Laurie Bauer

Compounds and multi-word expressions in English

1 Introduction

Compounds are traditionally defined as being, in the words of Lieber (2010: 43), “words that are composed of two (or more) bases, roots, or stems”. Multi-word expressions (also known as multi-word units or items, henceforth MWEs) can be defined as “lexical items which consist of more than one ‘word’ and have some kind of unitary semantic or pragmatic function” (Moon 2015: 120). Since all words (in the sense of ‘lexeme’, which is what I assume Lieber to mean in the cited passage) are lexical items, the first thing to note is that these two definitions overlap (pace ibid.: 121). Things called compounds, if they have ‘some kind of unitary semantic or pragmatic function’, which they can be argued always to have, are MWEs, although not all MWEs are compounds.

In this chapter, it will be argued that this fuzzy borderline between compounds and MWEs is real, that there is no generally accepted way of dividing compounds from MWEs, and that much of this derives from their common function as lexical items. Furthermore, there is no generally accepted way of dividing compounds from syntactic phrases, so that it follows that there is no generally accepted way of dividing MWEs from syntactic phrases. This situation arises partly from the data, and partly from the varying views of different scholars, who have tried to draw dividing lines in different places, thus illustrating the lack of commonality of opinion. Because this chapter focusses on the situation in English, the arguments affect English specifically, and may not all transfer to other languages. No attempt is made here to generalise to other languages; that is left for another chapter. The effect is, however, a claim that there is no agreed definition of a compound in English (and possibly not of an MWE, as noted ibid.).

2 The notion of ‘word’

Word must be one of the least well-defined technical terms in linguistics. There are innumerable discussions of why this is the case, and there is little point in adding to them here (cf., e.g., Bauer 2000; Dixon/Aikhenvald 2002; Hippisley 2015; Wray 2015). In some languages, some criterion or set of criteria can be used...
to define ‘word’ sufficiently well to allow a definition of a compound as a word to be meaningful. In English, it is less clear that this is true. Consider just three potential criteria, which are often used in other languages.

The first of these is stress. In other Germanic languages, stress is often used as a criterion for compoundhood. Any discussion of English in these terms, however, falls foul of examples like those in (1).

(1)  
| Forestress       | End-stress       |
|------------------|------------------|
| apple cake       | apple pie        |
| glass cupboard   | glass cupboard   |
| (‘cupboard in which glassware is kept’) | (‘cupboard made of glass’) |
| toy factory      | toy factory      |
| (‘factory that makes toys’) | (‘factory which is itself a toy’) |
| York Street      | York Avenue      |

In addition, Bauer (1983b) finds that speakers are inconsistent in assigning stress to (at least some) such expressions, and also notes (Bauer to appear) variable usage of stress in the speech of newsreaders. Kunter (2011) finds a reasonable minority of such forms show variable stress. While most authorities now see stress as not being a reliable guide to the status of such items as compounds (Giegerich 2004), this has not always been the case, so that some such expressions have seemed to be changing category from compound to non-compound in an apparently random fashion. Chomsky/Halle (1968), for example, use stress as definitional for compounds.

Spelling is, to some extent, linked with stress: railway is written as one word and has forestress, iron bar is written as two and has end-stress. Other factors are also involved, however: schoolgirl tends to be written as one word, while university student has to be written as two, despite parallel stress and semantic readings. Some of the examples in (1) equally show a distinction between stress and orthography. It is also well-known that English orthography is inconsistent when it comes to writing some compounds: rainforest, rain-forest and rain forest can all be found in dictionaries. Matters as difficult to quantify as house-style and fashion can influence such spellings. Spelling cannot be criterial for word status in English. Nonetheless, some scholars use it in this way, either by default (cf. Hall 1964: 134) or to make dealing with the computational analysis of written text possible (McEnery/Xiao/Tono 2006: 147).

As a third criterion, consider the notion that words allow for global inflection, but not for inflection which is internal and applies to some element within the word. If we consider a compound verb like badge-flash (see (2)), we can see how this works.
In (2) we see that *badge-flashed our way to the scene. Global inflection is possible, but not internal inflection. There are two problematic constructions in English in relation to this criterion. The first is illustrated by jobs growth, where the first element of the compound has an apparent plural. Pinker (1999) sees this as sufficient evidence to say that such constructions are phrasal, not words, others include these as compounds (and, hence, as words). The other awkward construction in this regard is the classifying genitive as in cat’s eye (‘reflecting road marker’ or ‘semi-precious stone’). Even if we ignore the question as to whether the s-genitive in English is inflectional or a clitic (cf. Bauer/Lieber/Plag 2013: 141f. for a brief summary), it is not clear whether such constructions count as single words. They are compound-like in many ways (Rosenbach 2006), though most scholars exclude them from the set of compounds. Corresponding expressions in other Germanic languages are generally thought of as compounds, although ‘uneigentlich’ (‘non-genuine, false’) compounds in Grimm’s terminology.

Other criteria for wordhood are frequently used in attempting to determine whether given constructions are words (compounds) or not. These include the fixed order of constructions, non-interruptibility of elements, lack of modification of internal elements, lack of coordination of internal elements, impossibility of referring back to individual elements by pronouns, including one, and listedness. Not only do such criteria not define a coherent set of items as words (Bauer 1998), they are often broken in derivatives, whose wordhood is not usually queried. These criteria will be referred to below, as required. The point here is that not only do the criteria for wordhood not fit compounds particularly well (cf. also Giegerich 2015; Bauer 2017), they do not allow agreement on what is or is not a compound in English. The border of compounding is vague partly because the border of wordhood is vague.

In what follows, a number of constructions will be considered in varying detail. Some of these constructions will be ones which some scholars see as compounds, others will be MWEs more loosely defined. The borderline between these two groups of construction will be shown to be non-principled, with different theoreticians making different decisions as to what is or is not a compound.

---

1 Karp, Marshall (2006): The rabbit factory. San Francisco: MacAdam, 179.
3 Formal constructions

3.1 N+N

There are several classes of N+N constructions in English, and while some of them are regularly considered to be compounds, many of them are equally regularly considered to be excluded from the category. We can illustrate some of the classes as in (3).

(3a) Doctor Johnson, Miss Havisham, King George
(3b) Elizabeth Taylor, John Lennon
(3c) President Donald Trump, Prime Minister Theresa May
(3d) beef Wellington, chicken Kiev
(3e) the category adjective, letter A, number nine, Model T
(3f) bank-box, bus-driver, car park, windmill
(3g) Oxford college, cutlery box
(3h) iron bar, copper wire, stone wall
(3i) the film ‘Jaws’, the year 1952
(3j) egg head, hatchback
(3k) father-daughter, hand-eye
(3l) Nelson-Marlborough, Daimler-Benz
(3m) murder-suicide, mind/brain
(3n) singer-songwriter, lawyer-poet
(3o) elm tree, tuna fish
(3p) salad-salad

Names are not usually counted as being compounds. Those in (3c) are generally seen as instances of apposition, and frequently have a pause and intonation break between the title and the name (unlike those in (3a) which would otherwise be parallel). Apposition is usually considered a syntactic construction rather than a lexical one, and so the examples in (3a–c) and also the examples in (3i) with common nouns, are excluded from compounds. However, at least those in (3a) and (3b) must be listed, since they denote individuals and have little semantic transparency. Those in (3a) appear to be left-headed (Doctor Johnson is a member of the class of people with title of doctor), while those in (3b) may not be – it is not clear whether it even makes sense to ask whether Elizabeth Taylor is a member of the set of Elizabeths or the set of Taylors, especially since asking such a question changes the category of both Elizabeth and Taylor from proper noun to common noun. The examples in (3e) may also be instances
of apposition, but it is less clear: Model T is something which deserves at least an encyclopedic entry, if not a lexical entry, and acts as a label for a class of objects in much the same way as the noun T-junction does. Model T, though, is left-headed. While headedness is not usually given as one of the criteria for compoundhood in English (though cf. Bauer/Lieber/Plag 2013), most of the items that are seen as clear cases of compounds in English are right-headed. The examples in (3d) are also left-headed, but here it seems even less likely that apposition is involved. These items are names of dishes and synchronically at least have little to do with any semantic content that might be derived from their second elements. They certainly fit the definition of compound given in Section 1 above, and they are listed.

The items in (3f) are the central examples of compounds (though including examples from rather different subsets), and those from (3h) are examples which are often thought of as syntactic, but for different reasons. For Giegerich (2015) the first word in these constructions is an adjective, for others they are syntactic because their orthography, stress and behaviour under coordination shows them to be so: copper and aluminium wire and copper wire and cable are both exceptional. The items in (3g) provide an intermediate step. For some scholars they are compounds, for others (e.g. Payne/Huddleston 2002) they are syntactic, because they fail at least one of the criteria for being words. For example, Oxford and Cambridge colleges is perfectly acceptable, as is four Oxford and three Cambridge colleges and cutlery and wine-glass boxes and assorted silver cutlery box.² Note that Payne and Huddleston have an overarching principle that any trace of syntactic behaviour makes something a syntactic structure rather than a principle that any hint of lexical behaviour makes something non-syntactic.

The items in (3i), as mentioned above, are appositional, and are usually excluded from the set of compounds, but they contrast with the set in (3o) which are usually included. Even so, they do not easily allow interruption, though they do allow coordination, where relevant, as in the movie and book Jaws,³ and they certainly allow submodification of just one element, as in the thrilling film “Jaws” or the thrilling film, the notorious “Jaws” (but note the necessity for the determiner in this last example, which may change the construction).

The items in (3j) are usually considered compounds, but exocentric compounds, often thought of as unheaded (cf. Carstairs-McCarthy 2002). The particular items listed here fit into the Sanskrit category of bahuvrihi compounds, which

² Internet: www.spotlightstores.com/party.party-decorator/room-table/decorating-accessories/amscan-assorted-silver-cutlery-box/p/BP80402188 (last access: 17 Nov 2017).
³ Internet: https://en.wikipedia.org/wiki/Frank_Mundus (last access: 17 Nov 2017).
others see as regular endocentric compounds interpreted through the figure of speech synecdoche (sometimes considered to be a type of metonymy) (Bauer 2016).

The types in (3k–p) are various kinds of coordinative compounds. Adams (2001: 3) excludes all of these from the set of compounds, apparently because they are unheaded. It should be noted that some of these are exocentric (for different reasons): hand-eye (in hand-eye coordination) is exocentric because it is used exclusively as a premodifier (and thus, possibly, an adjective), while Nelson-Marlborough is neither a hyponym of Nelson nor a hyponym of Marlborough. On the other hand, a singer songwriter is both a singer and a songwriter, and an elm tree is both an elm and a tree. We have already seen that examples like elm tree bear some resemblance to instances of apposition, another potential reason for not including them as compounds. Some scholars include items like that in (3p) as compounds, while others might see it as reduplication or even just repetition (a salad-salad is one which contains things typically found in a salad like lettuce and cucumber, as opposed, say, to a pasta salad).

3.2 A+N

Again, we can find many classes of construction involving adjectives and nouns. A+N compounds are usually distinguished from syntactic constructions by their stress (forestress) and, correspondingly, their orthographic unity, by the fact that the adjective cannot be submodified or graded, and by the fact that the adjective can be denied without contradiction. Thus blackbird is a compound by virtue of its stress, its orthography, the fact that we cannot have a blackerbird or a very-blackbird, and because This blackbird is brown is not a contradiction. The syntactic construction black bird differs from the compound blackbird in all of these respects. This distinction arises because black in black bird describes, while black in blackbird categorises.

If we look at intersective adjectives like black, heavy, silly etc. where a black bird represents the intersection of black things and birds, we discover that they are not always intersective. A red book may illustrate an intersective use of red, but a red squirrel does not: red squirrel behaves semantically like blackbird, not like black bird, despite different stress and orthography. Bauer (2004) points out that there is a difference in frequency between the forestressed words and the end-stressed expressions: the forestressed words are more frequent. This would seem to indicate that there are intersective adjectives used descriptively, intersective adjectives used to categorise, and intersective adjectives used with forestress to categorise. Most authorities distinguish the compounds with forestress from
the other two types, but we might equally distinguish the descriptive adjectives from the categorising ones.

If we now turn to relational adjectives like *canine, dental, parental, vernal*, they are not intersective. A *canine tooth* is not the intersection of canine things and teeth, but a kind of tooth related in some way to dogs (for fuller discussion cf. Giegerich 2015). Relational adjectives are rarely descriptive unless they are figurative or used predicatively, as in *his movements were feline, her attitude was vaguely parental.* The precise relationship between the adjective and the noun has to be discovered by considering the individual example, just as the relationship between nouns in N+N compounds has to be discovered by considering the individual example: a *windmill* uses wind power, but a *flour mill* grinds flour. Part of the result of this is that relational adjectives are by default categorising. Nevertheless, there are instances when they, too, can take forestress: consider for instance *dramatic society, mental hospital, primary school.* The reason for the forestress here is not clear. Neither is it clear whether things like *mental hospital* are compounds. Scholars disagree on whether A+N constructions with relational adjectives are compounds or not, but they certainly seem to fulfil a similar purpose. In some cases there are pairs with a modifying noun and a modifying adjective which may be nearly synonymous (*atom bomb, atomic bomb; language description, linguistic description*), while in other instances they contrast in meaning (*a civic centre is not the same as a town centre*). Speakers must know that it is *solar flare* but *sunspot*; there does not seem to be a way to predict such distinctions.

Expressions such as *attorney general, court martial, heir apparent,* where the adjective follows the noun it modifies, are usually of French origin, and follow the French order of noun and adjective. A few such as *postmaster general* are formed in English on a French pattern. The pattern does not seem to be productive, so in principle a full list of these can be given. There seems to be little reason to include such expressions among compounds, particularly since they are left-headed though most compounds are right-headed, but they are certainly MWEs.

### 3.3 Other word-classes +N

Examples of potential compounds formed with other word-classes in the modifying position are given in (4).

(4a)  *spoilsport, dreadnought*
(4b)  *call girl, show-room*
(4c)  *uptown, downdraught*
The examples in (4a) probably imitate a Romance pattern which is no longer productive in modern English. However, a similar type is found with the order of the elements reversed: *prick-tease*, for example. The type in (4b) has verbs in modifying position, but is endocentric (*show room* is a hyponym of *room*). It is often the case that modifying verbs in forestressed constructions take the *-ing* form: *dining room, shooting party, walking stick*. These are then usually considered to have nominal first elements. The type in (4c) shows adverbs/prepositions/particles in modifying position. Things like *through-put* may also belong here formally, though they are probably nominalisations of phrasal verbs. Reverse ordered forms like *put-down* are also found. The type in (4d) shows alternatives in modifying position, the alternatives being, in these instances, verbs or adverbs. The type in (4e) shows apparently unlimited syntactic constructions in initial position. These expressions do not have to be idiomatic or even familiar. If these are compounds, though, and most scholars accept that they are, they allow syntactic structure within word-structure. The types in (4f–g) have already been mentioned, with plurals or genitives in the first element. In both cases there is often an alternative with an unmarked noun in the first position (*lineman* and *linesman* are synonymous; according to the OED *tailor’s tack* and *tailor tack* are synonymous). At the same time, a genitive first element can contrast with an unmarked first element, as illustrated in (5) (data from the OED).

(5)  

| English           | French                           |
|-------------------|----------------------------------|
| *dog-tooth* ‘check pattern’ | *dog’s tooth* ‘architectural feature’ |
| *dog show* ‘event’     | *dog’s show* (Aust) ‘no chance’   |
| *dog collar* ‘clerical garb’ | *dog’s collar* ‘collar of a dog’ |
| *duck-foot* ‘having webbed feet’ | *duck’s foot* ‘plant sp.’ |

---

4 Meynell, Lawrence (1978): Papersnake. London: Macmillan, 10.
5 Francis, Dick (2006): Under orders. London: Michael Joseph, 87.
3.4 Adjectival compounds

Adjectival compounds are common, with examples like *crime-prone*, *grass-green*, *sky-blue*, *word-final*, *work-shy*, and coordinative compounds are also found: *philosophical-historic*, *spicy-mild*. It can be argued (Bell 2014) that there are a number of exocentric compound adjectives in English which are exocentric by virtue of not containing an adjectival head: words like *day-to-day*, *fly-by-wire*, *overhead*, *through and through*, *pass-fail*. Some of these may look more like non-compound MWEs, but recall the definition given in Section 1 above, that compounds are ‘words that are composed of two (or more) bases’, and it can be seen that all of these fit the definition. If this just indicates that the definition is incomplete, then that is part of the message of this contribution. It must be noted, though, that corresponding structures in related languages would not be considered adjectives, and the question of their status arises peculiarly in English.

There is a set of adjectives which appears to arise from the participle-form of phrasal verbs: *down-sized*, *up-graded*, *out-grown*. Whether these are viewed as compounds may well depend on whether phrasal verbs are viewed as compounds (see below). They have a form made up of two bases, but those two bases are not independent at the point of adjective-formation. The same point can be made with relation to the corresponding denominal forms like *black-hearted*, *green-eyed*, which are not strictly formed as compounds in English, since their structure is [[black heart]ed], so that they are derivatives based on phrasal structures.

3.5 Verbal compounds

Verbal compounds are something of a discussion point in English word-formation, following Marchand’s (1969: 100) definitive declaration that “[v]erbal composition does not exist in Present-Day English”. The point is that many of the things that look like compounds, and that we might want to term compounds, are actually formed by back-formation (*to baby-sit*, *to horror strike*) or conversion (*to breath test*, *to cold shoulder*). The argument that these are not compounds follows the pattern of the argument on *hard-hearted* in the last section. Nevertheless, it is clear that there is an increasing number of genuine verbal compounds which are not formed by these means (Bauer/Renouf 2001; Bauer 2017). Recent examples are *air-quote*, *dry-burn*, and coordinative examples like *to blow dry*, *to stir-fry* (these are controversial examples of coordinative compounds, though some authorities included them).

English does have a number of V+V constructions which might be viewed as compounds or as serial verbs (and, if the latter, probably of syntactic not lexical
origin). These are most commonly found with verbs of motion as the first verb (go see, come buy) but go beyond that (I hope see you soon), especially in US English. Some such constructions can be the base of further derivation, which seems to imply listedness, if not other features of words (consider go-getter, jump-starter).

3.6 Compounds in minor word-classes

Whether there are compound prepositions is a matter of definition. Things like into, onto, throughout are written as one word, and are probably instances of frozen syntax. Instances like away from, because of, except for, off of (esp. US English), out of are certainly common collocations in text, but whether they are compounds or not is not clear.

3.7 Binomials

Binomials are pairs of words linked usually by and, occasionally by or. They are normally called binominals only if they are fixed collocations. Thus Monday or Tuesday would not be considered a binominal, but the examples in (6) would be.

(6) Abbot and Costello, bacon and eggs, bread and butter, cat and mouse, chalk and cheese, fish and chips, gin and it, kit and caboodle, kith and kin, life or death, milk and honey, salt and pepper, slap and tickle, sun and sand, whisky and soda; do or die, kiss and tell, make or break, put up or shut up, wine and dine; black and blue, free and easy, neat and tidy, sick and tired, spick and span; as and when, back and forth, far and away 'by a wide margin', far and near 'everywhere', now and again

There is quite a large literature on the order of the elements in binomials (for a good summary cf. Benor/Levy 2006), and the fixedness of the order. Binomials vary in the degree to which each element presupposes the other. In spick and span we cannot have either element without the other; black can easily occur without blue, but black and blue is a fixed expression whose implications go beyond the colours involved; chalk and cheese collocate only when illustrating how different two things can be; Abbott and Costello illustrates a collocation which was originally purely arbitrary, but became more fixed as the team became more established. They also differ in how easily they can be interrupted: bread and manuka honey is perfectly possible, but sick and really tired is no longer an example of the relevant collocation. Again, they differ in how easily the coordi-
nated items can be reversed. Eggs and bacon or bacon and eggs seem to be equally good (and scrambled can be added to eggs in either ordering), jam and bread is possible, if slightly unusual (it is found in a song in The Sound of Music, for instance), chips and fish is mainly used when the chips and the fish are referred to separately rather than as a single dish. Those binomials that have a figurative reading cannot in general be interrupted or reversed: bread and butter ‘main source of income’, salt and pepper ‘colour term’, far and away.

3.8 N+P+N constructions

N+P+N constructions like lady-in-waiting are frequently established MWEs, even though there are many N+P+N constructions which appear to be perfectly freely syntactic, as in piece of cheese. The problem of description is exacerbated in comparison with a language like French, where N+P+N constructions are often the translational equivalent of Germanic compounds. For instance, French chemin-de-fer, lit. way of iron, ‘railway’ is equivalent to Danish jernbane, lit. iron way, ‘railway’ (compare also German, Italian and other European languages), and French jus de fruits ‘juice of fruits’ is equivalent to English fruit juice. The French expressions are sometimes called ‘compounds’ (Spence 1969 calls them ‘prepositional compounds’), while an opposing view sees them as syntactic constructions that may become fixed (Bauer 2001). The English construction is not as widespread as the French one is (because English has more compounds), but there are plenty of examples (cf. (7)).

(7) lady-in-waiting, line-of-sight, man-about-town, man-at-arms, man-of-war, mother-of-pearl, pay-per-view, sense of humour, son-in-law, stock-in-trade, trial by jury

Part of the question here (and, incidentally, also in French) is the status of items with internal determiners, such as those in (8). Are they a different construction by virtue of having an NP (or DP) in second position, or are they a variant of the same construction?

(8) belle of the ball, birds of a feather, two bites of the cherry, a Jack of all trades, the man in the moon, the man of the moment, a pain in the neck, the time of your life, will of the wisp

We might also ask whether toponyms such as Burton-in-Lonsdale, Gatehouse-of-Fleet, Moreton-in-Marsh, Newcastle-under-Lyme, Newcastle-upon-Tyne, Walton-
on-Thames, Weston-super-Mare are part of the same construction type. Again, there is a variant with determiners: Stow-on-the-Wold, Stanford-in-the-Vale, Widecombe-in-the-Moor.

To the extent that DPs can form part of the construction, these forms look more syntactic. But even then, we do not appear to find random DPs: adjectival modification within that DP does not appear to occur in established constructions of this form, though forms like cat-in-the-new-moon or Marton-in-the-Blue-Mountains might appear to be possible. On the other hand, with non-established examples, such as by the light of the new moon, there is no problem with adjectival modification. A fortiori, post-nominal modification does not occur in established examples.

Klinge (2005: 366) claims that only the preposition of is particularly productive in such phrases. He uses this as an argument for the lexical nature of these constructions. This is hard to establish, since other prepositions are clearly in use in the more syntactic phrases, and it seems unlikely that the rules of production for the more syntactic and more lexical types are completely independent. A more likely explanation is that only relatively non-specific forms are frequent enough to become established in usage, and that of is the most frequent preposition.

Overall, the descriptive problem here seems to be similar to the descriptive problem with genitive first elements: the formal description of the construction includes expressions which are clearly listed (sometimes idiomatic) and others which appear to be produced productively, possibly by syntactic rules. Perhaps equivalently, this means that some such expressions are more word-like than others.

3.9 Phrasal verbs

Phrasal verbs are usually taken to be syntactic units in English, though many of them are figurative or idiomatic. Look up is literal when it means ‘raise your eyes towards the sky’, but idiomatic when it means ‘refer to’ (as in look up a word in the dictionary) or even ‘improve’ as in business is looking up. Put up is literal in put your hand up the pipe, figurative in to put someone’s back up (‘annoy’) and idiomatic in I can put you up in our spare room (‘accommodate’). Note that some phrasal verbs have two particles, as put up with ‘tolerate’, look up to ‘admire’, but this construction too can be literal, as in fall out of. Phrasal verbs have syntax-like behaviour in being interrupted by their direct objects, but are lexical to the extent that their meaning is not predictable from their elements.
3.10 Phrases as words

It might be claimed that some of the items mentioned above are simply syntactic phrases that have become more word-like, by a process usually called univerbation. Since univerbation is a diachronic process that proceeds by degrees, and since there are a number of different univerbation processes, there are many different kinds of expression which, even if they started out containing two or more words, are currently considered to be single words. Some examples are given in (9).

(9) altogether, attorney general, bullseye, dyed-in-the-wool, forget-me-not, thank you, touch and go, wannabe

Because these fit the rough definition of a compound given in Section 1, they are sometimes considered to be compounds. To the extent that the constituent words are transparent, they might be considered to be MWEs. (Note that bullseye fits into the type illustrated in (4g) except that it is written as a single word and the genitive is not overtly marked.) They might also be considered to be single unanalysable words, as is implied in the term univerbation. Such items span the borders of MWEs.

4 Functional categories

The last section looked at categories that are more or less formally defined; in this section other types of category are considered, including formation-types that lead to MWEs. These are grouped together as ‘functional’ categories, in the sense that they are not formal, but they are nonetheless a heterogeneous group. In particular, the first section below scarcely seems to be a category at all, but contrasts with other categories discussed later.

4.1 Literal interpretation

It may seem trivial that literal interpretations of such constructions exist. For example, Kim is good at music and maths contains a N+and+N construction whose interpretation follows from the construction in which the coordinated pair occurs and the meanings of the words involved. Such examples are typically non-word-like. Discussion of such types is frequently carried out under the heading of
'semantic compositionality' or 'semantic transparency', which may or may not be equivalent. It is clear that semantic transparency is a matter of degree rather than a matter of yes/no; it is less clear – despite a large literature – just what is compositional (cf. Wisniewski/Wu 2012 for a useful discussion). It must be made explicit, however, that even listed items may appear perfectly transparent. Consider such examples as copper wire, singer-songwriter, elm tree, whisky and soda, Burton-in-Lonsdale. Whether that is sufficient to make them compositional is partly a matter of definition. Some of these show some evidence of word-like behaviour: for instance, whisky and soda is not reversible to soda and whisky, singer-songwriter is not easily interrupted to give, for instance, singer-sad song writer, singer-incompetent songwriter.

4.2 Figurative interpretation

An expression may also be interpreted figuratively. This is not the place to discuss the various possible figures of speech, or the distinctions between them. Suffice it to say that a figurative interpretation is a pragmatic interpretation based on the literal meaning, but providing an interpretation which is not literal. Consider the established metaphor a dog’s breakfast. We could interpret that as ‘a morning meal for a dog’, that is literally, but its established meaning is ‘a mess’, and that involves pragmatically inferring that where a dog has eaten, things are not tidy. A king’s ransom means ‘a lot of money’, which is pragmatically inferred from the amount that would be required to ransom a king. To be on the ropes is a metaphor from boxing and means ‘to be in a desperate position’. As has been shown in a number of publications (e.g. Lakoff/Johnson 2003), figurative language is ubiquitous in everyday communication, and appears to be cognitively normal and effortless; indeed, it is often the sign of brain damage if a listener cannot interpret figures of speech.

4.3 Idiomaticity

Following Grant/Bauer (2004), a distinction is drawn here between figurative interpretation and idiomatic interpretation (called ‘core idioms’ by Grant/Bauer). On this reading, an idiomatic expression cannot be understood literally (it is not semantically transparent) nor in terms of the pragmatic inferences of figurative usage. The label is frequently used for a range of different structures, including examples like red herring ‘misleading clue’ (once figurative, but the figure is not recuperable in the current state of the language), kick the bucket ‘die’, chew the fat
‘hold a conversation’, not by a long chalk ‘fall far short’, be in fine fettle ‘be fit and healthy’ (fettle is now extremely rare except in this phrase). The important point about this, though, is that expressions of all kinds can be idiomatic, including compounds (consider blackmail, yellowhammer ‘bird sp.’) and phrasal verbs (consider put up with ‘tolerate’, pan out ‘conclude’ – perhaps once figurative, but not now recuperable).

A different type of idiom is the constructional idiom, a syntactic construction where the idiomatic semantics is provided by the construction, and the construction may be filled with varied lexical content (Booij 2002). An example from English is found in (10) (cf. also Philip 2008), where all the examples mean ‘not to be particularly intelligent’.

(10) to be a couple of sandwiches short of a picnic
to be a couple of shrimps short of a barbie
to be two pennies short of the full shilling
to be several cards short of a full deck (with a variant, not to be playing with a full deck)
to be a few French fries short of a Happy Meal
to be a beer short of a six-pack
to be a few cakes short of a birthday party
to be a couple of bricks short of a wall

Another type of idiomaticity may be culture-bounded idiomaticity. Svensson (2008) considers this, looking at what she terms ‘encyclopedic (non)compositionality’, which she illustrates with expressions such as The White House and to expect a baby, which may be understood literally but which have much greater implications in our society (cf. also Sabban 2008). Examples like this show that the line between literal/transparent and non-compositionality/transparency may be more awkward than is often assumed, but also that the line between figurative and idiomatic is not necessarily easy to perceive.

### 4.4 Quotations, proverbs and the like

Any language will have a large number of recognised expressions which, in some way, acknowledge the wisdom of past speakers of the language. Some of these are quotations (from traditional tales, from literary works, from songs, movies or TV shows, from religious sources) others are proverbial or even family sayings. Their length and structure is infinitely variable: in principle, an actor or literary scholar might know the whole of Hamlet by heart and quote from it freely. Quotations are
often abbreviated, mis-quoted or even alluded to. The proverb Too many cooks spoil the broth may be shortened, as perhaps It’s a case of too many cooks, or, if someone was complaining about the number of people involved in a project, someone else might conceivably ask, So how did the broth turn out? Quotations may often go unrecognised by hearers. Some examples are given in (11).

(11) eye of the needle, fisher of men, the salt of the earth (Biblical); the goose that lays the golden egg, the grand old duke of York, white rabbits (said on the first of the month) (folklore); this sceptered isle, pound of flesh, star-crossed lovers, strange bedfellows (Shakespeare); dim, religious light, a modest proposal, a truth universally acknowledged (other literary sources); the curate’s egg, famous last words, lies, damned lies and statistics (non-literary sources); the early bird, a gift horse, a watched pot (proverbial)

Also included here are established similes like those in (12).

(12) bald as a coot
black as coal/ink/jet/night
bold as brass
clean as a whistle
cool as a cucumber (cool here means ‘unruffled’)
daft as a brush
pure as driven snow
thick as two short planks (thick here means ‘stupid’)
white as milk/snow

4.5 Abbreviations

Initialisms and acronyms deserve a marginal place in this discussion, as they are a means by which MWEs turn into single words. In initialisms, an MWE becomes an orthographic word: FBI is a single orthographic entity, while its origin, Federal Bureau of Investigation is an MWE. In acronyms, the MWE turns into a new phonological and orthographic word: the MWE North Atlantic Treaty Organization turns into NATO (/netəʊ/). Although there is a rather old-fashioned spelling convention whereby some of these items may have their individual letters interrupted by full stops/periods (N.A.T.O.), the more modern orthography stresses the wordhood of the outcome. For the most successful acronyms, the original MWE becomes lost, and a new morpheme arises: scuba < self-contained underwater breathing apparatus.
Blends may be seen as a cross between compounds and abbreviations. In a blend, typically, the first part of the first word and the last part of the second word are telescoped together with some loss of phonological material. An example is *infotainment* < *information + entertainment* or *administrivia* < *administration + trivia*. Because blends can be seen as a type of compound, they are MWEs.

### 4.6 Rhyming slang

The essence of rhyming slang is that a word is replaced with a (usually two- or three-word) phrase which rhymes with the original. In this first stage, non-MWEs are deliberately replaced by MWEs. The word *kids* is replaced by *dustbin lids*, the word *stairs* is replaced with *apples and pears*. Note that there is no semantic link between the original word and the rhyming replacement, though occasional examples may be (or may be thought to be) jocularly appropriate, such as *trouble and strife* for *wife*. To make things more difficult, the rhyming word is then often deleted, so that *kids* becomes *dustbins* and *stairs* becomes *apples* and what was an MWE is now replaced by a polysemous lexeme. Although this is often termed ‘Cockney rhyming slang’ it is not restricted to London English. Not only is it also found, for instance, in Glasgow, Australia and New Zealand, but occasional expressions of rhyming slang creep unacknowledged in the vocabulary of the wider language community: *to do bird* (*bird lime* = *time* [in prison]), *let’s have a butcher’s* (*butcher’s hook* = *look*), *my old china* (*china plate* = *mate*), *use your loaf* (*loaf of bread* = *head*), *rabbit on* (*rabbit and pork* = *talk*). All of these retain the distinctly informal style level of the originals, and form new idiomatic MWEs.

All the examples provided above are established examples. But rhyming slang can also be used productively. One website cites *Jar Jar Binks for forty winks* (‘a snooze’), clearly postdating the relevant Star Wars movie, and not necessarily widely known.

### 4.7 Collocation

Collocations are sets of words which habitually occur together, even if they are perfectly transparent. A standard example concerns the way in which *dry* changes its meaning depending upon what it collocates with, as shown in (13).

(13)  
- *a dry cough* (not producing catarrh)
- *a dry lecture* (not interesting)
a dry state (where alcohol is not sold)
a dry wall (built without cement)
a dry wine (not sweet)
a dry wit (dead-pan)
dry eyes (without tears)
dry ground (not wet)
dry toast (not buttered)
dry weather (not raining)

Collocations are not always of the same strength. Sometimes the ability to predict one of the items in the collocation from the other is strong, sometimes it is weak. This can be measured in terms of the mutual information each element provides as to the identity of the other element(s) in the collocation (Xiao 2015). This may complicate the process of deciding what belongs in the lexicon in a theoretical sense, but does not interfere with the notion that more than just the individual word might have to be listed.

Note that while dry in dry ground can be submodified (very dry ground), and many of these expressions can be interrupted (a dry French wine, dry red-rimmed eyes) some of them seem to be more word-like (*very dry toast, dry battery does not appear to allow random insertions).

A particular kind of collocation is that provided by light verbs. It is make a difference, give a lecture, make a mistake, take the opportunity, take a shower, have a smoke. There does not seem to be any straightforward semantic reason for the selection of these light verbs, and speakers (including native speakers) will often use a different one from the one expected, and say things like do a mistake.

Another similar case is provided by adjectives that take complements, and then collocate with fixed prepositions, as in afraid of, averse to, different from/ than/to, proud of. The case of different, which becomes a matter of prescription, shows that the preposition is not always fixed, but generally speaking the preposition has to be seen as being chosen by the head adjective. This puts such constructions of the borderline between being lexical combinations and syntactic structures showing government.

4.8 Formulae

Formulae are the way things are said rather than the way they could be said (cf. also Sabban 2008). In many European languages, there is an expression which can be translated as ‘good day’ which is a greeting. In England, good day is a
farewell. In Australia and New Zealand, *good day* (with a phonetically very much reduced first syllable) is again a greeting. In the usage of young New Zealanders around the turn of the millennium, *spot you later*, and *laters* were farewells (Bauer/Bauer 2003). The fact that these are greetings and farewells (as opposed to other potential expressions which are not, such as *until we meet again*, *till the next time* or *soon*), with the corresponding increase in usage of these precise phrases, makes them into formulae. Corresponding to the rather old-fashioned *How do you do?* heard in England, *How are you doing?* can be heard in other parts of the English-speaking world, but as a day-to-day greeting rather than as a greeting on first introduction. *How is it going?* is an alternative possibility, but not *How does it go?* There are many perfectly grammatical possible ways of saying things that are never used, and those that are used, and their precise meaning, may be unexpected.

Formulae, then, are particular types of collocation, with high frequency in particular social environments. While they have syntactic structure (in the case of *How do you do?* a rather outmoded syntactic structure), some of them may be learned as listed, fixed expressions, or have the status of words (as with *good-bye*).

### 4.9 Lexicalisation

Lexicalisation is the process of becoming a lexical item. It depends on semantic shift (often called idiomatisation, e.g. Lipka 1994) and formal change. Although it may be difficult or impossible to measure degrees of lexicalisation, it is a matter of more or less not either/or. At the one end, the most lexicalised items like *lord* are historically derived from elements meaning ‘loaf ward’, and all internal structure and the meaning of the original elements has been lost. At the other end, we have freely produced syntactic constructions which are perfectly transparent in form and meaning. The terminology of lexicalisation is very variable, and various intermediate stages have been postulated (cf., e.g., Bauer 1983a). In formal terms, we find constructions whose elements are transparent, instances where the elements have undergone some phonetic erosion (e.g. *Christmas* which phonologically contains neither *Christ* nor *mass* any more), to constructions whose elements probably cannot be perceived without formal instruction (such as *dearth*, related to *dear*). Semantically, transparent elements may have to be interpreted figuratively (e.g. *hedgehog* or *fire dog*), or, even if appearing formally transparent, be semantically totally opaque (such as *blackmail* and *woodchuck*). It will be clear from these examples that various factors influence lexicalisation, but many MWEs are, almost by definition, somewhere on the lexicalisation spectrum.
5 Discussion

While this wide range of MWEs has to be recognised (however difficult they may be to systematise), there are a number of expressions which do not appear to be sufficiently lexical to fit in the category. Any study of n-grams will come up with expressions like in a, which collocate not because in a is constituent with its own meaning, but because all members of the category preposition are typically followed by determiner phrases, typically headed by determiners like a in initial position. The high number of such cross-constituent collocations has thus more to do with the productivity of syntax than anything lexical. Similarly, colligations, such as the fact that the verb construct is transitive, is not a matter of lexis but a matter of grammar (again, perhaps, a matter of government). It is true that construct a building is likely to be more frequent than construct a daisy, but this has as much to do with the nature of the world as with the nature of lexical items. While it makes little sense to suggest that construct demands in its complement something with a feature [+ constructible], as has been done on occasions, it makes rather more sense to say that pragmatically the need for a sentence which contains construct a daisy is likely to be extremely low (although it might be possible if people were decorating a kindergarten and making flowers out of recycled material to use as decorations). As McCawley remarked many years ago (McCawley 1971), if someone says my toothbrush is pregnant, it is unlikely to be their grammatical competence which is at fault.

The borderline between things which happen to collocate because they are syntactically likely to arise in similar contexts and what is lexical is not necessarily an easy one to draw. I tend to think that is like a is on the grammatical side, but Wikberg (2008: 136f.) makes a case for it on the basis that it is a formula used to introduce similes.

Borderlines like these, and one mentioned earlier between government and lexical structure, are potentially problematic, and the entire idea that there are such borderlines is worthy of further discussion. At one extreme we find a view, which we can characterise as essentially Chomskian, that virtually everything we produce is the result of free syntactic rules in operation. The other extreme position, and one worth arguing for, would be that there is no such thing as free syntax, but that everything is lexically-driven, with MWEs, fixed phrases and strongly restrictive constructions accounting for the fact that speakers do not say many things which might appear to be grammatical. I distrust extreme views, and suspect that there is some of each involved, but that the limits of each require careful motivation. It seems to me that a line like Carroll’s (1871) ‘Twas brillig and the slithy toves did gyre and gimble in the wabe shows that there must be some syntax separate from vocabulary items, while the range of MWEs discussed in the phra-
seological and constructional literature shows that much of what we say on a
day-to-day basis requires minimal independent syntax to be formed into perfectly
normal conversational turns. In saying that, I imply consciously that there may be
a difference between written and spoken language in this regard. All of these are
open questions.

Two questions have been ignored in this presentation. The first is frequency. It
might seem that MWEs must be frequent enough to be recognised by speakers,
but there are many constructions that are invented on the spur of the moment and
yet fit (at least some of) the criteria for recognising MWEs. Consider, for instance,
examples in (4e) and (10). Frequency is a correlate of lexicalisation, but low fre-
quency does not prevent something from being an MWE.

The second point to be considered is speaker accuracy. As was pointed out in
relation to light verbs, speakers are not always consistent in what they say, and
what start out as errors may spread and cause language change. This seems to go
beyond performance errors in the sense of Chomsky (1965). Listening to current
spoken English suggests that there is huge variation in complementation patterns
at the moment, something else that lies on the borderline between government
and lexical collocation (if these can be fully distinguished).

In this contribution, I have presented a sketch of some of the types of MWE
that can be found in English. The classification I have used is, however, not
exhaustive, and the various categories I have used are not mutually exclusive, so
that I consider the classification used here to be no more than an ad hoc frame-
work for discussion and not a typology. Various alternative classifications are
provided in Granger/Meunier (eds.) (2008), but while I see the value of these clas-
sifications, I do not think we are yet at a point where a typology of MWEs is possi-
ble. Partly, as I have tried to suggest above, this is because the very nature of
MWEs is pluricentric. There is no simple distinction between lexical and syntac-
tic, there is no simple distinction between compositional and non-compositional
or between lexicalised and non-lexicalised. Rather there is a host of expressions
which link to syntactic structure and to semantic structure (and, indeed, even to
phonological structure, although I have not discussed matters such as allitera-
tion and rhyme here) in multiple ways. Compounds are one type of MWE, which
may not easily be distinguished from other MWEs, because they are part of the
network and give rise to the same problems of description and interpretation that
other MWEs do.

That brings us back to the starting point of this contribution. It is hard to
define compounds because they overlap with other MWEs in sharing features of
wordhood, they overlap with syntax in that some things which have been called
compounds are viewed by others as syntactic, because some of them, at least, are
semantically transparent, and because some of the things that some scholars call
compounds arise from pieces of syntactic structure being frozen. While anyone is free to define compounds as they see fit, agreement on any definition which can determine which of the structures that have been canvassed here are really compounds seems a long way off.

References

Adams, Valerie (2001): Complex words in English. Harlow: Pearson.
Bauer, Laurie (1983a): English word-formation. Cambridge, UK: Cambridge University Press.
Bauer, Laurie (1983b): Stress in compounds. A rejoinder. In: English Studies 64. 47–53.
Bauer, Laurie (1998): When is a sequence of two nouns a compound in English? In: English language and linguistics 2. 65–86.
Bauer, Laurie (2000): Word. In: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (eds.): Morphology. An international handbook on inflection and word-formation. Berlin/New York: De Gruyter. 247–257.
Bauer, Laurie (2001a): Compounding. In: Haspelmath, Martin et al. (eds.): Language universals and language typology. Berlin/New York: De Gruyter. 695–707.
Bauer, Laurie (2004): Adjectives, compounds and words. In: Nordic Journal of English Studies 3, 1. 7–22.
Bauer, Laurie (2016): Re-evaluating exocentricity in word-formation. In: Siddiqi, Daniel/Harley, Heidi (eds.): Morphological metatheory. Amsterdam/Philadelphia: Benjamins. 461–477.
Bauer, Laurie (2017): Compounds and compounding. Cambridge, UK: Cambridge University Press.
Bauer, Laurie (to appear): Stressing about the news. In: New Zealand English Journal.
Bauer, Laurie/Bauer, Winifred (2003): Playground talk. Wellington: Victoria University.
Bauer, Laurie/Lieber, Rochelle/Plag, Ingo (2013): The Oxford reference guide to English morphology. Oxford: Oxford University Press.
Bauer, Laurie/Renouf, Antoinette (2001): A corpus-based study of compounding in English. In: Journal of English Linguistics 29. 101–123.
Bell, Melanie J. (2014): The English noun-noun construct: a morphological and syntactic object. In: Ralli, Angela (eds.): Morphology and the architecture of grammar. 59–91. (Internet: https://geertbooij.files.wordpress.com/2014/02/mmm8_proceedings.pdf, last access: 20.4.2018).
Benor, Sarah/Levy, Roger (2006): The chicken or the egg? A probabilistic analysis of English binomials. In: Language 82. 233–278.
Booij, Geert (2002): Constructional idioms, morphology and the Dutch lexicon. In: Journal of Germanic linguistics 14. 301–329.
Carroll, Lewis (1871): Through the looking glass and what Alice found there. London: Macmillan.
Carstairs-McCarthy, Andrew (2002): An introduction to English morphology. Edinburgh: Edinburgh University Press.
Chomsky, Noam (1965): Aspects of the theory of syntax. Cambridge, MA: MIT Press.
Chomsky, Noam/Halle, Morris (1968): The sound pattern of English. New York: Harper & Row.
Dixon, R. M. W./Aikhenvald, Alexandra Y. (2002): Word: a typological framework. In: Dixon, R. M. W./Aikhenvald, Alexandra Y. (eds.): Word. A cross-linguistic typology. Cambridge, UK: Cambridge University Press. 1–41.

Giegerich, Heinz J. (2004): Compound or phrase? English noun-plus-noun constructions and the stress criterion. In: English Language and Linguistics 8. 1–24.

Giegerich, Heinz J. (2015): Lexical structures. Edinburgh: Edinburgh University Press.

Granger, Sylviane/Meunier, Fanny (eds.) (2008): Phraseology. An interdisciplinary perspective. Amsterdam: Benjamins.

Grant, Lynn/Bauer, Laurie (2004): Criteria for re-defining idioms: Are we barking up the wrong tree? In: Applied Linguistics 25. 38–61.

Hall, Robert A. Jr (1964): Introductory linguistics. Philadelphia/New York: Blackwell.

Hippisley, Andrew (2015): The word as a universal category. In: Taylor, John R. (ed.). 246–269.

Klinge, Alex (2005): The structure of English nominals. Copenhagen: Copenhagen Business School. [Doctoral dissertation].

Kunter, Gero (2011): Compound stress in English. (= Linguistische Arbeiten 539). Tübingen: Niemeyer.

Lakoff, George/Johnson, Mark (2003): Metaphors we live by. Chicago: University of Chicago Press.

Lieber, Rochelle (2010): Introducing morphology. Cambridge, UK: Cambridge University Press.

Lipka, Leonhard (1994): Lexicalization and institutionalization. In: Asher, R. E. (ed.): Encyclopedia of language and linguistics. Vol 4. Oxford: Pergamon. 2164–2167.

Marchand, Hans (1969): The categories and types of present-day English word-formation. 2nd ed. Munich: Beck.

McCawley, James D. (1971): Where do noun phrases come from? In: Steinberg, Danny D./Jakobovits, Leon A. (eds.): Semantics. Cambridge, UK: Cambridge University Press. 217–231.

McEnery, Tony/Xiao, Richard/Tono, Yukio (2006): Corpus-based language studies. London/New York: Routledge.

Moon, Rosamund (2015): Multi-word items. In: Taylor, John R. (ed.). 120–140.

Payne, John/Huddleston, Rodney (2002): Nouns and noun phrases. In: Huddleston, Rodney/Pullum, Geoffrey K. (eds.): The Cambridge grammar of the English language. Cambridge, UK: Cambridge University Press. 323–524.

Philip, Gill (2008): Reassessing the canon: ‘Fixed’ phrases in general reference corpora. In: Granger/Meunier (eds.). 95–108.

Pinker, Steven (1999): Words and rules. London: Phoenix.

Rosenbach, Anette (2006): Descriptive genitives in English: a case study on constructional gradience. In: English language and linguistics 10. 77–118.

Sabban, Annette (2008): Critical observations on the culture-boundness of phraseology. In: Granger/Meunier (eds.). 229–241.

Spence, Nicol C. W. (1969): Composé nominal, locution et syntagme libre. In: Linguistique 2. 5–26.

Svensson, Maria Helena (2008): A very complex criterion of fixedness: Non-compositionality. In: Granger/Meunier (eds.). 81–93.

Taylor, John R. (ed.) (2015): The Oxford handbook of the word. Oxford: Oxford University Press.

Wikberg, Kay (2008): Phrasal similes in the BNC. In: Granger/Meunier (eds.). 127–142.

Wisniewski, Edward J./Wu, Jing (2012): Emergency!!!! Challenges to compositional understanding of noun-noun combinations. In: Hinzen, Wolfram/Machery, Edouard/
Werning, Markus (eds.): The Oxford handbook of compositionality. Oxford: Oxford University Press. 403–417.
Wray, Alison (2015): Why are we so sure we know what a word is? In: Taylor, John R. (ed.). 725–750.
Xiao, Richard (2015): Collocation. In: Biber, Douglas/Reppen, Randi (eds.): The Cambridge handbook of English corpus linguistics. Cambridge, UK: Cambridge University Press. 106–124.