

What is a compound?

The main criteria for compoundhood

Abdel Rahman Mitib Altakhaineh

Al Ain University of Science and Technology

Al Ain, UAE

Abstract

This study aims to identify the main cross-linguistic criteria for compoundhood discussed in the relevant literature, with a special focus on English, ranking them from the most reliable to the least. 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 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. free pluralisation of the non-head, compositionality, stress, possibilities for modification and coordination, ellipsis, orthography and the replacement of the second element by a pro-form. The study also proposes a definition for compounds that may be the most widely applicable. Finally, the study concludes with ranking the main criteria for compoundhood discussed in the study.

Key words: morphology; word-formation; compoundhood; phrasehood; derivation.

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 & Štekauer, 2009; Fàbregas & Scalise, 2012; Bauer et al, 2013; among others), some of which deserve serious consideration, while others are less plausible. Hence, this study 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 the main universal criteria that can identify compounds and propose a hierarchical structure of these criteria.

In addition, it suggests a new definition of compounding that is meant to be applicable cross-linguistically. The significance of this study stems from the fact that compounds are considered a relatively cross-linguistic word-formation process found in many languages, and how determining their definition contributes to our understanding of how languages work. In fact, in their corpus of 55 languages, Štekauer et al. (2008, cited in Scalise & Vogel, 2010: 1) note that 50 languages have compounds. Languages which they cite as lacking compounds include East Dangla, Karao, West Greenlandic, Diola Fogy 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 & Vogel, 2010: 1). Thus, proposing a cross-linguistic definition for compounds could be regarded as an area still worthy of further investigation.

The study proceeds as follows: section 2 discusses the main general criteria that have been suggested in the literature to distinguish compounds from phrases. Section 3 discusses the boundary between compounding and derivation. Finally, section 4 summarises the main points and provides a definition of compounding that may be used cross-linguistically.

2. Background to the study

2.1. What is a prototypical 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. This lack of acknowledgment of phrases being elements of compounds has called the need for a more comprehensive definition that could be applicable cross-linguistically. Therefore, other researchers undertook the task of finding such a definition. In this regard, somewhat more precise definitions of compounding have been suggested by Bauer (2001: 695) and Plag (2003: 135). Bau-

er (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.” Of the definitions discussed above, Plag provides the most concise, yet detailed, definition of a compound. Thus, his definition is my departure point to provide my definition that could be applicable cross-linguistically.

2. 2. *Compounds and phrases*

Several linguists (e.g. Katamba, 1993; Bauer, 2003; Booij, 2007; Lieber & Štekauer, 2009, 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) | <i>atom bomb</i> | <i>atomic bomb</i> |
| (2) | <i>verb paradigm</i> | <i>verbal paradigm</i> |
| (3) | <i>language development</i> | <i>linguistic development</i> |

These two constructions 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 |
|-----|-----------------------|--------------------|
| (4) | <i>city parks</i> | <i>urban parks</i> |
| (5) | <i>ocean/sea life</i> | <i>marine life</i> |

The adjectives in (1-3) are derived from the nouns used in the competing construction, e.g. *verbal* from *verb* and *linguistic* from *language*. This is arguably also the case in (4) and (5), 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”.

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
(6) <i>dog house</i>	<i>dog's house</i>
(7) <i>lawyer fees</i>	<i>lawyer's fees</i>
(8) <i>Sunday lunch</i>	<i>Sunday's lunch</i>

Due to this overlap between the two constructs, several linguists (e.g. Bauer, 2003; Katamba & Stonham, 2006; Lieber & Štekauer, 2009; Fàbregas & Scalise, 2012; Bauer et al, 2013, among others) propose criteria to distinguish between compounds and phrases cross-linguistically. However, the boundaries between compounds and phrases are not completely clear. Therefore, I compile these criteria to form a comprehensive list of eleven tests for compoundhood. In the next section, these criteria are applied to N + N combinations with special focus on English to decide whether such combinations are compounds or phrases. In doing so, this study seeks to answers to the following research questions:

1. What are the most reliable criteria for compoundhood?
2. What is the most widely applicable definition of a compound cross-linguistically?

3. The main distinguishing criteria between compounds and phrases

3.1. 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 & 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 *free trade* (though this has the variant *free-trade*). 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 7th Edition*) and *girl friend* (*Webster's Third New International Dictionary*). It may also be noted that some morphologists (e.g. Bauer, 1983; Booij, 2007; Lieber 2009) have different ways of writing the name of one of the topics that they study, with both *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.

3.2. 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, the 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 & 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 (Lieber & Štekauer, 2009; Bauer et al, 2013: 445). Variation even within and across people who speak the same dialect can also be 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-cap* 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*).

There are languages in which stress plays no role in distinguishing compounds from phrases, e.g. Arabic and Hebrew (see Altakhaineh, 2017; Siloni, 1997: 21; Borer, 2009: 493–494). In particular, Altakhaineh (2017) argues that stress plays no role in distinguishing between various N+N combinations, i.e. compounds and phrases, e.g. *'muṣallim lfiizyaa?* ‘the physics teacher’ vs. *'bayt lwalad* ‘the boy’s house’, respectively. Analysis shows that the default position of stress in N+N combinations in Modern Standard Arabic (MSA) and Jordanian Arabic (JA), is on the first element (Altakhaineh, 2017). Siloni (1997: 21), writing about Hebrew, argues that in the construction known as the ‘Nominal State Construct’ in which both compounds and phrases can be found, the stress always falls on the second

element.

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

3.3. Modification

Lieber & 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 & 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 & Scalise (2012: 120–121), argue that internal modification is found in English for both compounds and phrases, as in the following examples:

(9) *He sells [red balloons].*

(10) *[red balloons [sic]] seller*

This suggests that internal modification is an unreliable 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, (11) shows that the compound *limpia ventanas* ‘window cleaner’ does not allow the modification of the element *ventanas* ‘windows’.

(11) **un limpia ventanas grandes*
a polish windows big

The intended meaning of example (11) 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, 2009b: 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):

- (12) *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 & Valera, 2007) and Fongbe (Lefebvre & 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 is recursive in Romance languages. As in Italian, the compound in (13) can be made longer by adding a third element:

- (13)a) *bar pizzería*
bar pizzeria
 b) *bar pizzería discoteca*
bar pizzeria disco

On the other hand, in Romance languages, subordinative and attributive compounds are not recursive, with some rare exceptions (Štekauer et al, 2012; Bisetto, 2010). 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 & Scalise, 2012: 116), as in (14):

- (14) **uomo lupo rana*
man wolf frog

Example (14) is intended to mean ‘a werewolf that has some properties of frogs’, but such a meaning cannot be conveyed through an attributive compound. Examples of rare recursive compounds in Romance languages include the following, observed in Spanish by Štekauer et al. (2012: 98) and in Italian by Bisetto (2010: 25), respectively:

- (15) a) *limpia para brisas (limpiaparabrisas)*
cleans stops breeze
 ‘windscreen wipers’
 b) *porta stuzzica denti (porta-stuzzicadenti)*
carry pick teeth
 ‘toothpick holder’

While it has been shown that recursive compounding can be found in Romance languages, such occurrences are unusual and rare. Conversely, the subordinative compound in Germanic languages is well known for being highly recursive

(Fàbregas & Scalise, 2012: 116), as in (16):

- (16) *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 used as a convincing 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.

3.4. *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 & 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*).

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. This principle is normally used to differentiate between endocentric and exocentric compounds. Allen's ‘IS A’ condition is given in (17).

- (17) In a compound [[]A []B]C C ‘IS A’ B

This can be seen in the endocentric compounds in examples (18) and (19):

(18) *house boat* IS A *boat*

(19) *hand bag* IS A *bag*

This principle can be used to show that *egghead* and *pickpocket* are exocentric compounds, as in (20) and (21):

(20) *egghead* IS NOT A *head*

(21) *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 & 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 *white lie*, and *old hand* 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. However, being reliable in one language and unreliable in another, surely, does not make a certain criterion valid cross-linguistically. Hence, more criteria need to be identified to make a valid cross-linguistic distinction between compounds and phrases.

3.5. Displacement

Fàbregas & 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 (22–23):

- (22) **Truck is what he likes a* [_____ *driver*].
 (23) *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 I now turn to.

3.6. Insertion

Insertion (also known as adjacency) is discussed by Lieber & 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 & 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:

- (24) [_{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. In Arabic, Altakhaineh (2016a: 135) notes that adjacency is found reliable in distinguishing between phrases and compounds, i.e. an intervening element such as the demonstrative *haada* 'this' and the quantifiers *baʿd* 'some' cannot be inserted between the head and the non-head of compounds, whilst such insertion is allowed in phrases (see also Zibin & Altakhaineh, forthcoming).

It is possible to group the two separate criteria of insertion and displacement 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 by considering the ban against insertion it is

possible to posit 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 that this criterion has not been fully explored by some researchers (e.g. Bauer, 2009a (Danish); Don, 2009 (Dutch); Fradin, 2009 (French); Neef, 2009 (German) among others) despite the fact that adjacency seems to be reliable in these languages.

3.7. 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 referring expression in bold *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 (25), *these* modifies the second element, *accounts*:

(25) *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?* A 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 *Amadinejad supporter* (Bauer et al, *ibid*). Clearly, the first element of *Beatles fans* specifically refers to the band whose members are Lennon, McCartney, Harrison and Starr, while the first element of *Amadinejad 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, use of referentiality is a reliable test for compoundhood in Arabic and Hebrew. Examination of Arabic compounds shows that referentiality of the non-head can be used to distinguish between phrases and compounds, since the non-head of the former is always referential, whereas that of the latter is usually non-referential (Altakhaineh, 2016a: 105-106; Altakhaineh, 2016b: 280). In Hebrew, in the three nominal constructs, i.e. R-constructs, M-constructs and compounds, the non-head of R-constructs (possessive constructs) is always referential, whereas that of M-constructs and compounds is non-referential (Borer, 2009). Therefore, I would argue that this criterion needs to be regarded as one of the most reliable criteria for compoundhood.

3.8. 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 & Scalise, 2012). On the other hand, Fàbregas & Scalise (2012: 120) argue that coordination is possible in English for both compounds and phrases, for instance:

(26) *He drinks tea and coffee.* (phrase)

(27) *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 elliptical. 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:

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

In (28), 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 (28). 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.

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. 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 3.4). 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 & Scalise, 2012: 122). Thus, *rock ‘n’ roll* should be treated as a phrasal compound.

3.9. 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. Note that this criterion has limited scope, since some languages do not have pro-forms that can be used to replace the second element of both phrases and compounds, e.g. Hebrew (see Borer, 2009).

3.10. Ellipsis

Fàbregas & Scalise (2012: 120) argue that one of the elements of a phrase can undergo ellipsis as in (29), but not the internal elements of a compound:

- (29) *He drives a truck and he does it every day.*
(Fàbregas & 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 (30), Fàbregas & Scalise (2012) claim that ellipsis is not allowed:

- (30) **He is a truck driver and he does it every day.*
(Fàbregas & Scalise, 2012: 120)

Here, it is worth pointing out Fàbregas & 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 (29), I argue that verb replacement takes place, rather than ellipsis through replacing *drives a truck* by *does it*. The same applies to the example (30), 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 (30). The examples needed to illustrate this criterion would be of the following type:

- (31) *When he buys a car, he always buys the fastest ____.*

Example (31) shows that an instance of ellipsis in which the head of the phrase, i.e. *car* is omitted.

- (32) **There was only one cup and that was a tea ____.*

Example (32) 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 (33):

- (33) *tea and coffee cups*

The compound in (33) is likely to be interpreted as *tea (cups) and coffee cups*. It could be said that the first occurrence of *cups* is elliptical (Fàbregas & Scalise, 2012: 120).

Similarly, the compound *truck and bus drivers* can undergo ellipsis in the same way as in example (33). 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 3.8 for the two possible interpretations of these constructs). Additionally, cases of ellipsis in English NPs are rather restricted, since normally the pro-form *one* has to be used, as in:

(34) **When he buys a car, he always buys a fast* ____.

Example (34) 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 the criterion, i.e. replacement of the second element by a pro-form (discussed in 3.9), 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. The last criterion discussed in the relevant literature to make a distinction between compounds and phrases is inflection and linking elements, which is reviewed and evaluated in the next section.

3.11. *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 & 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 & Štekauer, 2009: 5). In Hebrew and Arabic, free pluralisation of the non-head is a reliable criterion to differentiate between phrases and compounds (Altakhaineh, 2016a: 135; Altakhaineh, 2016c: 8-9; Borer, 2009: 505-6). Specifically, Altakhaineh (ibid) states that “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 does not have any semantic effect”.

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

- (35) *poet-isa-s* *pintor-a-s*
 poet-F-PL *painter-F-PL*
 'women who are poets and painters'

In example (35), 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 (1982: 52) 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 & 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 phrases in that in the latter, the possessor has a determiner function, expanding nominals into noun phrases (Huddleston & 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 is

the same as the car of a smoker. Here, one may notice that a car of a smoker does not have a corresponding 's possessive construction (Huddleston & 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 (36-39):

- (36) *Liebe-s-brief* ‘love letter’ (*Liebe* ‘love’ + *s* ‘LE’ + *Brief* ‘letter’)
 (37) *Arbeit-s-anzug* ‘work suit’ (*Arbeit* ‘work’ + *s* ‘LE’ + *Anzug* ‘suit’)
 (38) *Liebe-s-lied* ‘love song’ (*Liebe* ‘love’ + *s* ‘LE’ + *Lied* ‘song’)
 (39) *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, 2009b: 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:

- (40) *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 & 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 marker 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.

From another perspective, Štekauer & 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 & 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:

- (41) *Konzertreise* 'concert tour' (*Konzert* 'concert' + *Reise* 'tour')

Examples from Dutch are (Booij, 1992: 37):

- (42) *grootvader* 'grandfather' (*groot* 'grand' + *vader* 'father')
- (43) *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.

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)
- (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*.¹

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) suggests that 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).

¹ 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 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 & 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*

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 & 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 *electofunk* 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: 223–225) mentions the examples below of *-ish* separated from the adjective it qualifies:

(50) *They have a pleasantly happy_i ending (well, t_i ish).*

(51) *Is everyone excited_i? I am- t_i ish.*

(52) *Can you swim well_i?: t_i 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”.² As a result, I would argue that neoclassical formations are to be regarded as compounds.

Drawing on the above discussion of the criteria used to distinguish compounds from other constructs, i.e. phrases and derivation; the following section concludes this study and suggests a cross-linguistic definition for compounds.

5. Conclusion

In this study, 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, see Bauer 2009a). Furthermore, some criteria are partially useful to distinguish between compounds and phrases, e.g. free pluralisation of the non-head, compositionality, modification, etc. It has also become evident that drawing a boundary between compounding and phrases is not an easy task. For this reason, 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 1 below.

² 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 1. Possible internal elements of compounds in English

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

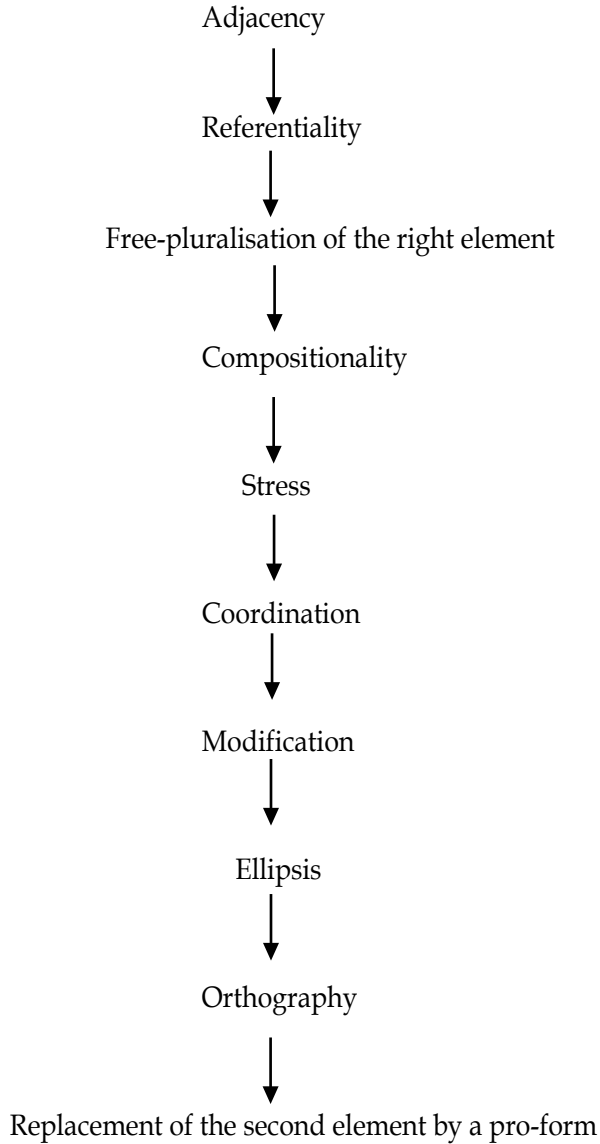
Taking the examples in Table 1 into consideration, the following definition of a compound, at least in English, can be suggested: a compound is a complex word that consists of at least two adjacent elements, in which each of these elements is either a word, combining form or a phrase, so that the whole 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, it can be considered a significant criterion when we are identifying compounding cross-linguistically. Therefore, I would propose the following general definition that could be used 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.

The typology in (53) ranks the main criteria for compoundhood discussed in this study based on their reliability:

(53) Compoundhood Criteria Ranking



Acknowledgements

I wish to thank the anonymous reviewers for their useful comments on an earlier draft of the paper. All remaining errors and inaccuracies are, of course, mine alone.

References

- Allen, Margaret Reece (1978). *Morphological Investigations in English*. PhD dissertation: University of Connecticut, Storrs.
- Altakhaineh, Abdel Rahman M. (2016a). Compounding in Modern Standard Arabic, Jordanian Arabic and English (Unpublished doctoral dissertation). Newcastle University, UK.
- Altakhaineh, Abdel Rahman M. (2016b). Identifying Arabic compounds other than the Synthetic Genitive Construction. *Acta Linguistica Hungarica* 63.3: 277–298.
- Altakhaineh, Abdel Rahman M. (2016c). Headedness in Arabic Compounds within the Synthetic Genitive Construction. *SAGE Open* 6.4: 1–16.
- Altakhaineh, Abdel Rahman M. (2017). Stress assignment in N+N combinations in Arabic. *International Journal of Applied Linguistics and English Literature* 6.6.
- Aronoff, Mark, Kirsten Fudeman (2005). *What is Morphology?* Oxford: Blackwell.
- Bauer, Laurie (1983). *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie (1998a). When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2: 65–86.
- Bauer, Laurie (1998b). Is there a class of neoclassical compounds in English and is it productive? *Linguistics* 36: 403–422.
- Bauer, Laurie (2001). Compounding. Haspelmath, Martin, Ekkehard König, Wulf Oesterreicher, Wolfgang Raible, eds. *Language Typology and Language Universals*. Berlin: Mouton de Gruyter, 695–707.
- Bauer, Laurie (2003). *Introducing Linguistic Morphology* (2nd edn.). Edinburgh: Edinburgh University Press.
- Bauer, Laurie (2005). The borderline between derivation and compounding. Dressler, Wolfgang U., Dieter Kastovsky, Oscar Pfeiffer, Franz Rainer, eds. *Morphology and its Demarcations*. Amsterdam: John Benjamins, 97–108.
- Bauer, Laurie (2009a). IE, Germanic: Danish. In Lieber, Rochelle, & Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 400–416.
- Bauer, Laurie (2009b). Typology of compounds. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 343–356.
- Bauer, Laurie, Antoinette Renouf (2001). A corpus-based study of compounding in English. *Journal of English Linguistics* 29: 101–23.
- Bauer, Laurie, Rochelle Lieber, Ingo Plag (2013). *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Bisetto, Antonietta (2010). Recursiveness and Italian compounds. *SKASE Journal of Theoretical Linguistics* 7.1: 14–35.
- Booij, Geert E. (1992). Compounding in Dutch. *Rivista di Linguistica* 4.1: 37–59.
- Booij, Geert E. (2007). *The Grammar of Words* (2nd edn.). Oxford: Oxford University Press.
- Booij, Geert E. (2009). Construction morphology and compounding. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 201–216.
- Booij, Geert E. (2012). *The Grammar of Words* (3rd edn.). Oxford: Oxford University Press.
- Borer, Hagit (2009). Afro-Asiatic, Semitic: Hebrew. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 386–399.
- Carstairs-McCarthy, Andrew (2002). *An Introduction to English Morphology*. Edinburgh: Edinburgh University Press.
- Chomsky, Noam, Morris Halle (1968). *The Sound Pattern of English*. New York: Harper and

Row.

- De Belder, Marijke (2013). Collective mass affixes: When derivation restricts functional structure. *Lingua* 126: 32–50
- Di Sciullo, Anna Maria (2009). Why are compounds a part of human language? A view from asymmetry theory. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 145–177.
- Dirven, René, Marjolijn Verspoor (1998). *Cognitive Exploration of Language and Linguistics*. Amsterdam: John Benjamins Publishing.
- Don, Jan (2009). IE, Germanic: Dutch. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 370–385.
- Donalies, Elke (2004). *Grammatik des Deutschen im Europäischen Vergleich*. Kombinatorische Begriffsbildung: Substantivkomposition. Mannheim: Institut für Deutsche Sprache.
- Dressler, Wolfgang (1986). Forma y función de los interfijos. *Revista Española de Lingüística* 16: 381–395.
- Fabb, Nigel (1998). Compounding. Spencer, Andrew, Arnold Zwicky eds. *The Handbook of Morphology*. Oxford: Blackwell, 66–83.
- Fàbregas, Antonio, Sergio Scalise (2012). *Morphology from Data to Theory*. Edinburgh: Edinburgh University Press.
- Fernando, Chitra (1996). *Idioms and Idiomaticity*. Oxford: Oxford University Press.
- Fernando, Chitra, Roger Flavell (1981). *On Idiom: Critical Views and Perspectives*. Exeter: University of Exeter Press.
- Fischer, Roswitha (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, Wolfgang (1975). *Wortbildung der Deutschen Gegenwartssprache*. Durchgesehene Auflage. Tübingen: Niemeyer.
- Fradin, Bernard (2009). IE, Romance: French. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 417–435
- Giegerich, Heinz J. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8: 1–24.
- Giegerich, Heinz J. (2005). Associative adjectives in English and the lexicon-syntax interface. *Journal of Linguistics* 41.2: 571–591.
- Huddleston, Rodney, Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hurford, James R. (2011). *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press.
- Jackendoff, Ray (2002). *Foundations of Language*. Oxford: Oxford University Press.
- Jackendoff, Ray (2009). Compounding in the parallel architecture and conceptual semantics. In Lieber, Rochelle, Pavol Štekauer eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 105–129.
- Johannessen, Janne Bondi (1996). Partial agreement and coordination. *Linguistic Inquiry* 27.4: 661–676.
- Jones, Daniel (1969). *An Outline of English Phonetics*. Cambridge: Heffner and Sons Ltd.
- Katamba, Francis (1993). *Morphology*. New York: St. Martin's Press.
- Katamba, Francis, John Stonham (2006). *Morphology* (2nd edn.). London: Macmillan.
- Kavka, Stanislav (2009). Compounds and idiomatology. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 19–33.

- Kiefer, Ferenc (2009). Uralic, Finno-Ugric: Hungarian. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 527–541.
- Kunter, Gero (2011). *Compound Stress in English. The Phonetics and Phonology of Prosodic Prominence*. Berlin: de Gruyter.
- Lefebvre, Claire, Anne-Marie Brousseau (2002). *A Grammar of Fongbe*. Berlin: Mouton de Gruyter.
- Lieber, Rochelle (1992). *Deconstructing Morphology. Word Formation in Syntactic Theory*. Chicago and London: The University of Chicago Press.
- Lieber, Rochelle (2005). English word-formation processes. Štekauer, Pavol, Rochelle Lieber, eds. *Handbook of Word-Formation*. Dordrecht: Springer, 375–428.
- Lieber, Rochelle (2009). IE, Germanic: English. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 357–369.
- Lieber, Rochelle (2010). *Introducing Morphology*. Cambridge: Cambridge University Press.
- Lieber, Rochelle, Pavol Štekauer (2009). Introduction: Status and definition of compounding. Lieber, Rochelle Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 3–18.
- Marchand, Hans (1960). *The Categories and Types of Present-Day English Word-formation: A Synchronic-Diachronic Approach*. Wiesbaden: Otto Harrassowitz.
- Neef, Martin (2009). IE, Germanic: German. Lieber, Rochelle Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 386–399.
- Norde, Muriel (2009). *Degrammaticalization*. Oxford: Oxford University Press.
- Olsen, Susan (2000). Compounding and stress in English: A closer look at the boundary between morphology and syntax. *Linguistische Berichte* 181: 55–69.
- Papafragou, Anna (1996). On metonymy. *Lingua* 99: 169–195.
- Plag, Ingo (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
- Plag, Ingo (2006). The variability of compound stress in English: Structural, semantic and analogical factors. *English Language and Linguistics* 10: 143–172.
- Preminger, Alex, T. V. F. Brogan (1993). *The New Princeton Encyclopedia of Poetry and Poetics*. Princeton: Princeton University Press.
- Rainer, Franz, & Soledad Varela (1992). Compounding in Spanish. *Rivista di Linguistica* 4: 117–142.
- Ralli, Angela (2009). IE, Hellenic: Modern Greek. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 453–463.
- Ralli, Angela (2010). Compounding versus derivation. Scalise, Sergio, Irene Vogel, eds. *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins Publishing, 57–76.
- Ralli, Angela (2013). *Compounding in Modern Greek*. London: Springer Science and Business Media.
- Rosenbach, Anette (2006). Descriptive genitives in English: A case study on constructional gradience. *English language and Linguistics* 10.1: 77–118.
- Rosenbach, Anette (2007). Emerging variation: Determiner genitives and noun modifiers in English. *English Language and Linguistics* 11: 143–189.
- Saeed, John I. (2003). *Semantics* (2nd edn.). Oxford: Blackwell.
- Scalise, Sergio, Irene Vogel (2010). Why compounding?. Scalise, Sergio, Irene Vogel, eds. *Cross Disciplinary Issues in Compounding*. Amsterdam: John Benjamins, 1–20.
- Selkirk, Elizabeth O. (1982). *The Syntax of Words*. Massachusetts: MIT Press.
- Siloni, Tal (1997). *Noun Phrases and Nominalizations: The Syntax of DPs*. Dordrecht: Kluwer Academic Publishers.

- Spencer, Andrew (2003). Does English have productive compounding? Booij, Geert E, Janet Ann DeCesaris, Angela Ralli, Sergio Scalise, eds. *Topics in Morphology. Selected Papers from 3rd Mediterranean Morphology Meeting*. Barcelona: Institut Universitari de Lingüística Aplicada, 329–341.
- Štekauer, Pavol, Salvador Valera (2007). *Universals, Tendencies and Typology in Word-Formation: A Cross-Linguistic Research*. Manuscript.
- Štekauer, Pavol, Salvador Valera, Lívia Kórtvélyessy (2012). *Word-Formation in the World's Languages: A Typological Survey*. Cambridge University Press.
- Szymanek, Bogdan (2009). IE, Slavonic: Polish. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 464–477.
- Zamponi, Raoul (2009). Arawakan: Maipure-Yavitero. Lieber, Rochelle, Pavol Štekauer, eds. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, 584–593.
- Zibin, Aseel, & Abdel Rahman M. Altakhaineh (forthcoming). An analysis of Arabic metaphorical and/or metonymical compounds: A cognitive linguistic approach. *Metaphor and the Social World*.

Dictionaries:

- Webster's Third New International Dictionary*. Merriam Webster USA.
- Concise Oxford Dictionary 7th Edition*. Oxford: Oxford University Press.
- Macmillan Dictionary*. Last accessed 11th March 2016, from: <http://www.macmillandictionary.com/dictionary/british/girlfriend>.

Author's address:

Abdel Rahman Mitib Altakhaineh
Assistant Professor of English Language and Linguistics
Room 21, Building T
Al Ain University of Science and Technology (AAU)
P.O. Box: 64141, Al Ain, UAE
Email: abdelrahman.takhaine@aau.ac.ae

Received: April 12, 2017

Accepted for publication: July 10, 2017