  
                  ‘the coffee cup’    (compositional compound)
- (128)      l-miʕallim                  tabaʕ                  l-fiizya  
                  the-teacher                          for                          the-physics  
                  ‘the physics teacher’    (compositional compound)

- |       |  |               |                      |                              |
|-------|--|---------------|----------------------|------------------------------|
| (129) | *l-ṣaruus<br>bride.FSG<br>'the mermaid'<br>lit. bride for the sea        | tabṣat<br>for | l-baḥir<br>the-sea   | (non-compositional compound) |
| (130) | *ʃ-ʃaṣar<br>the-hair<br>'the candy floss'<br>lit. the hair for the girls | tabaṣ<br>for  | l-banaat<br>the-girl | (non-compositional compound) |

Examples (127) and (128) show that it is possible to insert the possessive marker *tabaṣ* 'for' between the two elements of compositional compounds in JA. However, it is impossible to do so in non-compositional compounds, as shown in examples (129) and (130). Here, I would argue that the presence of the possessive marker *tabaṣ* 'for' in compositional compounds in JA does not change the fact that they are compounds. First of all, the meaning of *tabaṣ* 'for/of' in examples (127) and (128) does not denote possession. For instance, it is impossible to say that *lfindzaan* 'the cup' belongs to *lgahwih* 'the coffee' and *lmṣallim* 'the teacher' belongs to *lfiizya* 'the physics'. Secondly, other criteria, i.e. adjacency, referentiality, free pluralisation of the non-head and resumptive coordination confirm that compositional compounds are more like compounds than P-constructs. Finally, I would suggest that *tabaṣ* 'for/of' in compositional functions as a linking element, since it is semantically empty.

In sum, it seems that definiteness marked on the first element accompanied by the attachment of the possessive marker on the second element helps, as a criterion, in distinguishing between P-constructs and compounds in MSA. However, this criterion can partially distinguish between P-constructs and compounds in JA. In particular, it can only differentiate between P-constructs and non-compositional compounds due to the presence of the linking element *tabaṣ* 'for/of'.

#### 4.3.8 Cardinal numbers before the first element

The second language-specific criterion deals with the effect that cardinal numbers can have when they are added before the first/left element of P-constructs and compounds in MSA and JA; they trigger some changes to the second/right element of P-constructs and compounds. In P-constructs, the possessive marker *li-* 'for/of' has to be attached to the second element,

whereas in compounds, the possessive marker *li-* ‘for/of’ cannot be added and if there is a definite article marked on the second/right element, it disappears. Note that, in MSA, if the noun preceding cardinal numbers from 3-10 is marked with masculine gender, the cardinal numbers should be marked with the opposite gender, i.e. feminine and vice versa (see the gender markings on the number and the following noun). This is illustrated in the following examples from MSA (131-134):

- |       |    |   |                                      |                             |
|-------|----|---|--------------------------------------|-----------------------------|
| (131) | a. | buyuut<br>houses.M<br>‘the man’s houses’                | r-radʒul<br>the-man                  | (P-construct)               |
|       | b. | (*θalaaθ-at)<br>(*three-F)<br>‘the man’s three houses’  | buyuut<br>houses.M<br>the-man        | r-radʒul<br>the-man         |
|       | c. | θalaaθ-at<br>three-F<br>‘three houses for the man’      | buyuut<br>houses.M<br>for-the-man    | li-r-radʒul<br>for-the-man  |
| (132) | a. | markabaat <sup>33</sup><br>cars.F<br>‘the woman’s cars’ | l-marʔa<br>the-woman                 | (P-construct)               |
|       | b. | (*xams)<br>(*five.M)<br>‘the woman’s five cars’         | markabaat<br>cars.F<br>the-woman     | l-marʔa<br>the-woman        |
|       | c. | xams<br>five.M<br>‘five cars for the woman’             | markabaat<br>cars.F<br>for-the-woman | li-l-marʔa<br>for-the-woman |

Examples (131) and (132) show that when the cardinal numbers, i.e. *θalaaθat* ‘three’ in (131)

---

<sup>33</sup> The equivalent of the MSA lexical item *markabah* ‘car’ is *sayyarah* ‘car’ in JA.



and *xams* ‘five’ in (132) are added before the first/left element of the P-constructs, the possessive marker *li-* ‘for/of’ is attached to the second element. However, in examples (133) and (134) below, the addition of a cardinal number before the first element of compounds does not trigger the appearance of the possessive marker *li-*. Furthermore, this addition makes the definite article marked on the second element disappear.

- (133) a. fanaadziin l-qahwa (compound)  
 cups.M the-coffee  
 ‘the cups of coffee’
- b. ?arbaʕ-at fanaadziin (\*l-)qahwa  
 four-F cups.M (\*the-)coffee  
 ‘four cups of coffee’
- c. ?arbaʕ-at fanaadziin (\*li-l-)qahwa  
 four-F cups.M (\*for-the-)coffee  
 ‘four cups for coffee’
- d. ?arbaʕ-at fanaadziin qahwa  
 four-F cups.M coffee  
 ‘four cups of coffee’
- (134) a. raaʔid-i l-fadʕaaʔ (compound)  
 pioneers.M-GEN the-space  
 ‘the astronauts’  
 lit. the pioneers of space
- b. θalaaθ-at raaʔid-i (\*l-)fadʕaaʔ  
 three-F pioneers.M-GEN (\*the-)space  
 ‘three astronauts’  
 lit. three pioneers of space

- c.    θalaaθ-at      raaʔid-ii                    (\*li-l)-fadʕaaʔ  
       three-F        pioneer.M-PL.GEN   (\*for-the)-space  
       ‘three astronauts’  
       lit. three pioneers for space
- d.    θalaaθ-at      raaʔid-ii                    fadʕaaʔ  
       three-F        pioneer.M-PL.GEN   space  
       ‘three astronauts’  
       lit. three pioneers of space

Examples (133d) and (134d) are correct, since the possessive marker *li-* ‘for/of’ does not appear on the second element when the whole compound is preceded by a cardinal number. The same phenomenon is observed in JA with two modifications: the possessive marker *li-* ‘for/of’ is realised as *la-*, and the gender of numbers in JA do not follow the same rules in MSA. Specifically, the gender of cardinal number in JA is always masculine, especially in N + N combinations. The following examples represent data from JA:

- (135)    xams            banaatʕiil      la-l-walad                    (P-construct)  
           five.M        trousers.M      for-the-boy  
           ‘five trousers for the boy’
- (136)    sabʕ            ʃunatʕ          la-l-binit                    (P-construct)  
           seven.M      bags.F          for-the-girl  
           ‘seven bags for the girl’
- (137)    θalaθ            ʔasaatiðit      taariix                      (compound)  
           three.M      teachers.M      history  
           ‘three teachers of history’
- (138)    tisʕ            byuut            daradz                      (compound)  
           nine.M       houses.M       stairs  
           ‘nine stairwells’  
           lit. nine houses stairs

Interestingly, JA utilises two possessive markers in P-constructs: *la-* ‘for/of’, which appears in the case of cardinal numbers and *tabaʕ* ‘for’ which appears elsewhere (see examples 125-126).

#### **4.3.9 Summary**

In sum, there are some similarities and some differences between P-construct and compounds in MSA and JA on the basis of the criteria discussed above. The similarities and differences are summarised in Table 4.2.

Table 4.2. Similarities and differences between P-constructs and compounds in MSA and JA

Syntactic criteria	Compounds		P-constructs (possessive phrases)
	Compositional	Non-compositional	
Left element modification	Yes	Yes	Yes
Right element modification	Yes	No	Yes
Recursion	Yes	No	Yes
Adjacency	No	No	Yes
Coordination of the head with another noun	Yes	Yes	Yes
Coordination of the non-head with another noun	Yes	No	Yes
Resumptive coordination	No	No	Yes
Replacement of the second element by a pro-form	N/A	N/A	N/A
Ellipsis	Yes	No	Yes
Free pluralisation of the left element	Yes	Yes	Yes
Free pluralisation of right element	No	No	Yes
Definiteness of the first element, accompanied by the appearance of possessive marker <i>li-/tabaʕ</i> ‘for/of’	No (MSA)/ Yes (JA)	No	Yes
Cardinal numbers before the first element, accompanied by the appearance of possessive marker <i>li-/la</i> ‘for/of’	No	No	Yes

With regard to the definiteness criterion of the first element, it is worth pointing out that there is a noticeable difference between MSA and JA, that is the possessive marker *tabaʕ* ‘for/of’

appears in compositional compounds in JA, whereas its equivalent in MSA *li-* ‘for/of’ does not appear in either compositional or non-compositional compounds.

#### 4.4 Conclusion

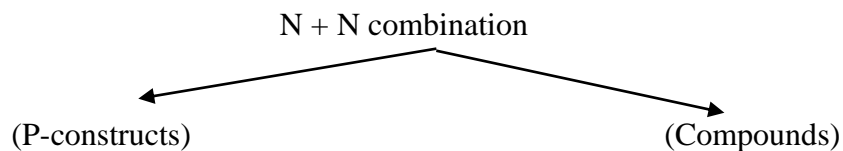
In this chapter, the semantic and syntactic criteria used to distinguish between compounds and phrases cross-linguistically, as discussed in chapter 2, have been applied to MSA and JA to distinguish between two types of N + N constructs, namely, P-constructs and compounds. With respect to the semantic criteria, compositionality can only identify non-compositional compounds. However, it cannot differentiate between P-constructs, on the one hand, and compositional compounds, on the other. Therefore, it is partially applicable. The most reliable criterion so far to distinguish between P-constructs and compounds is referentiality. It has been indicated that the non-head of P-constructs is referential, whereas the non-head of compounds (both compositional and non-compositional) is normally non-referential (except some non-heads that can have unique reference or are proper nouns).

Syntactically, adjacency has been found reliable in distinguishing between P-constructs and compounds. That is, an intervening element such as the demonstrative *haaḏa* ‘this’ and the quantifiers *baʕd* ‘some’ cannot be inserted between the head and the non-head of compounds whether compositional or non-compositional, whilst such insertion is allowed in P-constructs. Note, however, that the demonstrative *haaḏa* can be inserted between the two elements of the compound if the non-head has a unique reference. Concerning modification, it has been argued that it is partially applicable. That is, it can be used to distinguish non-compositional compounds from P-constructs, but it cannot make a distinction between P-constructs and compositional compounds. Regarding coordination, which is only partially applicable, the non-head of both P-constructs and compositional compounds can be coordinated, but such a process is not allowed in non-compositional compounds. Nevertheless, one type of coordination (in which the whole compound can be coordinated with another noun marked with a resumptive pronoun referring back to the non-head) can be used to distinguish between P-constructs and both compositional and non-compositional compounds. The possibility/impossibility of free pluralisation of the non-head has been found to be a good criterion, except for some few examples of compounds that have plural non-heads. However, the plurality of these exceptions do not have any semantic effect.

In addition to the general criteria used to distinguish P-constructs from compounds, I have suggested two language-specific criteria that are exclusive to Arabic, or potentially

Semitic languages in general. The first criterion is the definiteness of the first element; for this criterion, it has been noted that when the first element of a P-construct is marked with the definite article, the possessive marker *li-* ‘of/for’ has to be marked on the non-head in MSA. However, when the first element of either compositional or non-compositional compounds is marked with the definite article, the possessive marker *li-* does not appear in MSA. Unlike MSA, the possessive marker *tabaʕ* ‘for/of’ appears in compositional compounds in JA where I argued that it functions as LE. The second criterion is the appearance of the possessive marker *li-/la* ‘for/of’ when the first element is preceded by cardinal numbers. In P-constructs, when a cardinal number appears before the first element, it triggers the appearance of the possessive marker *li-/la* ‘for/of’. On the other hand, this does not apply to either compositional or non-compositional compounds in MSA and JA.

The typology in (139) can be proposed for the main differences between P-constructs and compounds in MSA and JA:



- Always compositional
- The non-head is referential
- The elements are separable
- Resumptive coordination (applicable)
- Free-pluralisation of right element
- Possessive marker appears when 1st N is definite
- Possessive marker appears when cardinal precedes 1st N

- Compositional and non-compositional
- The non-head is normally non-referential
- The elements are inseparable
- Resumptive coordination (N/A)
- Restricted-pluralization of right element
- Possessive marker does not appear when 1st N is definite
- Possessive marker does not appear when cardinal precedes 1st N

## Chapter Five: Identifying compounding in Arabic: Adj + N combinations

### 5.1 Introduction

Now that N + N compounds within SGCs in MSA and JA have been identified and differentiated from other nominal constructs, we turn to the analysis of the other types of compounds. As I noted in chapters 3 and 4, most compounds in MSA and JA consist of nouns, though other word classes including adjectives can be found. Any analysis will vary depending on the criteria for classification being examined. For instance, compounds can be classified according to the internal word classes of their elements, which include Adj + N combinations, Adj + Adj combinations and V + V combinations. In this chapter, I explore Adj + N combinations. Specifically, the general criteria discussed in the previous chapters to distinguish between N + N P-constructs and compounds are applied to Adj + N combinations in MSA to determine whether they are compounds or not.<sup>34</sup>

The chapter proceeds as follows: section 2 discusses the Adj + N combination, describing its structure and the internal word classes within. It also applies the cross linguistic criteria of compoundhood to Adj + N combinations to determine whether they are compounds or P-constructs. Section 3 provides a description of the adjectivehood criteria used to identify adjectives cross-linguistically in order to determine the syntactic category of the entire Adj + N combination. Section 4 determines whether Adj + N combinations are adjectives or nouns on the basis of the adjectivehood criteria discussed in section 3. It also proposes a language-specific criterion, involving adjacency and the order of elements within Adj + N combinations. Section 5 discusses certain types of Adj + N combinations within SGCs, claimed to be compounds in the relevant literature. Finally, section 6 concludes the chapter.

### 5.2 Adjective + Noun combinations: types and properties

#### 5.2.1 Overview

Compounds containing words that are not nouns can be found in many languages, e.g. English, Dutch, German, French, Spanish, Italian, etc. (Scalise and Bisetto 2009). In English, compound adjectives of the form Adj + Adj, e.g. *bitter-sweet*, Adj + N, e.g. *greenhouse* and N + Adj, e.g. *girl crazy* have been identified and investigated (Scalise and Bisetto 2009: 39). In MSA too,

---

<sup>34</sup> Adj + N combinations do not appear in JA. Other structures can be used to compensate for the loss of Adj + N combinations, and these structures will be discussed in section 5.2.3.10.

several Adj + N combinations are found. Examples of this type of combination are given in (1-4):

- |     |   |  |
|-----|---|--|
| (1) | [haadd <sub>ADJ</sub> ]<br>sharp<br>'a sharp-eyed person'     | [l-bas <sup>ʕ</sup> ar <sub>N</sub> ]<br>the-sight               |
| (2) | [laḏiiḏ <sub>ADJ</sub> ]<br>delicious<br>'the delicious food' | [t <sup>ʕ</sup> -t <sup>ʕ</sup> aʕaam <sub>N</sub> ]<br>the-food |
| (3) | [xafiif <sub>ADJ</sub> ]<br>light<br>'a funny person'         | [ḏ <sup>ʕ</sup> -ḏ <sup>ʕ</sup> ill <sub>N</sub> ]<br>the-shadow |
| (4) | [qawiyy <sub>ADJ</sub> ]<br>strong<br>'a brave person'        | [l-qalb <sub>N</sub> ]<br>the-heart                              |

Examples (1-4) have as their first element an adjective, i.e. *ḥaadd* 'sharp', *laḏiiḏ* 'delicious', *xafiif* 'light' and *qawiyy* 'strong', whereas the second element is a noun, i.e. *lbas<sup>ʕ</sup>ar* 'the sight', *t<sup>ʕ</sup>t<sup>ʕ</sup>aʕaam* 'the food', *ḏ<sup>ʕ</sup>ḏ<sup>ʕ</sup>ill* 'the shadow' and *lqalb* 'the heart'. The syntactic category of the output will be discussed in detail in section 5.4.

A number of scholars (e.g. Ryding 2005; Fassi-Fehri 1999, 2012; Al Mahmoud 2014) have discussed these combinations in MSA but without taking into account that they could be of different types, corresponding to the difference between the two possible constructs within SGCs, i.e. P-constructs and compounds. According to Ryding (2005: 221), Adj + N combinations are always phrases:

Sometimes an adjective or a participle with adjectival meaning will appear as the first term of a construct phrase instead of following the noun as a modifier. In these phrases the adjective remains in the masculine gender, but it may be singular or plural. These expressions are often set phrases and tend to be used with particular adjectives.



Ryding refers to the following Adj + N combinations as phrases:

(5)           muxtalif-u           l-mudun-i  
           various.M-NOM       the-cities.F-GEN  
           ‘the various cities’

(6)           qadiim-u           z-zamaan-i  
           old.M- NOM       the-time.M-GEN  
           ‘the olden times’

Ryding (2005: 221)

According to Ryding, in examples (5) and (6), *muxtalif lmudun* ‘various cities’ and *qadiim zzamaan* ‘the olden times’ are phrases, where *muxtalif* ‘various’ and *qadiim* ‘old’ cannot be marked for gender, i.e. \**muxtalifah* ‘various (feminine)’ and \**qadiimah* ‘old (feminine)’ even if they are followed by a noun marked for feminine gender, as in example (5). In other words, the default gender of the first element of Adj + N combinations is masculine. Similarly, Fassi-Fehri (1999: 115) discusses a type of nominal state construct known as the synthetic genitive in which a prenominal adjective, the head, is followed by a noun.<sup>35</sup> He refers to this type of Adj + N combination as the prenominal adjectival state construct (phrase). Examples can be seen in (7-9), adapted from Fassi-Fehri (1999: 115-116):

(7)           ʔakal-tu       laðiið-a           tʰ-tʰaʕaam-i  
           ate-I       delicious-ACC       the-food-GEN  
           ‘I ate the delicious food.’

(8)           ʔaqraʔ-u       dzadiid-a       l-kutub-i  
           read-I       new-ACC       the-books-GEN  
           ‘I read the new books.’

---

<sup>35</sup> Even though synthetic genitives behave like nominals rather than adjectives, Fassi-Fehri (1999: 115) argues that the adjective is the head, since it receives ‘external structural case’. This issue is discussed in detail in chapter 7.

- (9)       ʔbaħaθ-tu      ʕani   l-dzamiil-i                   l-wadʒh-i  
           looked-I       for    the-nice-GEN               the-face-GEN  
           ‘I looked for the one with a nice face.’

Fassi-Fehri (1999: 115-116) clearly considers examples (7-9) phrases. However, as in the case of N + N combinations discussed in chapters 3 and 4, there may well be two possible types of construct within what he refers to as the prenominal adjectival state construct. Examples (7) and (8) indeed seem to be different from example (9), at least in meaning. Examples (7) and (8) convey a phrasal meaning where the second elements *tʔtʕaʕaami* ‘the food’ and *lkutub* ‘the books’ are modified by the first elements *laðiið* ‘delicious’ and *dzadiid* ‘new’. On the other hand, example (9) means someone who has a nice face. On the basis of this difference, we might suspect that, like N + N combinations, Adj + N combinations too can be either compounds or phrases; as in the earlier discussion of N + N combinations. I will therefore posit a preliminary distinction between Adj + N phrases and Adj + N compounds; whether such a distinction is indeed tenable will be examined in detail in section 5.2.3.

Similar to Fassi-Fehri (1999), Al Mahmoud (2014: 2) posits that adjectives can be found prenominally, where they are neither marked for number, gender, case nor definiteness to agree with the noun they supposedly modify, without distinguishing between the two possible types of Adj + N combination. He illustrates with the following example:

- (10)       (ʔl)- dzamiil-u           l-wadʒh-i  
           def-pretty-nom        def-face-gen  
           ‘(the one with) the pretty face’

(Al Mahmoud 2014: 2)

Al Mahmoud (2014: 2) indicates that, in example (10), the definite article marked on (*al*) *dzamiilu* ‘(the) pretty’ is optional. He also suggests that it does not express definiteness agreement with the noun *alwadʒhi* ‘the face’. Al Mahmoud (2014) notes that this type of combination, which falls within the category of SGCs, is not very common and is probably more frequent in Classical Arabic (henceforth, CA). However, there is some doubt about the acceptability of examples like (9) and (10) suggested by Fassi-Fehri (1999: 115-116) and Al

Mahmoud (2014: 2)<sup>36, 37</sup> To test whether the adjective *dzameel* ‘beautiful’ is marked with the definite article *l-* is acceptable or not, I conducted an experiment, in which I asked 8 native speakers of JA, Saudi Arabic and Kuwaiti Arabic, who have a working knowledge of MSA, to judge the following sentence:

- (11)      \*raʔaytu      l-dzamiil-a      l-wadʒh-i      l-baariha  
              saw.I            the-pretty-ACC      the-face-GEN      the-yesterday  
              ‘I saw the one with the pretty face yesterday’

All the participants indicated that they found example (11) ungrammatical. They corrected the sentence by omitting the definite article marked on the adjective *ldzamiila* ‘the beautiful’. One may argue that the Adj + N combination in (11) was considered ungrammatical by the participants, because the first element of SGCs in Arabic is always indefinite. The definiteness of the second element spreads to the first element in SGCs, yielding a definite construction as a whole (see Fassi-Fehri 2012: 172). In fact, this is the main characteristic of SGCs that distinguishes it from other structures in MSA. Thus, the first element of the Adj + N combinations I present throughout the chapter is not marked with the definite article, based on the judgments I received from the native speakers of Arabic. Note, however, that cases where the definite article is marked on the adjective in Adj + N combinations can be found in the Quran, which represents CA, rather than MSA. In these cases, the construct behaves differently from SGCs in that the second element is not marked with the genitive case:

- (12)      wa-l-kaaḏʕimiin-a      l-ʔaydʕ-a  
              and-the-suppressive.PL-ACC      the-anger-ACC  
              ‘those who suppress their anger’  
(Aal-e-`Imran: 134)

In example (12) the second element *lʔaydʕa* ‘the anger’ is assigned accusative case, rather than genitive case. This may suggest that this combination is not SGC. Even if Fassi-Fehri (1999: 115-116) and Al Mahmoud (2014: 2) assume that the structure in (11) is used only in CA, it is

<sup>36</sup> Fassi-Fehri (1999) does not specify which variety of Arabic he is discussing.

<sup>37</sup> Al Mahmoud (2014: 1) does not distinguish between MSA and CA, he states that “.....the treatment of prenominal and postnominal adjectives in both classical and modern standard Arabic remains indistinguishable for the most part”.

still problematic to assume that this structure is SGC. Hence, I will not discuss the structure in which adjectives in Adj + N combinations are marked with the definite article any further, since this study is focused on MSA and JA, rather than CA.

Despite the disagreement on whether the adjective in Adj + N combinations in MSA can be marked with the definite article, the initial observation that examples (7) and (8) are different in meaning to example (9) still holds. Based on this observation, I provisionally argue that there are two possible types of construct within the prenominal adjectival state construct (Adj + N combinations) in MSA. Initially, examples (7) and (8) seem to exhibit similar behaviour to P-constructs, whereas example (9) appears to behave like a compound, based on the difference in meaning. Through applying the criteria discussed in chapters 3 and 4, I will determine whether my initial observation of the possibility of having two types of Adj + N combinations holds true throughout. Here, it should be noted that these criteria were developed specifically for N + N combinations. Some of them may therefore have to be discarded or modified for use with Adj + N combinations.

### **5.2.3 Applying the cross-linguistic criteria of compoundhood to Adj + N combinations in MSA**

In the following sub-sections, I apply the cross-linguistic criteria of compoundhood to identify compounds within Adj + N combinations in MSA, using only those compatible with Adj + N combinations.

#### **5.2.3.1 Orthography**

In examining the orthography of Adj + N combinations in MSA, it is clear that this offers no help to differentiate between possible constructs. The two elements of Adj + N combinations are usually written as two separate words. This can be illustrated with the following examples:

- |      |                   |             |               |
|------|-------------------|-------------|---------------|
| (13) | qadiim            | r-rasaaʔil  | (P-construct) |
|      | old               | the-letters |               |
|      | ‘the old letters’ |             |               |

(14) waasiŕ            l-ŕaynayn            (compound)  
 wide                the-eyes  
 ‘the one with big eyes’

(15) xafiif            ð<sup>ŕ</sup>-ð<sup>ŕ</sup>ill            (compound)  
 light                the-shadow  
 ‘an active/ lovable person’  
 lit. the one with the light shadow

In examples (13-15) and others, the elements of the Adj + N combination are written separately. This means that there is no difference between the two possible types of Adj + N combination with regard to orthography.

### 5.2.3.2 *Sandhi*

Looking at possible types of Adj + N combination, it appears that sandhi operates in all cases, as in the following examples:

(16) waasiŕ-u-l            mudun            (P-construct)  
 wide- NOM-the            cities  
 ‘the big cities’

(17) t<sup>ŕ</sup>awiil-u-l            lisaan            (compound)  
 tall-NOM-the            tongue  
 ‘a sharp-tongued person’  
 lit. the one with the tall tongue

(18) ŕaziiz-u-n            nafs            (compound)  
 dear-NOM-the            soul/spirit  
 ‘a proud person’  
 lit. the one with a proud spirit

Examples (16-18) show that the first segment of the second element is connected with the coda of the final syllable of the first element in all Adj + N combinations by the case marking segment *u*, i.e. *waa.si.**ful**.mu.dun* ‘the big cities’ in example (16), *tʰa.wii.**lul**.li.saan* ‘the sharp-tongued one’ in example (17) and *ʕa.zii.**zun**.nafʃ* ‘the proud one’ in example (18). The sandhi appears in the connection between the three elements in bold, i.e. the coda of the final syllable of the first element + the case marking segment + the first segment of the second element. Since sandhi operates in all Adj + N combinations, it fails to demonstrate the existence of any subtypes among them.

### 5.2.3.3 Stress

It has become clear in the previous chapters that the position of stress can be used as a criterion to distinguish between P-constructs and compounds in some languages. However, looking at the following examples, it seems that stress is assigned on the first element in both types of Adj + N combination, as discussed in chapter 3 in which the following examples were among the sentences included in the experiment:

(19)     dʒadiid        l-kutub                    (P-construct)  
           new            the-books  
           ‘the new books.’

(20)     baʕiid            n-naðʕar                    (compound)  
           far            the-sight  
           ‘a wise person’  
           lit. the one with the far sight

In examples (19) and (20), stress falls on the first element of both the compound and P-construct. The following examples are similar:

(21)     muxtalif        l-mayaadiin                    (P-construct)  
           various        the-fields  
           ‘the various fields’

- (22) laḏiiḏ      t<sup>ʕ</sup>-t<sup>ʕ</sup>aʕaam      (P-construct)  
delicious      the-food  
‘the delicious food’
- (23) naaʕim      s<sup>ʕ</sup>-s<sup>ʕ</sup>awt      (compound)  
soft      the-voice  
‘a soft voice’  
lit. the one with the soft voice
- (24) qawiyy      l-bunyah      (compound)  
strong      the-body  
‘a strong and healthy person’
- (25) ʕaziiz      n-nafs      (compound)  
dear      the-soul/spirit  
‘a proud person’  
lit. the one with the proud spirit

Examples (21-25) have stress on the first elements of both the P-constructs and compounds, i.e. *laḏiiḏ* ‘delicious’, *naaʕim* ‘soft’ *qawiyy* ‘strong’ and *ʕaziiz* ‘dear’, respectively. The lack of difference between the two types of Adj + N combination in terms of stress assignment indicates that this criterion fails to distinguish between P-constructs and compounds.

#### 5.2.3.4 Compositionality

Regarding compositionality, it seems that P-constructs are always compositional, whereas compounds are non-compositional. This can be illustrated with the following examples:

- (26) t<sup>ʕ</sup>awiil      l-ʔaʕdzaar      (P-construct)  
tall      the-trees  
‘the tall trees’

- |      |   |                           |                              |
|------|---|---------------------------|------------------------------|
| (27) | raθθ<br>shabby<br>'the shabby fabric'         | l-ʔaqmifah<br>the-fabric  | (P-construct)                |
| (28) | ḥasan<br>good<br>'a person with good manners' | l-xuluq<br>the-manner     | (non-compositional-compound) |
| (29) | qaliil<br>little<br>'a quiet person'          | l-kalaam<br>the-speech    | (non-compositional-compound) |
| (30) | qaliil<br>little<br>'a foolish person'        | l-ʕaql<br>the-mind        | (non-compositional compound) |
| (31) | ʕaziiz<br>dear<br>'a proud person'            | n-nafs<br>the-soul/spirit | (non-compositional compound) |

Examples (26) and (27) show that P-constructs are compositional, since the meaning of the whole construct is derived from the meaning of the internal elements. That is, when combined together, *raθθ* 'shabby' and *lʔaqmifah* 'the fabric' yield 'the shabby fabric'. This means that the meaning of the whole construct is predictable from the meaning of the internal elements. However, examples (28-31) show that the four compounds are non-compositional. Specifically, the meanings of both elements, i.e. *ḥasan* 'good' and *lxuluq* 'the manner' do not fully contribute to the meaning of the whole compound, i.e. 'a person with good manners'. A fully compositional meaning of this combination would be 'good manners', but that is not the meaning. Examples (28-31) exhibit a similar behaviour to that of exocentric compounds. Note, however, that there is a difference between examples (28) and (29) on the one hand, and examples (30) and (31), on the other. Specifically, the meanings of the whole compounds in (28) and (29) are semantically transparent, whereas the meanings of the compounds in (31) and (32) are semantically less transparent, i.e. 'a proud person', is not predictable from the meaning of its two elements, i.e. *ʕaziiz* 'dear' and *nnafs* 'soul/spirit'. However, there is insufficient scope



to discuss the differences between semantic compositionality and semantic transparency in the current study; thus, they will not be addressed any further.

In sum, it seems that compositionality can be used as a criterion in distinguishing between P-constructs and compounds, since P-constructs are always compositional, while Adj + N compounds are non-compositional.

### 5.2.3.5 Referentiality

Applying referentiality, the second element of P-construct Adj + N combinations is referential, as shown by the fact that the demonstrative *haaḏa* ‘this’ can be inserted between the two elements. This can be illustrated with the following examples:

(32)      naaṣim          haaḏihi          θ-θiyaab          (P-construct)  
              soft                these                the-clothes  
              ‘these soft clothes’

(33)      dzadiid          haaḏihi          l-kutub          (P-construct)  
              new                these                the-books  
              ‘these new books’

(34)      laḏiiḏ          haaḏa          t<sup>ʕ</sup>-t<sup>ʕ</sup>aṣaam          (P-construct)  
              delicious          this                the-food  
              ‘this delicious food’

However, the second element of Adj + N compounds is non-referential and the demonstrative *haaḏa* ‘this’ cannot be inserted between the two elements, as exemplified by (35) and (36):

(35)      saṣiid          (\*haaḏa)          l-ḥaḏ<sup>ʕ</sup>ḏ<sup>ʕ</sup>          (compound)  
              happy                (\*this)                the-luck  
              ‘this lucky person’  
              lit. happy this luck

- (36) xafiif (\*haaḏa) ḏ<sup>ᶜ</sup>-ḏ<sup>ᶜ</sup>ill (compound)  
 light (\*this) the-shadow  
 ‘this funny person’  
 lit. light this shadow

Thus, it seems that referentiality is the first criterion to differentiate between the two types of Adj + N combination in MSA. The second element of examples (32-34) is referential, whereas the second element of examples (35) and (36) is non-referential. In this respect, Adj + N combinations seem to behave similarly to N + N combinations within SGC, in which the second element of a compound is non-referential. We can say that there is one type of Adj + N combination, i.e. examples like (35) and (36), that fulfils the first compoundhood criterion, whereas the other type does not. This shows that there are indeed two types of Adj + N combinations within SGC.

Moving on to the syntactic criteria, which include adjacency, modification, coordination, the presence of inflection and linking elements, and finally adjacency and the order of elements, I start by examining adjacency.

### 5.2.3.6 Adjacency

Adjacency indeed seems to differentiate between Adj + N P-constructs and compounds. This is illustrated with the following examples:

- (37) baarid haaḏa ʃ-ʃaraab (P-construct)  
 cold this the-drink  
 ‘this cold drink’
- (38) dzadiid baʃd<sup>ᶜ</sup> l-kutub (P-construct)  
 new some the-books  
 ‘some new books’

- (39) t'awiil-u (\*haaḏih) l-qaamah (compound)  
 tall-NOM (\*this-F) the-figure  
 'this person with tall figure'  
 lit. tall this figure
- (40) ḥasan (\*baḥdʿ) l-ḡaxlaaq (compound)  
 good (\*some) the-manners  
 'some people with good manners'  
 lit. some good manners
- (41) ḥaqiil (\*haaḏa) ḏʿ-ḏʿiil (compound)  
 heavy (\*this) the-shadow  
 'this person is annoying'  
 lit. heavy this shadow
- (42) ʕaziiz-u (\*baḥdʿ) l-ḡanfus (compound)  
 dear-NOM (\*some) the-souls/spirits  
 'some proud people'  
 lit. dear some souls

In examples (37) and (38), the demonstrative *haaḏa* 'this' and the quantifier *baḥdʿ* 'some' can be inserted between the two elements, suggesting that these Adj + N combinations are P-constructs. Conversely, in examples (39-42), the demonstrative *haḏaa* 'this' and the quantifier *baḥdʿ* 'some' cannot be inserted between the two elements. In addition to referentiality, adjacency can therefore be used to distinguish between the two types of Adj + N combination in MSA.

### 5.2.3.7 Modification

With regard to modification, the (im)possibility of modification relies on the type of adjective used with Adj + N combination, i.e. prenominal or postnominal. Note that, in accordance with the normal positioning of modifiers in Arabic, such an additional adjective, if possible at all, would have to follow the noun:

- (43) waasiʕ      l-ʔamaakin      l-dzamiilah      (P-construct)  
large.M      the-places.F      the-beautiful.F  
‘the large beautiful places’  
lit. the large places beautiful
- (44) kariih      r-raaʔiha      (\*l-baʕiʕah)      (compound)  
foul.M      the-smell.F      (\*the-ugly.F)  
‘the foul ugly smell’  
lit. foul the smell ugly
- (45) ʕaziiz      l-qawm      (\*tʕ-tʕiwaal)      (compound)  
dear.M      the-people.M      (\*the-tall.M)  
‘the chief of the tall people’  
lit. dear the tall people

Examples (43-45) demonstrate that P-constructs allow other postnominal adjectives to modify their second elements, while such modification is impossible with compounds. The outcome in all cases is ungrammatical. However, it looks as if postnominal adjectives are allowed to modify the first element of a compound, though not that of a phrase, as in:

- (46) dzadiid      l-kutub      (\*l-mufiid)      (P-construct)  
new.MSG      the-books.MPL      (\*the-useful.MSG)  
‘the new useful books’  
lit. the new books useful
- (47) qaliil      l-kalaam      l-mutawaadʕiʕ      (compound)  
little      the-speech      the-modest  
‘a modest and taciturn person’

- (48)      qaliil            l-ṣaql            t-taafih            (compound)  
             little            the-mind            the-silly  
             ‘a foolish and silly person’  
             lit. the silly one with the little mind

Example (46) shows that the postnominal adjective cannot modify the first element of the P-construct *dzadiid lkutub* ‘the new books’, since an adjective cannot modify another adjective. However, examples (47) and (48) appear to demonstrate that the postnominal adjective can modify the first element of the compound. In fact, if examined closely, it seems that the postnominal adjective modifies an implicit head, which can be best described as ‘person/one’, rather than the first element. The headedness of Adj + N compounds is discussed in section 7.2.4.

On the other hand, prenominal adjectives are not allowed to modify the first element of either P-constructs or compounds, as shown in the following examples:

- (49)      (\*muriih)            sariif            l-muwaasʿala-at            (P-construct)  
             (\*comfortable.M)      fast.M            the-transportation-FPL  
             ‘the comfortable and fast transportations’
- (50)      (\*bafif)            kariih            r-raaʿiha            (compound)  
             (\*ugly.M)            foul.M            the-smell.F  
             ‘the foul ugly smell’
- (51)      (\*dzadiid)            ṣaziiz            l-qawm            (compound)  
             (\*new.MSG)      dear.MSG            the-people.MPL  
             ‘the new chief of people’  
             lit. the new dear people

Examples (49-51) show that prenominal adjectives are neither allowed to modify the first element of P-constructs nor that of compounds. Therefore, modification as a criterion partially differentiates between the two types of Adj + N combination, as discussed in examples (43-48).

### 5.2.3.8 Coordination

If a coordinator is inserted between the two adjectives in both P-constructs and compounds, the outcome is grammatical. Thus, it seems that the first element of P-constructs be coordinated with another adjective, as in the following examples:

- (52)      laḏiiḏ            wa      ʃahiyy            tʰ-tʰaʃaam      (P-construct)  
          delicious        and      appetizing        the-food  
          ‘delicious and appetizing food’

In example (52), the first element of the P-construct, i.e. *laḏiiḏ* ‘delicious’ can be coordinated with another adjective, i.e. *ʃahiyy* ‘appetising’. Nonetheless, such coordination is not permitted in compounds:

- (53)      ʃaziiz            (\*wa    kabiir) l-qawm            (compound)  
          dear            (\*and    big)    the-people  
          ‘the eminent chief of the people’  
          lit. dear and big the people

- (54)      waasiʃ            (\*wa    kaθiir) l-ħiilah            (compound)  
          wide            (\*and    plenty) the-ability  
          ‘a smart and resourceful person’  
          lit. wide and plenty the ability

The coordination of the first element of the compound with another adjective causes it to be ungrammatical, as shown in examples (53) and (54).

Concerning the second element, it appears that this can be coordinated in both P-constructs and compounds, as in (55-57):

- (55)      laḏiiḏ            tʰ-tʰaʃaam      wa      ʃ-ʃaraab            (P-construct)  
          delicious        the-food        and      the-drink  
          ‘the delicious food and drink’

(56) kabiir l-ʔanf wa l-famm (compound)  
 big the-nose and the-mouth  
 ‘the big nose and mouth’

(57) qawiyy l-qalb wa l-bunyah (compound)  
 strong the-heart and the-body  
 ‘a brave and strong person’  
 lit. the strong heart and body

Example (55) demonstrates that the second element of a P-construct, i.e. *tʔaʕaam* ‘the food’ can be coordinated with another noun, i.e. *ʃʃaraab* ‘the drink’. Examples (56) and (57) show that the second elements of the two compounds, i.e. *lʔanf* ‘the nose’ and *lqalb* ‘the heart’ can be coordinated with other nouns *lfamm* ‘the mouth’ and *lbunyah* ‘the body’, respectively. Therefore, the conclusion is that coordination can partially differentiate between P-constructs and compounds, since it can distinguish compounds from P-constructs when the first element is coordinated, but it cannot differentiate between them when the second element is coordinated.

### 5.2.3.9 Inflection and linking elements

The presence of inflection and linking elements can be used to distinguish compounds from P-constructs in languages that exhibit inflectional markings. Looking at the following examples, one can observe that the second element of Adj + N P-constructs is normally pluralised. In other words, the second element of P-constructs cannot be singular, i.e. *\*lkitaab* in (58a) vs. *lkutub* ‘the books’ in (58b) and *\*lʔuyniya* ‘the song’ in (59a) vs. *lʔayaani* ‘the songs’ in (59b):

(58) a. \*qadiim l-kitaab (P-construct)  
 old the-book.MSG  
 ‘the old book’

b. qadiim l-kutub (P-construct)  
 old the-book.MPL  
 ‘the old books’

- (59) a. \* dzadiid l-ʔuyniya (P-construct)  
 new the-song.FSG  
 ‘the new song’
- b. dzadiid l-ʔayaani (P-construct)  
 new the-song.FPL  
 ‘the new songs’

I would suggest that the second element of P-constructs should be plural to yield the partitive reading of the construct. In order to convey the meaning of ‘some of X’, X needs to be plural. In comparison, the second element of Adj + N compounds cannot be freely pluralised, as shown in (60) and (61):

- (60) a. tʿawiil l-qaamah  
 tall the-figure.FSG  
 ‘a tall person’
- b. \*tʿawiil l-qaamaat  
 tall the-figure.FPL  
 ‘tall people’  
 lit. the tall ones in figures
- (61) a. qaliil l-ʕaql  
 little the-mind.MSG  
 ‘a foolish person’  
 lit. the one with the little mind
- b. \*qaliil l-ʕuquul  
 little the-mind.MPL  
 ‘foolish people’  
 lit. the ones with the little minds

Examples (60) and (61) demonstrate that the second element of compounds cannot be freely pluralised, i.e. *ʕaql* ‘the mind’ in (61a) vs. \**ʕuquul* ‘the minds’ in (61b). However, there are



some exceptional cases where the second element actually must be pluralised, as in the following compounds:

(62) a. kabiir            l-masʔuul-iin  
           big                the-official-PL.GEN  
           ‘the senior (one) of the officials’

b. \*kabiir            l-masʔuul  
           big                the-official  
           ‘the senior one of the official’

(63) a. baarid            l-ʔaʕsʕaab  
           cold                the-nerves  
           ‘the cool-headed person’  
           lit. the one with the cold nerves

b. \*baarid            l-ʕasʕab  
           cold                the-nerve  
           ‘the cool-headed person’  
           lit. the one with the cold nerve

In examples (62) and (63), the second element of the compound must be plural. I would argue that it is ungrammatical for the second element to be singular by virtue of its meaning. For instance, in example (62), someone cannot be a head of one person; one can only be a head of many people. In example (63), *lʔaʕsʕaab* ‘the nerves’, has to be plural, since no one has only one nerve in his/her body.

Concerning the first element of P-constructs, analysis of examples of this type of construct shows that the first element cannot be pluralised:

(64) a. hadiiθ                    l-mabaani                    (P-construct)  
           modern.MSG            the-building  
           ‘the modern buildings’

- b.     \*ħadiiθ-uu           l-mabaani  
modern.MPL           the-building  
‘the modern buildings’
- (65) a.    laðiið               tʰ-tʰaʃaam           (P-construct)  
delicious.MSG       the-food  
‘the delicious food’
- b.     \*laðaaʔið           tʰ-tʰaʃaam  
delicious.MPL       the-food  
‘the delicious food’

Examples (64) and (65) show that the first element of a P-construct cannot be freely pluralised. The outcome of such a process is ungrammatical, i.e. *ħadiiθ* ‘modern’ in (64a) vs. *\*ħadiiθuu* ‘modern (plural)’ in (64b), and *laðiið* ‘delicious’ in (65a) vs. *\*laðaaʔið* ‘delicious (plural)’ in (65b). In addition, example (66) shows that the first element of a P-construct cannot be feminine, i.e. *dzadiid* ‘new’ in (66a) vs. *\*dzadiidat* ‘new (feminine)’ in (66b):

- (66) a.    dzadiid               l-kutub           (P-construct)  
new.MSG           the-books  
‘the new books’
- b.     \*dzadiid-at           l-kutub  
new-FSG           the-books  
‘the new books’

In fact, the first element of such constructs can only be masculine and singular, as shown in the previous examples (cf. Ryding 2005: 221-223).

In contrast, the first element of compounds can be freely pluralised. This is illustrated in the following examples:

- (67) a.    dʃaʃiif               l-basʃar           (compound)  
weak.MSG           the-sight  
‘a short/near-sighted person’

- b.      d<sup>ʕ</sup>uʕafaaʔ              l-basʕar  
             weak.MPL              the-sight  
             ‘short/near-sighted people’
- (68)    a.      ʕaziiz-at              n-nafs              (compound)  
             dear-FSG              the-soul/spirit  
             ‘a proud female’
- b.      ʕaziiz-aat              n-nafs  
             dear-FPL              the-soul/spirit  
             ‘proud females’

Unlike P-constructs, examples (67) and (68) demonstrate that the first element of compounds can be freely pluralised without resulting in ungrammaticality. In example (67b), the first element is pluralised, i.e. *d<sup>ʕ</sup>uʕafaaʔ* ‘the weak (plural)’, yet the construct is perfectly grammatical. Another difference between the two types of Adj + N combination is that the first element of compounds can also be feminine, i.e. *ʕaziizat* ‘dear (feminine)’ in (68a) vs. *ʕaziizaat* ‘dear (plural)’ in (68b). In sum, we can say that inflection as a criterion can distinguish between compounds and P-constructs with few exceptions, as discussed in examples (62) and (63).

### 5.2.3.10 Adjacency and the order of elements

A test that, to some extent, involves adjacency and the order of elements within Adj + N combinations can be suggested here to distinguish between Adj + N compounds and P-constructs in MSA. When the elements of P-constructs are reversed, i.e. N + Adj instead of Adj + N, the P-construct remains grammatical, but it loses its partitive reading, i.e. ‘some of X’. This is shown in the following example:

- (69)    a.      qadiim              l-ʔayaani              (P-construct)  
             old              the-song.FPL  
             ‘the old songs’  
             lit. the old of the songs

- b. l-ʔayaani      l-qadiimah      (P-construct)  
 the-song.FPL    the-old  
 ‘the old songs’

However, when the two elements of a compound are reversed, the meaning of the whole compound changes, losing its metaphorical and/or metonymic reading, as in:

- (70) a. dzamiil      l-wadʒh  
 beautiful      the-face  
 ‘the one with the beautiful face’
- b. l-wadʒh      l-dzamiil  
 the-face.MSG      the-beautiful.MSG  
 ‘the beautiful face’
- (71) a. qawiyy      l-qalb  
 strong      the-heart  
 ‘a brave person’  
 lit. the one with the strong heart
- b. l-qalb      l-qawiyy  
 the-heart.MSG      the-strong.MSG  
 ‘the physically strong heart’

Examples (70b) and (71b) show that the reversed order of the compound yields a different meaning. For instance, the compound in example (70a) means ‘the (one) with a beautiful face’, but in (70b), the face itself is described as beautiful. Note, also, that when the two elements of the compound are reversed, the outcome is no longer a compound. In other words, example (70b) is a phrase, since the adjective agrees with the noun it modifies in number, gender and definiteness. The same applies to example (71), where the meaning of (71a) is ‘a brave person’, while the meaning of (71b) is ‘the physically strong heart’.

Finally, all of the above Adj + N combinations, which appear in MSA, do not appear in JA. Instead, other constructions are used to convey the same meanings. The following examples are used to refer to *qadiim rrasaaʔil* ‘the old letters’ and *dzadiid lkutub* ‘the new books’ in JA, respectively:

(72)      gareet            r-rasaaʔil        l-gadiimeh  
              read.1sg        the-letters      the-old  
              ‘I read the old letters.’

(73)      baħibb            l-kutub            l-dzadiideh  
              love.1sg        the-books        the-new  
              ‘I love the new books.’

Examples (72) and (73) show that the adjective in JA always follows the noun, so they are like (70b) and (71b). The combination Adj + N, e.g. *qadiim rrasaaʔil* ‘the old letters’ does not exist in JA. Similarly, instead of MSA *tʿawiil lqaamah* ‘the tall person’, *qawiyy lqalb* ‘the brave person’ and *baarid ʔaʕsʕaab* ‘the cool-headed person’, the following examples are used in JA, respectively:

(74)      ʃofet            bint    tʿawiileh        fi-l-madraseh  
              saw.1sg        girl    tall                in-the-school  
              ‘I saw a tall girl at school’

(75)      haaðʕa z-zalameh    galb-oh        gawi  
              this    the-man        heart-his        strong  
              ‘this man is brave’

(76)      aħmad            ʔaʕsʕaab-oh    baarideh  
              Ahmad        nerves-his      cold  
              ‘Ahmad is a cool-headed person’  
              lit. Ahmad has cold nerves

Example (74) shows that the first element of the compound *tʿawiil lqaamah* ‘the tall person’ is kept, while the second element is dropped in JA. The adjective *tʿawiileh* ‘tall’ is also placed

after the noun, not before it. I would suggest that since Adj + N combinations do not appear in JA, different constructions are used, sometimes with the simple N + Adj order (see examples (75) and (76)). Specifically, the noun *galboh* ‘his heart’ in (75) is placed before *gawi* ‘strong’, with a resumptive pronoun marked on the noun *galboh* ‘his heart’ in (75), *ʔaʕsʕaaboh* ‘his nerves’ in (76), respectively. Note, also, that the sound /q/ is replaced with /g/ in *qawiyy* vs. *gawi* ‘strong’, since in JA the sound /q/ is realised as /g/.

We have established that, on the basis of the criteria of referentiality, adjacency, compositionality, the presence of inflection, and adjacency and the order of elements, two types of Adj + N combinations can indeed be distinguished in MSA: compounds and P-constructs. The P-constructs in all respects behave like phrases while the compounds have various properties pointing at word status for the entire Adj + N combination. Examining more closely now the status of these compounds, i.e. expressions like *dʕaʕiif lbasʕar* ‘someone who is short/near-sighted’, *tʕawiil lqaamah* ‘someone who has a tall figure’ and *ʕaziiz nnaʕs* ‘someone who is proud of himself’, I suggest that they are to be classified as bahuvrihi-compounds. That is, they should be interpreted as ‘the weak one in sight’, ‘the tall one in figure’ and ‘the proud one in spirit’, respectively.

As pointed out by Čermák (1997: 13), bahuvrihi is originally a Sanskrit term used for compounds, often of the Adj + N type, that have the morphological structure A + B but lack a true internal head. Concerning the head of bahuvrihi compounds, Čermák (1997: 13) maintains that the genuine head is located outside the compound itself. Simply put, bahuvrihi is a hyponym of an implicit or unexpressed semantic head. For instance, the bahuvrihi *hardhat* does not denote a special kind of *hat*, but refers to an individual who uses, possesses or is characterised by that kind of hat. Its semantic head is not explicitly expressed, rather it is implicitly understood, as being ‘person/one’. Čermák (1997: 13) notes, that as a consequence, it is not possible to analyse a bahuvrihi compound into its immediate elements; it is solely interpretable as predicated of an unrealised ‘third party’. The lack of a semantic head and the external nature of their reference means that bahuvrihis are structurally exocentric.

Based on these observations, Čermák (1997: 13) concludes that the presence of a zero head can be viewed as the main distinctive characteristic of bahuvrihis. Booij (2007: 80) concurs with Čermák (1997) on the fact that this kind of compound behaves as an adjective although there is no adjectival head. Ralli and Andreou (2012: 67) add to Čermák’s (1997) definition of bahuvrihi-compounds, stating that “Bahuvrihi or possessive compounds are composed of an adjective and a noun, and they denote someone who has something expressed by the noun that is modified by an adjective”. Examples of bahuvrihis in Present-Day English

are *redskin*, *redneck*, *paleface* and *blue-stocking*. Comparing the English compound *paleface* ‘the pale one in face’ with the MSA compound *tʿawiil lqaamah* ‘the tall one in figure’, it seems that both compounds are only interpretable as having an unrealised semantic head, i.e. ‘one’. In the case of JA equivalents (74-76) of bahuvrihis, such as *qawiyy lqalb* ‘the strong one in heart’, it is clear that the order in JA is N + Adj, with a resumptive pronoun marked on the first element. All in all, JA and MSA both allow simple N + Adj combinations; in addition, MSA - but not JA - has special Adj + N compounds, with distinctive meaning and structure.

#### **5.2.4 Summary**

The previous discussion shows that based on certain criteria, i.e. referentiality, adjacency, compositionality, inflection (free pluralisation and/or free gender marking of the first element) and adjacency and the order of elements (compounds lose their metaphorical and/or metonymic interpretation), there are two types of Adj + N combinations in MSA. The first type behaves in a similar way to P-constructs, whereas the second type is more in line with compounds. The further criteria of modification and coordination can partially distinguish between Adj + N compounds and P-constructs. In JA, Adj + N combinations do not appear; instead N + Adj combinations exist. Both N + Adj combinations behave like phrases, one in which the adjective agrees with the preceding noun in number, gender and definiteness, whereas the other one is characterised by the appearance of the resumptive pronoun on the first element.

With regard to syntactic category, the first/left element of the combinations examined in this chapter is clearly an adjective, whereas the syntactic category of the second element is a noun. However, the syntactic category of the resulting compound is not clear. Therefore, criteria that help to differentiate between different word classes need to be applied here to identify the syntactic category of the output of Adj + N compounds in MSA. The next sections deal with this issue.

### **5.3 Adjective-hood criteria**

#### **5.3.1 Overview**

There has been much debate about word classes and their identification criteria. For example, Wetzler (1996) argues that a clear set of general criteria for adjective-hood has not been provided yet, and may never be. He in fact claims that adjectives can be placed on a noun-verb continuum, because some languages have adjectives that are similar to nouns in terms of syntactic behaviour, whereas other languages have adjectives that are more syntactically

similar to verbs. Similarly, Haspelmath (2012: 109) proposes that the existence of cross-linguistic lexical categories should not be taken for granted. Specifically, he argues that adjectives in addition to other lexical categories need to be defined on “a language particular level”. In fact, Chafe (2012: 1) suggests that adjectives are more difficult to define compared to nouns and verbs, noting that verbs and nouns belong to open word classes and have the potential to be universal, whilst adjectives exhibit different characteristics cross-linguistically. Based on the previous discussion, it can be proposed that the adjective class is a debatable and potentially problematic lexical category as opposed to nouns and verbs, typically display common properties cross-linguistically. Even though the criteria for identifying certain words as adjectives are still problematic, several attempts have been made to deal with this elusive word class. The following section provides some of the criteria found in the relevant literature.

### **5.3.2 Cross-linguistic adjective-hood criteria**

Several scholars (e.g. Strang 1969; Quirk *et al.* 1985; Wetzler 1996; Baker 2003 among others) discuss the main criteria used to identify word classes cross-linguistically. In English, Quirk *et al.* (1985: 402) suggest four basic morphosyntactic criteria characterising adjectives, including modification by *very*; combining with suffixes to indicate comparison; and having the ability to modify a noun (attributive) and to be a predicative complement (predicative). If certain words meet all the above criteria, then they can be called “central adjectives”, whereas adjectives which only meet some of these criteria are known as “peripheral adjectives” (Quirk *et al.* 1985: 404). For example, the adjective *utter* satisfies only one criterion, i.e. the ability to modify nouns, such as *utter glory*. Thus, it is a peripheral adjective. Conversely, the adjective *happy* is a central adjective, because it meets all four criteria, i.e. *the happy girl, the girl was happy, the very happy girl* and *the happiest girl*. However, Quirk *et al.* (1985: 404) do not supply a semantic-based definition for adjectives; they have only identified them based on their syntactic behaviour. They claim that it is not possible to identify the lexical category of a particular word when it is examined in isolation, since the form of the word does not say much about its syntactic function. In addition to their ability to be combined with particular suffixes, Strang (1969: 133) argues that the final two criteria are the most central ones to identify adjectives, proposing that the other two criteria are less important, since many adjectives are not gradable.

From a cross-linguistic perspective, Baker (2003: 190) argues that adjectives, in contrast to nouns, are not referential. These two distinguishing criteria would, ideally, be



sufficient to characterise the behaviour of adjectives. In addition to these two defining criteria of adjectives, other derived properties of adjectives can be found. In particular, Baker (2003: 191, 230) suggests that there are three syntactic environments in which only adjectives can be used. The first environment is that adjectives can be directly attributive modifiers of nouns, while nouns and verbs cannot be. The second one is that adjectives can function as the complement of a degree head, e.g. *so*, *too*, etc. In contrast, neither nominal nor verbal projections can. The final environment is that adjectives can be resultative secondary predicates, e.g. *they beat the metal flat*. Conversely, nouns and verbs cannot be, e.g. *\*they beat the metal a sword* and *\*they polished the coin shine* (Baker 2003: 190). These environments show that adjectives in English do not form a natural class with either nouns or verbs.

Based on the works of the above researchers who examine adjective-hood criteria, we can try to formulate a set of criteria to identify adjectives in MSA. The discussion of these criteria is based on function, distribution and morphosyntactic features of the word. With regard to function, two basic ones can be distinguished, namely, attributive and predicative. These functions can be fulfilled by adjectives and adjective phrases (APs) alike. Attributive adjectives typically have a fixed position and directly modify a noun. In MSA, the difference between attributive and predicative adjectives is signalled by the presence vs. absence of the definite article. If the adjective is definite, then it is attributive, whereas if the adjective is indefinite, i.e. not marked with the definite article, then it is predicative. An example of a sentence with both an attributive and a predicative adjective is given in (77).

- (77)      l-bint-u                      l-dzamiilat-u                      ḏakiyyat-un  
             the-girl-NOM                  the-beautiful-NOM              smart-NOM  
             ‘the beautiful girl is smart.’

Example (77) shows that the difference in meaning between predicative and attributive adjectives triggers the presence of the definite article on the latter, i.e. the adjective *ldzamiilat* ‘the beautiful’. However, the adjective *ḏakiyyatun* ‘smart’ is predicative as demonstrated by the lack of the definite article. Along these lines, note that the adjectives in Adj + N combinations in MSA are always predicative as opposed to attributive and indeed the first element of these combinations in MSA is always indefinite (see section 5.2.1), as in (78-80):

- (78) a.    ʔuħibb-u      dʒadiid-a      l-kutub-i                   (P-construct)  
           love-I           new-ACC    the-books-GEN  
           ‘I love the new books.’  
           lit. I love some of the new books
- b.    ʔuħibb-u      (\*l-)dʒadiid-a      l-kutub-i  
           love-I           (\*the-)new-ACC      the-books-GEN  
           ‘I love the new books.’  
           lit. I love some of the new books
- (79) a.    haaða r-radʒul-u      qawiyy      l-qalb                   (compound)  
           this    the-man-NOM   strong      the-heart  
           ‘This man is brave.’
- b.    haaða           r-radʒul-u                   (\*l-)qawiyy    l-qalb  
           this            the-man-NOM               (\*the)-strong   the-heart  
           ‘this brave man’
- (80) a.    dʒalastu      maʕ    xafiif                   ð<sup>ʕ</sup>-ð<sup>ʕ</sup>ill                   (compound)  
           sat.I           with   light                   the-shadow  
           ‘I sat with the funny person’
- b.    dʒalastu      maʕ    (\*l-)xafiif               ð<sup>ʕ</sup>-ð<sup>ʕ</sup>ill  
           sat.I           with   (\*the-)light        the-shadow  
           ‘I sat with the funny person’

Examples (78-80) show that Adj + N combinations, whether P-constructs or compounds, occur in a predicative position not an attributive one. Hence, when the definite article in (78b), (79b) and (80b) is attached to the first element of the compound, the outcome is ungrammatical. This is because attaching the definite article to the first element of these two constructs changes them into another type of construct, which can only be found in CA (see section 5.2.1), not in MSA. Note, however, that the ability to occur in a predicative position is not necessarily a sign of adjectivehood. NPs can also be found in that position, as in:

- (81)      ʕaliyy            sʕadiiq-i  
             Ali                friend-my  
             ‘Ali is my friend’

Secondly, as far as distribution is concerned, we have seen that in English using adjectival modifiers such as *quite/more/most*, as in *quite/more/most awake* helps to identify words such as *awake* as an adjective. Additionally, use of the intensifier or degree modifier *very* is a reliable test for adjectivehood in English. Note, however, that this intensifier can only modify gradable adjectives, such as *beautiful, small and smart*, but not non-gradable adjectives, such as *dead, married* and *wooden*.

MSA has the intensifier *dziddan* ‘very’, which can be readily used with simple adjectives, as in (82).

- (82)      haaða            r-radʒul-u            qasʕiir-u-n            dziddan  
             this                the-man-NOM            short-NOM-INDF            very  
             ‘This man is very short.’

However, Adj + N combinations cannot be modified by *dziddan* ‘very’, as shown in (83):

- (83)      haaða            r-radʒul-u            tʕawiil-u            l-qaamah            (\*dziddan)  
             this                the-man-NOM            tall-NOM            the-figure            (\*very)  
             ‘This man is very tall.’

Example (83) shows that the adjectival modifier is incompatible with Adj + N combinations in MSA. This could be due to the fact that the adjectival modifier *dziddan* ‘very’ should not be separated from the adjective that precedes it. The above sentence is fully grammatical in the absence of the word *lqaamah* ‘the figure’.

- (84)      haaða            r-radʒul-u            tʕawiil-u-n            dziddan  
             this                the-man-NOM            tall-NOM-INDEF            very  
             ‘This man is very tall.’

Examples (82) and (84) show that the adjectival modifier *dziddan* ‘very’ immediately follows the adjective it modifies. No element is allowed to intervene. This explains why example (83)

is ungrammatical. That is, the second element, i.e. *lqaamah* ‘the figure’ intervenes between the adjective *tʿawiil* ‘tall’ and the adjectival modifier *dʒiddan* ‘very’. Furthermore, this could be due to the fact that Adj + N combinations are not adjectives.

The third criterion used to identify adjectives cross-linguistically is morphosyntax. It is common, though not universal, for languages to have the morphosyntactic category of agreement. In MSA, both attributive and predicative adjectives agree with the noun they modify in number, gender, definiteness and case. Thus, the element that has the same morphosyntactic features as the noun that follows it is definitely an adjective in MSA, as in:

- |      |   |                                    |                      |                           |
|------|---|------------------------------------|----------------------|---------------------------|
| (85) | r-radʒul-u<br>the-man.MSG-NOM<br>‘the strong man’     | l-qawiyy-u<br>the-strong.MSG-NOM   |                      |                           |
| (86) | r-radʒul-u<br>the-man.MSG-NOM<br>‘The man is strong.’ | qawiyy-u-n<br>strong.MSG-NOM-INDEF |                      |                           |
| (87) | haaḏihi<br>this.FSG                                   | l-marʔa-tu<br>the-woman.FSG-NOM    | ḥasan-at<br>well.FSG | l-xuluq<br>the-manner.MSG |
- ‘This woman is with good manners.’

Examples (85-89) show that the adjectives, i.e. *lqawiyyu* ‘the strong’ and *qawiyyun* ‘strong’ agree with the nouns they modify, i.e. *rradʒulu* ‘the man’ and *rradʒulu* ‘man’ in number, gender, definiteness and case. As far as Adj + N combinations are concerned, example (87) shows that the adjective *ḥasanat* ‘good’ in the Adj + N combination agrees with the noun it modifies, *lmarʔatu* ‘the woman’, in number and gender but not in definiteness. Lack of agreement in definiteness occurs because the first element of Adj + N combination is always indefinite. Note that the adjective *ḥasanat* ‘good’ in (87) does not agree with the following noun *lxuluq* ‘the manner’ in gender.

Through applying the previous criteria for adjective-hood in terms of function, distribution and morphosyntactic features to Adj + N combinations in MSA, it is not clear whether adding a prenominal adjective to a noun yields an adjective. Fassi-Fehri (1999: 115) discusses prenominal adjectives in MSA from a theoretical perspective and argues that these types of adjectives, together with the noun that follows them, constitute a noun phrase (NP),

rather than an adjective phrase (AP). He provides three pieces of evidence to support his argument. Following Fassi-Fehri (1999), the next section examines the syntactic category of the whole Adj + N combination. Here, note that if we assume that we are dealing with an NP, rather than AP, this indicates that the construction is right-headed. Since Arabic is a predominantly left-headed language, right-headed combinations are unexpected. This issue is discussed in detail in section 7.2, where the headedness of this combination is examined syntactically, semantically and morphologically.

#### 5.4 The syntactic category of Adj + N combinations in MSA

According to Fassi-Fehri (1999: 115), the adjective is the head of a type of nominal state construct known as “synthetic genitive” mentioned in section 5.2.1, (repeated here in (88) and (89) for convenience):

- (88)      ʔakal-tu      laḏiiḏ-a      tʰ-tʰaʕaam-i  
              ate-I            delicious-ACC      the-food-GEN  
              ‘I ate the delicious food.’  
              lit. I ate some of the delicious food

- (89)      ʔaqraʔ-u      dzadiid-a      l-kutub-i  
              read-I            new-ACC      the-books-GEN  
              ‘I read the new books.’  
              lit. I read some of the new books

(Fassi-Fehri 1999: 115)

Typically, the first element of this construction is assigned an external structural case, i.e. accusative, nominative or genitive based on the function of the whole construct in the sentence, whereas the second element always has genitive case. Unexpectedly, Fassi-Fehri (1999) notes that the behaviour of the construction as a whole is more like a noun, rather than an adjective. This argument is supported by three pieces of evidence. Firstly, this type of construction appears in a determiner phrase (DP) position, rather than an AP position. It can therefore replace other DPs, which are headed by nouns. For example, the following constructions are headed by a noun and they can be considered equivalents to the above examples, respectively:

- (90)      ʔakal-tu      t<sup>ʕ</sup>-t<sup>ʕ</sup>aʕaama      l-laḏiiḏ-a  
           ate-I           the-food-ACC      the-delicious-ACC  
           ‘I ate the delicious food.’
- (91)      ʔaqraʔ-u      l-kutub-a      l-dzadiid-at-a  
           read-I           the-books-ACC      the-new-F-ACC  
           ‘I read the new books.’

Although very similar in use and actual interpretation, examples (88) and (89) are different from (90) and (91) in that the former convey a partitive reading. For instance, in example (88), the subject, i.e. *I* only ate the delicious food, not any other type of food. However, in example (90), the subject, i.e. *I* eats the food which is described as delicious.

Secondly, the synthetic genitive phrase is definite, which indicates definiteness inheritance has occurred here. The definiteness of the phrase becomes apparent when it is modified by a definite relative (relative clauses that can only be used with definite nouns/NP in Arabic), as in examples (92) and (93):

- (92)      laḏiiḏ-u              t<sup>ʕ</sup>-t<sup>ʕ</sup>aʕaam-i              llaḏii              ʔakal-tu-hu  
           delicious-NOM      the-food-GEN              that              ate-I-it  
           ‘the delicious food that I ate’
- (93)      muxtalif-u      l-mayaadin-i              llatii      yaḥduḏu      haaḏa      fii-ha  
           various-NOM      the-fields-GEN              that      happens      this      in-them  
           ‘the various fields in which this happens’

However, adjectival state constructs do not have these characteristics. In addition to the fact that they only occur in AP positions, they do not trigger definiteness inheritance. For the adjectival state constructs to be definite, an adjectival head must be attached to a definite article, as in (94):

- (94)      ?baħaθ-tu      ʕan      l-dzamiil-i      l-wadʒh-i  
              looked-I      for      the-nice-GEN      the-face-GEN  
              ‘I looked for the one with a nice face.’

Nevertheless, the adjectival head of the prenominal adjectival state construct cannot take a definite article. Hence, marking by a definite article yields an ungrammatical construction, as discussed by Fassi-Fehri (1999: 116):

- (95)      a.      baħaθ-tu      ʕan      waafir-i      l-iħtiraam-i  
              looked-I      for      plentiful-GEN      the-respect-GEN  
              ‘I looked for the plentiful respect.’
- b.      \*baħaθ-tu      ʕan      l-waafir-i      l-iħtiraam-i  
              looked-I      for      the-plentiful-GEN      the-respect-GEN  
              Intended to mean: ‘I looked for the plentiful respect.’

Thirdly, as discussed before in section 5.3.2, the fact that the prenominal adjectival state construct cannot be modified by a degree adverbial, i.e. *dziddan* ‘very’ provides additional evidence that this construction is nominal, rather than adjectival. This can be illustrated with the following example adapted from Fassi-Fehri (1999: 116):

- (96)      a. ?ukinn-u      la-hu      l-iħtiraam-a      l-waafir-a      dzidd-an  
              entertain-I      for-him      the-respect-ACC      the-plentiful-ACC      very-ACC  
              ‘I have a very plentiful respect for him.’
- b. \*?ukinn-u      la-hu      waafir-a      l-iħtiraam-i      dzidd-an  
              entertain-I      for-him      plentiful-ACC      the-respect-GEN      very-ACC  
              ‘I have plenty of respect for him.’

Example (96a) shows that a typical adjectival state construct can be modified by a degree adverbial, i.e. *dziddan* ‘very’. In other words, adjectival state constructs are compatible with adverbs. However, example (96b) shows that adjectival state constructs cannot be modified by an adverbial (as we saw in example 83). This shows that the latter construction behaves more

like a noun, rather than an adjective, since adjectives can be modified by adverbs. In addition, even when the degree adverbial *dziddan* ‘very’ is moved closer to the adjective, the outcome will still be ungrammatical, as in:

- (97)      ?ukinn-u      la-hu    (\*dzidd-an)    waafir-a      l-iħtiraam-i  
entertain-I      for-him    (\*lot-ACC)    plentiful-ACC    the-respect-GEN  
‘I have plenty of respect for him.’

That the syntactic category of the output of Adj + N combinations is a noun is also evident from sentences like (98) and (99). The following examples show that the prenominal adjectival state construct can also be a complement inside a prepositional phrase (PP), which again means that it behaves like an NP:

- (98)      ?istamtaŕ-tu    bi-laðiið-i                      t<sup>ʕ</sup>-t<sup>ʕ</sup>aŕaam-i  
enjoyed-I      with-delicious-GEN    the-food-GEN  
‘I enjoyed the delicious food.’

- (99)      ?usaafir-u      maŕ    qawiyy-i      l-qalb-i                      wa    laa    ?axaaf  
travel-I      with    strong-GEN    the-heart-GEN                      and    not    fear.I  
‘I travel with the brave man and I do not feel afraid.’

In sum, based on this section and the argument presented by Fassi-Fehri (1999), I argue that the outcome of all Adj + N combinations in MSA is in fact a noun that behaves like other NPs, not an adjective.

### 5.5 Other compound adjectival expressions within SGC

Ryding (2005: 274) suggests that there are some compound adjectival expressions, i.e. N + Adj within SGCs in MSA, stating that “They occur primarily as adjective *’iDaafas*, or, for negative concepts, as adjectives in construct with the noun *ghayr*”. Some examples that start with *ghayr* (henceforth, *yayr*, based on the convention of the study) ‘non-, un-, in-, other than’, as discussed by Ryding (2005: 223, 275), are shown below:



(100)	<i>ɣayr</i> not 'unsuitable'	<i>munaasib</i> suitable
(101)	<i>ɣayr</i> not 'indirect'	<i>mubaaʃir</i> direct
(102)	<i>ɣayr</i> not 'unusual'	<i>ʕaadi</i> usual

Ryding (2005: 223, 274-5) claims that the word *ɣayr* in examples (100-102) should be treated as a noun, since it can be either nominative, accusative or genitive, based on the function of the whole construct in the sentence. Additionally, it never carries the definite article *l-* like other nouns when they appear as the first elements in SGCs. Contrary to Ryding (2005: 274-5), I would argue that *ɣayr* in examples (100-102) is to be analysed as a prefix, rather than a noun, since *ɣayr* has some properties making it more like a prefix, so its noun-status is questionable and therefore the compound status of the combination is not very clear. Specifically, *ɣayr* cannot stand alone unless it is attached to an adjective.

Another type of construct mentioned in Ryding (2005: 274) is the one that starts with the adjective *mutaʕadid* 'numerous'. The following examples illustrate this type:

(103)	<i>mutaʕaddid</i> multi 'multilateral'	<i>l-ʔatʕraaf</i> the-sides
(104)	<i>mutaʕaddid</i> multi 'multi-use'	<i>l-ʔistixdaamaat</i> the-uses
(105)	<i>mutaʕaddid</i> multi 'multinational'	<i>l-dʒinsiyyaat</i> the-nationalities

According to Ryding (2005: 274), examples (103-105) show a type of compound, in which the first element is fixed, whereas the second one is changeable. The adjective *mutaṣaddid* ‘numerous’ occurs in all these expressions as the first element, having the same sense, i.e. ‘multi’. In considering the status of this element, a comparison with English may be useful. While Marchand (1969: 100) treats English elements such as *over-*, *out-* and *under-* as first elements of compounds, e.g. *overfly*, I suggest that the first element of this construct is to be considered as a prefix on the basis of Lieber’s (2009) argument. Specifically, Lieber (2009: 366) considers the above elements, *over-*, *out-* and *under-*, prefixes, since they differ semantically from their preposition counterparts, by adding a meaning of excess that is lacking in independent prepositions. Bauer *et al.* (2013: 336) suggest that morphemes such as *over-*, *out-* and *under-* should be regarded as prefixes, rather than the first elements of compounds. These morphemes differ in the range of meanings they denote when used in compounds in comparison to their meanings when they are used as prepositions. Lieber’s (2009) argument seems to apply to the Arabic adjective *mutaṣaddid* ‘numerous’, as in the following examples:

(106)      waḏʿaaʔif      haaḏa      l-dzihaaz      mutaṣaddida  
                  functions      this      the-device      numerous  
                  ‘The functions of this device are numerous.’

(107)      ṡtaraytu      dzihaazan      mutaṣaddida      l-waḏʿaaʔif  
                  bought.I      device      multi      the-functions  
                  ‘I bought a multifunctional device.’

Examples (106) and (107) show that the senses of *mutaṣaddid* differ based on the function it serves in the sentence. In example (106), *mutaṣaddid* denotes ‘numerous’ when it acts as an adjective, whereas in example (107), it denotes ‘multi’ when it functions as a prefix. Note that the second element in these compounds is always plural, i.e. the adjective *mutaṣaddid* ‘multi’ functions like *many*. Therefore, it requires the second element to be plural, e.g. *\*mutaṣaddid lwaḏʿiifa* ‘multifunctional (singular)’. This behaviour is not normally exhibited by adjectives. Specifically, adjectives can be followed by both singular and plural nouns.

Therefore, both *yayr* ‘not’ and *mutaʕaddid* ‘multi-’ should be treated as prefixes. Hence, the expressions in which they combine with a noun, which are called compounds by Ryding (2005), are actually best treated as derived words.

## 5.6 Conclusion

In this chapter, I have discussed Adj + N combinations, describing their structure and the internal word classes within them. By applying four robust criteria, i.e. referentiality, adjacency, compositionality and inflection on the first/second element, which we used in chapter 3 to distinguish between P-constructs and compounds in N + N combinations in MSA, I have been able to show that there are two types of Adj + N combinations in MSA. One of these combinations behaves in a similar way to P-constructs, while the other behaves like a bahuvrihi-compound. Regarding referentiality, inserting the demonstrative *haaḏa* ‘this’ between the two elements shows that the second element of a P-construct is referential, whereas the second element of a compound is non-referential. In addition, the insertion of the demonstrative between the two elements of a P-construct means that an element can intervene between the two parts of the P-construct in MSA. Conversely, no element can intervene between the two elements of the compounds. Concerning compositionality, Adj + N compounds seem to denote a person that cannot be detected from both elements of the compound. In other words, compounds are non-compositional, whereas P-constructs are compositional. With regard to the inflection on the first/second element, free pluralisation and/or free gender marking of the first element are possible with Adj + N compounds. In addition, the second element of compounds is normally singular with few exception, while that of P-construct is plural.

Nevertheless, other criteria such as orthography, sandhi and stress fail to distinguish between the two types of construct, other criteria such as modification and coordination can partially distinguish between P-constructs and compounds.

Based on criteria for adjective-hood and Fassi-Fehri’s (1999) arguments, it has become apparent that output of Adj + N combinations behaves more like a noun than an adjective. I suggest that there is an implicit head, i.e. ‘one’ that determines the syntactic category of the Adj + N output, as will be discussed in detail in section 7.2.

Finally, contrary to Ryding (2005), I have argued that *yayr* ‘not’ and *mutaʕaddid* ‘multi’ are best treated as prefixes.

## Chapter Six: Identifying compounds in Arabic: combinations other than SGC

### 6.1 Introduction

We have seen that compounds can be categorised into different types based on the syntactic category of their internal elements. As noted in chapters 3, 4 and 5, most compounds in Arabic are examples of SGC and the syntactic category of the internal elements is N + N or Adj + N. However, there are certain N + N combinations that are not SGCs. Additionally, other closed sets of compounds may include verbs, adjectives and particles. In JA, it could be argued that there is also a group of V + V combinations, which have similar characteristics to V + V compounds in English. This chapter investigates such further 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 chapter identifies several further types of compounds on the basis of the syntactic category of their internal elements, e.g. V + V, Adj + Adj, etc. Finally, this chapter shows that reduplicated items and some types of numeral are best treated as compounds. We begin the discussion of all this with an analysis of N + N combinations other than SGCs.

### 6.2 Noun + Noun combinations other than SGCs

Various examples of N + N combinations other than SGCs can be found in Arabic. They are illustrated in (1) and (2).

(1)      s'abaaħa                      masaa?  
            morning                      evening  
            'all day long'

(2)      layla                      nahaar<sup>38</sup>  
            night                      daytime  
            'twenty-four seven'

---

<sup>38</sup> This is the form in MSA. In JA, it is phonologically realised as *leel nhaar* 'twenty-four seven'.

In examples (1) and (2), the internal elements of the combinations, *sʿabaaħa* ‘morning’, *masaa?* ‘evening’, *layla* ‘night’ and *nahaar* ‘daytime’, are all nouns. The syntactic category of the output is therefore most plausibly also taken to be a noun, though the function of these combinations is an adverbial of time, as in examples (3) and (4):

(3)      yadrus              tʿ-tʿullaab      sʿabaaħa      masaa?  
             study              the-students      morning      evening  
             ‘The students study all day long.’

(4)      yaſmaal              l-ſummaal              layla      nahaar  
             work              the-employees              night      daytime  
             ‘The employees work twenty four seven.’

The adverbial function of the combinations in examples (3) and (4) 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 can be noted that the first and second N have to be adjacent, and neither the first nor the second elements are referential. Any insertion would result in ungrammaticality, as shown in (5) and (6):

(5)      sʿabaaħa              (\*wa)              masaa?  
             morning              (\*and)              evening  
             ‘all day long’  
             lit. morning and evening

(6)      layla              (\*wa)              nahaar  
             night              (\*and)              daytime  
             ‘twenty four seven’  
             lit. night and daytime

Therefore, the constructs in (1) and (2) are to be treated as compounds.

### 6.3 Noun + Adjective combinations

Ryding (2005: 59-60) suggests that N + Adj combinations in MSA, as in the following examples, are best regarded as phrasal constructs:

- |     |    |   |  |
|-----|----|---|--|
| (7) | a) | l-walad<br>the-boy.MSG<br>'the tall boy'    | t <sup>ʕ</sup> -t <sup>ʕ</sup> awiil<br>the-tall.MSG |
|     | b) | l-ʔawlaad<br>the-boy.MPL<br>'the tall boys' | t <sup>ʕ</sup> -t <sup>ʕ</sup> iwaal<br>the-tall.MPL |
|     | c) | walad<br>boy.MSG<br>'a tall boy'            | t <sup>ʕ</sup> awiil<br>tall.MSG                     |
|     | d) | ʔawlaad<br>boy.MPL<br>'tall boys'           | t <sup>ʕ</sup> iwaal<br>tall.MPL                     |

In all of (7a-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), as shown in (7a-d).

However, while examples like (7a-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:

- |     |               |             |
|-----|---------------|-------------|
| (8) | l-baħr        | l-ʔaħmar    |
|     | the-sea.MSG   | the-red.MSG |
|     | 'the Red Sea' |             |

- (9)      l-baħr                      l-ʔabyad<sup>ʕ</sup>                      l-mutawassit<sup>ʕ</sup>  
             the-sea.MSG                  the-white.MSG                  the-middle.MSG  
             ‘the Mediterranean Sea’  
             lit. the white middle sea
- (10)     l-muħiit<sup>ʕ</sup>                      l-haadii  
             the-ocean.MSG                  the-calm/quiet/pacific.MSG  
             ‘the Pacific Ocean’
- (11)     l-qaamuus                      l-muħiit<sup>ʕ</sup>  
             the-dictionary.MSG      the-comprehensive.MSG  
             ‘the Comprehensive Dictionary’

Similar to example (7), the syntactic category of the combinations in examples (8-11) is the same as that of the first/left element. For instance, example (9) *lbaħr lʔabyad<sup>ʕ</sup> lmutawassit<sup>ʕ</sup>* ‘the Mediterranean sea’ is a noun phrase even though the elements *lʔabyad<sup>ʕ</sup>* ‘the white’ and *lmutawassit<sup>ʕ</sup>* ‘the middle’ are adjectives. However, examples (8-11) are different from example (7) in several other respects. Firstly, although the adjective in examples (8-11) agrees with the noun in number, gender and definiteness, this agreement is purely morphosyntactic. In other words, if the morphosyntactic features of the noun change, and in turn, those of the adjectives follow suit, the meaning of the result will be unacceptable. This can be seen in the following example:

- (12)     \* l-qaawaamiis              l-muħiit<sup>ʕ</sup>a  
             the-dictionary.MPL      the-comprehensive.MPL  
             ‘the Comprehensive Dictionaries’  
             lit. the dictionaries the comprehensives

Example (12) shows that if the adjective *lmuħiit<sup>ʕ</sup>a* ‘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 (7).

Secondly, unlike ordinary N + Adj sequences as in (13), the elements of examples (14-16) are inseparable in the sense that no element can intervene between them, as shown below:

- (13) l-walad          l-dzamiil          t<sup>ʕ</sup>-t<sup>ʕ</sup>awiil  
the-boy.MSG    the-beautiful.MSG    the-tall.MSG  
‘the tall beautiful boy’
- (14) l-baħr          (\*l-waasiŋ)          l-ʔabyad<sup>ʕ</sup>          l-mutawassit<sup>ʕ</sup>  
the-sea.MSG    (\*the-wide.MSG)    the-white.MSG    the-middle.MSG  
‘the wide Mediterranean Sea’  
lit. the wide white middle sea
- (15) l-baħr          l-ʔabyad<sup>ʕ</sup>          (\*l-waasiŋ)          l-mutawassit<sup>ʕ</sup>  
the-sea.MSG    the-white.MSG    (\*the-wide.MSG)    the-middle.MSG  
‘the wide Mediterranean Sea’  
lit. the white wide middle sea
- (16) l-qaamuus          (\*l-dzadiid)          l-muħiit<sup>ʕ</sup>  
the-dictionary.MSG    (\*the-new.MSG)    the-comprehensive.MSG  
‘the new Comprehensive Dictionary’  
lit. the dictionary the new the comprehensive

In example (13), the adjective *ldzamiil* ‘the beautiful’ can be inserted between the two elements of the phrase *lwalad t<sup>ʕ</sup>t<sup>ʕ</sup>awiil* ‘the tall boy’. But no such insertion is allowed in examples (14-16). If the adjective *lwaasiŋ* ‘the wide’ is inserted either between the internal elements *lbaħr* ‘the sea’ and *lʔabyad<sup>ʕ</sup>* ‘the white’ or between *lʔabyad<sup>ʕ</sup>* ‘the white’ and *lmutawassit<sup>ʕ</sup>* ‘the middle’, the result is not acceptable. The same applies to example (16). Note that the adjectives found in examples (14) and (15) do not behave as normal adjectives in terms of agreement, i.e. number (see example 12). This could be due to that fact that examples (8-11) 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: 214-15) points out that since adjectives in



Dutch in Adj + N combinations can be marked with the final inflectional ending *-e* (pronounced as schwa), it is evident that Dutch Adj + N combinations are phrases. However, some Adj + N combinations like this can nevertheless 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, between 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 grouped under ‘phrasal compounds’.

#### 6.4 Adjective + Adjective combinations

In MSA and JA, there exists a closed set of Adj + Adj coordinative compounds (see chapter 7.3 for more detail) that are characterised by a semantic relationship between their internal elements in which the meaning of the whole compound is a combination of both elements. Examples of this type can be seen in (17) and (18):

(17)    ħaamid<sup>s</sup>        ħilw  
           sour            sweet  
           ‘sweet-and-sour’

(18)    ħilw                murr  
           sweet            bitter  
           ‘bitter-sweet’

In examples (17) and (18), the syntactic category of the output is the same as that of the internal elements, i.e. adjective. The meaning of the whole compound in (17), ‘sweet-and-sour’, is derived from both elements *ħaamid*<sup>6</sup> ‘sour’ and *ħilw* ‘sweet’. The two elements of the compound must be adjacent. Examples like (19a, b) are ungrammatical:

- (19) a.    *ħilw*            (\**wa*)            *murr*  
           sweet            (\*and)            bitter  
           ‘bitter and sweet’
- b.    *ħilw*            (\**ħaamid*<sup>6</sup>)        *murr*  
           sweet            (\*sour)            bitter  
           ‘bitter, sour and sweet’

Example (19a, b) shows 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 most compounds in examples (17) and (18). Specifically, *murr* ‘bitter’ denotes an opposite meaning to the adjective *ħilw* ‘sweet’.

## 6.5 Particle + Adjective combinations

Similar to the constructs whose first element is *yayr* ‘not’ (see section 5.5), Ryding (2005: 100) also discusses examples as in (20) and (21), which she refers to as compounds.

- (20)    *laa*            *faqaari*  
           no            spine  
           ‘invertebrate’
- (21)    *laa*            *markaziyyah*  
           no            centralisation  
           ‘decentralisation’

However, I would argue that, similar to *yayr* ‘not’ (cf. section 5.5), *laa* ‘no’ is to be considered a prefix. This would explain why *laa* ‘no’ cannot stand on its own, as in the following examples:

(22)    laa            silki  
           no            wired  
           ‘wireless’

(23)    laa            ʔaxlaaqi  
           no            moral  
           ‘immoral’

Examples (22) and (23) 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 those of English, e.g. *un-*, *il-*, *im-*, *in-*, etc.

Overall, contrary to Ryding (2005), *laa* ‘no’ should be treated as a prefix. Therefore, examples (20-23) are instances of derived words, rather than compounds.

## 6.6 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-Rajihī 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’. These verbs are:

(24)    maa            zaala  
           not            ceased  
           ‘didn’t cease/continue to be’

(25)    maa            bariha  
           not            left  
           ‘didn’t leave’

(26)    maa            nfakka  
           not            be disconnected  
           ‘didn’t get disconnected’

(27)    maa            fatiʔa  
           not            ended  
           ‘didn’t 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 (24-27). In other words, it cannot stand alone to convey that sense.

### 6.7 Verb + Verb compounds or serial verbs

The category of V + V combinations seems to be absent from MSA, though examples of it appear in JA, as in (28) and (29):

(28)    ali        gaam            waggaf        lamma ʔabuu-h    ʔidza  
           Ali    rose.3SG        stood up.3SG when    father-his        came. 3SG  
           ‘Ali rose and stood up when his father came.’

(29)    aħmad        ʔaxað        ħaka        fi        t-talafo:n  
           Ahmad        took.3SG.M    talked.3SG.M in        the-telephone  
           ‘Ahmad picked up the phone and talked on the phone.’

Examples (28) and (29) are typical cases of V + V combinations in JA. The intervention of another element between the internal elements of V + V combination is prohibited in these examples. This may suggest that these combinations could be regarded as compounds. However, the criterion of non-separability has been suggested to identify serial verbs. In particular, Aikhenvald (2006: 1) suggests that serial verbs are a sequence of juxtaposed verbs that serve as a single predicate, without any overt intervening element, e.g. a marker of subordination, coordination or syntactic dependency. This type of construction denotes one single event and shares core and other arguments. The next section investigates the criteria for verb serialisation and applies them to JA V + V combinations in order to determine whether these V + V combinations are best analysed as compounds or as serial verbs.

## **6.7.1 Serial verb criteria**

### ***6.7.1.1 Overview***

Several linguists (e.g. Foley and Van Valin 1984; Foley and Olson 1985; Crowley 1987, 2002; Durie 1988; Lord 1993; Muysken and Veenstra 1995; Aikhenvald 2006; Bisang 2009; Van Breugel 2014 among others) discuss criteria that can be used to identify serial verb construction cross-linguistically. Pragmatically, the literature suggests that serial verbs express the results of causing events, add noncore arguments like instrument or beneficiary to the clause or add motion components to events (Foley and Van Valin 1984). It has also been noted that serial verbs are characterised by certain morphosyntactic properties (Durie 1988: 3). Firstly, both verbs stand on an equal footing, meaning that neither one of them is dependent on the other. Secondly, serial verbs share one or more core arguments. Thirdly, there is no morphological or intonational marker of a clause boundary separating them, meaning that they are one unit. Finally, the verbs do not have independent scope for mood, aspect, tense, negation or illocutionary force, that is, these features are shared by verbs in a serial verb construction (Durie *ibid*). Muysken and Veenstra (1995: 293-301) expand the list of serial verb criteria. A serial verb construction should have: (1) only one expressed subject; (2) one expressed direct object; (3) one specification for tense/aspect (only on the first verb, or on both verbs/but semantically one specification/, or only on the second verb); (4) only one possible negator; (5) no intervening coordinating conjunction; (6) no intervening subordinating conjunction: and (7) no intervening pause possible. In addition to the previous criteria, some researchers (e.g. Aikhenvald 2006: 1; Bisang 2009: 796) suggest that serial verb construction usually express a single event. Here, I compile these criteria to form a list of five tests for serialisation where criteria 5, 6 and 7 are combined under one criterion, i.e. no intervening element can appear between the two verbs. In the next section, these criteria are applied to V + V combinations in JA to decide whether such combinations are compounds or serial verbs.

### ***6.7.1.2 Applying verb serialisation criteria to V + V combinations in JA***

In testing the potential serial verb status of the relevant V + V combinations in JA, let us first consider the criterion of inseparability. We indeed find that no intervening coordinating

conjunction, subordinating conjunction or pause is allowed in JA.<sup>39</sup> The two verbs appear to be closely tied together in a single predicate. Examples are (30) and (31):

- (30)    ʕali    gaam<sup>40</sup>       waggaf        lamma ʔabuu-h       ʔidza  
           Ali    rose.3SG       stood up.3SG when father-his       came. 3SG  
           ‘Ali rose and stood up when his father arrived.’

- (31)    ʕali    gaam        ħaka        fi    l-leel  
           Ali    rose.3SG       talked.3SG in    the-night  
           ‘Ali sleep-talked at night.’

In some languages, if the first of the two serial verbs is transitive, an object noun phrase (NP) can appear between the two verbs (Givón 1979; Foley 1997). Examples of this case are presented below (cited in Givón 1979: 13-15):

- (32)    wó    lá    shnáknu       bā ya       (Gwari, Hyman, 1971)  
           he    take pot       break  
           ‘he has broken the pot.’

- (33)    mo    fi    àdá        gé    igi    (Yoruba, Stahlke, 1970)  
           I    take machete cut wood  
           ‘I cut the wood with the machete.’

Examples (32) and (33) show that the objects of the transitive verbs *lá* ‘take’ and *fi* ‘take’, i.e. *shnáknu* ‘pot’ and *àdá* ‘machete’ intervene between the serial verbs *lá* ‘take’ and *bā ya* ‘break’

---

<sup>39</sup> Some speakers of JA allow the coordinating conjunction, i.e. *wa* ‘and’ between the two verbs in some V + V combinations. These can be analysed as straightforward examples of coordination of Vs or VPs. The V + V examples given in (30) and (31) are ones where *wa*-insertion would be rare or impossible, suggesting they form a different construction. In this regard, speculations about the possibility to view some types of the English multi-verb sequences as serial verbs did not lead to conclusive results. A sequence like ‘go get the book’ is excluded from the list of potential candidates since, as Crowley (2002: 11) argues, the same meaning can be conveyed using coordination, e.g. *go and get the pen* or subordination, e.g. *go to get the pen*.

<sup>40</sup> The verb *gaam* ‘rose’ can be also used as defective verb. In this regard, *gaam* is both used as an emphatic marker that means ‘did’ when it is followed by a past tense verb, e.g. *ali gaam liʕib fatʔbool gabil saadʕa* ‘Ali did play football an hour ago’. It also functions as a present perfect maker that means ‘he has been doing’ when it is followed by a present tense verb, e.g. *ali gaam yiʕab fatʔbool* ‘Ali has been playing football’. There is indeed layering, since both the lexical meaning and grammatical meaning exist side by side. In this study, only the lexical meaning of the verb *gaam*, i.e. ‘rose’ is used.

in (32) and *fī* ‘take’ and *gé* ‘cut’ in (33). This suggests that an intervening element can be inserted between the two serial verbs at least in some languages. Similarly, Foley (1997: 382) acknowledges the possibility of an intervening conjunction, e.g. the sequential marker *mpi* in Yimas, between serial verbs. Specifically, Foley suggests that serial verb constructions can be viewed as strings of adjacent verb stems, normally without an overt conjunction, which have at least one core argument in common.

Interestingly, in JA, even when the first verb is transitive, no object NP appears between the two elements. This is illustrated with the following examples:

- (34) waliid            misik            fataħ            kull    ūlab l-pepsi  
 Walid            grabbed.MSG opened.MSG all    cans the-Pepsi  
 ‘Walid grabbed and opened all Pepsi cans.’

- (35) saayig            t-taxi            fayyal                            ħarrak            s-siyyaarah  
 driver            the-taxi            kick-started.MSG moved.MSG the-car  
 ‘The taxi driver kick-started and moved the car.’

In examples (34) and (35), the object is placed after the V + V combination and, semantically, it is shared by the two verbs. In (34), the QP *kull ūlab l-pepsi* ‘all cans of Pepsi’ denotes the object which the subject, i.e. *Waliid*, grabbed and opened. In (35), the NP *ssiyyaarah* ‘the car’ is used to denote the object which the subject, i.e. *saayig ttaxi* ‘the taxi driver’, kick-started and moved. In fact, if the objects, i.e. *kull ūlab l-pepsi* ‘all cans the-Pepsi’ and *ssiyyaarah* ‘the car’, are placed between the two verbs, the sentences will be ungrammatical, as in:

- (36) \*waliid            misik            kull    ūlab l-pepsi            fataħ  
 Walid            grabbed.MSG all    cans the-Pepsi            opened.MSG  
 ‘Walid grabbed and opened all Pepsi cans’  
 lit. Walid grabbed all Pepsi cans opened.

- (37) \*saayig t-taxi            fayyal                            s-siyyaarah            ħarrak  
 driver the-taxi            kick-started.MSG the-car            moved.MSG  
 ‘The taxi driver kick-started and moved the car.’  
 lit. the taxi driver kick-started the car moved.

Therefore, it can be proposed that the verbs in V + V combinations in JA are inseparable. No element can be inserted between the two elements of the construct even if the first element is a transitive verb and would be in other cases followed by an object.

Secondly, the meanings of the two serial verbs together often constitute a single complex event. For instance, the verbs in *She took the book and came* seems to denote the same complex event denoted by the verb *bring* ‘get something and take it to your destination’. This can be seen in V + V combinations in JA, as in (38-40):

(38) mhammad tʿaar safar ʕala ʔlurdun  
 Mohammad flew.3SG travelled.3SG to Jordan  
 ‘Mohammad flew to Jordan.’

(39) l-xaatim dʿaaʕ ʔixtafa  
 the-ring lost.3SG got away.3SG  
 ‘The ring disappeared.’

(40) guum ʔingaliʕ min hoon  
 rise.2MSG go.2MSG from here  
 ‘Just get the hell out of here!’

In examples (38-40), the two verbs in the V + V combination denote a single complex event or action. For instance, in example (38), *tʿaar safar* ‘flew and travelled’ denotes the same action denoted by the verb *fly* ‘travel by plane’. In example (39), *dʿaaʕ ʔixtafa* ‘lost and got away’ denotes the same action denoted by the verb *disappear*. Finally, in example (40), the V + V combination *guum ʔingaliʕ* ‘rise and go’ denotes the same complex action denoted by the verb *leave*.

Thirdly, the two finite verbs in a verb serialization construction must have the same subject, which is essential to support the argument that, together, the two verbs in the construction make up one clause. Examples from JA to illustrate are given in (41) and (42):

(41) ʕali ʔaxað haka fi t-talafo:n gabil saaʕah  
 Ali took.MSG talked.MSG in the-telephone before hour  
 ‘Ali talked on the telephone an hour ago’



- (42)    ʕali   wigiʕ            tzahlag       fi-l-madrased  
           Ali    fell.MSG       slipped.MSG   in-the-school  
           ‘Ali slipped and fell at the school’

The subject, *Ali*, is shared by the two verbs *ʔaxað* ‘took’ and *ħaka* ‘talked’ in (41) and *wigiʕ* ‘fell’ and *tzahlag* ‘slipped’ in (42). This is demonstrated by the fact that both verbs are marked with the same morphosyntactic features, i.e. MSG, which refer to one subject, i.e. *Ali*.

A fourth criterion for serialisation is that there is only one marker of negation for the whole serial verb construction. Examples showing this in JA are (43) and (44):

- (43)    a.    s<sup>ʕ</sup>-s<sup>ʕ</sup>ahin    maa   wigiʕ                    nkasar       mbaarih  
           the-plate    not    fell down.3SG        broke.3SG    yesterday  
           ‘The plate did not fall down and break yesterday.’

- b.    \*s<sup>ʕ</sup>-s<sup>ʕ</sup>ahin    maa   wigiʕ                    \*maa   nkasar       mbaarih  
           the-plate   not    fell down.MSG        \*not   broke.MSG    yesterday  
           ‘The plate did not fall down and break yesterday.’  
           lit. the plate not fell not broke yesterday.

- (44)    a.    ʕali    maa    gaam                waggaf        lamma jaafn-i  
           Ali    not    rose.3SG        stood up.3SG   when   saw.he-me  
           ‘Ali did not get up when he saw me.’

- b.    ʕali    gaam            (\*maa)        waggaf        lamma jaafn-i  
           Ali    rose.3SG        (\*not)        stood up.3SG   when   saw.he-me  
           ‘Ali did not get up when he saw me.’  
           lit. Ali rose not stood up when he saw me.

Examples (43a and 44a) show that the two verbs share one negative marker, i.e. *maa* ‘not’. Note that the negative marker is placed before the first element of the V + V construct, allowing it to negate the whole clause. In contrast, example (43b) shows that the two serial verbs cannot be marked separately with the negative marker, i.e. *maa* ‘not’. Example (44b) demonstrates that the negative marker, i.e. *maa* ‘not’ cannot be inserted between the two serial verbs.

Finally, a fifth criterion that identifies serial verbs, making them behave as one clause, is sharing the same grammatical categories, e.g. tense, aspect, mood, etc. This means that the two verbs in a serial verb construction are not marked separately for these categories. Consequently, such categories are either identically marked on each verb or they appear just once but are shared by both verbs. The following examples show how this criterion applies to V + V combinations in JA, where the same grammatical categories are marked on each verb separately:

- (45) waliid            gaam            maʃa            fi            l-leel  
 Walid            rose.3SG            walked.3SG            in            the-night  
 ‘Walid sleepwalked last night.’  
 lit. Walid rose (and) walked at night.

- (46) s<sup>ʕ</sup>-s<sup>ʕ</sup>aħin            wiggif                            nkasar            mbaariħ  
 the-plate            fell down.3SG            broke.3SG            yesterday  
 ‘The plate fell down and broke yesterday.’

The verbs in examples (45) and (46) are marked similarly for tense. In (45), both the first verb and the second verb are marked for past tense (*gaam* ‘rose’ and *maʃa* ‘walked’); the whole V + V combination is marked for past tense. The same applies to (46), in which each verb is marked for past tense and, in turn, the whole V + V construct is marked for past tense. Looking at these examples, it seems that the fifth criterion of serial verb constructions applies to V + V constructions in JA too, since both verbs are marked for the same tense. Here, it is important to point out that both verbs in V + V combinations in JA should be marked with tense, since there is no verb base or stem that can be used independently (cf. the next section). Interestingly, a few examples of V + V + V combinations can be found in JA in (47):

- (47) guum            ʃayyil                            ħarrik            s-siyyarah  
 rise.MSG            kick-start.MSG            move.MSG            the-car  
 ‘Stand up, start and move the car.’

Example (47) usually occurs in the imperative form, where someone asks the addressee to stand up, turn the engine on and move the car.

Now that the verb serialisation criteria have been applied, it seems that V + V combinations in the examples above in JA behave similarly to serial verbs. Although these facts in themselves are certainly interesting (and have not been noted in the literature on JA or -- it appears -- on other varieties of Arabic), the analytic question of which criteria differentiate between V + V compounds and serial verbs remains unanswered. While at this point the jury is still out on V + V combinations in JA, the compoundhood of V + V combinations in English has already been established (Miller 2014: 56). Therefore, in the next section, the above-mentioned criteria are applied to what have been established in the literature as V + V compounds in English in order to determine the properties of this construct.

### ***6.7.1.3 Applying verb serialisation criteria to V + V compounds in English***

Taking English into account, it seems that applying the above criteria to what have been always treated as V + V compounds, does not provide a clear picture of the difference between serial verbs, on the one hand, and compounds, on the other. Examples of V + V compounds in English are *test-release*, *dry-burn*, *stir-fry*, *kick-start*, *corkscrew*, *freeze-dry*, *crash-land*, *spin-dry*, *sleep-walk*, *sleep-talk*, *drip-dry*, *shrink-wrap* and *force-feed* (Bauer and Renouf 2001: 110; Payne 2011: 100, 330; Miller 2014: 56 among others). There is no definitive solution concerning the differences/similarities between serial verbs and compounds, which means that applying the above-mentioned criteria is essential. This is illustrated in Table 6.1.

Table 6.1. The application of the criteria characterising serial verb construction to English V + V compounds

No.	Criterion	Example
1.	No element can be inserted between the two verbs	The government is <i>kick-starting</i> the economy by different means. *the government is kicking and starting the economy by different means
2.	The two verbs denote a single complex event	Alex <i>stir-fried</i> the vegetables (a complex event of stirring and frying that occur at the same time).
3.	The two finite verbs in a verb serialisation construction must have the same subject	The plane has <i>crash-landed</i> in the next field (both verbs, i.e. <i>crash</i> and <i>land</i> have the same subject, i.e. <i>the plane</i> ).
4.	Only one marker of negation for the whole construction	This machine does not <i>spin-dry</i> this kind of clothes (one negator for the two verbs)
5.	The two verbs share the same grammatical categories (e.g. tense, aspect, mood, etc.).	Dad <i>sleepwalked</i> last night (i.e. Dad walked while sleeping, all happened last night). It seems that the two verbs share the tense, even though it is realised only on the second verb.

Table 6.1 shows that the criteria used to identify serial verb constructions apply to V + V compounds in English. The only difference between V + V compounds in English and V + V combinations in JA is that in the latter, the two verbs are marked separately, though identically, for tense. In fact, Donohue (2003: 126) suggests that:

...serial verbs come in two forms: there can be two constructions with two fully inflected verbs, constrained so that they meet the above criteria, or, in some languages, the two verb roots can be adjacent, sharing a single set of inflectional affixes.

Donohue (ibid) points out that some linguists (e.g. Crowley 1987) have called the latter form, the one in which both verbs share a single set of inflexional affixes, ‘compounding’. However, this single-affix-set requirement is not conclusive and may not apply to all languages if they have inflectional markings. For instance, in JA, verbs always have to be marked for tense, since no stem can be used on its own. Specifically, the root *ktb* denotes ‘write’, but it can only ever surface in an inflected form, such as *katab* ‘he wrote’, *buktub* ‘he is writing’, *btuktub* ‘she is writing’, etc. Similarly, Crowley (2002: 18) suggests a structural continuum with V + V compounds at its maximum pole, whereas coordinate clauses are at its minimum pole, and serial verbs are in-between. Again, Crowley does not explain the grounds for proposing such a

continuum. In addition, he possibly has not taken into consideration that there are languages to which this classification does not apply, e.g. JA, which does not have uninflected verbs. Contrary to Crowley (2002), Van Breugel (2014: 367) suggests that verb serialisation in Atong is actually verb compounding, since the putative serial verbs in this language are combined to form one phonological word.

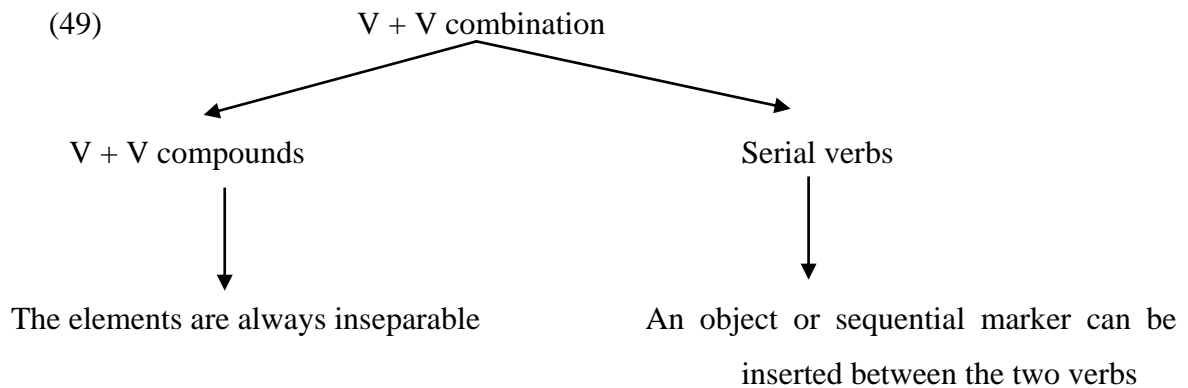
Altogether, it is clear that the criteria found in the literature to identify compounds and serial verbs show a great deal of overlap. There could therefore be doubts about how useful it is to try and classify V + V combinations in a language as either compounds or serial verbs. Nevertheless, it could be argued that there is a difference. This is due to the fact that serial verbs in some languages accept the insertion of an object between them if they are transitive (cf. Givón 1979 as in examples (32) and (33)) or the insertion of a sequential marker (cf. Foley 1997), violating the adjacency criterion. When no such insertion is allowed, we could say that the two elements form a V + V compound. This is supported by JA examples (34) and (35) in which the object has to be placed after the second verb, not in-between. Here, it is worth pointing out that despite the fact that the sentence in (48a) is grammatically correct, it is still marked. JA speakers tend to use example (48b), since it has a less complicated structure than example (48a).

- (48) a) ?waliid misik kull ʃulab l-pepsi wu fataħ-hin  
 Walid grabbed.MSG all cans the-Pepsi and opened-them  
 ‘Walid grabbed all Pepsi cans and opened them.’
- b) waliid misik fataħ kull ʃulab l-pepsi  
 Walid grabbed.MSG opened.MSG all cans the-Pepsi  
 ‘Walid grabbed and opened all Pepsi cans.’

#### 6.7.1.4 Summary

This section has investigated V + V combinations in JA. In particular, I applied verb serialisation criteria drawn from the literature to V + V combinations in JA. The result was that JA indeed has serial verbs. When it comes to the question whether these serial verbs can be classed as compounds, things are more difficult. The analysis shows that the distinction between V + V compounds and serial verbs is not clear-cut. Nevertheless, I concluded that V

+ V compounds are different from serial verbs based on the adjacency criterion, as illustrated below:



Whether this distinction is useful also for V + V combinations in other languages must await further investigation.

## 6.8 Reduplication

Another category of compounding mentioned by traditional Arab grammarians writing about MSA (e.g. Al-Rajihī 2000: 75) involves reduplicated words. The internal elements of such compounds are normally nouns. For example, in (50-54), the words *layl* ‘night’, *nahaar* ‘daytime’, *yawm* ‘day’ and *bayt* ‘house’ are all nouns, while *bayn* ‘between’ is a preposition.

(50)    *layla*        *layl*  
           night        night  
           ‘every night’

(51)    *nahaara*        *nahaar*  
           daytime      daytime  
           ‘all day long’

(52)    *yawma*        *yawm*  
           day            day  
           ‘daily’

(53) bayna bayn  
 between between  
 ‘in-between’

(54) bayta bayt  
 house house  
 ‘close in distance’

Examples (50-54) show that the meaning of these reduplicated words can be compositional or non-compositional. In examples (50-53), the meanings of the whole compound could be predicted from the meanings of the internal elements, whereas example (54) denotes a meaning that is unpredictable from the elements of the compound.

Regarding the function of these compounds, reduplicated compounds seem to exclusively function as time or place adverbials. That is, the compounds, i.e. *yawm yawm* ‘daily’ or *bayt bayt* ‘close in distance in reference to a building’ have an adverbial function, as in examples (55) and (56):

(55) yaʕmal muḥammad yawma-yawm  
 work Mohammad daily  
 ‘Mohammad works daily.’

(56) yaʕiif muḥammad wa ʕaliyy bayta-bayt  
 live Mohammad and Ali close in distance  
 ‘Mohammad and Ali live close to each other.’

Reduplicated compounds can be found in JA too, as in examples (57-59):

(57) kθiir kθiir  
 much much  
 ‘very much’

(58) ʃway ʃway  
 little little  
 ‘slowly’

- (59)    *basʿiitʿah*      *basʿiitʿah*  
          simple          simple  
          ‘an expression denoting threat’

In examples (57-59) the internal elements of the replicated words *kθiir* ‘much’, *ʃway* ‘little’, and *basʿiitʿah* ‘simple’ are all adjectives. Regarding the function of the whole reduplicated compounds, it seems to vary in JA. Examples (57) and (58) can be used in an adverbial position, while example (59) functions as an interjection, as in (60) and (61) respectively:

- (60)    *miʃaan*                      *ʔallah suug*                      *ʃway ʃway*  
          for.the sake                  Allah drive.you                  little little  
          ‘For Allah’s sake, drive slowly.’

- (61)    *basʿiitʿah*      *basʿiitʿah*      *rah ʔaħki la ʔabuuy*  
          simple          simple          will tell.I to father.my  
          ‘Just you wait, I will tell my father.’<sup>41</sup>

Note that these reduplicated items do not accept insertion of any other elements, as shown in (62) and (63):

- (62)    *layla (\*maʃ/fii)*      *layl*  
          night (\*with/in)      night  
          ‘every night’  
          lit. night with/in night

---

<sup>41</sup> The meaning of *basʿiitʿah basʿiitʿah* 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.



- (63)    *ʃway*    (\**wu*)            *ʃway*  
           *little*    (\**and*)            *little*  
           ‘slowly’  
           lit. *little and little*

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 (2001: 69) states that whole word reduplication can be considered 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 (2001) argument here looks plausible, since the internal elements are meaningful words that can stand alone. Henri (2012: 215) agrees with Fabb (2001) that reduplication 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. However, 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. However, this option of multiple compounding is a rather recent one in the language, being attested only after c.1800, (e.g. Biber and Gray 2011: 237). With regard to the second point, many undoubted compounds have the same category as their components, as in *bookshop*, *bitter-sweet* and *stir-fry*. So the reduplicated compounds in Arabic in (50-59) 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 though they were simple lexical items.

Bauer *et al.* (2013: 463, 490) too 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 (Bauer *et al.* *ibid.*). Based on the above discussion, it seems that several researchers acknowledge that full reduplicated items are compounds.

Another characteristic could be proposed to treat 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 *fway fway* ‘slowly’, where the meaning of the combination cannot be straightforwardly deduced from that of the individual elements. It has been argued in chapters 2 and 4 that if a word is non-compositional, it has to be treated as a compound; this is a characteristic that is also shared by reduplicated items in the above examples.

Note that similar 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 compositional, since its meaning is derived from the meaning of the individual words. In contrast, the meaning of the whole compound *fway fway* ‘slowly’ is non-compositional, because its meaning (‘slowly’) is rather different from the meaning of its internal elements (with *fway* meaning ‘little’).

## 6.9 Numerals

Several linguists (e.g. Al-Rajih 2000, Al-Humaydi 2005: 243; Booij 2010b; Hurford 2011; Al-Hariiri 2013: 175 among others) have examined the structure of numerals in various languages, suggesting that a subset of these 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 (64):

- (64) a. *xamsah* ‘five’  
 b. *xamsat* *ʕafar*<sup>42</sup> ‘fifteen’  
 c. *xamsah* *wa* *xamsuun* ‘fifty five’ (lit. five and fifty’)  
 d. *miʔah* *wa* *xamsah* ‘one hundred and five’ (lit. hundred and five’)

In MSA, all numerals above 10 are complex expressions. For example, unlike examples (64c, d) whose two elements are separated by a coordinating conjunction, the numeral *xamsat ʕafar* ‘fifteen’ in (64b) looks like a compound, consisting of two separate elements, *xamsah* ‘five’ and *ʕafar* ‘ten’. Therefore, Al-Rajih (2000: 75-76) considers cardinal numerals from eleven to nineteen to be compounds. Below is the full sequence from 11 to 19:

- (65) *ʔahad/ ʔiθnaa/ θalaaθat/ʔarbaʕat/xamsat/sittat/sabʕat/θamaaniyat/tisʕat ʕafar*  
 one/two/three/four/five/six/seven/eight/nine ten  
 ‘eleven/twelve/thirteen/fourteen/fifteen/sixteen/seventeen/eighteen/nineteen’

The N + N combinations in example (65) can be regarded as compounds, since they are inseparable; no element can be inserted between the internal elements of the compound. For example, it is ungrammatical to say:

- (66) *xamsat* (\**wa*) *ʕafar*  
 five (\*and) ten  
 ‘fifteen’  
 lit. five and ten

The Arabic numerals in (65) seem to be quite similar to Dutch, English and German numerals. Therefore, I analyse Arabic numerals with special focus on Dutch numeral as analysed by Booij (2010b).

As Booij (2010b: 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

<sup>42</sup> In JA, *tʕaʕif* ‘-teen’ is used instead of *ʕafar* ‘ten’ in numerals between eleven and nineteen. Note that *tʕaʕif* ‘-teen’ cannot stand on its own and *ʕafar* ‘ten’ is used in JA to mean ten, e.g. *ʕafar ʕyaal* ‘ten boys’.

above the number twelve are complex expressions. For instance, the numeral *vijf-tien* ‘fifteen’ in Dutch has the shape of a compound, because it consists of the two lexemes *vijf* ‘five’ and *tien* ‘teen’. It also has the stress pattern of Dutch compounds, with the main stress on the first element (Booij 2010b: 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 element of *vijftien*, i.e. *tien* ‘ten’ does not have the features characterising semantic heads (Booij 2010b: 88). This is accounted for by viewing this special type of compounds as being historically derived from (asyndetic) coordination (Booij *ibid*). 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. The members of these compounds are verb stems plus nouns which refer to body parts. The left member is a verb stem and the whole compound is a verb as follows (Don 2009: 379):

- (67) *schuddebuik* ‘lit. shake-belly’ ‘shake with laughter’ (*schud* ‘shake’+ *buik* ‘belly’)
- (68) *reikhals* ‘lit. reach-neck’ ‘reach anxiously’ (*reik* ‘reach’+ *hals* ‘neck’)
- (69) *stampvoet* ‘lit. stamp-feet’ ‘stamp with rage’ (*stamp* ‘stamp’+ *voet* ‘feet’)

Here, it is worth noting that, in Dutch left-headed compounds in (67-69), the non-head is always an argument of the verb. Second, 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.<sup>43</sup>

Investigating other types of numeral, Booij (2010b: 85) notes that examples (70) and (71) have the appearance of phrases due to being formed by means of coordination with the conjunction *en* ‘and’:

---

<sup>43</sup> Some researchers are still debating whether *white collar* is an adjective or a noun.

- (70) een-en-vijftig ‘51’ ‘one-and-fifty’  
 (71) honderd (en) vijf ‘105’ ‘one hundred (and) five’

Examples (70) and (71) demonstrate that these numerals appear like phrases.<sup>44</sup> However, they can serve as bases of word-formation, especially for the formation of ordinal numerals by adding the suffixes *-ste* and *-de* (Booij 2010b: 85), as in (72a, b):

- (72) a. een-en-vijftig-ste ‘one-and-fifty-th, fifty-first’  
 b. honderd(-en)-vijfde ‘hundred (and) fifth’ Booij (2010b: 85)

As a result, numerals in examples (72a, b) are best treated as words, or it can be claimed that morphological operations may take phrases as their bases (Booij 2010b: 85). The same seems to apply partially to MSA numerals through adding the prefix *l-*, as in the following example:<sup>45</sup>

- (73) a. *l-xamsah* ‘fifth’ (lit. the-five’)  
 b. *l-xamsat* *ʕaʕar* ‘fifteenth’ (lit. the-five ten’ )  
 c. *l-xamsah wa l-xamsuun* ‘fifty-fifth’ (lit. the-five and the-fifty’)  
 d. *l-miʔah wa xamsah* ‘one hundredand fifth’ (lit. the-hundred and five’)

Examples (73a, b and d) show that by adding the prefix *l-* to the first element, cardinal numbers change into ordinal numbers. Example (73c) 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 (73a, b) raise the question as to what extent their formation is morphological or syntactic. This issue needs further investigation; therefore, the argument will rest here for the time being.

---

<sup>44</sup> According to Booij (2010b), it seems that there are two differences between examples (70) and (71); firstly, it is possible to delete the conjunction in (71). Secondly, in (70), the conjunction *en* is pronounced as [ən], whereas in (71) it must be pronounced as [ɛn].

<sup>45</sup> The prefix *l-* normally functions as a definite article in Arabic. However, here it does not; it changes the number from cardinal to ordinal.

## 6.10 Summary

In sum, with regard to the classification of compounds in terms of the syntactic category of their internal elements, there are five putative types in MSA, and four in JA. In the former, V + V combinations are absent, whereas Prep + Prep and Adj + N combinations are missing from the latter. It seems that Arabic in general and MSA in particular do not have a wide diversity in the internal word classes of compounds. Table 6.2 shows the components of possible syntactic categories of compounds in MSA and Table 6.3 shows those used in JA. MSA has a small number of compounds with word classes other than N + N and Adj + N combinations.

Table 6.2. All possible combinations of compounds in terms of word class in MSA

First element	First element	Examples	Gloss
Noun	Noun	<i>s'abaaha masaa?</i> 'twenty four-seven' <i>findzaan lqahwa</i> 'the coffee cup' <sup>46</sup>	<i>s'abaaha</i> 'morning' + <i>masaa?</i> 'evening' <i>findzaan</i> 'cup' + <i>lqahwa</i> 'the coffee'
Adjective	Noun	<i>haadd lbas'ar</i> 'sharp eyed' <sup>47</sup>	<i>haadd</i> 'sharp' + <i>lbas'ar</i> 'the sight'
Noun	Adjective	<i>lbaħr l?abyad<sup>s</sup> lmutawassit<sup>s</sup></i> 'the Mediterranean sea'	<i>lbaħr</i> 'the sea' + <i>l?abyad<sup>s</sup></i> 'the white' + <i>lmutawassit<sup>s</sup></i> 'the middle'
Adjective	Adjective	<i>haamid<sup>s</sup> hilw</i> 'sour-sweet'	<i>haamid<sup>s</sup></i> 'sour' + <i>hilw</i> 'sweet'
Preposition	Preposition	<i>bayna bayn</i> 'in-between'	<i>bayna</i> 'between' + <i>bayn</i> 'between'

Table 6.3. All possible combinations of compounds in terms of word class in JA

First element	First element	Examples	Gloss
Noun	Noun	<i>leel nhaar</i> 'twenty four-seven' <i>findzaan lgahwa</i> 'the coffee cup'	<i>leel</i> 'night' + <i>nhaar</i> 'morning' <i>findzaan</i> 'cup' + <i>lgahwa</i> 'the coffee'
Noun	Adjective	<i>lbaħr l?abyad<sup>s</sup> lmutawassit<sup>s</sup></i> 'the Mediterranean sea'	<i>lbaħr</i> 'the sea' + <i>l?abyad<sup>s</sup></i> 'the white' + <i>lmutawassit<sup>s</sup></i> 'the middle'
Adjective	Adjective	<i>haamid<sup>s</sup> hilw</i> 'sour-sweet'	<i>haamid<sup>s</sup></i> 'sour' + <i>hilw</i> 'sweet'
Verb	Verb	<i>guum ?ingali<sup>s</sup></i> 'fuck off and leave'	<i>guum</i> 'rise' + <i>?ingali<sup>s</sup></i> 'fuck off and leave'

Table 6.2 and 6.3 show that in terms of the syntactic category of the internal elements, there are five types in MSA, and four in JA. The available syntactic categories for compounding in

<sup>46</sup> This is an instance of SGC, as explained in chapters 3 and 4.

<sup>47</sup> This is an instance of SGC, as explained in chapter 5.

MSA are nouns, adjectives and prepositions. On the other hand, JA has nouns, adjectives and verbs.

On the basis of the internal word classes of compounds in 23 languages (which do not include Arabic), Scalise and Vogel (2010: 10) propose that there are 110 compound types. In these types, the hierarchy of the preference of the output is the same as that of the syntactic categories of the input, as shown below:

(74) N > Adj > V > Adv > Prep

This order means that regarding the relative frequency of the various syntactic categories of compound types, a clear hierarchy can be identified (Scalise and Vogel 2010: 10). A corresponding hierarchy for compounding in MSA would look as in (75):

(75) N > Adj > Prep

The hierarchy in (75) means that nouns are more likely to be the internal element in compounds in MSA, followed by adjectives, and a few cases of prepositions.

The order of the categories in MSA and JA is the same, the only difference is that the two varieties of Arabic do not form compounds with all available categories. In JA, V + V combinations exist, as in:

(76) N > Adj > V

The hierarchy in (76) means that, in JA, nouns are more likely to be the internal element in compounds, followed by adjectives and verbs. In comparison with the hierarchy (75) proposed by Scalise and Vogel (2010: 10), it seems that the order in JA is almost the same as the one preferred universally with respect to the syntactic categories N > Adj > V. The same applies to MSA which shares the first two preferences N > Adj with the one preferred universally. By the same token, Dressler (2006) indicates that compound nouns tend to be more frequent than compound verbs. He also notes that the same applies to endocentric compounds and exocentric compounds, where the former is more frequent than the latter. The reason why some languages exhibit a pattern of compounds which seems to contradict the general preferred types is still unanswered (Dressler *ibid*).

The most frequent word combinations in compounds found by Scalise and Vogel (2010: 12) are given in (77).

- (77) N + N  
Adj + N  
Adj + Adj  
N + Adj  
V + N  
N + V  
V + V  
Adv + N  
Adj + V  
Adv + Adj

These types, namely, N + N, Adj + N, Adj + Adj, etc. are quite similar to the ones found in Arabic as discussed in the previous sections.

With respect to reduplications, in Arabic these items are to be considered 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, whereas in JA they can be adverbials or interjections. Finally, Arabic numerals from eleven to nineteen, as argued previously, are compounds, whereas the rest needs further investigation.

Although we have now looked at all the main types of compounds in Arabic, there is one important issue that we have skirted over so far, i.e. that of headedness of the compounds; this will be taken up in the next chapter. In addition, the next chapter applies the universal classification of compounding, proposed by Scalise and Bisetto (2009), to Arabic data.



# **Chapter Seven: Headedness and compound classification in MSA and JA**

## **7.1 Introduction**

How to pinpoint the position of the head in various syntactic and morphological structures cross-linguistically has been heavily debated in the relevant literature (Zwicky 1985; Hudson 1987; Bauer 1990; Polinsky 2012; Arcodia 2012 among others). The debate, in part, centres on the criteria used to determine the head of a structure. In addition, the types of the head, whether semantic, syntactic or morphological, are a topic for debate among linguists (Allen 1978; Bauer 2009b; Lieber 2010; Scalise and Fàbregas 2010). Due to the fact that compounding is one of the most common word-formation processes cross-linguistically, determining the position of the head is of substantial importance. In this chapter, I explore the notion of headedness in Arabic compounds, taking into account the existing discussion in the relevant literature. I also classify compounds in Arabic in order to determine whether they conform to Scalise and Bisetto's (2009) taxonomy of English compounds, which is based on headedness. If they do, this would support the universality of Scalise and Bisetto's (2009) taxonomy and show that compounds may exhibit similar behaviour cross-linguistically.

The chapter proceeds as follows: section 2 provides an overview of the concept of headedness, discussing the main proposed criteria of headedness. It then uses the criteria to identify the position and type of the head in Arabic compounds. Section 3 describes the various types of compounds in Arabic on the basis of Scalise and Bisetto's (2009) classification. Finally, section 4 summarises the main findings.

## **7.2 Headedness**

### **7.2.1 Overview**

In an important study of the general notion of headedness, Zwicky (1985: 2) indicates that “[t]he intuition to be captured with the notion HEAD is that in certain syntactic constructs one constituent in some sense ‘characterises’ or ‘dominates’ the whole”. According to Zwicky (1985), determining the head relies on the idea that in any syntactic construct, one element governs or dominates the rest of the elements within that construct. However, there has been considerable debate on the definition of the head (Zwicky 1985; Croft 1995; Arcodia 2012; Polinsky 2012; among others). Some scholars argue that a unanimous definition of what

exactly a head is may be attainable (e.g. Hudson 1987), whilst others are not that optimistic (e.g. Polinsky 2012: 348). The difficulty of giving a definition of the head stems from the problems facing linguists when dealing with syntactic constructs. For instance, if the phrase *the dog* is interpreted to mean a dog classified as a definite object, then the determiner, i.e. *the*, appears to be the governing element. On the other hand, if the same phrase is interpreted as a definite object, which is classified as a dog, the governing element may be the noun, i.e. *dog*. Arcodia (2012: 367) notes that similar problems are also found in morphology; these are discussed in detail in section 7.2.2.

In spite of such difficulties, there is a certain amount of consensus about the headedness of many syntactic and morphological constructs. In fact, languages are often divided into two main types in terms of the position of the head. A language is considered head-final when the head element is usually or always placed in a final position, whereas head-initial languages tend to place the head element in an initial/left position. Using this criterion, Johannessen (1996) suggests that Arabic is a head-initial language. The same point is made by Fender (2008: 106, 112), for colloquial forms of Arabic as well as MSA. However, neither Johannessen (1996) nor Fender (2008: 106, 112) discusses the position of the head in Arabic compounds.

With regard to English compounds, Williams (1981: 248) claims that the head of a complex word in English is always the right element, formulated in his famous right-hand head rule. Later, Selkirk (1982) proposes that the location of the head in general is a parameter, i.e. it can be either the left or the right element of the word in a language. In a sample of thirty-six languages, excluding Arabic, the overall preference in nominal compounds is for right-headedness (Bauer 2001: 697). However, in many languages such as Vietnamese and Mandarin, both left-headed and right-headed compounds can be found, which means that the parametric approach is insufficient to account for the position of the head (Booij 2010a: 100).

The next section aims to shed more light on the definition of headedness, both cross-linguistically and in Arabic. It provides a discussion of previous work on headedness in the relevant literature, by examining the criteria discussed by several scholars on what makes a certain element the head of a construct, and then applies these criteria to compounds in Arabic. The aim is to determine the position of the head in Arabic compounds and to identify its properties.

### 7.2.2 Headedness criteria in the previous literature

Many criteria relevant to determining the head of a particular construct are discussed by Zwicky (1985). He proposes that the notion HEAD needs to capture the intuition that, in certain syntactic constructs, one element will dominate the rest (Zwicky 1985: 2). He examines the following eight criteria in order to identify an element as a syntactic head:

1. The head is the semantic argument, which means that the element called ‘the head’ has a meaning that acts as an argument to a functor (predicate modifier or connective). From a formal semantics viewpoint, a functor is “a sign that attaches to one or more expressions of given grammatical kind(s) to produce an expression of a given grammatical kind” (Quine 1982: 129).
2. The head is the element with which other constituents must agree, i.e. it is the determinant of concord.
3. The head is the element which is marked with morphosyntactic features that indicate the syntactic relations between the construct as a whole and other syntactic units in a sentence. In other words, the head bears the inflections.
4. The head is the element that selects its sisters, i.e. it is the subcategorizand.
5. The head is the governor, which means that it can determine or select the morphological form of its sister on the tree. For instance, in a V + NP construct, the governor V assigns a morphological case to its sister NP.
6. The head is the element which has the same distribution as that of the whole construct, i.e. it is the distributional equivalent.
7. The head is the obligatory element, in the sense that if it is removed, the whole construct must be recategorised.
8. From a dependency theory perspective, the head is the element on which other elements rely in a dependency analysis.

Publication of this list in Zwicky (1985) 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 (1935),

Marchand (1969: 214), Lyons (1977: 294), Williams (1981: 248), among others. Bauer's (1990: 2–3) useful summary of past research on headedness criteria is provided below:

1. A phrase is a hyponym of its head. Hudson (1987) calls this a 'kind of' relation.
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: 2–3) criteria may not be valid to identify the head in a compound (Arcodia 2012: 368). In particular, criteria 3 (the head is the governor) and 7 (the head is lexical) are not applicable to English compounds (Arcodia 2012: 368). 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 formed in compliance with 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 section 7.2.3, I apply the above criteria to N + N compounds within SGC in MSA and JA. In section 7.2.4, these criteria are applied to Adj + N compounds within SGC in MSA and JA. Finally, section 7.2.5 applies these criteria to other types of compound, namely, Adj + Adj and N + N other than SGC.

## **7.2.3 Applying headedness criteria to N + N compounds in MSA and JA within SGC**

### ***7.2.3.1 The semantic criterion***

The semantic criterion of headedness is in essence simple: it states that the head of a compound

is the element that determines the semantic category of the whole compound, making a compound a hyponym of its head (Lieber 2010: 178). For instance, in English, the word *pole* in *flagpole* is the head, since a *flagpole* is a hyponym of *pole* (Bauer 2009b: 348). This principle was originally proposed by Allen (1978: 11), who refers to it as the ‘IS A’ condition, as in:

(1) In a compound [ [ ]<sub>A</sub> [ ]<sub>B</sub> ]<sub>C</sub> where B is the head, C ‘IS A’ B

This condition suggests that the whole compound denotes a subclass of the concept that the head denotes. In Arabic N + N compounds, the left element is usually the head, since it denotes a hypernym of the whole compound, as in examples (2-4):

(2) *xaatam*            *l-ʔalmaas*  
       ring                the-diamond  
       ‘the diamond ring’

(3) *muʕallim*        *l-fiizyaaʔ*  
       teacher            the-physics  
       ‘the physics teacher’

(4) *mudiir*            *l-madrasah*  
       principal            the-school  
       ‘the school principal’

On the basis of Allen’s (1978) principle, it is clear that:

(5) *xaatam lʔalmaas*        ‘the diamond ring’ IS A *xaatam* ‘ring’

(6) *muʕallim lfiizyaaʔ*        ‘the physics teacher’ IS A *muʕallim* ‘teacher’

(7) *mudiir lmadrasah*        ‘the school principal’ IS A *mudiir* ‘principal’

Examples (5-7) show that the left element is a superset of the whole compound, identifying them as endocentric compounds (see section 1.2.3). However, there are also exocentric compounds and those do not denote a type of the left element, as shown in examples (8) and (9):

- (8)      bayt              d-daradʒ  
             house              the-stairs  
             ‘the stairwell’
- (9)      ʕaruus              l-baħr  
             bride              the-sea  
             ‘the mermaid’  
             lit. the sea bride

In examples (8) and (9), *bayt d-daradʒ* ‘stairwell’ is not a type of *bayt* ‘house’ and *ʕaruus l-baħr* ‘mermaid’ is not a type of *ʕaruus* ‘bride’. Since Arabic compounds can be either endocentric as in (2-4) or exocentric as in (8) and (9), we have to recognise that the ‘IS-A’ criterion cannot always be applied. But when it can, it always points to the left-hand element being the head in N + N combination within SGC.

### 7.2.3.2 *The syntactic criteria*

The first syntactic criterion addresses the notion of subcategorization in relation to headedness. In this respect, Zwicky (1985: 5) points out that:

In some constructions, one slot has a special status in that the items that can fill that slot must be listed in the lexicon, while its sister constituents are not so constrained. These are instances in which one constituent is SUBCATEGORIZED with respect to its ability to occur with a particular set of sister constituents.

The notion subcategorizand has been subject to a wide debate, especially in relation to Determiner Phrases (DPs). For instance, Arcodia (2012: 373) notes that it is not clear whether, in the phrase *these black boxes*, the noun *boxes* or the demonstrative *these* is to be considered the subcategorizand (cf. Zwicky 1985: 5–6). It may be suggested that the determiner is the subcategorizand, since it is well known that determiners are lexically subcategorized; they can combine with singular count nouns (e.g. *cat*), plural count nouns (e.g. *cats*), or mass nouns (e.g. *sugar*). In fact, currently, several scholars (Abney 1987; Siloni 1997; Choi 2014) in the generative tradition would regard the determiner (*these*, a demonstrative) as the head of the previous example *these black boxes*, which is termed DP, rather than NP.

With respect to compounding, in the endocentric English compound *hairstyle*, the right-hand element, i.e. *style* is the subcategorizand, since it (i.e. the head) selects the non-head (Arcodia 2012: 373). Arcodia (2012: 373) adds that “the definition of head in endocentric compounding is closer to that of syntactic head”; this would indicate that this definition works with neither exocentric compounds nor coordinating compounds.

In Arabic, the left element of the compound selects a set of elements to accompany it, as in (10-12):

(10)      muʕallim      l-fiizyaaʔ  
              teacher      the-physics  
              ‘the physics teacher’

(11)      muʕallim      l-fann  
              teacher      the-arts  
              ‘the arts teacher’

(12)      muʕallim      t-taariix  
              teacher      the-history  
              ‘the history teacher’

In examples (10-12), the left element *muʕallim* ‘teacher’ selects the type of words that can occur with it. In the above examples, these are subjects which the teacher teaches, i.e. *lfiiizyaaʔ* ‘the physics’, *lfann* ‘the arts’ and *ttaariix* ‘the history’. Therefore, the head *muʕallim* ‘teacher’ is the subcategorizand, since it selects its sisters, which are in this case school subjects. Note that this selection is due to the fact that the left element *teacher* has the verb *teach* inside it. However, if we take the example *sikkat lhadiid* ‘the iron railway’ in (13), the right element is a modifier which sometimes cannot be selected by the head, i.e. when it is an adjunct. This means that the criterion does not work in this case.

(13)      sikkat      l-hadiid  
              rail      the-iron  
              ‘the iron railway’

The second syntactic criterion is that the head is the governor (excluded by Arcodia 2012, since it does not apply to English compounds). In MSA compounds (JA has no case markings), the left element can have any case, such as nominative, accusative or genitive on the basis of the function of the whole construct in the sentence, whereas the right element is always in the genitive case, as exemplified by (14-16):

- (14)      qaabala-ni      muṣallim-u      l-fiizyaa?-i  
             met-me          teacher-NOM      the-physics-GEN  
             ‘The physics teacher met me.’
- (15)      qaabalt-u      muṣallim-a      l-fiizyaa?-i  
             met-I            teacher-ACC      the-physics-GEN  
             ‘I met the physics teacher.’
- (16)      taḥaddaθt-u    maṣ    muṣallim-i    l-fiizyaa?-i  
             talked-I          with    teacher-GEN    the-physics-GEN  
             ‘I talked with the physics teacher.’

In example (14), the left element *muṣallimu* ‘the teacher’ has nominative case (since it is the subject), in (15) it has accusative case *muṣallima* ‘the teacher’ (being the object) and in (16) it has genitive case *muṣallimi* ‘the teacher’ (as required by the preposition *maṣ*). On the other hand, the right element is always genitive in all examples. This can be accounted for by saying that the head N governs its sister, assigning it genitive case.

The third syntactic criterion is that the head is the element which has the same distribution as that of the whole construct. For example “...V is the distributional equivalent of V + NP, since the distribution of V + NP is roughly the same as the distribution of Vs like *write* and *vanish*...” (Zwicky 1985: 12). For English, the head is the element that determines the syntactic category of a compound such as *high school*, which is syntactically a noun like *school* (Lieber 2010: 178). However, it is very difficult to identify which element of the compound is responsible for determining the syntactic category of the whole N + N compounds in both English and Arabic, since both elements are nouns. Examples from Arabic are:



(17) malik            l-yaabah  
king                the-jungle  
'the lion'

(18) findzaan        l-qahwah  
cup                the-coffee  
'the coffee cup'

The fourth syntactic criterion to identify the head is the obligatoriness of an element within the compound. This means that if such an element is omitted, the outcome will be ungrammatical depending on the nature of the sentence. For instance, Bauer (2009b: 348) suggests that the word *pole*, in the compound *flagpole*, is obligatory, because *pole* can be used, but not *flag*, without changing the meaning. In Arabic, the left element of a compound is obligatory, as in the following examples:

(19) a) raʔayt-u        muʕallim        l-fiizyaaʔ  
saw-I            teacher        the-physics  
'I saw the physics teacher.'

b) raʔayt-u        l-muʕallim / raʔayt-u        muʕallim-a-n  
saw-I            the-teacher / saw-I        teacher-ACC.INDF  
'I saw the teacher.' / 'I saw a teacher.'

c) \* raʔayt-u        l-fiizyaaʔ  
saw-I            the-physics  
\* 'I saw the physics.'

(20) a) kasart-u        findzaan        l-qahwa  
broke-I        cup            the-coffee  
'I broke the coffee cup.'

- b)      kasart-u          l-findzaan/kasart-u      findzaan-a-n  
              broke-I          the-cup/broke-I          cup-ACC-INDF  
              ‘I broke the cup.’/ ‘I broke a cup.’
- c)      \*kasart-u          l-qahwa  
              broke-I          the-coffee  
              ‘I broke the coffee.’

Looking at examples (19) and (20), the obligatory element is clearly the left element in the compounds above. Examples (19b and 20b) show that the left elements can stand on their own, whereas (19c and 20c) demonstrate that the right elements cannot be used on their own, resulting in an ungrammatical sentence. Note, however, in the case of non-compositional compounds, both elements are equally important to convey the meaning of the whole compound; neither of them can be deleted, as shown in (21):

- (21)      a)      qaabalt-u          raaʔid          l-fadʕaaʔ  
              met-I              pioneer          the-space  
              ‘I met the astronaut.’
- b)      ʔqaabalt-u      r- raaʔid /      qaabaltu      raaʔid-aan  
              met-I              the-pioneer /      met.I          pioneer- INDF.ACC  
              lit. I met the pioneer.’ / ‘I saw a pioneer.’
- c)      \*qaabaltu          l-fadʕaaʔ  
              met.I              the-space  
              lit. I met the space.

Example (21b) demonstrates that the first element can be used on its own, but the meaning of the whole compound is totally lost. On the other hand, example (21c) shows that, in addition to losing the meaning of the whole compound, the second element does not normally stand on its own. Note, however, that the grammaticality vs. the ungrammaticality of what remains after the first element of the compound is deleted relies heavily on the context of use. This can be illustrated with the following example:

- (22)
- |    |                               |             |             |
|----|-------------------------------|-------------|-------------|
| a) | ʔuħubb-u                      | muʕallim    | t-taariix   |
|    | love-I                        | teacher     | the-history |
|    | ‘I love the history teacher.’ |             |             |
| b) | ʔuħubb-u                      | l-muʕallim  |             |
|    | love-I                        | the-teacher |             |
|    | ‘I love the teacher.’         |             |             |
| c) | ʔuħubb-u                      | t-taariix   |             |
|    | love-I                        | the-history |             |
|    | ‘I love the history.’         |             |             |

In example (21), deleting the left element results in an ungrammatical sentence, whereas deleting the right element yields a grammatical sentence even though the meaning of the whole compound is lost. On the other hand, in example (22), if either of the elements is omitted, the sentence would remain grammatical. This means that this criterion does not yield a clear result here. The head alone will always be acceptable in any context where the entire compound is acceptable, whereas the non-head alone is only sometimes acceptable in the same contexts. For instance, *muʕallim ttaariix* ‘the history teacher’ denotes someone like the left element, i.e. *muʕallim* ‘teacher’ rather than the right element, i.e. *ttaariix* ‘the history’, which denotes ‘a subject to be studied at an academic institution’. Therefore, it seems that this criterion is closely tied to the first criterion, in which the whole compound is a hyponym of the head.

Another interesting observation with regard to the obligatoriness criterion is that, in certain contexts, the left element which is supposed to be the head on the basis of criteria 1, 2 and 3, can be omitted. This observation has been noted in particular contexts such as those related to news headlines. Examples of such cases are illustrated below:

(23) fi haal ʔasʕsʕarra l-ʔaʕyaan ʕala mawqifih min  
 in case insisted the-senators on stand.MSG from  
 t-taʕdiilaat...  
 the-amendments...

‘In case the senate insisted on its stand regarding the amendments...’

(Al-Ghad, 16<sup>th</sup> Sep 2015)

(24) n-nuwwaab yuqirr ʕadadan min  
 the-parliament members passed.MSG.PERF a number of  
 mawaadd qaanuun t-tanfiiḏ  
 articles law the-implementation  
 ‘the parliament has passed a number of articles pertaining to the  
 implementation law.’

(Al-Distour, 15<sup>th</sup> Sep 2015)

(25) l-muʕallimiin tatawaʕʕad alḏneeboot maʕ qtiraab l-fasʕl  
 the-teachers threaten.MSG Althneeboot with approach the-term  
 d-diraasiyy  
 the-academic

‘the teachers union threatens *Althneeboot* (the Minister of Education) with the  
 approach of the academic term’

(Khaberni, 19<sup>th</sup> Jan 2015)

In examples (23-25), the left elements *madʕlis* ‘council’ *madʕlis* ‘council’ and *niqaabah* ‘union’ of the compounds *madʕlis lʔaʕyaan* ‘the senate’, *madʕlis nnuwwaab* ‘the parliament’ and *naqaabat lmuʕallimiin* ‘the teachers union’ are deleted, respectively. Only the right element remains in order to refer to the institution which the members, i.e. senators, parliament members and teachers work in/represent. The deletion can be observed if one examines the cross-referencing marked on the element which follows the compounds. In example (23), the singular resumptive pronoun on the word *mawqifih* ‘stand’ refers to the singular ellipted antecedent *madʕlis* ‘council’. In (24) and (25), the resumptive pronouns marked on the verbs, i.e. *yuqirr* ‘has passed’ and *tatawaʕʕad* ‘threatens’ refer to the ellipted antecedents *madʕlis* ‘council’ and *niqaabah* ‘union’, respectively.

The phenomenon in which the people stand for or refer to the whole institution in which they work/ represent is called metonymy. In the context of newspapers, one may find instances in which one element of the compound is deleted, provided that the other element which is not deleted can compensate for the loss of the omitted element. Another interesting aspect of these metonymic compounds is that the left element is the one which is omitted, rather than the right element. This may indicate that the right element is indispensable, whereas the left element is, at least in relation to metonymic compounds. The deletion of the left element may take place for verbal economy purposes. That is, instead of repeating the two elements throughout the news article, it would be more economical if the whole compound is mentioned at the beginning only once, then in the remaining sections, the author could delete one of the elements. Note, however, that such ellipsis is not possible with other compounds, as in:

- (26)      \*muʕallim    l-fiizyaaʔ    yaaʔib laakin      l-kiimyaaʔ    ʔaadʕir  
               teacher        the-physics    absent but            the-chemistry    present  
               Intended: ‘the physics teacher is absent but the chemistry teacher is present’

In English compounds too, deletion of the head sometimes takes place in some cases. Bauer *et al.* (2013: 479) note that several compounds that consist of two elements lose the right element, becoming metonymic expressions, e.g. *chair* for *chairperson*, *business* for *business class*, *Tasman* for *Tasman Sea*, the *Tate* for *the Tate Gallery* and *vacuum* for *vacuum cleaner*. They also suggest that the semantic outcome of this type of ellipsis in compounds appears to be similar to the metonymy that occurs in syntactic ellipsis, e.g. the *House* for the House of Representatives or that which takes place without ellipsis, e.g. *Washington* for the government in Washington.

### 7.2.3.3 *The morphological criteria*

The first morphological criterion to determine the head in a compound is the ‘morphosyntactic locus’. In Arabic compounds, the morphosyntactic locus seems to be the left element. Firstly, pluralisation has always been used to identify the head (Bauer 2009b: 348). An example from English is *schoolboy*, which consists of *school* and *boy* and has the plural *schoolboys*, based on the head *boys*. In Arabic, the left element of the compound is the one marked for number and gender, as in examples (27) and (28):

(27) a) waraaq-at l-ʔimtihaan  
 paper-FSG the-exam  
 ‘the exam paper’

b) ʔawraaq l-ʔimtihaan  
 paper.FPL the-exam  
 ‘the exam papers’

(28) a) ʕarus l-baħr  
 bride.FSG the-sea  
 ‘the mermaid’  
 lit. the sea bride

b) ʕaraaʔis l-baħr  
 bride.FPL the-sea  
 ‘the mermaids’  
 lit. the sea brides

Examples (27) and (28) demonstrate that plural *ʔawraaq* ‘papers’ and *ʕaraaʔis* ‘brides’ are the heads of the compounds *ʔawraaq lʔimtihaan* ‘the exam papers’ and *ʕaraaʔis lbaħr* ‘the mermaids’ on the basis of the left element. In Arabic N + N compounds, the right element normally appears in the singular form, with a few exceptions where the right element of the compound is always plural, as in examples (23-25) and the compound *raʔiis lwuzaraaʔ* ‘the prime minister’ (see example 29). These exceptions are discussed in detail in section 4.2.6.

In addition, the morphological gender can be used to demonstrate how the morphological locus can be determined, as in the following Arabic compounds:

(29) a) raʔiis l-wuzaraaʔ  
 president.MSG the-minister.MPL  
 ‘the male prime minister’

b) raʔiis-at l-wuzaraaʔ  
 president-FSG the-minister.MPL  
 ‘the female prime minister’

- (30) a)  $\zeta$ aabir                      s-sabiil  
 passer by.MSG                  the-path  
 ‘the male passer-by’
- b)  $\zeta$ aabir-at                      s-sabiil  
 passer by-FSG                  the-path  
 ‘the female passer-by’

The left element of the above examples is the morphosyntactic locus, since it inflects for gender. In example (29) and (30), the feminine form is *raʔiisat lwuzaraaʔ* ‘the female prime minister’ and *\zetaaabirat ssabiil* ‘the female passer by’, with no changes to the right element of the compound.

However, in some compounds, the gender and plurality tests fail to determine the head, as in examples (31) and (32) below:

- (31)  $\zeta$ ayn-u                      s-samaaʔ-i  
 eye-NOM                      the-sky-GEN  
 ‘the sun’
- (32) yawm-u                      l-hisaab-i  
 day-NOM                      the-judgment-GEN  
 ‘judgment day’

For semantic reasons, these compounds cannot be marked either for number or gender. In such cases, a further morphological property, i.e. case, can be used to determine the position of the head, as follows:

- (33)  $\zeta$ ayn-u                      s-samaaʔ-i                      dʒamiil-at-un                      l-yawm  
 eye.FSG-NOM                  the-sky.FSG-GEN                  beautiful-FSG-NOM                  the-day  
 ‘The sun is beautiful today.’

The adjective *dzamiilatun* ‘beautiful’ agrees with the word *ʕaynu* ‘eye’ in case, both being nominative. Thus, the case test here identifies *ʕaynu* ‘eye’ as the head. If the head was the noun *ssamaaʔi* ‘the sky’, the adjective *dzamiilah* should inflect for genitive case, but this is ungrammatical, as shown in (34):

- (34)      \*ʕayn-u                      s-samaaʔ-i                      dzamiil-at-in                      l-yawm  
             eye.FSG.NOM                      the-sky.FSG-GEN                      beautiful-FSG-GEN                      the-day  
             ‘The sun is beautiful today.’

The case test criterion thus looks fairly reliable for determining the morphosyntactic locus of compounds in MSA. Note that the case test cannot be applied in JA, since it has no case marking system. Therefore, if the plurality and gender tests are also inapplicable, the morphological locus criterion cannot be used to determine the head in JA.

The second morphological criterion, pertaining to the head being lexical rather than phrasal (Zwicky 1985: 5), has been excluded by Arcodia (2012), since it does not apply to English compounding. However, this criterion does apply to Arabic compounds, since the left element is lexical, whereas the right element can be phrasal, as in (35-37) and many of the earlier examples:

- (35)      mawqif                      l-ħaaḥilah  
             stop                              the-coach/bus  
             ‘bus stop’
- (36)      saaʕ-at                      s<sup>ʕ</sup>-s<sup>ʕ</sup>ifr  
             hour-FSG                      the-zero  
             ‘the last hour’
- (37)      bayt                              d-daradʒ  
             house                              the-stairs  
             ‘the stairwell’

In examples (35-37) the left elements, *mawqif* ‘stop’, *saaʕah* ‘hour’ and *bayt* ‘house’, are lexical items, whereas the right elements *ħaaḥilah* ‘the coach’, *s<sup>ʕ</sup>s<sup>ʕ</sup>ifr* ‘the zero’ and *ddaradʒ* ‘the stairs’ consist of the determiner *l-* ‘the’ and the lexical items *ħaaḥilah* ‘coach/bus’, *s<sup>ʕ</sup>ifr*



‘zero’ and *daradʒ* ‘stairs’. Simply put, the right elements are determiner phrases (DPs), consisting of DET + N. This criterion thus seems to lend support to most of the above criteria, by which the left element is usually shown to be the head in Arabic compounding.

To sum up, seven criteria have been applied to compounds in Arabic in order to identify the head. Despite a few limitations of some of the criteria, it is clear that the head is the left element in Arabic N + N compounds. The next section discusses the types of head in Arabic Adj + N compounds in relation to the above mentioned criteria.

#### 7.2.4 Applying headedness criteria to Adj + N compounds in MSA within SGC

In this section, the headedness criteria discussed above are applied to Adj + N compounds in MSA to identify the position of the semantic, syntactic and morphological heads. Semantically, Allen’s (1978: 11) condition does not apply to Adj + N compounds, since adjectives do not form super/subsets. With regard to the syntactic criteria, the first criterion concerning subcategorization, does not apply either. Adjectives cannot be shown to be heads, since they do not select their sisters, i.e. nouns. Secondly, the left element is the governor; it can be nominative, accusative or genitive based on the function of the whole compound in the sentence, whereas the second element is always genitive, as in (38):

- (38)
- |    |  |          |                |
|----|--|----------|----------------|
| a) | qaabala-ni                               | ḥasan-u  | l-xuluq-i      |
|    | met-me                                   | good-NOM | the-manner-GEN |
|    | ‘The one with good manners met me’       |          |                |
| b) | ʔaḥtarim-u                               | ḥasan-a  | l-xuluq-i      |
|    | respect-I                                | good-ACC | the-manner-GEN |
|    | ‘I respect the one with good manners.’   |          |                |
| c) | ʔaʕmal-u                                 | maʕ      | ḥasan-i        |
|    | work-I                                   | with     | good-GEN       |
|    |  |          | the-manner-GEN |
|    | ‘I work with the one with good manners.’ |          |                |

Thirdly, concerning the distributional equivalent criterion, the left element *qawiyy* ‘strong’ is an adjective, while *lbunyah* ‘the body’ is a noun. The lexical category of the whole compound is a noun. This may suggest that the head is the right element, which determines the syntactic

category of the whole compound. However, examples (39) and (40) may suggest that the adjective *qawiyy* ‘strong’, which is the left element, is nominalised:

(39)      ḡaṣmal-u      maṣ      qawiyy      l-bunyah  
              work-I            with      strong      the-body  
              ‘I work with the strong one in body.’

(40)      taḥaddayt-u      qawiyy      l-bunyah  
              challenged.I      strong      the-body  
              ‘I challenged the strong one in body.’

However, I would argue that the adjective *qawiyy* ‘strong’ has not been nominalised; it remains an adjective, modifying a silent noun. This argument is supported by Günther’s (to appear) analysis of English and German data. Günther argues that the adjectives in noun phrases like *the rich*, *the poor*, *the impossible*, etc. have not been nominalised but are still adjectives which modify an implicit noun.

Before discussing such implicit heads, Günther (to appear) points out a difference between phrases like *the innocent*, *the guilty* and *the impossible*, which have a default reading as referring to people or abstract concepts (called the “Human/Abstract Construction” or “Nounless Noun Phrases”), as in (41), and ellipted noun phrases as in (42):

(41)      Just as the innocent should not be punished, so the guilty should be made to pay.

(42)      The fact remains, however, that the challenger whose record is 19 wins and nine defeats has lost four of his last six fights and six of his last eight.

(Günther to appear)

In example (41), the noun phrases *the innocent* and *the guilty* do not have antecedents, but the ellipted noun phrase in (42) does. To account for the difference, Günther (to appear) assumes that in the Human Construction in (41) the silent noun, i.e. *one*, has a generic personal reading, while in cases like (42) the silent noun refers to a specific person or entity.

In support of these ideas, Günther (to appear) provides some counterevidence against the nominalisation analysis of the adjectives in *the rich*, *the poor*, etc. The first piece of evidence is that these adjectives are unable to carry plural markings as nouns normally do, i.e. *\*the riches*, *\*the poors*, etc. The second piece of evidence is that these adjectives, in both English and German, exhibit adjectival morphology, e.g. comparative in (43) and superlative in (44) (Günther to appear):

(43) a. ... the wealthier have an obligation to help the weaker and the poorer.

b. *Die Reichen werden immer reicher, die Ärmere immer ärmer.*

‘The rich are constantly getting richer, the poorer are constantly getting poorer.’

(44) a. New aid to the poorest is given as grants, not loans.

b. *das Unglaublichste zu denken und das Unmöglichste für möglich zu halten*

‘to think of the most unbelievable and to consider the most impossible to be possible’

Note that Günther does not mention the fact that the German adjectives have N case. I would suggest that N case is marked on the adjective when the head noun is implicit (see example (38) from Arabic). This part of the analysis is not developed by Günther; thus, it requires further investigation.

The third piece of evidence is that, in English, the silent noun *one* in nounless noun phrases can in fact be inserted into the construction, indicating that a nominal position must be available in the structure (Günther to appear). Interestingly, if the silent noun appears, it agrees with the adjective in case, number and gender in German and in number in English, i.e. *the innocent ones* (Günther to appear).

Examining the Arabic compounds in line with Günther’s (to appear) analysis shows that the lack of an antecedent discussed by Günther (to appear) can also be observed in examples (39) and (40), where the Adj + N compounds have a default reading referring to a person. Similar to Günther’s (to appear) examples from English and German, if the silent noun *one/person* appears in Arabic, it normally has a generic reading in the sense that it only refers to a human being without specifying who he/she is. Note, however, that the context plays a

role in identifying the intended referents. For instance, *the poor* in the sentence *I help the poor in my village* does not have a generic reading. Similarly, the following sentence from Arabic does not have a generic reading, since the context assigns a specific referent to the compound:

- (45) raʔay-tu      waasiʕ-at-a    l-ʕaynayn      maʕ    sadʕiiq-i      khalid  
 saw-I            wide-F-ACC    the-eyes        with    friend-my      Khalid  
 ‘I saw the girl with wide eyes with my friend Khalid.’

Additionally, the silent noun agrees with the adjective in case, number and gender, when it is pronounced, as in (46):

- (46) a)    qaabalt-u      ʕaxsʕ-an                      tʕawiil-a      l-qaamat-i  
 met-I            person.MSG-ACC            tall.MSG-ACC    the-figure-GEN  
 ‘I met a person with a tall figure.’
- b)    qaabalt-u      ʔaʕaaxsʕ-an                    tʕawiil-ii<sup>48</sup>      l-qaamat-i  
 met-I            person.MPL-ACC            tall-MPL.ACC    the-figure-GEN  
 ‘I met people with a tall figure.’
- c)    qaabalt-u      fataat-an                      tʕawiil-at-a      l-qaamat-i  
 met-I            girl.FSG-ACC                tall-FSG-ACC    the-figure-GEN  
 ‘I met a girl with a tall figure.’
- d)    qaabalt-u      fatayaat-in                    tʕawiil-aat-i<sup>49</sup>    l-qaamat-i  
 met-I            girl.FPL-ACC                tall-FPL-ACC    the-figure-GEN  
 ‘I met girls with a tall figure.’

Example (46) shows that the implicit noun *person/girl* can be realised in Adj + N compounds in MSA and the adjective has to agree with it. Note that adjectives in MSA can carry the plural marker quite generally; thus, the plurality test is not applicable.

<sup>48</sup> Originally, the adjective is *tʕawiiil* ‘tall (plural)’, but the *-n* is deleted from the left element when it is followed by a noun. Note that *-ii* is the case marking of both the accusative and the genitive in masculine sound plural in MSA.

<sup>49</sup> The accusative case in the feminine sound plural in MSA is realised in the same way as the genitive with *-i*.

The other test of adjectivehood proposed by Günther (to appear), namely, exhibiting adjectival morphology, such as comparative and superlative does not apply to the adjective in Adj + N compounds, as in:

- (47)      ʔuhibbu        l-ʔaqwaa        qalb-a-n                min    bayni ʔaqraan-i  
           love.I            the-strongest   heart-ACC-INDF       from   among peers-my  
           ‘I love the bravest among my peers.’  
           lit. I love the strong one in heart among my peers.

In particular, when we apply the comparative and superlative test to Adj + N compounds (see example (47) a case of superlative morphology), the structure of the compound changes. Specifically, the definite article is attached to the first element and the second element is assigned accusative case. Therefore, I will disregard this test.

All in all, it seems that Adj + N compounds in MSA can be analysed as being headed by an empty noun, i.e. *one/person*, and it is this noun that determines the syntactic category of the whole compound. Thus, the whole Adj + N compound is a noun on the basis of the syntactic category of the implicit head. In addition, it seems as though the adjective acts on behalf of the implicit noun when it is not realised, bearing the case, number and gender markings.

The third syntactic criterion is obligatoriness of the head. The obligatory element is clearly the left element:

- (48)      a)      ʔuhibbu        qawiyy        l-bunyah  
                   love.I            strong        the-body  
                   ‘I love the strong and healthy person.’
- b)      ʔuhibbu        l-qawiyy  
                   love.I            the-strong  
                   ‘I love the strong person.’
- c)      \* ʔuhibbu        l-bunyah  
                   love.I            the-body  
                   lit. I love the body.

Example (48b) shows that the right element can be deleted, whereas example (48c) demonstrates that the left element cannot, suggesting that the left element is the obligatory one.

With regard to the morphological criteria, firstly, it has already been established that the left element is the morphosyntactic locus (see example 46). Secondly, it is clear that the left element is lexical, whereas the second element is phrasal, because it consists of the lexical items marked with the obligatory determiners *l-* ‘the’ or *-n* ‘a/an’, as in (49-51):

(49)      ṣaziiz            l-qawm  
           dear            the-people  
           ‘the chief of people’  
           lit. the dear of people

(50)      saṣiid            l-ḥaḏḏḥ  
           happy           the-luck  
           ‘the lucky person’

(51)      dṣaṣiif          l-qalb  
           weak.MSG      the-heart  
           ‘the coward male’

To sum up, the left element of the Adj + N compounds in MSA is the implicit head *one*. Its absence seems to trigger case marking on the adjective, a phenomenon that is also found in German.

### 7.2.5 Applying headedness criteria to compounds in MSA and JA other than SGC

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

(52)      ḥilw-un            murr-un  
           sweet-NOM      bitter-NOM  
           ‘bitter-sweet’

- (53)      haamid<sup>ʕ</sup>-un              ħilw-un  
sour-NOM                      sweet-NOM  
‘sweet-and-sour’
- (54)      s<sup>ʕ</sup>abaaħ-a              masaaʔ-a  
morning-ACC                  evening-ACC  
‘all day long’
- (55)      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 (52) and (53), since adjectives do not form super/subsets. However, examples (52) and (53) clearly denote that the meanings of the whole compounds are a mixture of both elements. With regard to (54) and (55), *s<sup>ʕ</sup>abaaħ masaaʔ* ‘all day long’ is neither *s<sup>ʕ</sup>abaaħ* ‘morning’ nor *masaaʔ* ‘evening’, indicating that both elements in this type of compound have semantically equal status. Here, it is worth pointing out that there is debate on whether these compounds are semantically double-headed (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 three following paragraphs, I examine the concept of semantic headedness in coordinative compounds and its implications for the Arabic examples (52-55).

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 semantically-equal status. Examples from English are *blue-green* and *washer-dryer* (Booij 2007: 81). 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 some compounds can be semantically headless (exocentric) based on his analysis of examples from Classical Tibetan, which he cited from Beyer (1992: 105). The head of the following examples is something like ‘property’, as in:

- (56)      *mtho-dman*    ‘height’              (*mtho* ‘high’ + *dman* ‘low’)

(57) *srab-mthug* ‘density’ (*srab* ‘thin’ + *mthug* ‘thick’)

The semantic head of example (56) 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 (56) and (57) the elements are both adjectives, yielding nouns, so I propose that neither adjectives can be the head. Applying Haspelmath’s (2002) analysis to Arabic compounds, it is clear that (52) and (53) are semantically double-headed, since the two elements are semantically equal and can be paraphrased with *and*. In addition, the two elements of examples (52) and (53) are adjectives, yielding an adjective.

In comparison with the Tibetan compounds, the Arabic examples (54) and (55) are syntactically different, since the latter consist of two nouns, yielding a noun. Additionally, the compounds in examples (54) and (55) can be used with adverbial function and this is obviously also the case with the single nouns in examples (59) and (60) below. What is still special about (54) and (55) is that, while the word *s‘abaaħ* ‘morning’ or *masaaʔ* ‘evening’ can be used either as a noun or with adverbial function, the compounds of the two nouns in (54) and (55) can only be used with 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 relation*. Furthermore, the meaning of the compounds in (54) and (55) does not denote a combination of the two elements. For instance, if we coordinate the elements of the compound in (55), yielding a sentence like (58), the meaning and structure are different from that of (55):

(58) 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 (1935: 235), who notes that the copulative compound *bitter-sweet* ‘bitter and sweet at the same time’ is endocentric, since the compound, like its coordinated elements, *bitter* and *sweet*, has the function of adjective. However, the plant-name *bitter-sweet* is exocentric, since the grammatical function of the whole compound, as a noun, is different from the two adjective elements (Bloomfield 1935: 235). That is, examples (52) and (53) are endocentric, whereas (54) and (55) 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.* (2013: 443) 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.

Resuming with the cross-linguistic criteria of headedness, syntactically neither element selects the other in examples (52-55). Concerning which of the elements is the governor, it seems that both of the elements share the same case marking, i.e. nominative in (52) and (53), and accusative in (54) and (55). As far as the distributional equivalent is concerned, both elements share the same syntactic category, i.e. adjectives in (52) and (53), and nouns in (54) and (55). Regarding the final syntactic criterion, namely obligatoriness, both elements are obligatory; if one of them is deleted, the compound loses its meaning, as shown in (59) and (60):

(59)      taftaḥu          l-maktabah      s‘abaah-an  
               open                the-library      morning-ACC  
               ‘The library opens in the morning.’

(60)      taftaḥu          l-maktabah      masaa?-an  
               open                the-library      evening-ACC  
               ‘The library opens in the evening.’

Examples (58) and (59) can never denote ‘all day long’.

Regarding the first morphological criterion, both elements are marked equally for case, number and gender. 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 (*ibid*) explains that the syntactic category of the reduplicated word can be the same as that of the root/stem, as in (61) and (62), or they can differ, as in (63) and (64):

(61)      ọmọ (N)          ‘child’       $\longrightarrow$       ọmọọmọ (N) ‘grandchild(ren)’  
               (62)      ñlá (Adj)          ‘big’         $\longrightarrow$       ñláńlá (Adj) ‘very big’

- (63) dára (V) ‘be good’ → dáradára (Adj) ‘very good’  
 (64) fẹlẹ (V) ‘be thin’ → fẹlẹfẹlẹ (Adj) ‘very thin’

Comparing data from Yorùbá to reduplicated compounds in Arabic, the syntactic category of the reduplicated words seem to be similar to that of the stem, as in examples (49-58) given in chapter 6, some of which are repeated here for the readers’ convenience:

- (65) kθiir (Adj) ‘much’ → kθiir kθiir (Adj) ‘very much’  
 (66) layla (N) ‘night’ → layla layl (N) ‘every night’  
 (67) nahaara (N) ‘daytime’ → nahaara nahaar (N) ‘all day long’  
 (68) bas<sup>s</sup>iit<sup>s</sup>ah (Adj) ‘simple’ → bas<sup>s</sup>iit<sup>s</sup>ah bas<sup>s</sup>iit<sup>s</sup>ah (Adj) ‘an expression denoting threat’

Note that the function of the reduplicated compounds in (65-68) is diverse. Specifically, example (65) consists of two adjectives and functions as an adjective. The reduplicated compounds in (66) and (67), which are comprised of two nouns, are always used with an adverbial function. The reduplicated compound in (68), which is formed from two adjectives, functions as an interjection. This suggest that example (65) is best treated as double-headed, whereas examples (66-68) are headless.

Finally, with regard to V + V compounds, the semantic criterion does not apply, since verbs do not form super/subsets. Syntactically, neither element selects the other in examples like (69-70). Concerning which of the elements is the governor, this criterion does not apply, since verbs do not carry case. As far as distributional equivalence is concerned, both elements have the same syntactic category, i.e. they are verbs in (69) and (70):

- (69) waliid            misik            fataħ            kull    ʕulab    l-pepsi  
 Walid            grabbed.MSG    opened.MSG    all    cans    the-Pepsi  
 ‘Walid grabbed and opened all Pepsi cans.’

- (70) guum fayyil ħarrik s-siyyaarah gabil  
 rise.MSG kick-start.MSG move.MSG the-car before  
 mayiidzi ʃ-ʃurtʕi  
 arrives the-policeman  
 ‘Stand up, start and move the car before the policeman arrives.’

Concerning the obligatoriness criterion, the meaning of the whole construction in examples (63) and (64) is focused on opening and moving, rather than grabbing and rising/starting, respectively. It is clear that that one must grab something in order to open it. Therefore, it seems that the first element *misik* ‘grabbed’ is less important than the second element. This can be shown with (71) and (72), which are identical to the relevant parts of (69) and (70), except that only the final verb of the compound is retained.

- (71) waliid fatah kull ʃulab l-pepsi  
 Walid opened.MSG all cans the-Pepsi  
 ‘Walid opened all Pepsi cans.’

- (72) ħarrik s-siyyaarah gabil mayiidzi ʃ-ʃurtʕi  
 move.MSG the-car before arrives the-policeman  
 ‘Move the car before the policeman arrives.’

Examining the morphological criteria, both elements are in the past tense. The two elements of V + V compounds are lexical, sharing the same syntactic category, i.e. verb. The conclusion is that in (69) and (70), the last element is the head.

All in all, Adj + Adj combinations, N + N other than SGC and reduplicated words can be double-headed or headless, whereas V + V compounds seem to be right-headed.

### 7.2.6 The implications of the headedness criteria

In the previous section, we silently assumed that every compound has one single head, identified by all 7 properties simultaneously. However, there have been suggestions that there are different types of heads, identified by different properties. Several scholars (e.g. Allen 1978: 11; Bauer 2009b: 348; Lieber 2010: 178; Scalise *et al.* 2009: 49-50; Scalise and Fàbregas 2010: 124) have suggested that two main types of head, namely semantic and syntactic heads,

can be distinguished. Here, it is worth pointing out that the seven criteria of headedness discussed in section 7.2.3 can also be used to identify types of head. For example, Haspelmath (2002: 88) argues that a compound has a semantic head when the whole compound denotes a hyponym of either of its elements, such as *sea bird*, *houseboat*, *sailboat*, *school bus*, *handbag* and *handbrake*, where *sea bird* is a hyponym of *bird* and similarly for the other cases. Concerning the syntactic head, Scalise *et al.* (2009: 49-50) indicate that a compound has a ‘formal’ head when (1) the head determines the class of the whole compound; or (2) when the head carries all the inflectional markers; or (3) a combination of both. For example, *green* is an adjective, but the syntactic category of the whole compound *greenhouse* is N, since the formal head is a N. In *school buses*, the head *buses* is inflected for number. Finally, in *blue skies*, the formal head *skies* determines the syntactic category of the whole compound and carries the plural suffix *-s*.

This approach is taken even further by Scalise and Fàbregas (2010: 124), who discuss three types of head: (1) the syntactic head, which determines the syntactic category of the whole compound (e.g. *sky blue* is an adjective like its syntactic head *blue*); (2) the semantic head, which determines the semantic category (as in *bookshop*, which is a type of *shop*); and (3) the inflectional head, which carries the inflectional markers (such as the plural *-s* in *mother(s)-in-law*). When dealing with the semantic head, it is important to re-state the difference between endocentric and exocentric compounds. Endocentric compounds are semantically headed, such as *sailboat* which is a type of *boat*, while exocentric compounds are not semantically headed, such as *egghead* which is not a hyponym of *head* (Booij 2007: 81). In particular, the syntactic head is connected to the distributional equivalence and subcategorization criteria, the semantic head is related to hyponymy and the inflectional head is a reflection of the morphosyntactic locus criterion. The normal expectation would be that all three types of head coincide in one and the same element of the compound. This is certainly the case in English compounds like *blackbird*, where *bird* is the semantic, syntactic and inflectional head simultaneously; *blackbird* is a type of *bird*, it is a N like *bird* and the plural is *blackbirds*. However, according to Arcodia (2012: 366), these types of head do not necessarily coincide, such as *California beauty and white collar* (see Scalise *et al.* 2009: 61; Scalise and Fàbregas 2010: 125).

In Arabic, the semantic head in endocentric compounds is the left element, as exemplified by (73):

- (73)      minfaar      l-xaʃab  
              saw            the-wood  
              ‘the wood saw’

The compound *minfaar lxaʃab* ‘the wood saw’ is a hyponym of *minfaar* ‘saw’; thus, the head of the compound is *minfaar* ‘saw’. However, the semantic relationship is not always clear in case of non-compositional compounds, as in example (74):

- (74)      bint            l-ʕayn  
              daughter      the-eye  
              ‘the tear’

In example (74), *bint lʕayn* ‘the tear’ is not a type of *bint* ‘daughter’. As pointed out by Borer (2009), who discusses the notion of semantic head in another Semitic language, i.e. Hebrew, the ‘IS A’ condition of Allen (1978: 11) cannot always be applied. While a phrase IS A modified version of its head, such a semantic connection does not always obtain in Hebrew compounds; compare (75) and (76), where (75) is a phrase and (76) is a compound:

- |      |    |                   |                  |                      |                              |
|------|----|-------------------|------------------|----------------------|------------------------------|
| (75) | a. | <i>beyt</i>       | <i>mora</i>      | IS                   | A <i>bayit</i>               |
|      |    | house             | teacher          | IS                   | A house                      |
|      | b. | <i>šomer</i>      | <i>mexoniyot</i> | IS                   | A <i>šomer</i>               |
|      |    | guard             | cars             | IS                   | A guard                      |
| (76) | a. | <i>beyt sefer</i> | ‘school’         | IS NOT (necessarily) | A <i>bayit</i> <sup>50</sup> |
|      |    | house book        | ‘school’         | IS NOT (necessarily) | A house                      |
|      | b. | <i>yošev roš</i>  | ‘chairman’       | IS NOT               | A <i>yošev</i>               |
|      |    | sitter head       | ‘chairman’       | IS NOT               | A sitter                     |

Borer (2009) uses examples like (76) to argue that some compounds in Hebrew are semantically headless. In particular, the whole compound is not a hyponym of either of its

<sup>50</sup> Here, one may argue that *bayit* ‘house’ is a building/place; thus, it does have a semantic head. This case may require a grammaticality judgment task by Hebrew speakers and is thus not pursued here any further.

elements. Thus, looking at examples (74) from Arabic and (76) from Hebrew, it seems that some compounds lack a semantic head. However, in the approach sketched above, they can still have a syntactic and/or morphological head.

With respect to the syntactic head of a compound, it is the element that determines the syntactic category of the whole compound (Scalise and Fàbregas 2010: 124). In Arabic, the left element is the syntactic head, as in (77):

- (77)      ʔuħibb-u      qawiyy-at      l-qalb  
           love-I           strong-FSG      the-heart  
           ‘I love the brave female.’  
           lit. I love the one with a strong heart.

Example (77) shows that the implicit left element *one* is the syntactic head of the whole compound. The syntactic category of the compound *qawiyyat lqalb* ‘the brave female one’ is a noun like the implicit head *one*, not like the adjective *qawiyyat* ‘strong’.

Finally, Scalise and Fàbregas (2010: 124) state that the morphological head carries the inflectional features of a word, such as gender, number and case. In Arabic, the element which carries the inflectional features, namely, number and gender is the left element. This is illustrated with the following examples:

- (78)      a.      muʕallim      l-kiimyaaʔ  
                   teacher.MSG   the-chemistry  
                   ‘the chemistry teacher’
- b.      muʕallim-at           l-kiimyaaʔ  
                   teacher-FSG           the-chemistry  
                   ‘the female chemistry teacher’

- |    |                                 |                             |
|----|---------------------------------|-----------------------------|
| c. | muṣallim-uu<br>teacher-MPL      | l-kiimyaa?<br>the-chemistry |
|    | ‘the chemistry teachers’        |                             |
| d. | muṣallim-aat<br>teacher-FPL     | l-kiimyaa?<br>the-chemistry |
|    | ‘the chemistry female teachers’ |                             |

The left element in (78) bears the inflectional markers of the compound as a whole and it is lexical rather than phrasal, as discussed in 7.2.3.

### 7.2.7 Summary

Taking all the previous points into consideration, several provisional generalisations can be made with regard to headedness in Arabic compounds:

1. In Arabic, all compounds within SGC are left-headed in accordance with the semantic, syntactic and morphological criteria.
2. The semantic, syntactic and morphological heads always coincide in Arabic compounds within SGC.
3. The most reliable test to determine the morphological head in MSA is the case test, which provides a better test than number and gender.
4. Adj + Adj compounds, N + N compounds other than SGC and reduplicated compounds can be either semantically double-headed or semantically headless.
5. V + V compounds seem to be right-headed.

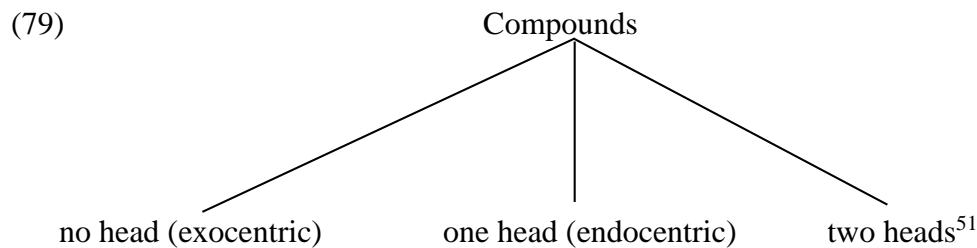
## 7.3 Arabic compounds within the cross-linguistic compound taxonomy of Scalise and Bisetto (2009)

### 7.3.1 Overview

Having discussed the position and types of head in Arabic compounds in the previous section, a natural further step now is to classify compounds on the basis of the types of head. In this section, I analyse compounds in Arabic on the basis of a taxonomy proposed by Scalise and Bisetto (2009) which is comprehensive, meant to be universally applicable, and based on fully

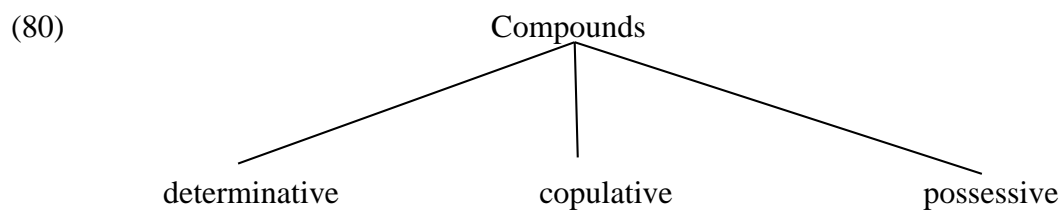
explicit criteria. By way of introduction to this taxonomy, I will briefly review some classifications suggested by other morphologists (Fabb 1998; Olsen 2001; Haspelmath 2002 and Booij 2005), since they formed the basis on which Scalise and Bisetto's (2009) more elaborate taxonomy was designed.

According to Fabb (1998), compounds can be classified into three types on the basis of the presence vs. absence of the head, as depicted in the following diagram:



Examples of the three types of compounds suggested by Fabb (1988) are: (1) exocentric, e.g. *redhead*, *pickpocket* and *flat-foot*; (2) endocentric, e.g. *handbag*, *bookshop* and *table lamp*; and (3) two heads, e.g. *producer-director* and *painter-poet*.

Olsen (2001) classifies compounds into three types, as in the following diagram:



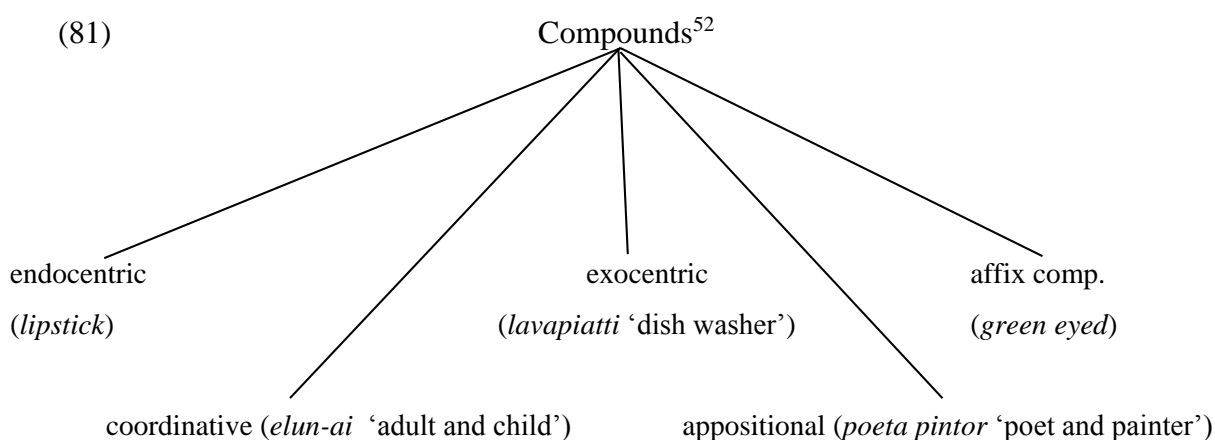
Olsen's (2001) classification seems to replicate that of Fabb (1998). It is based on the relationship between the two elements of the compounds. In a determinative compound, e.g. *tea cup*, the two elements are in a modifier-head relationship (this is Fabb's endocentric type), whereas in a copulative compound, e.g. *poet-doctor*, the relationship between the elements is based on asyndetic syntactic coordination (as in Fabb's double-headed type). Finally, in a possessive compound, e.g. *greybeard*, the possessive relationship is between the whole compound and the missing head, rather than a relationship between the two elements (this is Fabb's exocentric type).

---

<sup>51</sup>This type includes co-ordinate, appositional and dvandva compounds.



According to Haspelmath (2002), there are five types of compounds, as in the following diagram:

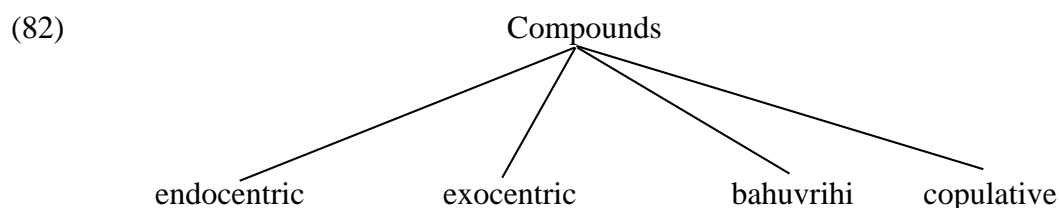


Haspelmath's classification (i.e. endocentric, exocentric, coordinate, appositional and 'affixed compounds') depends mainly on the presence vs. absence of a semantic head. The difference between endocentric and exocentric compounds relies on the presence versus the absence of a hyponym relationship between the elements. In addition, Haspelmath (2002: 89) distinguishes between coordinative compounds and appositional compounds, both of which have more than one semantic head. In coordinative compounds, the compound elements refer to several referents, whilst in appositional compounds, both elements have the same reference (Haspelmath 2002: 89). A new type suggested by Haspelmath (ibid) is referred to as an 'affixed compound', which he defines as a type of compound that involves two stems and one affix, e.g. *green eyed*. However, Haspelmath (2002: 89-90) suggests that coordinative compounds can be exocentric, using some examples from Classical Tibetan without a clear distinction between endocentric vs. exocentric coordinative compounds. Scalise and Bisetto (2009: 42) note that "Haspelmath's classification (endocentric, exocentric, appositional, coordinate, and affixed) seems to overlook the fact that affixed compounds, as also (additive) coordinates like *adult-child*, are exocentric, whereas appositives are endocentric".

Finally, Booij (2005) classifies compounds into four types as outlined in the diagram below. Again, the distinction between endocentric vs. exocentric compounds is based on the presence vs. the absence of a semantic head. Booij's (2007: 81) classification also includes copulative compounds, e.g. *blue-green* and *washer-dryer*, which he treats as semantically

<sup>52</sup> The Italian compound *lavapiatti* 'dishwasher' consists of *lava* 'washes' and *piatti* 'dishes'. Regarding, the Korean compound *elun-ai* 'adult and child', it is made up of *elun* 'adult' and *ai* 'child'. Finally, the Spanish compound *poeta pintor* formed from the two words *poeta* 'poet' and *pintor* 'painter'.

headless.<sup>53</sup> Finally, Booij (2007: 80) considers bahuvrihi compounds, e.g. *baldhead* a separate type, even though many scholars regard them as a subtype of exocentric compounds.



These classifications are the basis upon which Scalise and Bisetto's (2009) taxonomy is established. However, Scalise and Bisetto (2009) address some problematic issues which they have identified in the previous classifications. For instance, an examination of the classifications shows that the notions of endocentricity and exocentricity are not allowed to expand across classes (Scalise and Bisetto 2009: 41). For example, in Haspelmath's classification (endocentric, exocentric, coordinate, appositive and 'affixed compounds'), the fact that both affixed and coordinate (additive) compounds, e.g. *adult-child* are exocentric, whereas appositional ones are endocentric seems to be obscured (Scalise and Bisetto 2009: 41). Additionally, although Fabb's classification is consistent in the sense that it uses a single criterion, i.e. number of heads, it is too general to capture the variety of attested compounds. Olsen's classification seems to have the advantage of using the notion of determinative, rather than the notion of coordinative compounds. But, in addition to these two classes, Olsen proposes a class of exocentric possessive compounds, a class clearly based on a different criterion. This mixture has a disadvantage, since it obscures the classification of determinative and copulative compounds in terms of endocentricity and exocentricity. Additionally, there is a difference between determinative and copulative compounds on the one hand, and possessive compounds on the other. In the former, the relationship is between the two components, while in the latter, it is between the whole compound and the absent head (Scalise and Bisetto 2009: 42).

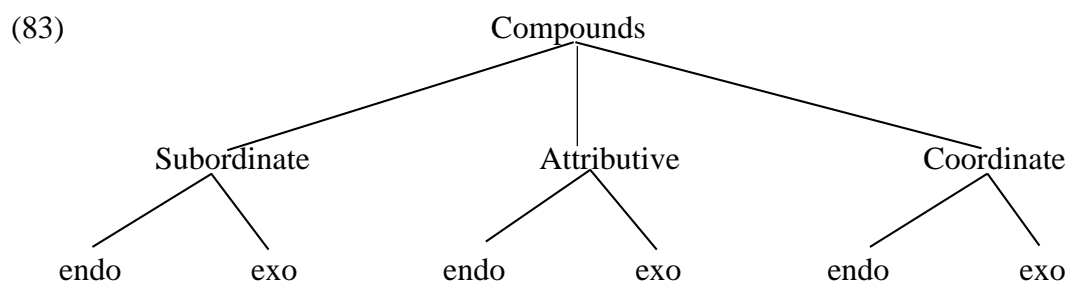
### 7.3.2 Bisetto & Scalise's (2005) classification

Bisetto and Scalise's (2005) classification, an early version of Scalise and Bisetto (2009), suggests that there are syntactically three different types of compounds. The core of the

---

<sup>53</sup> Based on Bloomfield's (1935: 235) analysis of copulative compounds, examples like *blue-green* and *washer-dryer* are endocentric.

proposed classification is the grammatical relation between the elements of the compound. Three grammatical relations are found: subordination, attribution and coordination. This is the first level of classification. Then, they are further divided on the basis of a semantic criterion; endocentricity vs. exocentricity, resulting in the following taxonomy:



Attributive compounds occur when the non-head acts as a modifier of the head, as in *snail mail*, which is metaphorically a kind of *mail* that moves like a *snail*. Some endocentric examples are *keyword*, *swordfish*, *blue cheese*, *backyard*, and some exocentric examples are *greybeard* and *loudmouth*. Note that the first element may express many relationships with the head. In subordinate compounds, one element of the compound can be interpreted as a complement of the other one, usually its object. Examples where one of the elements is derived from a verb are *truck driver* ‘drive the truck’, *lion tamer* ‘tame the lion’, *hand holding* ‘hold hands’, *food shopping* ‘shop for food’, *meal preparation* ‘prepare meals’ and *cost containment* ‘contain costs’. Examples in which one of the elements is not derived from a verb and still show head-complement relation are *love story* ‘story about love’ and *coffee cup* ‘cup of coffee’. Some exocentric examples of subordinate compounds are *pickpocket* and *killjoy*. Finally, coordinate compounds occur when the first element of the compound does not modify the second; instead the two are equal. Examples of endocentric coordinative compounds are *doctor-patient*, *blue-green*, *producer-director* and *prince consort*, and some examples of exocentric ones are *mother-child* and *mind-brain*. Coordination can in fact: (1) connect two individual elements without referring to either of them as a separate entity, as in *mother-child*; or (2) express two properties associated with an entity, e.g. *producer-director* (Scalise and Bisetto 2009: 36).

However, there are some limitations to Bisetto and Scalise’s (2005) classification. For example, Montermini (2008: 169) asks whether we should label a compound like *swordfish* as subordinative, ‘fish with a sword’, or attributive, ‘fish like a sword’. In addition, the feasibility of a distinction between attributive and intersective (coordinative) compounds has been doubted by researchers such as Montermini (2008: 165-166). In particular, the distinction

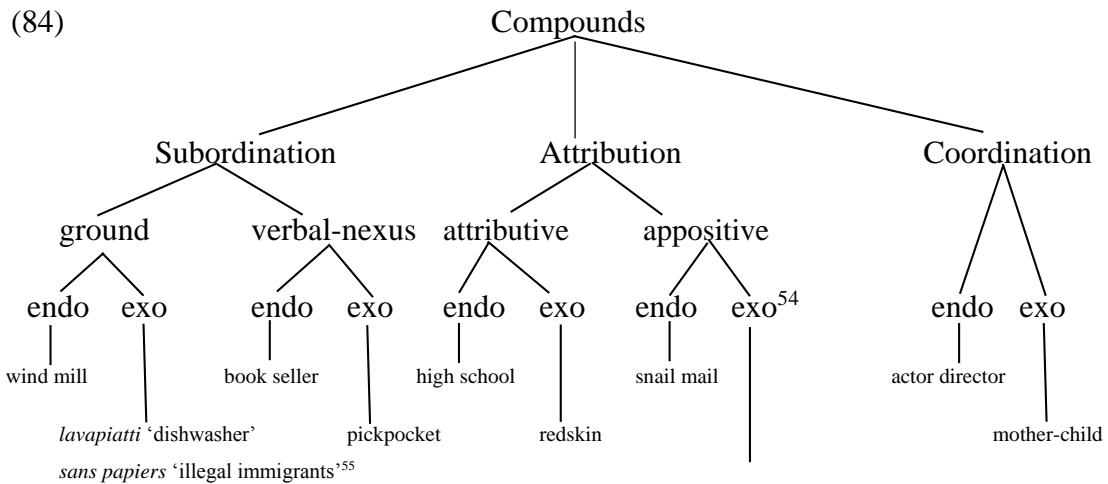
between the two types needs to be made more systematic, especially in relation to the semantic contribution within the compound. For instance, the head and modifier provide an unbalanced semantic contribution in the attributive type; e.g. *snail mail* is a type of *mail*, yet it is not a *snail*, whereas the two elements of the compound have a more equal status in the coordinative type, e.g. a *producer-director* is both a *producer* and a *director*. Yet there are cases where the relation is less clear-cut. For example, Bisetto and Scalise (2005: 328) consider the compound *ape man* an attributive compound, but one can wonder: could *ape man* be regarded as a coordinative compound? In other words, *ape man* could have two possible meanings: (1) a man that looks like an ape (attributive reading); or (2) a man who is also an ape (coordinative reading).

### 7.3.3 Scalise and Bisetto's (2009) classification

Scalise and Bisetto (2009) revised their classification of compounds of 2005 on the basis of further data and analysis. A level has been added to capture the different types of semantic/interpretive relations that exist between the two elements of the compound in each category. Starting with subordinate compounds, these are divided into ground compounds and verbal-nexus compounds. The former, also called root compounds, do not have a verb accompanied by one of its arguments; rather they depend on the semantics of the elements to interpret the meaning of the compounds, e.g. *wind mill*. On the other hand, verbal-nexus subordinate compounds, e.g. *bookseller*, are identified by a verb-argument relation between the constituents, or possibly a verb-adjunct relation (cf. Selkirk 1982). An example of an adjunct relation between the elements of a compound would be the compound *tree eater* used in the meaning 'someone eating on a tree' (Scalise and Bisetto 2009: 50-51).

In comparison with their classification of 2005, a new label is proposed to express a different attribution relation, enabling a separation between two types of semantic relations, i.e. attributive and appositive. In attributive compounds, the head is modified by the non-head, expressing a property of the head whether it is an adjective or a verb. In this way, the non-head element expresses a quality of the head element, e.g. *redskin* and *high school*. In appositive compounds, the non-head element expresses a property of the head element using a noun, an apposition, which serves as an attribute, e.g. *snail mail*.

Taking all this together, Scalise and Bisetto's (2009: 49-50) new classification of compounds looks as follows.



As can be seen, English does not have examples for all categories and some of the examples given are therefore Italian and French. Scalise and Bisetto (2009) indeed claim that the classification has cross-linguistic validity. The next sub-sections apply it to Arabic compounds.

### 7.3.3.1 Subordinate compounds

In Arabic, the following examples can be considered subordinate endocentric compounds:

(85)      daliil            l-mustaxdim  
           guide            the-user  
           ‘user guide’

(86)      qaa?id            l-hamlah  
           leader            the-campaign  
           ‘the campaign leader’

(87)      t‘aahuunat        l-hawaa?  
           mill            the-air  
           ‘the wind mill’

<sup>54</sup> Based on the CompoNet database of 23 languages, developed at the University of Bologna, Scalise and Bisetto (2009: 49-50) note that exocentric appositive compounds are not easy to track, with possible exceptions like the Norwegian *kryssord* ‘crossword = crossword puzzle’ and the Chinese *rén shé* ‘people snake = illegal immigrant’. Here, one may wonder why the equivalent of the Norwegian compound *kryssord* ‘crossword = crossword puzzle’ in English, i.e. *crossword* is not regarded as an example of exocentric appositive compound.

<sup>55</sup> The French compound *sans papiers* ‘illegal immigrants’ literally translates as ‘without documents’, consisting of *sans* ‘without’ and *papiers* ‘documents’.

Examples (85-87) show that the whole compound is a hyponym of the head, indicating that these compounds are endocentric. On the other hand, in examples (88) and (89), the whole compound is not a hyponym of the head:

(88)      qaat<sup>ʕiʕ</sup>      t<sup>ʕ</sup>-t<sup>ʕ</sup>ariiq  
               crosser      the-road  
               ‘the bandit’

(89)      liʕb      ʕyaal  
               playing      children  
               ‘child’s play’

Specifically, the compounds *qaat<sup>ʕiʕ</sup> t<sup>ʕ</sup>ariiq* is ‘someone who stops people on the road to rob them’ and *liʕb ʕyaal* ‘child’s play’ are not hyponyms of the heads, i.e. *qaat<sup>ʕiʕ</sup>* ‘crosser’ and *liʕb* ‘play’, respectively.

In terms of the difference between ground and verb nexus, *t<sup>ʕ</sup>aaḥuunat lhawaaʔ* ‘the wind mill’ and *liʕb ʕyaal* ‘child’s play’ represent subordinate ground compounds. They are interpreted as *t<sup>ʕ</sup>aaḥuunat lilhawaaʔ* ‘mill powered by wind’ and *liʕb ʕyaal* ‘child’s play’ as a game for children. On the other hand, *daliil lmustaxdim* ‘user guide’ and *qaaʔid lḥamlah* ‘campaign leader’ are good examples of verb nexus. They are interpreted as ‘a booklet to guide users’ and ‘someone who leads a campaign’, respectively.

There are also many verbal nexus compounds in Arabic in which the argument is the subject of the action, not the object or complement. This can be seen in the following examples:

(90)      wus<sup>ʕuul</sup>      l-malik  
               arriving      the-king  
               ‘the arrival of the king’

(91)      zaʔiir      l-ʔasad  
               roaring      the-lion  
               ‘the roaring of the lion’

In these examples, the subject/doer of the action in (90) is *lmalik* ‘the king’ and in (91) is *lʔasad* ‘the lion’. It is clear that the crux of these interpretations is the subject.

### 7.3.3.2 *Attributive compounds*

According to Scalise and Bisetto (2009), attributive-appositive compounds exhibit a modifier-head relationship, either directly or metaphorically. In Arabic, the following is an example of attributive endocentric compounds:

- |      |                    |          |
|------|--------------------|----------|
| (92) | sikkat             | l-ḥadiid |
|      | rail               | the-iron |
|      | ‘the iron railway’ |          |

Example (92) demonstrates that the compound *sikkat ḥadiid* ‘the iron railway’ is a subset of its head, i.e. *sikkat* ‘railway’. Conversely, in examples (93-96), the compounds are not a subset of their heads, as in:

- |      |                           |                 |
|------|---------------------------|-----------------|
| (93) | raʔs                      | l-maal          |
|      | head                      | the-money       |
|      | ‘the capital’             |                 |
|      |                           |                 |
| (94) | ṡaruus                    | l-baḥr          |
|      | bride                     | the-sea         |
|      | ‘mermaid’                 |                 |
|      |                           |                 |
| (95) | ṡaziiz                    | n-nafs          |
|      | dear                      | the-soul/spirit |
|      | ‘the proud one in spirit’ |                 |
|      |                           |                 |
| (96) | maaʔ                      | l-wadʒh         |
|      | water                     | the-face        |
|      | ‘the dignity’             |                 |

Examples (93-96) are exocentric, since the compounds *raʔs lmaal* ‘the capital, money’, *ʕaruus lbaħr* ‘mermaid’, *ʕaziiz nnaʕs* ‘the proud one in spirit’, *maaʔ l-waɖʒh* ‘dignity’ do not denote a subset of the head elements *raʔs* ‘head’, *ʕaruus* ‘bride’, *ʕaziiz* ‘dear’ and *maaʔ* ‘water’, respectively.

In terms of attributive and appositive on the second layer, *ʕaziiz nnaʕs* ‘the proud person’ is an example of attributive compounds, in which the non-head expresses a quality of the implicit head *one/person*, realised as an adjective, i.e. *ʕaziiz* ‘dear’. In contrast, *sikkat lħadiid* ‘the iron railway’ and *raʔs lmaal* ‘the capital (money)’ are examples of appositive compounds, in which the non-head expresses a property of the head, represented as a noun (apposition), i.e. *sikkat* ‘railway’ and *raʔs* ‘head’.

An interesting point to mention here is related to the productivity of exocentric attributive compounds in Arabic. Bauer *et al.* (2013: 478) note that exocentric attributives in English are very productive, and they are all metaphorical or metonymic in nature. Some examples, in which the second element is the word *head* are: *air head*, *block head*, *meat head*, *dick head* and *egg head*. Some other examples are *hardtop* ‘car’, *turtleneck* ‘garment’, *underground* ‘railway’ and *house-warming* ‘party’. However, examples of the same type of compounds in Arabic are rare, and to the best of my knowledge, no new compounds have been recently coined in this category.

### 7.3.3.3 Coordinate compounds

Coordinate compounds occur when the first element of the compound does not modify the second; instead the two are equal, such as *doctor-patient*, *blue-green*, *producer-director* and *prince consort*. MSA has only a very small number of coordinate compounds, such as *sʕabaah masaaʔ* ‘lit. morning evening, all day long’, *layla nahaar* ‘lit. night day, twenty four seven’ and *ħaamidʕ ħilw* ‘lit. sour-sweet’. In MSA, coordinating compounds are syntactically conditioned that is, the lexical category of the whole compound is either a noun or an adjective. They are also semantically conditioned, since they contain antonyms. JA has some more N + N coordinative compounds than MSA, as in<sup>56</sup>:

---

<sup>56</sup> The two elements in examples (97-99) are active participles, but they do not have a one-to-one equivalent in English.



- (97) l-walad      **t'aayih**      **t'aaliṣ**      ṣala      d-daradʒ  
the-boy      is going down      is going up      on      the-stairs  
‘The boy keeps going up and down the stairs’
- (98) mhammad      **raayih**      **dʒaay**      ʔala      d-dukkan  
Mohammad      going      coming      to      the-minimarket  
‘Mohammad keeps going to and coming from the minimarket’
- (99) ali      **maakil**      **jaarib**      **naayim**      ṣinna  
Ali      eating      drinking      sleeping      at.ours (house)  
‘Ali is relaxing at our house (in the sense that he is not doing anything).’  
lit. Ali is eating, drinking and sleeping at our house.
- (100) ali      **ṣatʔaal**      **batʔaal**  
Ali      unemployed      quitter  
‘Ali is jobless.’

Examples (97-100) are coordinative compounds in bold found in JA. Example (93) is an instance of a recursive coordinative compound that consists of three elements, i.e. *maakil* ‘eating’, *jaarib* ‘drinking and *naayim* ‘sleeping’. Interestingly, coordinative compounds can also be recursive in English. For instance, Bauer *et al.* (2013: 444, 479-80) note that the number of elements in coordinative compounds is not limited to two nouns. For instance, compounds like *actor-producer-director* are entirely possible and three-member compounds can also be found in French (Bauer 2003: 43), as in (101):

- (101) *bleu-blanc-rouge*  
blue-white-red  
‘the French flag’

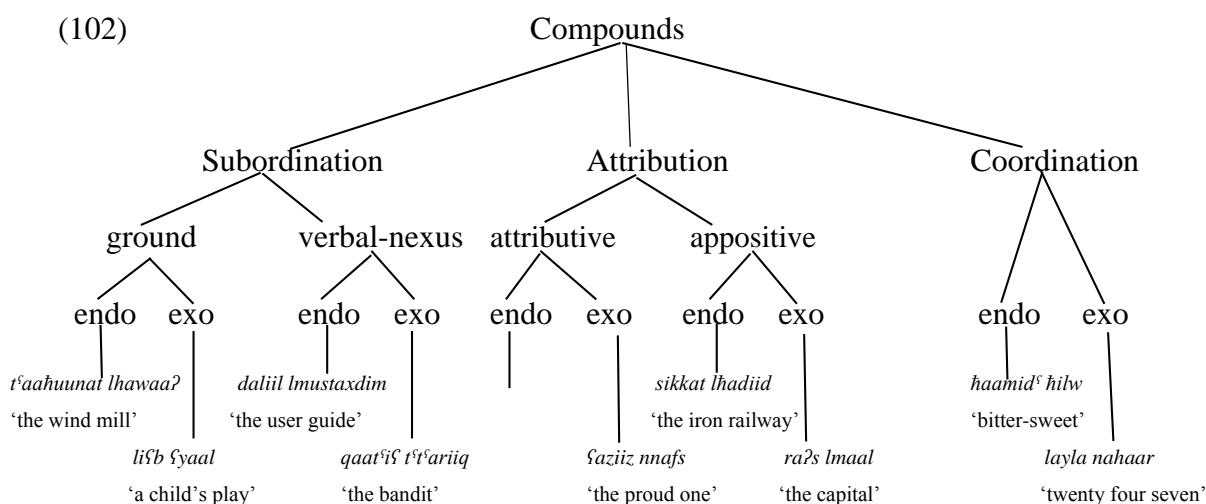
Altogether, while the subordination and attribution categories of Scalise and Bisetto (2009) are very well represented in Arabic, the coordination category is much more limited. The following table is a summary of the main coordinative compounds found in Arabic.

Table 7.1. Coordinating compounds in Arabic

Variety	Compound	First element	Second element	Third element	Gloss	Meaning
MSA	s <sup>ʕ</sup> abaah masaaʔ	s <sup>ʕ</sup> abaah ‘morning’	masaaʔ ‘evening’	-	morning- evening	all day long
MSA + JA	hilw murr	hilw ‘sweet’	murr ‘bitter’	-	bitter-sweet	bitter-sweet
MSA + JA	haamid <sup>ʕ</sup> hilw	haamid <sup>ʕ</sup> ‘sour’	hilw ‘sweet’	-	sour-sweet	sour-sweet
MSA + JA	layla nahaar <sup>57</sup>	layla ‘night’	nahaar ‘daytime’	-	night-day	twenty-four seven
JA	t <sup>ʕ</sup> aayih t <sup>ʕ</sup> aaliʕ	t <sup>ʕ</sup> aayih ‘going down’	t <sup>ʕ</sup> aaliʕ ‘going up’	-	going down and going up	going up and down
JA	raayih dzaay	raayih ‘going’	dzaay ‘coming’	-	going and coming	going and coming
JA	maakil ʕaarib naayim	maakil ‘eating’	ʕaarib ‘drinking’	naayim ‘sleeping’	eating, drinking and sleeping	doing nothing
JA	ʕat <sup>ʕ</sup> ʕaal bat <sup>ʕ</sup> ʕaal	ʕat <sup>ʕ</sup> ʕaal ‘unempolyed’	bat <sup>ʕ</sup> ʕaal ‘quitter’	-	unemployed and quitter	jobless

### 7.3.4 Summary

The following tree diagram repeats Scalise and Bisetto’s (2009) taxonomy with some illustrative Arabic examples:



We may conclude that Scalise and Bisetto’s (2009) taxonomy is indeed applicable to the Arabic language, which supports the universality of this taxonomy.

<sup>57</sup> In JA, it is pronounced as *leel nhaar*.

## **7.4 Conclusion**

To sum up, it has been confirmed that compounding in Arabic N + N and Adj + N combinations within SGC is left headed, regardless of the types of head. In Adj + N combination, a silent noun is responsible for the syntactic category of the whole construct. This silent noun is the head, which fits in with the fact that Arabic is a predominantly head-initial language. Semantically, compounds other than SGC can be either double-headed or headless. In addition, Arabic compounds show the universal applicability of Scalise and Bisetto's (2009) classification. Finally, note that a difference between MSA and JA is that recursive coordinative compounds occur only in the latter.

## **Chapter Eight: Conclusion**

### **8.1 Introduction**

This chapter summarises the main points discussed in this study. It also makes recommendations for further studies.

### **8.2 Main findings about compounding in MSA and JA**

Firstly, after investigating compounding in MSA and JA, a compound can be defined as a complex word that consists of at least two adjacent words, and where the non-head is normally non-referential. Note that this definition seems also applicable to Hebrew compounds on the basis of Borer's (2009) analysis. Furthermore, this definition fits with the working definition suggested for compound in English in section 2.5, which is:

a complex word that consists of at least two adjacent elements, in which each of these elements is either a word, combining form or a phrase, so that the whole compound is a combination of these elements.

There is, however, a subtle difference between English and Arabic compounds, in that the elements of the latter are in all cases attested words which can stand on their own, whereas those of the former can either be a word, a combining form or a phrase. Therefore, I would propose the following general definition to identify compounds cross-linguistically, incorporating the idea of non-referentiality:

A compound is a complex word that consists of at least two adjacent elements, where the non-head is normally non-referential. Each of these elements is either a word, combining form or a phrase, so that the whole compound is a combination of these elements.

Secondly, it has become apparent that the four most general and reliable criteria to distinguish between compounds and P-constructs (phrases) in MSA and JA are adjacency, referentiality, the impossibility of free pluralisation of the non-head and the impossibility of resumptive coordination. Essentially, the elements of compounds are inseparable, the non-head is normally non-referential and cannot be freely pluralised, and the whole compound cannot be coordinated

with another noun marked with a resumptive pronoun referring back to the non-head. In contrast, the elements of P-constructs are separable, the non-head can freely be referential and can be freely pluralised, and the whole phrase can be coordinated with another noun marked with a resumptive pronoun referring back to the non-head. In addition, I have suggested two language specific criteria that are exclusive to Arabic, or potentially Semitic languages in general. The first criterion is concerned with the definiteness of the head. It has been observed that in MSA when the head of a P-construct is marked with the definite article, the non-head has to be marked with the possessive marker *li-* ‘of/for’. However, when the head of a compound (whether compositional or non-compositional) is marked with the definite article, the possessive marker *li-* does not appear in MSA. Unlike MSA, in JA the possessive marker *tabaʕ* ‘for/of’ appears in compositional compounds. The second criterion also involves the appearance of the possessive marker *li-/la* ‘for/of’, but concerns cases where the head is preceded by a cardinal number. In P-constructs, when a cardinal number appears before the head, it triggers the appearance of the possessive marker *li-/la* ‘for/of’. On the other hand, this does not apply to either compositional or non-compositional compounds in both MSA and JA.

Other criteria which have been shown to be partially applicable are compositionality, modification and coordination. Specifically, compounds can be either compositional or non-compositional, the non-head of compositional compounds can be modified, whereas that of non-compositional ones cannot, and the non-head of compositional compounds can be coordinated, while that of non-compositional compounds cannot. Conversely, P-constructs are always compositional and the non-head can be both modified and coordinated.

Criteria which fail to distinguish between compounds and P-constructs in Arabic are orthography, sandhi and stress. Even though stress does not distinguish between compounds and P-constructs, the Praat analysis that was carried out yielded worthwhile results. Praat analysis showed that the default position of the stress in SGC in MSA and JA is on the first element contrary to what has been suggested in the relevant literature (cf. Siloni 1997 and Alexiadou *et al.* 2007). There is only one systematic exception, which is phonetically conditioned: in N + N combinations with assimilated geminates on the word boundary, a secondary stress or perhaps double stress is assigned.

Thirdly, I have shown that all cases of Adj + N compounding in MSA are bahuvrihis, since they denote a person in possession of the entity denoted by the compound. It has also become apparent that the output of Adj + N compounding behaves more like a noun than an adjective. I have argued that such constructions have a silent N head, i.e. ‘one/person’, which

determines the syntactic category of Adj + N compounds in MSA (cf. Günther (to appear) for a similar analysis of nominalised adjectives in English).

Fourthly, in terms of the syntactic category of their internal elements, there are five main types of compounds in MSA, i.e. N + N, Adj + N, N + Adj, Adj + Adj and Prep + Prep. In contrast, there are only four in JA, i.e. N + N, V + V, N + Adj and Adj + Adj. Further investigation showed that reduplications and numerals from eleven to nineteen are also to be treated as compounds, in both MSA and JA.

Fifthly, compounding in Arabic N + N and Adj + N combinations within SGC is left-headed, regardless of the types of head. Compounds other than SGC can be either semantically double-headed or headless. Specifically, reduplicated compounds, Adj + Adj and N + N combinations can be double-headed or headless, while V + V compounds seem to be right-headed.

Finally, Arabic compounds show the wide applicability of Scalise and Bisetto's (2009) classification. In Arabic compounds, cases of both endocentric and exocentric examples have been found. In addition, Arabic exhibits a wide range of subordinate, attributive and coordinate compounds. With regard to the specific class of coordinative compounds, I established a hitherto unnoted difference between MSA and JA: only the latter allows recursion in this class.

### **8.3 Recommendations for further research**

The findings of the present study highlight a few specific areas where further research could be fruitfully carried out. For one thing, it has shown that cross-linguistic studies are required to identify criteria to distinguish between verb serialisation and V + V compounding. For Arabic in particular, more research on all linguistic features of serial verb constructions in Arabic is needed, including the pragmatics of verb serialisation in various contexts in order to identify their socio-pragmatic functions.

It can also be noted that the syntax and semantics of definiteness vs. indefiniteness in MSA and other Arabic dialects is worthy of further investigation. Although proper nouns are definite by default, they can be marked with the so-called indefinite marker, i.e. nunation in MSA. Another intriguing aspect of definiteness is its function to transform cardinal numbers into ordinal ones and attributive adjectives into predicative ones.

Furthermore, the first/second language acquisition of Arabic compounds is an area worthy of investigation (cf. Dressler *et al.* 2010 and Gagné and Spalding 2010). This may

reveal details of the process of acquiring these complex words and their properties which were identified in the preceding chapters.

Finally, I recommend a full cognitive investigation of metaphoric and metonymic relations in Arabic compounding in line with Conceptual Metaphor Theory (Lakoff and Johnson 2003) and Blending Conceptual Theory (Turner and Fauconnier 2002). This analysis can illustrate the conceptual relationship between the internal elements of the compounds (cf. Jackendoff 2009) and whether their transparency plays a role in their mental processing.

## References

- Abney, S. P. 1987. *The English Noun Phrase in its Sentential Aspect*. PhD dissertation: Massachusetts Institute of Technology, Massachusetts.
- Abu Abbas, K. 2008. 'Introducing weight sensitive EDGEMOST'. *SKY Journal of Linguistics* 21: 11-36.
- Aikhenvald, A. 2006. 'Serial verbs constructions in a typological perspective'. In Aikhenvald, A. Y. and Dixon, R. M. (eds.) *Serial Verb Constructions: A Cross-Linguistic Typology*. Oxford: Oxford University Press. 1-68.
- Al Mahmoud, M.S. 2014. 'Prenominal and postnominal adjectives in Arabic: A proposed analysis'. *SAGE Open*: 1-8.
- Al-Distour Newspaper. Last accessed 21<sup>st</sup> February 2016, from: <http://www.addustour.com/17704>.
- Alexiadou, A., Haegeman, L. and Stavrou, M. 2007. *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter.
- Al-Ghad Newspaper. Last accessed 21<sup>st</sup> February 2016, from: <http://www.alghad.com/prints/893381->
- Al-Hariiri, M. 2013. 'The structure of names in Arabic language'. *Al Hijaz International Refereed Journal for Islamic & Arabic Studies* 2: 168-193.
- Al-Humaydi, A. 2005. 'Iasmaa Imurakkabah: anwaaha wa iraabha. diraasah nahwiyya [Compound nouns: Types and parsing. a syntactic study]'. *Ad Diriyah Journal* 29: 243-264.
- Al-Jahaawi, A. 1982. 'Zahirat al-tanwen fi al-lughah al-arabiyeh'. Riyaad: Dar Al- Rifaa.
- Allen, M. R. 1978. *Morphological Investigations in English*. PhD dissertation: University of Connecticut, Storrs.
- Al-Rajihi, A. 2000. *altat'biiq nnaħwii* [Applied Syntax]. Alexandria: Dar Al-Maarifah Al-Jamiya.
- Amer, W. M. and Menacere, K. 2013. 'The challenges of translating English compounds into Arabic: For better or for worse'. *Babel* 59(2): 224-243.
- Andersen, H .1986. 'Sandhi'. Andersen, H. (ed.) *Sandhi Phenomena in the Languages of Europe*. Berlin, New York, Amsterdam: Mouton de Gruyter. 1-8.
- Arcodia, G. F. 2012. 'Constructions and headedness in derivation and compounding'.



*Morphology* 22(3): 365-397.

- Aronoff, M. and K. Fudeman. 2005. *What is Morphology?* Oxford: Blackwell.
- Attia, M. A. 2006. 'Accommodating multiword expressions in an Arabic LFG grammar'. In Pahikkala, S. (ed.) *Advances in Natural Language Processing*. Springer Berlin Heidelberg. 87-98.
- Baker, B. 2008. *Word Structure in Ngalakgan*. Stanford CA: CSLI.
- Baker, M. 1992. *In Other Words: A Coursebook on Translation*. London: Routledge.
- Baker, M. C. 2003. *Lexical Categories: Verbs, Nouns and Adjectives*. Cambridge: Cambridge University Press.
- Bamakhramah, M. 2009. *Syllable Structure in Arabic Varieties with a Focus on Superheavy Syllables*. PhD dissertation: Indiana University.
- Bauer, L. 1978. *The Grammar of Nominal Compounding*. Odense: Odense University Press.
- Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, L. 1990. 'Be-heading the word'. *Journal of Linguistics* 26: 1-31.
- Bauer, L. 1998a. 'When is a sequence of two nouns a compound in English?' *English Language and Linguistics* 2: 65-86.
- Bauer, L. 1998b. 'Is there a class of neoclassical compounds in English and is it productive?' *Linguistics* 36: 403-422.
- Bauer, L. 2001. 'Compounding'. In Haspelmath, M., Ekkehard, K., Wulf., Ö and Wolfgang, R. (eds) *Language Typology and Language Universals*. Berlin: Mouton de Gruyter. 695-707.
- Bauer, L. 2003. *Introducing Linguistic Morphology*. 2<sup>nd</sup> ed. Edinburgh: Edinburgh University Press.
- Bauer, L. 2005. 'The borderline between derivation and compounding'. In Dressler, W. U, Kastovsky, D, Pfeiffer, O. E. and Rainer, F. (eds) *Morphology and its Demarcations*. Amsterdam: John Benjamins Publishing. 97-108.
- Bauer, L. 2008. 'Exocentric compounds'. *Morphology* 18: 51-74.

- Bauer, L. 2009a. 'IE, Germanic: Danish'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 400-416.
- Bauer, L. 2009b. 'Typology of compounds'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 343-356.
- Bauer, L. 2010. 'The typology of exocentric compounding'. In Scalise, S and Vogel, I. (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 167-175.
- Bauer, L., Lieber, R. and Plag, I. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Bauer, L. and Renouf, A. 2001. 'A corpus-based study of compounding in English'. *Journal of English Linguistics* 29: 101-23.
- Bell, M. J. and Plag, I. 2012. 'Informativeness is a determinant of compound stress in English'. *Journal of Linguistics* 48(3): 485-520.
- Bell, M. J. and Plag, I. 2013. 'Informativity and analogy in English compound stress'. *Word Structure* 6(2): 129-155.
- Benczes, R. 2006. *Creative Compounding in English: The Semantics of Metaphorical and Metonymical Noun-Noun Combinations*. Amsterdam: John Benjamins Publishing.
- Beyer, S. 1992. *The Classical Tibetan Language*. New York: State University of New York.
- Biber, D. and Gray, B. 2011. 'Grammatical change in the noun phrase: The influence of written language use'. *English Language and Linguistics* 15(2): 223-250.
- Bisang, W. 2009. 'Serial verb constructions'. *Language and Linguistics Compass* 3(3): 792- 814.
- Bisetto, A. and Scalise, S. 2005. 'The classification of compounds'. *Lingue e linguaggio* 4(2): 319-332.
- Bloomfield, L. 1935. *Language*. London: Allen and Unwin.
- Boersma, P. and Weenink, D. 2015. Praat: Doing phonetics by computer (Version 5.4.08) [computer program]. Last accessed 15<sup>th</sup> May 2015, from: <http://www.praat.org>.
- Bolinger, D. 1961. 'Contrastive accent and contrastive stress'. *Language* 37: 83-96.

- Booij, G. 1992. 'Compounding in Dutch'. *Rivista di Linguistica* 4(1): 37-59.
- Booij, G. 2002. 'Constructional Idioms, Morphology, and the Dutch Lexicon'. *Journal of Germanic Linguistics* 14: 301-327.
- Booij, G. 2005. *The Grammar of Words*. Oxford, Oxford University Press.
- Booij, G. 2007. *The Grammar of Words*. 2<sup>nd</sup> ed. Oxford: Oxford University Press.
- Booij, G. 2009. 'Construction morphology and compounding'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 201-216.
- Booij, G. 2010a. 'Compound construction: Schemas or analogy? A construction morphology perspective'. In Scalise, S. and Vogel, I. (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 93-107.
- Booij, G. 2010b. 'Constructions and lexical units: An analysis of Dutch numerals'. In Olsen, S. (ed) *New Impulses in Word Formation*. Hamburg: Helmut Buske Verlag. 81-100.
- Booij, G. 2012. *The Grammar of Words*. 3<sup>rd</sup> ed. Oxford: Oxford University Press.
- Borer, H. 2009. 'Afro-Asiatic, Semitic: Hebrew'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 386-399.
- Borgwaldt, S. R. and Benczes, R. 2011. 'Novel object naming in Hungarian and German'. In Schönefeld, D. (ed.) *Converging Evidence: Methodological and Theoretical Issues for Linguistic Research*. Amsterdam: John Benjamins Publishing.
- Campbell, N. 2002. 'Recording techniques for capturing natural every-day speech'. In *Proc Language Resources and Evaluation Conference (LREC-02)*. Las Palmas, Spain. 2029–2032.
- Carstairs-McCarthy, A. 2002. *An Introduction to English Morphology*. Edinburgh: Edinburgh University Press.
- Čermák, J. 1997. 'A prow in foam: The Old English bahuvrihi compound as a poetic device.' *Prague Studies in English*: 13-31.

- Chafe, W. 2012. 'Are adjectives universal? The case of Northern Iroquoian'. *Linguistic Typology* 16(1): 1-39.
- Choi, J. 2014. *Pronoun-Noun Constructions and the Syntax of DP*. PhD dissertation: The University of Arizona, Arizona.
- Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.
- Creissels, D. 2004. 'Bambara'. In Arnaud, P. J. L. (ed) *Le nom composé: données sur seize langues*. Lyon: Presses Universitaires de Lyon. 21-46.
- Croft, W. 1995. 'What's a head?' Rooryck, J. and Zaring, L. (eds) *Phrase Structure and the Lexicon*. Amsterdam: Springer. 35-62
- Croft, W. and Cruse, D. A. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.
- Crowley, T. 1987. 'Serial verbs in Paamese'. *Studies in Language* 11(1): 35–84.
- Crowley, T. 2002. *Serial Verbs in Oceanic: A Descriptive Typology*. Oxford: Oxford University Press.
- Davis, S. 2011. 'Geminates'. In van Oostendorp, M., Ewen, C.J., Hume, E and Rice, K. (eds) *The Blackwell Companion to Phonology*. Oxford: Wiley-Blackwell. 837- 859.
- De Belder, M. 2013. 'Collective mass affixes: When derivation restricts functional structure'. *Lingua: International Review of General Linguistics* 126: 32-50.
- De Jong, K. and Zawaydeh, B.A. 1999. 'Stress, duration, and intonation in Arabic word-level prosody'. *Journal of Phonetics* 27(1): 3-22.
- Di Sciullo, A. M. 2009. 'Why are compounds a part of human language? A view from asymmetry theory'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 145-177.
- Dirven, R. and Verspoor, M. 1998. *Cognitive Exploration of Language and Linguistics*. Amsterdam: John Benjamins Publishing.
- Don, J. 2009. 'IE, Germanic: Dutch'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 370-385.
- Donalies, E. 2004. *Grammatik des Deutschen im Europäischen Vergleich*. Kombinatorische

- Begriffsbildung: Substantivkomposition. Mannheim: Institut für Deutsche Sprache.
- Donohue, M. 2003. 'Morphological templates, headedness, and applicatives in Barupu'. *Oceanic Linguistics*: 111-143.
- Dressler, W. U. 1986. 'Forma Y Funcion de los Interfijos'. *Revista Española de Lingüística* 16: 381-395.
- Dressler, W. U. 2006. 'Compound Types'. In G. Libben and Jarema, G. (eds) *The Representation and Processing of Compound Words*. Oxford: Oxford University Press. 23-44.
- Dressler, W. U., Lettner, L. E. and Korecky-Kröll, K. 2010. 'First language acquisition of compounds—with special emphasis on early German child language'. In Scalise, S. and Vogel, I. (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 323-344.
- Durie, M. 1988. 'Verb serialization and "Verbal-Prepositions" in Oceanic Languages'. *Oceanic Linguistics* 27: 1-23.
- Fabb, N. 1998. 'Compounding'. In Zwicky, A. and Spencer, A. (eds) *The Handbook of Morphology*. Oxford: Blackwell. 66-83.
- Fàbregas, A. and Scalise, S. 2012. *Morphology from Data to Theory*. Edinburgh: Edinburgh University Press.
- Faraclas, N. 2013. *Nigerian Pidgin*. London: Routledge.
- Fassi-Fehri, A. 1993. *Issues in the Structure of Arabic Clauses and Words*. Dordrecht: Kluwer Academic Publishers.
- Fassi-Fehri, A. 1999. 'Arabic modifying adjectives and DP structures'. *Studia Linguistica* 53(2): 105-154.
- Fassi-Fehri, A. 2012. *Key Features and Parameters in Arabic Grammar*. Amsterdam: John Benjamins Publishing.
- Fender, M. 2008. 'Arabic literacy development and cross-linguistic effects in subsequent L2 literacy development'. In Koda, K. and Zehler, A. M. (eds) *Learning to Read across Languages: Cross-Linguistic Relationships in First-and Second-Language Literacy*

- Development*. New York: Routledge. 101-124.
- Fernando, C. 1996. *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Fernando, C. and Flavell, R. 1981. *On Idiom: Critical Views and Perspectives*. Exeter: University of Exeter Press.
- Fischer, R. 1998. *Lexical Change in Present-Day English: A Corpus-Based Study of the Motivation, Institutionalization, and Productivity of Creative Neologisms*. Tübingen: Gunter Narr Verlag Tübingen.
- Fleischer, W. 1975. *Wortbildung der Deutschen Gegenwartssprache*. Durchgesehene Auflage. Tübingen: Niemeyer.
- Foley, W. A. 1997. 'Polysynthesis and complex verb formation: The case of applicatives in Yimas'. In Alsina, A., Bresnan, J. and Sells, P. (eds) *Complex Predicates*. Stanford: CSLI. 355-395.
- Foley, W. A. and Olson, M. 1985. 'Clausehood and verb serialization'. In Nichols, J. and Wodbury, A. (eds) *Grammar inside and outside the Clause*. Cambridge: Cambridge University Press. 17-60.
- Foley, W. A. and Van Valin, R. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Fradin, B. 2009. 'IE, Romance: French'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 417-435
- Gaeta, L. and Ricca, D. 2009. 'Composita solvantur: Compounds as lexical units or morphological objects'. *Italian Journal of Linguistics* 21(1): 35-70.
- Gagné, C.L. and Spalding, T. L. 2010. 'Relational competition during compound interpretation'. In Scalise, S. and Vogel, I. (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 287-300.
- Giegerich, H. J. 2004. 'Compound or phrase? English noun-plus-noun constructions and the stress criterion'. *English Language and Linguistics* 8: 1-24.
- Giegerich, H. J. 2005. 'Associative adjectives in English and the lexicon-syntax interface'. *Journal of Linguistics* 41 (2): 571-591.
- Girju, R., Moldovan, D., Tatu, M. and Antohe, D. 2005. 'On the semantics of noun

- compounds'. *Computer Speech & Language* 19(4): 479-496.
- Girl friend. 2008. *Webster's Third New International Dictionary*. Merriam Webster USA.
- Girl-friend. 1982. *Concise Oxford Dictionary 7th Edition*. Oxford: Oxford University Press.
- Girlfriend. *Macmillan Dictionary*. Last accessed 11<sup>th</sup> September 2015, from: [http://www.macmillandictionary.com/dictionary/british/girlfriend\\_](http://www.macmillandictionary.com/dictionary/british/girlfriend_)
- Givón, T. 1979. 'Prolegomena to any sane creology'. In Hancock, I. (ed.) *Readings in Creole Studies*. Amsterdam: John Benjamins Publishing. 3-35.
- Goossens, L. 1990. 'Metaphonymy: The interaction of metaphor and metonymy in expressions for linguistic action'. *Cognitive Linguistics* 1(3): 323-342.
- Gross, G. 1996. *Les expressions figées en français*. Gap: Ophrys
- Group, P. 2007. 'MIP: A method for identifying metaphorically used words in discourse'. *Metaphor and Symbol* 22(1): 1-39.
- Günther, C. to appear. 'The rich, the poor, the obvious: Arguing for an ellipsis analysis of "adjectives used as nouns"'. In Leung, A. and van der Wurff, W. (eds) *The Noun Phrase in English: Past and Present*. Amsterdam: John Benjamins Publishing.
- Hammond, M. 1999. *The Phonology of English*. Oxford: Oxford University Press.
- Haspelmath, M. 2002. *Understanding Morphology*. London: Hodder.
- Haspelmath, M. 2012. 'How to compare cajor word classes across the world's languages' *UCLA Working Papers in Linguistics: Theories of Everything* 17(16): 109-130.
- Hayes, B. 1989. 'Compensatory lengthening in moraic phonology'. *Linguistic Inquiry* 20: 253-306.
- Hayes, B. 1995. *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- Henri, F. 2012. 'Attenuative verbal reduplication in Mauritian: A morpho-semantic approach'. In Aboh, E. O, Smith, N. and Zribi-Hertz, A. (eds) *The Morphosyntax of Reiteration in Creole and Non-Creole Languages* Amsterdam: John Benjamins Publishing. 203-234.

- Heselwood, B. and Watson, J. C. E. 2013. 'The Arabic definite article does not assimilate'. *Leeds Working Papers in Linguistics and Phonetics* 18: 34–53.
- Heyvaert, L. 2009. 'Compounding in Cognitive Linguistics'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 233–254.
- Huddleston, R. and Pullum, G. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hudson, R. A. 1987. 'Zwicky on heads'. *Journal of Linguistics* 23(01): 109–132.
- Hume, E., Muller, J. and van Engelenhoven, A. 1997. 'Non-moraic geminates in Leti'. *Phonology* 14: 371–402.
- Hurford, J. R. 2011. *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press.
- Jackendoff, R. 2002. *Foundations of Language*. Oxford: Oxford University Press.
- Jackendoff, R. 2009. 'Compounding in the parallel architecture and conceptual semantics'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 105–129.
- Ji, H., Gagné, C. L. and Spalding, T. L. 2011. 'Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds'. *Journal of Memory and Language* 65(4): 406–430.
- Johannessen, J. B. 1996. 'Partial agreement and coordination'. *Linguistic Inquiry* 27(4): 661–676.
- Jones, D. 1969. *An Outline of English Phonetics*. Cambridge: Heffner and Sons Ltd.
- Jongen, R. 1986. 'Les phénomènes de sandhi dans un dialecte bas-francique méridional'. In Andersen, H. (ed.) *Sandhi Phenomena in the Languages of Europe*. Berlin: Mouton de Gruyter, 117–143.
- Joseph, B. D. and Janda, R. D. 1988. 'The how and why of diachronic morphologization and demorphologization'. In Hammond, M. and Noonan, N. (eds) *Theoretical Morphology*. San Diego: Academic Press. 193–210.



- Katamba, F. 1993. *Morphology*. New York: St. Martin's Press.
- Katamba, F. and Stonham, J. 2006. *Morphology*. 2<sup>nd</sup> ed. London: Macmillan.
- Kavka, S. 2009. 'Compounds and idiomatology'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 19-33.
- Khaberni News Agency. Last accessed 21<sup>st</sup> February 2016, from: <http://www.khaberni.com/more-140763-1>.
- Khattab, G. and Al-Tamimi, J. 2014. 'Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure'. *Laboratory Phonology* 5(2): 231-269.
- Kiefer, F. 2009. 'Uralic, Finno-Ugric: Hungarian'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 527-541.
- Kouloughli, D. 2007. 'Sur la valeur du tanwiin: Nouvelle contribution a letude du systeme determinatif de l'arabe'. *Arabica* 54: 94-131.
- Kunter, G. 2011. *Compound Stress in English. The Phonetics and Phonology of Prosodic Prominence*. Berlin: de Gruyter.
- Labov, W. 1972. *Sociolinguistic Patterns*. Oxford: Blackwell.
- Lakoff, G. and Johnson, M. 2003. *Metaphors We Live By*. 2<sup>nd</sup> ed. Chicago: The University of Chicago Press.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford: Stanford University Press.
- Lathey, E. 1986. 'Pragmatic classification of idioms as an aid for the language learning'. *IRAL* 24(3): 217-233.
- Lefebvre, C. and Brousseau, A. 2002. *A Grammar of Fongbe*. Berlin: Mouton de Gruyter.
- Lieber, R. 1992. *Deconstructing Morphology. Word Formation in Syntactic Theory*. Chicago and London: The University of Chicago Press.
- Lieber, R. 2005. 'English word-formation processes'. In Štekauer, P. and Lieber, R. (eds) *Handbook of Word-Formation*. Dordrecht: Springer. 375-428.
- Lieber, R. 2009. 'IE, Germanic: English'. In Lieber, R. and Štekauer, P. (eds) *The Oxford*

- Handbook of Compounding*. Oxford: Oxford University Press. 357-369.
- Lieber, R. 2010. *Introducing Morphology*. Cambridge: Cambridge University Press.
- Lieber, R. and Štekauer, P. 2009. 'Introduction: Status and definition of compounding'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 3-18.
- Lipiński, E. 2001. *Semitic Languages: Outline of a Comparative Grammar*. Leuven: Peeters Publishers.
- Lipka, L. 1977. 'Lexikalisierung, Idiomatisierung und Hypostasierung als Probleme einer Synchronischen Wortbildungslehre'. In Brekle, H. E. and Kastovsky, D. (eds) *Perspektiven der Wortbildungsforschung*. Bonn: Bouvier. 155–164.
- Lord, C. 1993. *Historical Change in Serial Verb Constructions*. Amsterdam: John Benjamins Publishing.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Marchand, H. 1960. *The Categories and Types of Present-Day English Word-formation: A Synchronic-Diachronic Approach*. Wiesbaden: Otto Harrassowitz.
- Marchand, H. 1969. *The Categories and Types of Present-Day English Word-formation: A Synchronic-Diachronic Approach*. 2<sup>nd</sup> ed. München: Beck.
- Marelli, M. and Luzzatti, C. 2012. 'Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency'. *Journal of Memory and Language* 66(4): 644-664.
- Mathieu-Colas, M. 1996. 'Essai de typologie des noms composés français'. *Cahiers de lexicologie*: 71–125.
- McCarthy, J. 1979. 'On stress and syllabification'. *Linguistic Inquiry* 10 (3): 443-465.
- Miller, D. G. 2014. *English Lexicogenesis*. Oxford: Oxford University Press.
- Montermini, F. 2008. 'La composition en italien dans un cadre de morphologie lexématique'. In Amiot, D. (ed.) *La composition dans une perspective typologique*. Arras: Artois Presses Université, 161-187.

- Muysken, P. and Veenstra, T. 1995. 'Serial verbs'. In Adrends, J., Muysken, P. and Smith, N. (eds) *Pidgins and Creoles: An Introduction*. Amsterdam: John Benjamins Publishing. 289-301.
- Neef, M. 2009. 'IE, Germanic: German'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 386-399.
- Newman, P. 2000. *The Hausa Language: An Encyclopedic Reference Grammar*. London: Yale University Press.
- Norde, M. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Odden, D. 2005. *Introducing Phonology*. Cambridge: Cambridge University Press.
- Olsen, S. 2000. 'Compounding and stress in English: A closer look at the boundary between morphology and syntax'. *Linguistische Berichte* 181: 55–69.
- Olsen, S. 2001. 'Copulative compounds. A closer look at the interface between morphology and syntax'. In Booij, G. and Marle, J. (eds) *Yearbook of Morphology 2000*. Dordrecht: Kluwer. 279-320.
- Owens, J. 2013. 'A house of sound structure, of marvelous form and proportion: An introduction'. In Owens, J. (ed.) *The Oxford Handbook of Arabic Linguistics*. Oxford: Oxford University Press. 2-23.
- Papafragou, A. 1996. 'On metonymy'. *Lingua* 99: 169-195.
- Payne, T. E. 2011. *Understanding English Grammar: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.
- Plag, I. 2006. 'The variability of compound stress in English: Structural, semantic and analogical factors'. *English Language and Linguistics* 10: 143-172.
- Plag, I., Kunter, G. and Schramm, M. 2011. 'Acoustic correlates of primary and secondary stress in North American English'. *Journal of Phonetics* 39(3): 362-374.
- Plag, I., Kunter, G., Lappe, S. and Braun, M. 2008. 'The role of semantics, argument structure, and lexicalization in compound stress assignment in English'. *Language* 84 (4): 760-794.

- Polinsky, M. 2012. 'Headedness, again'. *UCLA Working Papers in Linguistics, Theories of Everything* 17 (40): 348-359.
- Preminger A. and Brogan, T. 1993. *The New Princeton Encyclopedia of Poetry and Poetics*. Princeton: Princeton University Press.
- Quine, W. 1982. *Methods of logic*. 4<sup>th</sup> ed. Cambridge Mass: Harvard University Press.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Rainer, F. and Varela, S. 1992. 'Compounding in Spanish'. *Rivista di Linguistica* 4: 117-142.
- Ralli, A. 2009. 'IE, Hellenic: Modern Greek'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 453-463.
- Ralli, A. 2010. 'Compounding versus derivation'. In Scalise, S and Vogel, I (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 57-76.
- Ralli, A. 2013. *Compounding in Modern Greek*. London: Springer Science and Business Media.
- Ralli, A. and Andreou, M. 2012. 'Revisiting exocentricity in compounding: Evidence from Greek and Cypriot'. In Kiefer, F., Ladányi, M and Siptár, P (eds) *Current Issues in Morphological theory: (Ir)regularity, Analogy and Frequency*. Amsterdam: John Benjamins Publishing. 65-81.
- Ringgen, C. and Robert M. V. 2011. 'Geminates: Heavy or long?' In Charles E. C. and Eric, R (eds) *Handbook of the Syllable*. Leiden: Brill. 155–170.
- Roelofs, A. and Baayen, H. 2002. 'Morphology by itself in planning the production of spoken words'. *Psychonomic Bulletin & Review* 9(1): 132-138.
- Rohrer, T. 1997. 'Conceptual blending on the information highway: How metaphorical inferences work'. In Liebert, W. A., Redeker, G. and Waugh, L. R. (eds) *Discourse and Perspective in Cognitive Linguistics*. Amsterdam: John Benjamins Publishing. 185-204.
- Rosenbach, A. 2006. 'Descriptive genitives in English: A case study on constructional gradience'. *English language and Linguistics* 10(1): 77-118.

- Rosenbach, A. 2007. 'Emerging variation: Determiner genitives and noun modifiers in English'. *English Language and Linguistics* 11: 143-189.
- Rutter, B. 2008. *Acoustic Properties of Repair Sequences in Dysarthric Conversational Speech: An Interactional Phonetic Study*. PhD dissertation: University of Louisiana, Lafayette.
- Ryding, K. C. 2005. *A Reference Grammar of Modern Standard Arabic*. Cambridge: Cambridge University Press.
- Saeed, J. 2003. *Semantics*. 2<sup>nd</sup> ed. Oxford: Blackwell.
- Saiegh-Haddad, E. 2003. 'Linguistic distance and initial reading acquisition: The case of Arabic diglossia'. *Applied Psycholinguistics* 24(03): 431-451.
- Scalise, S. and Vogel, I. 2010. 'Why compounding?'. In Scalise, S and Vogel, I (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 1-20.
- Scalise, S. and Bisetto, A. 2009. 'The classification of compounds'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 35-53.
- Scalise, S. and Fábregas, A. 2010. 'The head in compounding'. In Scalise, S and Vogel, I (eds) *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing. 109-126.
- Scalise, S., Fábregas, A., and Forza, F. 2009. 'Exocentricity in compounding'. *Gengo Kenkyu [Journal of the Linguistic Society of Japan]* 135: 49-83.
- Selkirk, E. 1982. 'The syntax of words'. *Linguistic Inquiry Monographs* 7. Massachusetts: MIT Press.
- Siegel, S. 2002. 'The role of perception in demonstrative reference'. *Philosophers' Imprint* 2(1): 1-21.
- Siloni, T. 1997. *Noun Phrases and Nominalizations: The Syntax of Dps*. Dordrecht: Kluwer Academic Publishers.
- Spencer, A. 2002. *Phonology: Theory and Description*. Cambridge: Blackwell.
- Spencer, A. 2003. 'Does English have productive compounding?' In Booij, G., DeCesaris, J.,

- Ralli, A. and Scalise, S. (eds) *Topics in Morphology. Selected Papers from 3rd Mediterranean Morphology Meeting*. Barcelona: Institut Universitari de Lingüística Aplicada. 329-341
- Štekauer, P. and Valera, S. 2007. *Universals, Tendencies and Typology in Word-Formation: A Cross-Linguistic Research*. Manuscript.
- Štekauer, P., Valera, S. and Körtvélyessy, L. 2012. *Word-Formation in the World's Languages: A Typological Survey*. Cambridge: Cambridge University Press.
- Strang, B. M. H. 1969. *Modern English Structure*. London: Arnold.
- Szymanek, B. 2009. 'IE, Slavonic: Polish'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 464-477.
- Táiwò, O. 2009. 'Headedness and the structure of Yorùbá compound words'. *Taiwan Journal of Linguistics* 7(1): 27-52.
- Talmy, L. 1988. 'Force dynamics in language and cognition'. *Cognitive Science* 12(1): 49-100.
- Turner, M. and Fauconnier, G. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Van Breugel, S. 2014. *A Grammar of Atong*. Leiden: Brill.
- Watson, J. 2002. *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.
- Waugh, L. R. 1994. 'Degrees of iconicity in the lexicon'. *Journal of pragmatics* 22(1): 55-70.
- Wetzer, H. 1996. *The Typology of Adjectival Predication*. New York: Mouton de Gruyter.
- Williams, E. 1981. 'On the notion 'lexically related' and head of a word'. *Linguistic Inquiry* 12: 245-274.
- Zamponi, R. 2009. 'Arawakan: Maipure-Yavitero'. In Lieber, R. and Štekauer, P. (eds) *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 584-593.
- Zwicky, A. 1985. 'Heads'. *Journal of Linguistics* 21: 1-30.

## Appendix: Information for participants and reading texts in MSA and JA

### A) Information for participants

السادة/السيدات الأفاضل،

أود أن أَدعوكم للمشاركة في دراسة تتمحور حول علم الصرف في اللغة العربية الفصحى واللهجة الأردنية. ستكون المشاركة في هذه الدراسة اختيارية، فبإمكانكم التراجع أو التوقف عن التسجيل في أي وقت دون أي عواقب ودون الحاجة للتفسير. أنتم غير مطالبين بذكر أسمائكم أو أيه معلومات شخصية أخرى خاصة بكم. إن الشخص الوحيد الذي لديه صلاحية الاستماع للتسجيلات هو الباحث فقط. سأقوم بتسجيل أصواتكم لمدة تتراوح بين 3-5 دقائق. قد لا تستطيعون الإستفادة بصورة مباشرة من المشاركة في هذه الدراسة، ولكن مشاركتكم قد تساعد على تحسين فهمنا لعلم الصرف في اللغة العربية الفصحى واللهجة الأردنية. مشاركتكم في هذه الدراسة مهمة للغاية فلکم منا كل التقدير.

إذا كان لديكم أية أسئلة حول هذه الدراسة أو إذا كنتم ترغبون في الحصول على ملخص للنتائج، فلا تترددوا في التواصل معي عبر البريد الإلكتروني الخاص بي: [a.r.m.s.altakhaineh@newcastle.ac.uk](mailto:a.r.m.s.altakhaineh@newcastle.ac.uk) أو الخاص بالمشرفين على رسالتي: الأستاذة الدكتورة ماغي تالرمان Maggie Tallerman : [maggie.tallerman@ncl.ac.uk](mailto:maggie.tallerman@ncl.ac.uk) أو الدكتور وليام فان

در ورف William van der Wurff : [w.a.m.van-der-wurff@ncl.ac.uk](mailto:w.a.m.van-der-wurff@ncl.ac.uk)

اقرأ الفقرات التالية بصوت واضح واذكر جنسيتك في بداية التسجيل.

B) The reading text in Modern Standard Arabic (MSA)

في مدرستنا يشربُ مُعَلِّمُ الرِّياضياتِ البعيدُ النَّظرِ القهوةَ كُلَّ صباحٍ في مقصفِ المدرسة. ويعملُ هو ليلَ نهارٍ على تطويرِ خِطَّةِ المُعَلِّمِ وكتابِ الطَّالبِ التَّعليميِّ بناءً على قرارِ وزيرِ التَّربيةِ والتَّعليمِ ورئيسِ الوزراءِ. تبحثُ المدرسةُ الآنَ عن سِجَلٍ مصحَّحٍ كان قد فُقدَ الأُسبوعَ الماضي. إضافةً إلى، قابلتِ المدرسةُ مُعَلِّمَ صَفٍّ ليعملَ في المدرسة. تُدرِّسُ ذلك، المدرسةُ جَدِيدَ الكُتُبِ في جميعِ المراحلِ التَّعليميَّةِ.

في هذا الوقتِ من العامِ، يدرِّسُ الطُّلابُ صباحَ مساءٍ حتَّى يستطيعوا النَّجاحَ في امتحانِ التَّانويةِ العامَّةِ. قبلَ بدايةِ الامتحانِ، يتأكَّدُ مدقُّ الكُرَّاسَةِ من عدمِ إحضارِ الطُّلابِ لأَيَّةِ موادٍّ إضافيَّةِ. ويُنصَحُ الطُّلابُ بشربِ كأسِ عصيرٍ وتناولِ فطيرةٍ تَفَّاحٍ قبلَ الإمتحانِ. وصفتِ الطَّالباتُ امتحانَ الكيمياءِ لمُعَلِّمِ الفيزياءِ والذي يعملُ كمنسِّقِ علومٍ في المدرسةِ بأنَّه حلَّوٌّ مرٌّ بسببِ سهولةِ الأسئلةِ من جانبٍ وقلةِ الوقتِ من جانبٍ



آخر. بعد انتهاء الامتحان، وجد مديرُ القاعةِ كتابَ فتاةٍ وقُبَّعةَ طالبٍ وقلمَ مُعَلِّمٍ على الأرضِ فغضِبَ وطلبَ من عاملِ النظافةِ إزالتهم وإعادتهم إلى سيارَةِ المديرِ.

C) The reading text in Jordanian Arabic (JA)

في مدرستنا بشرب معلّم الرياضيات اللّي نظره بعيد القهوة كلّ يوم الصّبح في مقصف المدرسة. وبعمل هو ليل نهار عشان يطوّر خطة المعلّم وكتب الطلاب حسب قرار وزير التربية والتّعليم ورئيس الوزراء. بتدور المدرسة هسه عن سجلّ مصحّح ضاع الأسبوع الماضي. وكمان قابلت المدرسة معلّم صفّ عشان يشتغل في المدرسة. المدرسة تبعتنا بتدرّس الكتب الجديدة في كل الصّفوف.

هسه، بدرس الطلاب ليل نهار عشان ينجحوا في التّوجيهي. قبل ما يبيلش الامتحان، مدقق الكراسية بيتأكد انو ما حد جاب أي كتب إضافية. لازم الطلاب يشربوا كاسة عصير ويوكلوا كيكة تفّاح قبل الإمتحان. حكّت البنات لمعلّم الفيزياء اللّي يشتغل منسق علوم في المدرسة انو امتحان الكيمياء كان حلو مرّ عشان

الأسئلة كانت سهلة بس الوقت كان قليل. بعد ما خلص الامتحان، لقي مدير القاعة كتاب بنت وطاقيّة ولد وقلم معلّم مرميات على الأرض فعصّب وطلب من عامل النظافة يقيمهم ويرجعهم لسيّارة المدير.

## **D) Information for participants and reading text translated into English**

### **1) Information for participants**

Dear participants,

You are invited to participate in a study about morphological processes in MSA and JA. Participation is entirely voluntary; you may withdraw your consent or discontinue participation at any time without any consequences or any explanation. This study will not record your names or other identifying information; participation is therefore anonymous. Only I will listen to the recordings. Participation will take up approximately 3 to 5 minutes of your time. You may not benefit directly from taking part in this study, but your participation may help to improve our understanding of how some morphological processes operate in MSA and JA. Your participation is highly appreciated and is required for completion of this study.

If you have any questions about this study or you would like to have a summary of the results, please feel free to contact me at [a.r.m.s.altakhaineh@newcastle.ac.uk](mailto:a.r.m.s.altakhaineh@newcastle.ac.uk) or my supervisors, Professor Maggie Tallerman, at [maggie.tallerman@ncl.ac.uk](mailto:maggie.tallerman@ncl.ac.uk) or Dr William van der Wurff, at [w.a.m.van-der-wurff@ncl.ac.uk](mailto:w.a.m.van-der-wurff@ncl.ac.uk).

Please read the following paragraphs in a clear voice and mention your nationality at the beginning of the recording.

### **2) The reading text**

At our school, the far-sighted maths teacher drinks coffee every morning in the school's canteen. He is working day and night to develop the teacher's plan and the student's educational book in accordance with the regulations issued by the Minister of Education and the Prime

Minister. At present, the school is looking for a marker's record which was lost last week. Also, the school has already interviewed a primary school teacher to work at the school. The school only teaches the new (of the) books at all stages.

At this time of year, students study twenty-four seven (day and night) in order to pass the secondary school examination. Before the exam starts, the notes inspector makes sure that the students have not brought any extra material with them. The students are advised to drink a glass of juice and to have an apple pie before the exam. Female students described the chemistry exam to the physics teacher, who works as a science coordinator of the school, as bitter-sweet due to the ease of the questions on the one hand, and the lack of time on the other. Following the exam, the head invigilator found a girl's book, a boy's cap and a teacher's pen on the floor so he was angry. Consequently, he asked the cleaner to remove them and return them back to the head invigilator's car.