When is a sequence of two nouns a compound in English?[1]

LAURIE BAUER
Victoria University of Wellington
(Received 9 July 1997; revised 14 November 1997)

Constructions of noun + noun have been treated in two distinct ways in the literature: either they have been treated as compounds, or they have been treated as noun phrases with modifiers which happen to be nouns. Sometimes it is assumed that there are two distinct classes, which can be neatly distinguished. In this paper it is argued that the criteria which are usually assumed to distinguish between these two construction types do not draw a clear and consistent distinction between a syntactic and a morphological construction. Many of the criteria instead are indirect measures of listedness, which, it is argued, is not sufficient to show morphological status. Accordingly, it is claimed that the criteria to which reference is generally made do not allow us to distinguish between a class of noun + noun compounds and a class of noun + noun syntactic constructions. Rather the two should be treated as variants of a single construction (possibly morphological, possibly syntactic), at least until such time as a suitable coherent distinction can be properly motivated.

1 Introduction

A compound is usually defined as being a word (in the sense of lexeme) that is made up of two other words (in the sense of lexeme) (see Bauer, 1978: 48–52 for some sample definitions). This, however, avoids an important question: how do we know whether a sequence of two words forms a new lexeme or simply a syntactic construction? The question framed in the title to this article is one which has caused great division among writers on English, though the problems involved are not always recognized. We may divide the scholars who have answered this question into two camps: the splitters and the lumpers. The splitters see two classes of noun + noun sequence in English: syntactic constructions consisting of nouns with nominal modifiers, and compounds. The lumpers see a single class, usually identified as a class of compounds. I have been among the lumpers (see e.g. Bauer, 1978, 1983a), at least insofar as any division can be correlated with a difference in stress, so I am scarcely a disinterested party in the debate.

Despite the fact that some might consider this question to have been settled some twenty years ago by Levi (1978), and despite the fact that there is a great deal of consensus among scholars of English word-formation on this topic, I have twice recently found myself having to defend the position that there is no clear division between these two sets of constructions. In attempting to justify this stance, I have had to consider rather more potential criteria than were discussed by Levi (1978). This fact, along with the fact that Levi's arguments are apparently not well known, makes it worthwhile revisiting the topic here. That many linguists do believe there to be a distinction is shown, among other things, by the fact that the questionnaire for the Lingua / Croom Helm / Routledge Descriptive Grammar series allows authors to deny the existence of compounds while affirming the existence of nominal modifiers to nominal heads. While it may be that in some languages there is ample justification for such a distinction, none of the relevant grammars provides definitions to indicate how the distinction is being drawn. We also find books like Fabb (1994: 87) where *boat train* is said to be a noun, but *Liverpool boat train* is said to be a syntactic construction. Thus, however much this question has been considered in the past, I can assure readers who might believe otherwise that every one of the criteria discussed here has been suggested to me in the last two years or so as being a possible criterion for distinguishing compounds from syntactic constructions, and that I am not simply setting up a straw man. It seems that at least some morphologists and at least some syntacticians have conflicting assumptions in this area, and these deserve to be aired.

Splitters do not always make clear precisely what the criteria are for making a division between syntactic constructions and compounds. Frequently a single criterion is stated, and it is assumed that other criteria support the division, without argument. An interesting case in point is Chomsky & Halle (1968: 156), where it is recognized that a distinction on the basis of stress between *Fifth Avenue* and *Fifth Street*, elsewhere (e.g. Chomsky & Halle, 1968: 21) attributed to a distinction between a noun phrase and a compound, is not obviously reflected in syntactic behaviour. The problem is put aside as being irrelevant to phonological theory, but it is extremely relevant to morphological theory.

Bauer (1978: 89–95) discounts such distinctions on the grounds that 'the stress criterion is inconsistent; [. . .] there are no syntactic correlates; and [. . .] there are no semantic correlates'. This will be considered further below, but it should be noted that it is, in effect, a lack of evidence for the contrary position which leads to the conclusion that there is a single class, not positive evidence in its favour.

I shall begin by considering some possible criteria for distinguishing between syntactic constructions and compounds. I shall then ask whether these various possible criteria correlate with each other. It should be noted at the outset, however, that I am taking on an impossible task in this paper: I am trying to prove a negative. I take it as given that I cannot prove conclusively that there is no difference between two sets of constructions. All I can do, and what I shall attempt to do here, is suggest that criteria which have been advanced in the past as the basis for drawing a

distinction do not have the effect that has been claimed for them. In other words, I shall argue not that it is impossible to draw a distinction between two classes of construction in this area, but that we do not appear to have sufficient reason to divide the class of noun + noun sequences into two separate classes whose behaviour can be distinguished in a consistent manner.

2 Possible criteria

In this section I shall consider six possible criteria which might be used to distinguish between syntactic constructions and compounds. In each case I shall conclude that the criterion does not provide a coherent way of distinguishing between morphological and syntactic constructions.

2.1 Compounds are listed

Viewing compounds but not syntactic constructions as listed is perhaps rather more a lexicographical approach than a linguistic approach. Yet many linguists seize upon one aspect of listedness – namely idiomaticity – and use that as a criterion for compound status. Consider, for example, the definition given by Jespersen (1942: 137):

> we have a compound if the meaning of the whole cannot be logically deduced from the meaning of the elements separately

(see Bauer, 1978: 44–5, 52–4 for similar comments from other scholars, and Levi, 1978: 44 for some discussion of such claims). Di Sciullo & Williams (1987) argue at great length that listedness is not a property of words, *per se*, although more words than sentences are listed because of the nature of these linguistic objects. They set up parallel hierarchies, where each step in the hierarchy is defined in terms of the previous step, and where the saturation of listedness follows the same steps (Di Sciullo & Williams, 1987: 14):

Hierarchy of definition	*Hierarchy of listedness*
morpheme	all the morphemes are listed
word	'most' of the words are listed
compound	many of the compounds are listed
phrase	some of the phrases are listed
sentence	four or five of the sentences are listed

The hierarchy of listedness is expressed in very vague terms (and the scare quotes round the word *most* do not help in this regard), but is important in that it denies an absolute correspondence between listedness and word status. Listed phrases include idioms; listed sentences include fixed locutions such as *How do you do?*; unlisted words include those which are productively formed. Consider the following simple examples:

(1) 'How about a drinkette?' said Helen.
(Sam Llewellyn, *Death Roll*. London: Michael Joseph, 1989: 121)
(2) the multiverse is full of little dimensionettes, playstreets of creation.
(Terry Pratchett, *Pyramids*. London: Corgi, 1989: 233)

It seems unlikely that either *drinkette* or *dimensionette* is a familiar word, any more than this sentence is a familiar sentence, and in neither case is the construction a listed one (see Di Sciullo & Williams, 1987: 7–15). Yet the status of *drinkette* and *dimensionette* as words is not in doubt. So we have listed items which are not words and words which are not listed items.

Moreover, it would seem that it is historical accident that some words are listed while others are not. Given the possibility of *raper* (in (3)), it is surely accident that the listed English word is *rapist*.

(3) Hates a raper, does your average villain.
(William Garner, *Paper Chase*. London, etc.: Grafton, 1988: 87)

I have avoided compounds in the discussion here, but what is true of derivatives and phrases is presumably also true of compounds, whether they are viewed as syntactic constructions or not. That is, listedness in itself is totally irrelevant in deciding whether or not something is a syntactic construction or a morphological construction, and thus cannot be used to set off compounds from anything else.

However, the potential use of listedness as a criterion raises a significant problem. In the same way that new derived words are produced by productive processes (as illustrated in (1), (2), and (3)), the process for making items which behave like compounds is productive (see e.g. Gleitman & Gleitman, 1970; Downing, 1977). Seeing compounds as exclusively listed items would appear to be a denial of this productivity. Langacker (1987: 29) discusses what he terms the 'rule/list fallacy', and claims that it is possible for something which is produced by rule nevertheless also to be listed. If we accept that rule does not exclude list, the inverse, that list does not imply rule, seems equally valid. One implication might then be that syntactic constructions were produced by rule, and compounds would arise only from the listing of syntactic constructions. This, however, would lead to a contradiction. On the one hand, syntactic constructions are said to be different types, with compounds defined as being listed. On the other hand, compounds all derive from syntactic constructions, and so are of the same fundamental type, and they differ only marginally in whether or not they are listed. This means that any claim that there are two distinct sets of constructions here and that listedness is the (or even a) fundamental distinction between the two is probably self-destructive, as well as counter to the general widely accepted notion of the productivity of compounding.

2.2 Compounds are written as a single word

It is, of course, extremely dangerous to assume that orthography has any linguistic value, since the spoken language is, in theory at least, independent of any

orthography. Nevertheless, orthography might reflect strong linguistic intuitions, intuitions which themselves are based on nonorthographic factors, and the question is thus worth considering.

It is well known that English orthography is extremely inconsistent in dealing with noun + noun collocations. Norms do not even operate on single collocations, so that *girl friend, girl-friend* and *girlfriend* are all found in different English dictionaries (with *Chambers English Dictionary* 1988 distinguishing the first and third, but otherwise the same meaning attached to all three: Bauer, 1988: 101). Furthermore, there are three options here, and it is not clear whether hyphenation should be taken as a 'single word' solution or a 'two word' solution, since it is so obviously a compromise. With the orthography depending so clearly on 'the taste and fancy of the speller' (as Samuel Weller would put it) or on house-style (as a publisher might put it), it cannot represent a consistent linguistic judgement about the nature of the constructions.

Given an example like *girl friend, girl-friend, girlfriend*, it might appear that writing the two words separately is a conservative option while writing them together is an innovative option, and that there is a diachronic process of change in orthography based on the degree of item-familiarity shown by the whole construction. If this were all there was to say on the matter, orthography could well be expected to correlate with compound frequency and the age of the compound (the length of time for which it has been listed). (Given what was said about listedness above, this would still not necessarily show anything about the status of the construction as a lexeme.) However, there is evidence that varying orthography is not simply a matter of item-familiarity. *The Oxford Dictionary of New Words* (Tulloch, 1991), for example, lists *daisy wheel, daisy-wheel* and *daisywheel* as alternative spellings for the same item without there being any evidence for a diachronic progression. The same work lists the neologism *airside* written as a single orthographic word, without any evidence that it has ever been written in any other way. *College degree* is an item-familiar construction which is nevertheless written as two words. It seems that, to the extent that one spelling becomes relatively standardized, it is by convention rather than because of any linguistic principles at work.

This is not to say that no regularities can be observed in the orthography or that there is absolutely no consistency in the writing of such constructions. On the whole, for example, long words are written separately, independent of the stress pattern, while short words are more likely to be written together. But if there is a distinction which depends on the number of letters in a word, this is clearly not primarily a structural linguistic distinction, but one which depends on factors such as ease of perception for the reader. Even if the length of words is a matter of syllables rather than letters (and thus more obviously linguistic), it is not clear that a preferred analysis would be one in which the number of syllables in a word could determine the construction to which strings containing that word would belong. This is one possible analysis of comparatives in English, so the case is not unique, but the analysis of comparatives is also controversial.

2.3   Compounds have fore-stress

The terminology of fore-stress or unity stress or compound stress no doubt leaves something to be desired, but the general principle involved here is simple enough: a compound made up of two words carries the major stress on the first of the two elements; a syntactic construction made up of two words carries the major stress on the second of the two elements. (Some authorities prefer to talk in terms of level stress rather than final stress, and a few distinguish the two, although it is my belief that level stress will be resolved into one of the two other patterns under an intonational nucleus.) Thus *'apple cake* is a compound, while *apple 'pie* is a syntactic construction.

There are a number of different problems with a distinction based on stress. First, we may question the accuracy and consistency of judgements assigning two-word combinations to one set or the other. Experimental evidence suggests not only that groups of speakers do not provide a consistent answer to where the stress falls in such collocations, they are not even individually consistent in reporting their own behaviour (Bauer, 1983b). If we look away from self-reporting to reporting by a single observer, we find the same kind of vacillation (Vos, 1952). To illustrate something of the problem, consider the marking for a few noun + noun constructions from four published sources:

Item	Chambers	COD	EPD	LPD
churchwarden	level	initial	final	final
kitchen garden	level	stress unmarked	final ~ initial	final
master key	initial	stress unmarked	initial	initial ~ final
nightwatch	initial	not listed	final ~ initial	not listed
steam-roller	level	initial	initial ~ final	initial
wage-earner	level	stress unmarked	initial	initial

Chambers: Chambers English Dictionary (Cambridge University Press, 1988); *COD: Concise Oxford Dictionary* (Oxford University Press, 9th edition, 1995); *EPD: English Pronunciation Dictionary* (Dent, 14th edition, 1977); *LPD: Longman Pronunciation Dictionary* (Longman, 1990).

In providing this brief list, I do not mean to imply that there is no agreement at all, or that disagreement is equally likely with all constructions or with all types of construction. I do, however, wish to show that there is considerable disagreement, and that it is therefore not clear that there is a 'right answer' to the question 'What is the stress pattern on this noun + noun collocation?' Accordingly, there is not

necessarily a clear division here between something which is a compound by this criterion and something which is a syntactic construction.

Some authors (Fudge, 1984: 144–9; Ladd, 1984; Liberman & Sproat, 1992) have argued that stress patterns are associated with particular semantic relationships between the two nouns involved. 'B made of A' as in *stone wall* calls forth phrase stress, while 'B used for A' as in *pruning shears* calls forth compound stress. As far as I am aware, no precise and reliable set of such relationships has ever been established, although important trends have been noticed. But even if such correlations were absolutely perfect, it is not clear that they would indicate a distinction between a lexical construction and a syntactic one, since it is not clear that the 'made of' relation is in any sense less lexical than the 'used for' relation. The two types could just as easily be seen as prosodically distinguished sub-types of a single construction. Given the failure of such attempts to provide coherent pictures, I see even less reason to assume that stress reflects a distinction between lexeme and syntactic construction.

The next attack on the notion that stress distinguishes between compounds and syntactic constructions is based on data which, as far as I know, were first pointed out by Lees (1960: 120). Lees noted that some apparently parallel constructions show reliable differences in stress pattern. He cited *Madison 'Avenue* versus *'Madison Street* and *apple 'pie* versus *'apple cake* and commented that 'all composites in *-street* and *-cake* are compounds, while all in *-avenue* and *-pie* are invariably nominal phrases'. Where the examples with *cake* and *pie* are concerned, it is perhaps worth pointing out that this appears to be independent of status as listed (compare *apple cake* and the unexpected *turnip cake*, and *apple pie* with the unexpected *turnip pie*), and also independent of the precise semantic relationship between the elements (*apple cake* contains apples, *birthday cake* is for a birthday, *pancake* is made in a pan, *angel cake* is fit for angels, etc., but all have fore-stress). Such examples (which can be multiplied) make it seem as though at least in some cases the stress pattern in noun + noun collocations is lexically conditioned. If we say that stress differences indicate different constructions, then the implication is that these nouns are determining the syntactic constructions in which they can occur by lexical conditioning. The closest parallel case that has been suggested to me is that *give* permits dative shift but *donate* does not, but there is considerable discussion in the literature that suggests that this distinction is predictable and not simply a matter of lexical conditioning (see e.g. Pinker, 1989: 45ff).

It is true that not all noun + noun constructions show lexical conditioning of stress in this way. In some cases lexicalization may cause variation from this general pattern (*'snowball* contrasted with *rubber 'ball*); in some cases particular semantic relationships are associated with one or the other stress pattern, and this may take precedence (thus *rubber ball* is a typical member of a group of material + object constructions (or 'B made of A') which take double stress, so that *glass 'case* 'case made of glass' can be distinguished from *'glass case* 'case for displaying, etc. glasses', which conforms to the usual pattern with *case*). There are also instances which

appear totally inexplicable, such as the distinction in stress (where maintained) between 'town house and country 'house.

Finally, it should be pointed out that initial and final stressed noun + noun collocations are indistinguishable in terms of standard criteria for wordhood (Bauer 1978: 93–4).

2.4   The first element in a compound is syntactically isolated

The lexicalist hypothesis states that elements within the word should not be available to the syntax (that words are, in another terminology, the atoms of syntax). Just as we do not expect, according to this hypothesis, roots in derivatives to be pluralizable or even items to which reference can be made, so we should not expect elements within compound words to be pluralizable, picked out by anaphora or independently submodifiable. While this, in principle, applies to either the first or the second element in a compound, my experience is that intuitions tend to be a little hazy in applying this to head elements. For example, in (4), speakers may not know whether the adjective or the plurality refers simply to the second element of the compound or to the compound as a whole.

(4) The colourful houseboats are moored on every canal.

Accordingly, and in line with most of the literature on this subject, this criterion will be considered with reference to the modifying element of compounds only. It should, however, be noted that there are certain cases where the letter of the lexicalist hypothesis is broken, even with derivatives. Consider

(5) . . . what sharply distinguishes Chomskyan practice from that of his structuralist
         forebears is . . .   (N. Vincent, Zero, in Asher, 1994: 5082)

In (5) his refers back to Chomsky, even though Chomsky was the root in a derivative. This kind of breach is well-attested (see Liberman & Sproat, 1992: 173–4). Ward, Sproat & McKoon (1991) review the literature on these and other similar cases, and argue at some length that there is no GRAMMATICAL problem with the use of anaphora like that in (5), though there may be PRAGMATIC conditions which render it less likely. They cite many attested examples of the phenomenon, including

(6) So, I hear you're a real cat-lover. How many do you have now?
                                                  (TV game show, cited in Ward et al., 1991: 471)

It is thus at least arguable that this particular type of syntactic isolation is totally irrelevant for distinguishing compounds from syntactic constructions; but even if there are those who wish to disagree with Ward et al., there is no evidence that this criterion distinguishes between morphological and syntactic constructions.

While it is the case that plural-marking on the first element in a noun + noun collocation is rare, it is far from unknown, and some linguists have expressed the opinion that the phenomenon is becoming more rather than less common (Mutt, 1967; Dierickx, 1970). The difficulty with assigning games mistress to one construc-

tion and *art mistress* to another is that nothing else seems to correlate with this distinction, so that it becomes purely a matter of definition. Moreover, it is not always clear that one or the other type must be used: *drug courier* and *drugs courier* seem synonymous, take the same stress pattern, and appear to be quite free variants of each other. If the two are always substitutable, there would seem to be no good reason to assign them to separate construction types, except on the basis of the variable <s>.

In longer compounds, the <s> marking is sometimes used to show the immediate constituent structure in the compound. Contrast the attested distinction between *[[[British Council] jobs] file]* and *[British Council] [job file]* (cited in Bauer, 1978: 40). If the <s> can be used in this way, it suggests that plurality is not all that is at stake here, and that a so-called 'plural' attributive may not be a genuine breach of the lexicalist hypothesis.

Where adjectival modification of the first element of a noun + noun compound is concerned, the facts are far from clear. The general assumption is that it cannot occur. Consider, for example, the collocation *river-bed*. A *broad river-bed* will be interpreted as a broad bed for a river rather than as a bed for a broad river. And if an adjective is used which must refer to the river and not the bed, e.g. a *fast-flowing river-bed*, most speakers will deny the acceptability (possibly the grammaticality) of the construction. Given that, it is not clear how to deal with the attested form *swollen river-bed* (Radio New Zealand, 6 p.m. News, 6 Nov. 1988): is it simply ungrammatical (in which case, why was it used in the neutral context of a news bulletin?), is it rule-breaking creativity (same question), or is it, in fact, normal language use? Or consider the collocation *walking stick*. At one level it seems difficult to modify the first element of this construction: *unsteady walking stick, slow walking stick* either sound unlikely, or are interpreted with the adjective modifying the construction as a whole. Yet with the parallel construction *lending right*, we find *public lending right* listed in *COD9*, where *public* must modify *lending* and not *right*. Other attested examples include *big-ticket items, light-rail system, Serious Fraud Office, instant noodle salad*. While it may be the unmarked case to find the adjective modifying the collocation as a whole, such examples show that the interpretation where the adjective modifies the first element only is perfectly possible. It is therefore not at all clear whether the ban on adjectival modification of the first element in a compound can really be upheld as a definite fact in English grammar.

The other side of this particular coin is that, if this criterion is to be used to distinguish two sets of constructions, we expect to find noun + noun syntactic constructions where the default reading is for an adjectival modification to apply to the first noun only. It has been suggested to me that material + object constructions fit this pattern, and that *Chinese jade figure* and *Swedish steel blade*, for instance, are interpreted as *[Chinese jade] figure, [Swedish steel] blade*. While I agree that this interpretation is a possible one, given an appropriate context, it seems to me that the default interpretation is still *Chinese [jade figure], Swedish [steel blade]*, and thus

that no difference is established between such constructions and *river-bed* and *lending right* discussed above.

A further point that needs to be considered is the possibility of constructions whose bracketing is [[noun + noun] noun]. In these cases, it could be argued, the first noun of a noun + noun collocation gets independent premodification. If such structures are all compounds, this makes no difference to the line of argumentation. However, if these are syntactic constructions, then it is probably the case that all noun + noun collocations allow independent premodification of the first element. However, such an argument is circular, since it uses its own conclusion as a premise to argue for the conclusion that has already been presupposed. Instances of this type are, therefore, no help on either side of the debate.

2.5   Compounds do not permit co-ordination

If we consider a word like *buttercup*, it does not seem possible to co-ordinate anything with either the first or the second element: **bread and buttercups, *buttercup and saucer* do not seem possible. On the other hand, co-ordination does seem possible in the construction *steel bar: iron and steel bars, steel bars and weights*. This might, then, be considered a criterion for distinguishing between two types of construction.

The precise facts about co-ordination with noun + noun constructions have never, as far as I am aware, been properly researched. Claims about what happens in this area accordingly have to be treated with a certain amount of care. I shall present here some constraints which seem to me to hold in such instances of co-ordination, though a proper empirical study would be welcome.

The first rather broad constraint seems to be that co-ordination is possible only where the nouns co-ordinated are in the same domain. *Cat and dog shows* is unobjectionable, *antique and dog shows* sounds odd. The difficulty of co-ordination with *buttercup* is thus that because of the idiomatic (or at least nonliteral) reading of both elements in *buttercup*, there can be nothing in the same domain to co-ordinate with it. Were *honeycup* also a flower, it might be possible to talk of *butter and honeycups*, but in the absence of such parallels, co-ordination with *buttercup* becomes impossible. This is, however, nothing to do with the compound status of *buttercup*, but rather to do with the idiomatic status of the lexeme and with the historical accident that no other type of flower has been named in the same way. Consider (7).

(7) He has been breeding buttercups of different colours for some years now. His most successful hybrid is a buttercup with a lambent golden brown colour, that he is marketing under the name of honeycup. He hopes that his butter and honeycups will prove a big success in the next few years.

The second constraint is that co-ordination of noun + noun constructions can take place only when the semantic relationship between the elements co-ordinated is

parallel. In *steel bars and weights* the relationship of 'B made of A' is constant, and the co-ordination is possible. But although *steel bar* and *steelworks* are both possible, *steel bar and works* is not. Perhaps more tellingly, consider *windmill* and *flourmill*. Here both words are established, and both have (to the extent that it is maintained) the same stress pattern. Yet

   (8) We saw a landscape dotted with wind- and flour-mills.

does not seem likely, because the semantic relationship between the two elements is different in each case, and this contrasts with (9), where there is a parallel relationship, and co-ordination is easy.

   (9) We saw a landscape dotted with wind- and water-mills.

In effect, this is a constraint banning zeugma from normal language use, and it applies far more generally than just to noun + noun constructions, as shown in (10).

   (10) (a) He's a good father and husband / ? accountant
        (b) ? He ran up the road and the flag.
        (c) ? She left in a hurry and in a taxi / in tears.

   The result of these constraints – which are no more than one might consider to be required by normal constraints on co-ordination – is that it is extremely difficult to find anything which might co-ordinate with one of the elements in some noun + noun constructions, and the more idiomatic the construction is (and thus the more likely it is to be listed) the less easy it is to find possible items to co-ordinate. But this is not necessarily a result of there being two different constructions, one of which allows co-ordination and one of which does not; rather it can be seen to be the result of general constraints on co-ordination and varying degrees of lexicalization.

   Unfortunately, co-ordination within noun phrases is more complex than even these examples show. Consider first sentences like the constructed example (11).

   (11) All of the chemistry and most of the physics teachers have agreed.

The striking thing about (11) is that although we might classify *physics teachers* as a noun + noun construction, the coordination seems to be between *All of the chemistry teachers* and *Most of the physics teachers*, where the determiners are associated with the head noun *teachers*, and not with the modifying noun. In (11), however, the determiner sequence occurs with the modifying noun *chemistry* and not with the head noun belonging to that modifier. This seems to imply a constituent analysis *[[All of the chemistry] teachers]*, although *[All of the [chemistry teachers]]* would be expected on semantic grounds. In other words, this is a bracketing paradox, at least on the surface. What is crucial for the discussion here, though, is that this bracketing paradox will not be solved by saying that *chemistry teacher* is a syntactic construction rather than a compound, or vice versa. Whatever the analysis of constructions like *chemistry teacher*, there will still be a bracketing paradox.

   Next consider a sentence like (12) (also constructed):

(12) The Australian and New Zealand High Commissioners met in London.

There are two points to note in (12). The first is that a nominal modifier can be co-ordinated with an adjectival modifier. This seems to be allowed because of the functional equivalence of *Australian* and *New Zealand* in (12), rather than because of any formal match in categories. However, this can be interpreted as meaning that *New Zealand* in (12) acts like an adjective, and that any noun phrase with *New Zealand* as a modifier must be a syntactic construction rather than a compound. Not all adjectives can be co-ordinated with nominal modifiers as in this example. But whether an adjective of the appropriate class can be co-ordinated with a particular modifying noun or not seems to depend on the (non-)existence of appropriate lexicalized expressions rather than anything structural in the noun + noun colloca-tion. If it is not possible to co-ordinate an adjective with *movie* in *movie star*, it is because there is no adjective of the appropriate class that forms a fixed collocation with *star* and stands in the appropriate semantic relationship to it. Where such things are found, there is no problem: consider *medical and life insurance*, for instance, or the following examples from Bache (1978) and the Wellington Corpus of Written New Zealand English (Bauer, 1993).

(13) visual, auditory, gustatory or pain perception (cited by Bache, 1978: 86)
(14) sensory and motor ability (ibid.)
(15) marketing their active and leisure wear (H30 274)
(16) dental and food technology (H02 216)
(17) affected by economic and business conditions (A31 154)

The second point that needs to be discussed in relation to (12) is the occurrence of the adjective *High* between the nominal modifier and the head. Again, this makes it look as though the nominal modifier and the head are not as closely bound as would be expected, and thus that this might be a syntactic construction. But if it is a normal syntactic construction, the order of the adjectives *Australian* and *High* is very unusual, and requires an explanation. The explanation that *High Commis-sion(er)* is a listed noun phrase which accepts modification only globally is a justification for the unexpected ordering of modifiers in a syntactic construction, but is an equally good justification for the occurrence of *High* between *New Zealand* and *Commissioner*. In fact, examples like these, and also examples like *Ministry documents* and *Ministry of Education documents*, suggest that it might be possible to improve on the definition of a compound, to make it a combination of two (or more) listemes to form a new lexeme. There are some disadvantages with such a suggestion, but it would explain a number of apparent oddities in the construction of compounds.

2.6   The head in a phrase can be replaced by *one*

If we go back to our steel bars, we can say

(18) I told you to bring me a steel bar but you have brought me an iron one.

The construction with a modifier like *iron* or *steel* is exactly the same as if we had used adjectives like *red* or *blue*. Replacement of the head noun in a construction by *one* looks like a syntactic operation, and might be expected to be impossible within a (lexical structure like a) compound. It is certainly the case that we cannot carry out this type of operation with a compound like *buttercup* (*?I asked for a buttercup, but you brought me a margarine one*), but as was pointed out above, *buttercup* has non-compositional semantics, and the same force which prevents co-ordination with the *butter* element in *buttercup* militates against replacing the *cup* element with *one*. Instead we have to look for noun + noun constructions with rather more literal readings.

It turns out that coordination involving these is rare. In the million words of the Wellington Corpus of Written New Zealand English, there are only seven clear examples. Jespersen (1914: 317–18) lists rather more examples, commenting on the recency of the construction. Perhaps for these two reasons – that the construction is still gaining in popularity, and that it remains rather unusual – it is not always clear what is and is not possible. Consider the following examples:

(19) (a) There were mills dotted all over the landscape, watermills and wind ones.
    (b) There were mills dotted all over the landscape, windmills and water ones.
    (c) It was an industrial landscape, with flour mills and sugar ones.
    (d) It was an industrial landscape, with sugar mills and flour ones.
    (e) There were all kinds of mills there, including flour ones.
    (f) There were all kinds of mills there, including water ones.

(20) (a) Do you want a table-spoon or a tea one?
    (b) Do you want a teaspoon or a table one?
    (c) Do you want a serving spoon or a table one?
    (d) Do you want a table-spoon or a serving one?
    (e) Do you want a teaspoon or a jam one?
    (f) Do you want a dessert spoon or a soup one?

(21) (a) He wanted a riding horse, as neither of the carriage ones would suffice.
    (b) He wanted a rocking horse, because the clothes one was too flimsy.
    (c) He wanted a carriage horse, since he said the cart one was not fine enough.
    (d) He wanted a race-horse, because a saddle one wouldn't go fast enough for him.
    (e) Shall we look at the shire horses or the plough ones?
    (f) He made both saw-horses and clothes ones.

(22) (a) I wanted a sewing machine, but he bought a knitting one.
    (b) I wanted a sewing machine, but he bought a washing one.

It is not clear in (19)–(22) which if any of the sentences is acceptable (although (21a) is attested), and whether there are differences of acceptability between the sentences. It seems likely to me that these sentences do differ in acceptability, which implies that any distinctions being made are not entirely of a grammatical nature. This is particularly true of the distinction in acceptability between (22a) and (22b), which does not seem to correlate with any other linguistic difference between the two collocations. In my list of attested examples, most of the examples where *one* is used

Table 1 *Non-correlation of criteria*

Collocation	Orthography	Stress	Listed	Pre-mod.	Co-ord 1	Co-ord 2	*one*
dinner plate	P	C	C	C	P	C	P
finger nail	P	C	C	P	P	C	C
gimlet eye	P	P	P	C	C	C	C
housewife	C	C	C	C	C	P	C
love-force	H	C	P	P	P	P	C
origami bird	P	P	P	C	C	P	P
pepper salad	P	P	P	P	P	C	P
rescue worker	P	C	C	C	P	P	C
sewing machine	P	C	C	C	P	P	P
truck tyre	P	C	P	P	P	P	P

have even/final stress (where this can be certain) and most of them are written as two words, but we have already seen that these criteria are not coherent in determining two categories of noun + noun collocation.

3   Correlating the criteria

I have tried to argue above that none of the possible criteria gives a reliable distinction between two types of construction. The implication is that any distinction drawn on the basis of just one of these criteria is simply a random division of noun + noun constructions, not a strongly motivated borderline between syntax and the lexicon. However, this type of argument may fail if all of the criteria, insufficient as they are in isolation, correlate with each other to define two major types. In this section I shall therefore consider the extent to which there is a correlation between the various criteria discussed above.

Consider first a simple table with information given for a number of noun + noun collocations and the various criteria that have been mentioned above. Co-ordination is split in two, to allow for co-ordination of first elements (*iron and steel bars*) or co-ordination of second elements (*iron bars and ingots*). 'C' in a table entry says that this item operates like a compound with regard to this particular criterion, 'P' that this item operates like a phrase (syntactic construction). Table 1 presents the relevant information, with the proviso that some of the judgements may be controversial (see the earlier discussion about the difficulty of deciding with *one*, for instance, or the variability in stress judgements).

Those noun + noun collocations in table 1 which are not item-familiar are attested. The items given are not randomly selected. Rather they are chosen to show different patternings with the more syntactic tests (the last four). Ten of the sixteen possible combinations of results for these four criteria are shown. Some of the missing ones are almost certainly just gaps in my data, though some of them may be systematically excluded (for example, *one* is not likely to be possible with many

collocations which do not allow co-ordination of the first elements). This is despite the fact that the last four criteria attempt to measure more or less the same thing: syntactic isolation or involvement. The first three criteria can also be seen as measuring the same thing: lexicalization or lack of it. There is no great agreement here, either. We even find forms like *mankind* which are orthographically a single word, but have stress on the second element. While there are items like *buttercup* and *myrtle crown* which are consistently 'C' or 'P' (respectively), these are not necessarily the majority of forms. Moreover, it is clear from the attested combinations that it is not the case that some criteria simply establish subsets of other criteria.

However, the answers provided above come from selected data. It might be that if the data were not carefully selected, a different pattern would emerge. To test this, the following experiment was carried out. A book was chosen at random from my shelves (Kate Millet's *Sexual Politics*), and a die was shaken to give a random number between one and six. The number that came up was five. Then the first noun + noun collocation on every fifth page was noted, until a sample of fifty had been collected. In a second, parallel experiment, noun + noun collocations were chosen from journalistic texts (*Time*, New Zealand edition, for 21 April 1997 and *The UK Mail*, the international edition of the British *Daily Mail*, for 18 June to 24 June 1996), again giving a total of fifty examples, the first noun + noun collocation from alternate pages being chosen. The seven tests were then applied to the examples collected. Results for each of the tests were compared pairwise to see if the two agreed or not. The null hypothesis is that any two measures have half a chance of agreeing, and the alternative hypothesis is that their probability of agreement is more than half. The results are provided in figures 1 and 2, where the number of agreements in each cell out of a possible fifty is given. Results which show a significant test result on a goodness-of-fit test at the 5 per cent level are marked with an asterisk, those which show a significant test result at the 1 per cent level are marked with two asterisks.

Now, while these figures do not show a complete lack of correlation between the various tests, neither do they show any really strong agreement among the tests in general. They appear to show that the source of data may have an effect on the degree to which the various tests give similar results, although this may be a result of relatively low numbers. However, the places where there is most obviously correlation are between the spelling and the possibility of co-ordinating the first elements, and between listedness and the syntactic tests. On the face of it, both of these are surprising, since both appear to show a correlation between something which is generally thought to be mainly a matter of item-familiarity and something which might be deemed syntactic behaviour. It has already been pointed out that the spelling variable often depends on things such as word-length, which would not be expected to correlate with syntactic behaviour.

I would like to suggest that these figures are unconvincing evidence in either direction, but certainly fail to show the overwhelming agreement between criteria we might expect if there were genuinely two different types of construction. I would also

	Ortho-graphy	Stress	Listed	Pre-mod.	Co-ord 1	Co-ord 2	one
Orth.		25	28	25	45**	35**	32*
Stress			32*	29	23	28	22
Listed				33*	27	35**	28
Pre-mod.					26	31	23
Co-ord 1						35**	31
Co-ord 2							29
one							

Figure 1   Number of agreements, test results, data from *Sexual Politics*

	Ortho-graphy	Stress	Listed	Pre-mod.	Co-ord 1	Co-ord 2	one
Orth.		23	28	28	38**	31	29
Stress			24	21	24	33*	31
Listed				39**	30	34*	32
Pre-mod.					30	30	29
Co-ord 1						29	31
Co-ord 2							31
one							

Figure 2   Number of agreements, test results, journalistic data

suggest that the correlations with listedness are probably correlations with idiomaticity rather than with listedness *per se*.

4 Prototypes?

Given that we have two labels (compound, syntactic construction) and a number of criteria which distinguish between these, albeit in an inconsistent manner, it might be argued that what we have are two distinct categories which are not clear-cut categories but are prototypical categories. Consider for example the following passage from Koskenniemi (1991: 64), which appears to make this kind of point:

> So far, no universally accepted criteria of compound status have been proposed. In particular, sequences of substantive + substantive can in borderline cases be taken either as a compound or an attributively used substantive qualifying another substantive.

This kind of argument is regularly used about the distinction between inflection and derivation, for instance (Plank, 1994, presents a good example). This notion is lent some credence by the fact that we have hierarchies such as that presented in section 2.1 above, which seem to be hierarchies of degree of lexicalization. Similarly, Bybee (1985) presents a continuum of degrees of fusion in which inflection marks the least fusional end of the scale and separate lexemes mark the most fusional end, a continuum into which compounding can be fitted. What we have to recognize is that the distinction between the two types of noun + noun collocation is not a matter of fusion or listedness of the types illustrated by these hierarchies. Listedness is orthogonal to the syntactic/morphological distinction which is being sought here, and fusion is linked to scope and to generality, neither of which is clearly at issue in any attempt to distinguish compounds from syntactic constructions.

If there are two prototypical categories we would expect to find absolutely clear-cut differences in the prototypical cases. Even if we make the assumption that 'B made of A' constructions like *iron bar* are prototypical syntactic constructions and idiomatized expressions like *buttercup* are prototypical compounds, the whole line of argumentation that has been presented here is that it is difficult to support these distinctions in terms of morphological and syntactic behaviour. Moreover, if we take Plank's (1994) discussion of inflection and derivation as a model, we might expect to find a large number of criteria agreeing on where the line should be drawn in individual cases. As has been seen above, this is not the case here.

5 Wordhood

In Bauer (1978, 1983a) I argued that unity-stressed and double-stressed noun + noun collocations do not differ in terms of general criteria for wordhood. Here this argument needs to be taken further, since stress is not the only possible criterion which may be taken to distinguish compounds and syntactic constructions.

Any definition of the concept 'word' is controversial, and it is questionable whether much progress in this field has been made since the Sixth International Congress of Linguists (Lejeune, 1949). Since we are discussing the word as a morphosyntactic unit here (as a semantic unit we would need to discuss listemes rather than words), internal cohesion is generally accepted as providing the major criterion for wordhood. This can be divided into positional mobility (the word is moved as a unit), uninterruptability (external items cannot be freely inserted within a word) and internal stability (elements within the word do not show contrastive ordering and have a fixed order) (see e.g. Lyons, 1968: 202–4). For various reasons, the crucial one of these three is uninterruptability (and this is the only one used by some other linguists).

Since we are here concerned with morphosyntactic structure rather than semantic structure, it is possible for a phrasal verb such as *look up* (the answer) to be defined as being made up of two words by this criterion, because it is regularly possible to interrupt the construction with 'rearranged material from elsewhere within a given sentence' (Matthews, 1972: 98 fn. 2) – at least if we take it that a noun phrase realized as a pronominal direct object starts life to the right of the *up* element. In the same way, there appear to be many noun + noun constructions which can be interrupted in a similar way. We have already seen an isolated instance of this in *New Zealand High Commissioner*. The phenomenon is more widespread (my thanks to Andrew Carstairs-McCarthy for drawing this to my attention). Consider, for instance, the following examples from the Wellington Corpus of Written New Zealand English:

(23) to establish district high schools (J57 134)
(24) Recent government economic policy has been dominated . . . (H01 151)
(25) The 323 LX includes a rear hatch remote release (E02 230)
(26) radio religious broadcasting (D08 267)

In most cases we seem to be faced with an adjective + noun construction which might plausibly be viewed as listed (as was suggested in section 2.5). If a compound was defined as a lexeme made up of two or more listemes (independent of the status of the listemes, so there is no claim that *religious broadcasting,* for instance, is a lexeme), this would account for many such examples. But it would not account for all of them. Thus, while *old-fashioned car* might be listed enough to account for *toy old-fashioned car*, no such excuse can be made for *toy old-fashioned cash register.* The ordering of pre-modifiers in the English noun phrase is a vexed question (see e.g. Bache, 1978), but in the terminology of Bybee (1985), we can perhaps say that it appears to be related to relevance: modifiers closer to the head noun are more relevant than more distant modifiers, and more distant modifiers have scope over all modifiers which are closer to the head than they are. Given this, the order of modifiers can be manipulated by the speaker as required to fit the constraints. In the vast majority of cases this will lead to nouns being the most relevant (and closest) modifiers. Occasionally, where a listeme, which may be idiomatic, has an adjectival

modifier, the most relevant modifier will no longer be the nominal modifier. Sometimes, there will be a reason for the most relevant modifier to be something other than a noun when a nominal modifier is also present. The question which is of interest here is whether the constructions in which this can happen are of a different type from the constructions in which it appears not to be possible. If they are all being determined by the same set of principles, there is no reason why a distinction of this kind should be indicated.

Now, it is true that it appears to be impossible to interrupt *buttercup* in any way at all. The reasons for this have been seen earlier, and are to do with the idiomaticity of that item rather than its grammatical structure. Nonidiomatic, forestressed, orthographically cohesive noun + noun constructions can be interrupted in the same way where it is pragmatically useful. Consider (27).

(27) I was told to cut down on the amount of snuff I was taking. So I got a special snuff-box made for me. I had to put a coin in every time I wanted a pinch of snuff. I thought it would make the snuff so expensive that I would stop taking it, but I've discovered that my snuff money-box is just saving up for the next tin of snuff!

What we appear to have is not 'free' interpolation of modifiers, but very limited interpolation of modifiers when a new relevant combination is required. Consequently, this kind of interruption fails to show that noun + noun constructions are not morphosyntactic words.

6   Discussion and conclusions

We have seen that there is a common perception that noun + noun constructions fall into two categories in English: the morphological (compounds) and the syntactic (syntactic constructions). A number of criteria have been considered, and it has been suggested that most of these fail to give a distinction which really reflects a morphological versus syntactic division, and that the various criteria do not correlate with one another to provide a coherent division into two categories. It has been suggested that what the criteria really represent is two distinct things: first there are those criteria which say something about degree of lexicalization; second there are those criteria which say something about degree of syntactic availability of the individual elements in the collocation. The first set of these ought not to be relevant to drawing a distinction between morphology and syntax, since the lexicalized/ nonlexicalized distinction is orthogonal to the morphology/syntax one. The second set, though, looks as though it should be relevant.

I would like to conclude that this second set of criteria would be relevant ones if there were indeed two categories to distinguish. However, the criteria delimit different subsets of noun + noun collocations in ways which do not seem to be related to grammatical behaviour, but rather to be related to social history: we can have *shoulder and handbags* because both *shoulder bags* and *handbags* exist; we cannot have *handbags and trays* because there is nothing in our society lexicalized as

a *hand-tray* which is in the same domain as a handbag and with the same semantic relationship between the elements as is shown in *handbag*.

Unfortunately, this leaves the contrast illustrated in (28) and (29) unexplained.

(28) I went on a tour round the 'toy factory where they make Lego.
(29) We played with the toy village: there was a toy church, a toy post office, even a toy 'factory where all the dolls worked.

If a single class of compounds is to be claimed, then the possibility of contrast within the class will have to be explained. If I wish to deny the existence of two categories, I have to explain the existence of contrasts within the single category. I think this can be done in terms of the 'avoid synonymy' principle. It is a well-known finding in diachronic lexical studies that near-synonyms tend to become specialized, and thus distinguished. What we have in the case of (28) and (29) seems to me to be the same phenomenon, though perhaps on a slightly larger scale. That being the case, though, the distinction is nothing to do with the two types belonging to different constructions, although if there were sufficient such examples they could lead to constructions being distinguished.

It seems likely to me that the desire to analyse two sets of constructions here arises from what are largely irrelevant origins: the fact that English happens to write some noun + noun collocations as one word and others as two and the fact that English has – albeit variably – two available stress patterns for such collocations. I suspect, but cannot prove, that these things arise from the fact that English is a Germanic language strongly influenced by a Romance one (namely French), and that the current situation is a blend of conventions from the two sources. While such a blend could give rise to two distinct sets of constructions, I remain unconvinced that there is hard evidence that it does so.

But if there is only one construction, the question remains open as to whether it is a morphological one or a syntactic one. In previous work I have called such constructions 'compounds', perhaps influenced by the Germanic model (in Bauer, 1978, I was comparing 'compounds' in English and Danish). In many ways, it would be just as satisfactory to call all such constructions N-bars, and this would mark the parallels between noun + noun constructions and non-predicate adjective + noun constructions commented on by Levi (1978). I have avoided comment on adjective + noun constructions in this paper, because they complicate the picture considerably. A consideration of the parallels and distinctions between noun + noun and adjective + noun constructions would be another paper. The major factor that makes the compound solution seem the more attractive one is the parallel with forms like *hovercraft*. Given that *hover* is a verb (in the same sense that *butter* in *buttercup* is a noun), *hovercraft* is not a noun + noun construction, and its analysis has not been discussed in this paper. It is, though, generally recognized as a compound. I know of no analyses in which *hover* is seen as the modifying element in an N-bar. It seems to me that the parallels between *hovercraft* and *landing craft* or *water-craft* or *surface*

craft are such that they should probably all be seen as belonging to the same general construction type.

The conclusion I reach here, that there is no strong evidence for a distinction between two fundamental types of noun + noun construction, is much stronger than the one I put forward in Bauer (1978, 1983a), where I considered only stress as a possible distinguishing criterion between two classes. Nevertheless, I believe that this conclusion is justified by the available evidence, and that until such time as we find that there is a consistent way of drawing a distinction, Occam's razor should encourage us to work with a single category. Recall, however, that this conclusion does not rule out the possibility of a case for two types of construction subsequently being made, since – as was stated at the outset – I cannot prove a negative. If such a case is subsequently made, it will have to take into account the objections and problems that have been discussed here.

Author's address:
School of Linguistics and Applied Language Studies
Victoria University of Wellington
PO Box 600
Wellington
New Zealand
laurie.bauer@vuw.ac.nz

References

Asher, R. E. (1994). *Encyclopedia of language and linguistics.* Oxford: Pergamon Press.
Bache, C. (1978). *The order of pre-modifying adjectives in present-day English.* Odense: Odense University Press.
Bauer, L. (1978). *The grammar of nominal compounding.* Odense: Odense University Press.
Bauer, L. (1983a). *English word-formation.* Cambridge: Cambridge University Press.
Bauer, L. (1983b). Stress in compounds: a rejoinder. *English Studies* **64**: 47–53.
Bauer, L. (1988). *Introducing linguistic morphology.* Edinburgh: Edinburgh University Press.
Bauer, L. (1993). *Manual of information to accompany the Wellington corpus of written New Zealand English.* Wellington: Department of Linguistics, Victoria University.
Bybee, J. (1985). *Morphology.* Amsterdam and Philadelphia: Benjamins.
Chomsky, N. & M. Halle (1968). *The sound pattern of English.* New York: Harper & Row.
Dierickx, J. (1970). Why are plural attributives becoming more common? In Dierickx, J. & Y. Lebrun (eds.), *Linguistique contemporaine.* Bruxelles. 39–46.
Di Sciullo, A.-M. & E. Williams (1987). *On the definition of word.* Cambridge, MA: MIT Press.
Downing, P. (1977). On the creation and use of English compound nouns. *Language* **53**: 810–42.
Fabb, N. (1994). *Sentence structure.* London and New York: Routledge.
Fudge, E. (1984). *English word-stress.* London, etc.: Allen & Unwin.
Gleitman, L. R. & H. Gleitman (1970). *Phrase and paraphrase: some innovative uses of language.* New York: Norton.
Jespersen, O. (1914). *A modern English grammar on historical principles,* part II: *Syntax,* vol. 1. Copenhagen: Munksgaard; London: George Allen & Unwin.

Jespersen, O. (1942). *A modern English grammar on historical principles*, part IV: *Morphology*, vol. 1. Copenhagen: Munksgaard; London: George Allen & Unwin.

Koskenniemi, I. (1991). The creation and use of compound words in Jacobean tragedy. *Neuphilologische Mitteilungen* **92**: 63–73.

Ladd, D. R. (1984). English compound stress. In Gibbon, D. & H. Richter (eds.), *Intonation, accent and rhythm*. Berlin and New York: de Gruyter, 253–66.

Langacker, R. W. (1987). *Foundations of cognitive grammar*, vol. 1: *Theoretical prerequisites*. Stanford: Stanford University Press.

Lees, R. B. (1960). *The grammar of English nominalizations*. The Hague: Mouton. 5th edn 1968.

Lejeune, M. (ed.) (1949). *Proceedings of the Sixth International Congress of Linguists*. Paris: Klincksieck.

Levi, J. N. (1978). *The syntax and semantics of complex nominals*. New York, etc.: Academic Press.

Liberman, M. & R. Sproat (1992). The stress and structure of modified noun phrases in English. In Sag, I. A. & A. Szabolcsi (eds.), *Lexical matters*. Stanford: CSLI. 131–81.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.

Matthews, P. H. (1972). *Inflectional morphology*. Cambridge: Cambridge University Press.

Mutt, O. (1967). Some recent developments in the use of nouns as premodifiers in English. *Zeitschrift für Anglistik und Amerikanistik* **15**: 401–8.

Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge, MA, and London: MIT Press.

Plank, F. (1994). Inflection and derivation. In Asher, R. E. (ed.), *Encyclopedia of language and linguistics*, vol. 3: 1671–8. Oxford, etc.; Pergamon.

Tulloch, S. (1991). *The Oxford dictionary of new words*. Oxford and New York: Oxford University Press.

Vos, A. L. (1952). Stress in English compound nouns. Unpublished Diploma Dissertation, Department of Linguistics, Edinburgh University.

Ward, G., R. Sproat & G. McKoon (1991). A pragmatic analysis of so-called anaphoric islands. *Language* **67**: 439–74.