

# A Corpus-Based Study of Compounding in English

LAURIE BAUER

*University of Liverpool and Victoria University of Wellington*

ANTOINETTE RENOUF

*University of Liverpool*

It is long established that corpus-based studies force the linguist-analyst to come face-to-face with a number of phenomena that might easily be overlooked in an armchair-type study. In this article, we demonstrate the validity of this truism once again in a study of English compounding patterns. We report here on a study of word-formation patterns in words from a large corpus of British newspaper English. In this article, we consider only new compound formations, and we show that considering real data can cause problems for the theoretician of word-formation and for the descriptive grammarian alike. Not only do we report on patterns that are not described in the major handbooks (Adams 1973; Bauer 1983; Marchand 1969), but we show that some of the patterns that are being used productively in the English of the early 1990s break principles that are laid down as absolute in some of the theoretical works.

## The Data

The source of data for this study was a large collection of journalism, from the U.K. daily broadsheet newspaper the *Independent*, covering the period from 1988 to 1998. The total text amounted to more than 360 million words. Analytical tools developed in the AVIATOR project (Renouf 1993) and the ACRONYM project (Renouf 1996; Collier and Pacey 1997) were used to extract the new words occurring in each quarter of this ten-year period. Renouf and Baayen (1998) had carried out a manual classification of the neologisms in the last quarter of 1995, around

AUTHORS' NOTE: Some of the material in this article has been presented at seminars at the universities of Auckland, Freiburg, Heidelberg, Leeds, Munich, Paris VII, and Reading, and we are grateful to the participants at those presentations as well as to the referees for *JEngL* for their comments and feedback, many of which have led to improvements (in the forms of additions and omissions) to the current text.

nine million words of text. In our current study, the neologistic compounds we focus on are the lower-case initial ones identified in that analysis. These amount to approximately 3,000 formations, both hyphenated and unhyphenated.

A word is deemed to be new when it appears for the first time in the chronologically stored, cumulative database. Many words classed as neologisms by this criterion will indeed be new: ad hoc coinages for the delectation of news readers or new terms denoting new real-world concepts, things, people, and events. Others may simply be rare, previously dormant but now reviving, unusual spellings or plurals, possessives of known words, or typographical errors.

## Problems in the Description of English Compounds

### The Headedness Rule and Its Exceptions

One of the basic principles of compounding in English is that English compounds are right-headed. This is fundamental, not only to Williams's (1981, 248) right-hand head rule (which has become notorious for its overstatement of the principle) but to most treatments of English compounds. For example, Allen's (1978, 105) ISA condition states that in English, compounds are hyponyms of their right-hand element and that they behave like their right-hand elements. Marchand (1969, 54) speaks of the determinant/determinatum structure of English compounds. For Lieber (1983, 253), it is the percolation of category features from the right-hand pair of sister nodes that gives a compound its category. Thus, given a compound like *godchild* in which the right-hand element is *child*, we predict that it denotes a subcategory of *child* rather than a subcategory of *god*, we predict that it makes its plural in the same way as *child* does, we predict that it is a noun because *child* is a noun (and the category of *god* is irrelevant to this question), and we predict that *god* modifies *child* in some way. There are a number of well-known and mainly extremely marginal exceptions, discussed briefly below.

The major class of exceptions is clearly provided by exocentric compounds. There are several types, some better described than others.

1. The first major subtype is made up of Bahuvrihi compounds like *redskin* (a person, a potato, or an apple) or *egghead* (a person), where the entity denoted has the feature described by the compound attributed to it (explaining the alternative terminology, "possessive" compounds). Although these compounds have denotata which are external to the structure of the compound, the remaining compound is nonetheless right-headed: that which a *redskin* has is a skin, which may be further described as red. They might be viewed as compounds of a type whose underlying form, here  $[[\text{red} + \text{skin}(\text{ned})] [(\text{potato})]]$ , is rarely found on the surface of a perfectly regular structure.

2. The second subtype is made up of lexicalized compounds such as *pickpocket* (Marchand 1969, 380-81). Again there is no overt head (contrast the modern construction *pocket-picker*, which demands the overt head *-er*), but the structure is harder to comprehend from a synchronic viewpoint. There appears to be some kind of incorporation of the direct object into the verb, but the order is unexpected. The type appears to be fundamentally a Romance one (consider French *porte-manteau*, Italian *portamantello*, etc.), which had a brief life within English and has now vanished again. Given the lack of productivity of this type, perhaps we need not give it too much attention here.
3. A third major subtype is made up of various types of compound (or apparent compound) involving particles (a label we will use to avoid commitment as to whether these things are really adverbs, prepositions, or some separate class). *Input* and *put-down* represent two contrasting patterns, differing in order and in productivity. Although both are nouns, there is no noun in their makeup. These are probably best seen as being derived from the relevant verb + particle grouping (*put in*, *put down*, respectively) with inversion in the one case and a stress shift in the other indicating their status as nouns. We might treat this second type as equivalent to words like *import* or *discount*, which (in many varieties at least) have verbal and nominal forms distinguished by the position of the stress.
4. There are also a handful of compounds, usually used as premodifiers but sometimes used independently, whose syntactic functions are not easily equated with the word-classes of their individual elements. Some of these are discussed in Bauer (1983, 202-13), for example, *passlail* (test), *nosebleed*, *before-tax* (profits), and *quick-change* (artiste). Some of these, but not all, may be seen as lexical uses of syntactic groups; others appear to be genuine instances of word-formation with unexpected grammatical characteristics.

A second type is made up of phrasal items, which we might characterize as being "compound like" in that they are listemes (Di Sciullo and Williams 1987) and perceived as single words by naive speakers. Some of these are simple lexicalizations of syntactic structures. Others are overtly left-headed.

1. Compound phrases that are lexicalizations of a head noun with a following prepositional phrase complement: *lady-in-waiting*, *mother-in-law*, and so on. These are hyponyms of the initial element and according to standard reference works mark plural on the first element, so are clearly left-headed. They are left-headed because they are not yet fully integrated into the lexical system but retain traces of their syntactic origins. The alternation between *spoonsful* and *spoonfuls* (with the latter being the innovative form) shows that reinterpretation of such constructions in line with lexical principles is possible (see Katamba 1993, 317). Indeed, the same comment could be

made with reference to *mother-in-law*, which is often pluralized as *mother-in-laws*.

2. French loan constructions, which retain French word order: *attorney general*, *notary public*. Again these are clearly left-headed, but the reason is clear: these constructions are not fully integrated into the English system. This particular construction does not appear to be productive.
3. A miscellaneous group of expressions with no overt head that appear to be lexicalizations of phrasal structures (see Bauer 1983,207): *has-been*, *love-lies-bleeding*, *forget-me-not*, *gofer*. As is shown by *gofer* ('person told to go for this or go for that') or *twofer* ('two for the price of one'), new forms are created using this technique, but they are by their very nature unpredictable formations.

There are two other major left-headed patterns:

4. Verb + particle constructions such as *pass by* are left-headed in that they take their inflections on the left-hand element and are hyponyms of their lefthand element. Derived from these we also get constructions such as *passerby* (Katamba 1993,316) and *passing-by*, which take their left-headedness from that of the root verb. While *passer-by* is well enough established not to show variability, other formations may be treated as right-headed by speakers in that they add the *-erto* the right-hand end of the word to give forms like *paper-over-er* (not in our sample but in the total corpus), *blower-upper* (see Bauer 1983, 71), or even *blower-upperer* (see Round 1998). These nonstandard deverbal forms seem to show that native speakers of English are unhappy with forms that are not unambiguously right-headed in English.
5. In rather conservative English, there is a small class of compounds with *who* in the first element that requires that form to inflect for its role in the sentence: *whoever*, *whomever*, *whose-ever* (and similarly, forms of *whoso* and *whosoever*) (Bauer 1994,1531).

Given the small number of such cases and the recessive nature of some of the types, we would not expect to find any great proliferation of types that break with the normal pattern of right-headedness. It is thus surprising that a number of such patterns are found in our data. These are discussed in the next section, which opens the report on our study.

#### Classes of Wrong-Headedness

Our first observation concerns those items that are used as premodifiers but whose morphological makeup does not show this. The type is not new: it was discussed with examples such as *passlail* (test), *nosebleed*, *before-tax* (profits), above, and the whole question is discussed in some detail in Bauer (1983,210-12).

The main point that needs to be made about this type of construction is that theorists usually ignore it when dealing with the patterns of compounding.' While we would not wish to claim that the patterns illustrated below are of major numerical importance in the formation of English compounds, they are frequent enough in our data to require some treatment. A few examples will suffice to show the kinds of things that are found.

"He has a very *kick-arse* attitude," one says. "He'll say, 'right, here's the strategy. Go do it.' "

The turkey doesn't need to roast forever, either. Current culinary creed favours cooking it on a higher heat for a shorter time rather than the 10-hour *slow-cook* marathon that usually means some poor soul-invariably Mumhas to stagger out of bed at 5 a.m.

The Biro hit the shops of Britain 50 years ago. In the run-up to Christmas 1945, this was no el cheapo *chuckaway* item costing a few pence but a luxury purchase. At 55 shillings (pounds 2.75), it cost the weekly wage of a secretary.

It is perhaps worth drawing attention to one rather peculiar subtype that appears not to have been noted previously. This is the type that is treated grammatically as being headed and yet fails the hyponymy test because membership of the class is specifically denied within the word. These are usually derivatives with prefixes such as *non-*, *anti-*, *pseudo-*, and so on, but an occasional compound is observed to creep into the same grouping:

But a number of *no-drug* behaviours are regarded by some authorities as addictive, including gambling and playing computer games.

The semantics of this type of formation is complex, and we will do no more here than make a few points about what is going on. In calling someone a *nonperson*, for example, we are superficially denying that they belong to the category of person; on another level, we are using the element *person* to indicate that 'person' is an appropriate categorization, and we are using *non-* to deny that the person denoted has sufficient of the prototypical characteristics to fit neatly within the categorization while using it to attribute a specific quality, here of political unacceptability, to that kind of person. "That person is a nonperson" is thus not (or no longer) a contradiction in terms, although it appears that it should be. (For further discussion of other aspects of the semantics of *non-* in particular, see Algeo 1971.) With the more ad hoc formation *no-drug* cited above, the function of *drug* appears to indicate to the reader the appropriate lexical domain, or general topic area, within which some specific deviance occurs, while the choice of the prefix *no-* seems to avoid the se

mantics that have come to be associated with *non-* and to meet the discoursal and grammatical requirements of the preceding context:

"the perception of addiction is that there has to be a drug," he says. "But a number of *no-drug* behaviours are regarded by some authorities as addictive."

That is, *no-* licenses the exact repetition of the topic word for the purposes of lexical cohesion while being a grammatically more conventional negativizer for a noun.

Higginbotham (1985, 567) discusses similar structures briefly, suggesting that *person* or *drug* in the examples given above are instances of "mention" rather than "use." If by this he is trying to identify their metalinguistic function, we agree, our interpretation of this use being that it is a matter of discourse strategy (we know that the person being discussed is a person, so to term them a *nonperson* breaks a Gricean principle and has to be interpreted as being informative).

A much more serious type of unheaded construction in our data-if only numerically-is provided by those compounds that have the form of nouns but are used adverbially. Some examples are given below:

in the first stage, the legs are drawn up under the body. They are 'cocked,' ready to kick outwards and backwards, *frog-fashion*. Kick with flexed feet, using the soles like paddles.

It's just that rather than wring their hands (see Rudolph or Trapped in Paradise or Santa Claus: The Movie-or better yet, don't) they'll wring your neck instead. *Turkey-style*. That'll teach you to take the birth of Christ in the spirit intended.

He is a wild gesticulator, arms going *windmill-style* to emphasise a point.

Eventually the manager, Charlotte Hindle, suggested that the distributor in Geneva, which does not have a shop window, hang a banner, *bunting-style*, in the street.

These items are more frequently used as premodifiers, and the adverbial usage raises the question of whether they are actually perceived as being adjectival, rather than simply as nominal, premodifiers. If they are, then we may have a situation where exocentric compound adjectives are being created, and the pattern may well be new. If not, then it may simply be the adverbs that are exocentric.

Clearly, their independent lexical meanings aside, *-fashion* and *-style* have a peculiar ability to be paraphrased as "in the manner of an X" (unlike say *-type* or *-like*, which are semantically similar) and thus to seem quite affix-like. There is some in

dications in the total corpus that they are grammaticalizing in such combining formations, *-style* more so than *-fashion*.

We will not discuss compound verbs or adjectives here since we have sections devoted to them below, but we note that some such types are also apparently unheaded from this point of view. See Renouf and Baayen (1998) and Baayen and Renouf (forthcoming) on the proliferation of certain similar types, which appear to be in the process of creating new affixes.

Finally, we find instances of a very common type of compound premodifier that is left-headed. This type involves compounds whose second element is *-only*, as in the examples below:

it provided a level of functionality that *data-only* services cannot provide. I hate all those *dry-clean-only* labels, such a waste of time and money.

the most rounded and effective of the night's *quartet-only* pieces.

Bates insisted on a *success-only* fee and the business relationship broke down.

Those desiring excitement and a result would have done better to attend a *singles-only* karaoke rather than a league game at Highbury.

These bikers are safety-conscious, perhaps married, often *leisure-only* riders. .

The market for *paging-only* products and services is unlikely to be immediately threatened.

Wafer-thin, organic brown flour, low-salt bran scones with a scraping of soya-based sour cream and a smidgin of *'fruit-only*, . no-sugar jam.

People wanting to watch a video on their computers will do so because it is part of a film or a game, and should be happy with a *software-only* solution.

If they all switched to other brands, such as the older *oestrogen-only* pill the number contracting thrombosis would fall.

men she had come to *sex-only* arrangements with had been unable to deal with not being in control of the situation.

In each of these cases, it seems to us that it is the left-hand element that is obligatory in the construction and provides a superordinate term for the compound as a whole: *oestrogen-only pill* is a subtype of *oestrogen pill*, and the construction *only pill* could not be interpreted in the same way. Furthermore, whatever the part of

speech of the premodifier, it is not the same part of speech as *only*, while it might be the same as *oestrogen*. All of this matches standard arguments for the left-headedness of the construction. Yet it has also been suggested to us that these constructions are not left-headed; rather than *oestrogen-only* denoting a subtype of *oestrogen*, it denotes a subtype of exclusivity-something that is, in this case, exclusively *oestrogen*.

There are thus two problems here: either we have a left-headed structure and it is not clear why we might have a left-headed compound, or we have principles for the determination of headedness that can be used to derive conflicting answers for the same data. In the one case, an explanation is desirable; in the other, a refined set of criteria, based on a better understanding of what headedness means, seems to be indicated.

One potential solution to the problem this raises has been suggested to us. It is that in constructions like *oestrogen-only pill*, we do not have a compound premodifier at all but a piece of syntax that has been captured to be a premodifier, in the same way that pieces of syntax are captured in examples like the following:

an if-you-really-want-to-know sneer (Ripley, Mike, *Angels in Arms*, New York: St Martin's Press, 1991, 10).

There is some evidence to support this. First, in an example like *dry-clean-only labels*, the premodifier can clearly be seen as a direct quotation from the label, and direct quotations are easily imported in this way. The second argument is more subtle and involves the semantics of the construction. If we say *It is only oestrogen*, the phrase is ambiguous between "it is merely oestrogen" and "it is exclusively oestrogen." If we want to make it clear that only the exclusive reading is possible, we have to say *It is oestrogen only*. The premodifiers, which demand the exclusive reading, always have the appropriate syntactic structure to guarantee this reading. The headedness of the structure derives from this rather unusual syntactic fact and is parasitic on it.

While it seems entirely plausible that this syntactic pattern is the origin of the type, it is not necessarily clear to us that the pattern has not now become one of word-formation. There are two arguments to suggest that this might be so. The first is the frequency of the pattern, despite the relative infrequency of the syntactic pattern. If syntactic patterns are simply being captured randomly, then we might expect to find more forms on the pattern *only-oestrogen* than on the actually attested pattern *oestrogen-only* since *only NP* occurs in syntax far more frequently than *NP only*. The second is that in syntax, the pattern arises with plural nouns preceding the *only* as often as with singular nouns (the comment is based on an analysis of relevant *only* use in the Wellington Corpus of Written New Zealand English), while in the hyphenated examples we cite above, the only possibly plural noun (and even

that is a little dubious) is *data*. Thus, it looks as though the grammatical conditions for the two may be slightly different.

### Compound Verbs

The next type to be discussed here is compound verbs. The general assumption in the literature is that the majority of compound verbs in English arise from back-formation or conversion (Adams 1973, 105). Marchand (1969,100) goes so far as to deny the existence of verbal compounding in English as a process, though Adams (1973,109) could be read as implying that the process is beginning to take hold. Marchand's comment may be seen as an uncharacteristic failure to distinguish between a synchronic structure and the diachronic process leading to that structure. Selkirk (1982,47) allows for the construction of compound verbs, but only for those containing a particle and a verb. While these are the most frequent type in our data, they are not the only type. Lieber (1992,58) says that there are no left-headed compound verbs in English (but what about the phrasal verbs?) and virtually no right-headed ones, the few exceptions being back-formations from synthetic compounds.

In many cases, it is difficult to tell whether a compound verb arises through back-formation, zero-derivation, or some third process. Consider the following:

It would be nice to see the new drive to *up-skill* the nation taking this into account but I doubt it will.

Is *up-skill* a case of back-formation from *up-skilling*, or is it an independent formation? It may not be possible to tell. However, as far as we can see, the following must be an independent formation:

And it must be said, OJ [Simpson] *outsoaped* the soaps.

The interesting thing about this construction—apart from its apparent status as a type of genuine compound verb—is that it is derived from a particle and a noun, with no trace of a verb. It is thus not a conventionally headed construction (see above). The type is not listed in Bauer (1983) but is not new; Hamlet speaks of "out-Heroding Herod." It is not clear here whether there is a covert reference to Shakespeare, but the pattern is quite productive (ninety-three instances) among the particle-proper name constructions across our ten-year database, less so with common nouns (six).

More interesting, but rarer, are those compound verbs that appear to be formed without back-formation or conversion. The various types are all listed in Bauer (1983), but what is interesting is to find them still being used to create new words. We may not be seeing any great increase in the use of verbal compounds created

without back-formation or conversion just yet, but the types show enough productivity for us to be able to deny Marchand's (1969) assertion that verbal compounding does not exist in English. Various types are presented below.

Plans have just been unveiled to install a new type of mobile toilet for climbers on Mount Fuji. According to reports it uses a kerosene heater to *dry-burn* human excrement, considerably reducing the volume of such waste.

This year something else came out of the oyster lunches: plans to *test-release* an unnamed French film in the UK in a dubbed version.

He slapped its body, he *thumb-strummed* zig-zag bristling, funk-rock riffs.

The Woking-based company Bourne, Europe's largest promotional product firm, is about to launch no fewer than 3,000 different mousemat fragrances, and will *custom-produce* other smells on request.

Another dilemma is whether to *slow-bake* or cook them at a high temperature.

Microbics grows the bacteria in *fermenters*, *freeze-dries* them and then stores measured amounts in plastic vials.

'We wanted to *hardwire* marketing and sales together, which unfortunately eliminates my role as head of central marketing.'

He also *part-runs* a unit trust company, Portfolio, which invests wholly in other investment funds.

He *mock-whispers* that his ideal relationship is the one between Stephen Rea and Jaye Davidson in The Crying Game.

The type illustrated above by *part-run* is quite common, though usually only in the past-participle form. This might, accordingly, be a case of back-formation. Other examples raise other points, but our purpose here is merely to show that compound-verb formation is alive in English. Note that our data include examples with nouns, verbs, and adjectives in the first element and with variable meaning relationships between the elements, just as we would expect to find with compounds.

## Compound Adjectives and Premodifiers

The major difficulty with compound adjectives is in recognizing them. The main problem is with the definition of an adjective, though, as we shall see, even the compound form may present problems. The definition of an adjective in English is a

problem in any case, the various criteria that are used for defining the category not always matching up (consider *former* and *afraid* as simple examples of this problem). The question arises most obviously for compound adjectives in those compound constructions that are used as premodifiers to nouns.

There are many of these that are not at all controversial. They are compound in form, have an overt adjective (or, equivalently, perhaps, a participle) in head position, permit submodification by adverbs, and so on. Some examples from our data will illustrate these patterns, which are well established.

Determined this time to avoid embarrassment (*sic*), criticism and sniping from the *rumour-sodden* TV industry, the ITC thought it had covered all the bases.

... a way of actually coming to some sort of practical agreement over the children being able to see both of us in a reasonably *confrontation-free* atmosphere.

The book also seeks to show that in a predominantly *evil-seeming* world, goodness endures and triumphs.

Although it is perhaps worth noting the extreme productivity of the *N-free* and *N-friendly* patterns in our data, there is nothing here to cause descriptive problems. The vast majority of the headed adjectival compounds found in our data contain participles in the head position. In other cases, the patterns listed in Bauer (1983, 209-10) are found with premodifying nouns (*tax-shy*, *forms-compatible*), adjectives (*Shakespearean-comic*, *folky-acoustic*, *Croat-Serbian*) and adverbs (*nearmonopolistic*, *under-mighty*), and some sound-motivated forms (*cryptic-mystic*, *gloopy-bloopy*).

At the other end of the scale, we find things that look like compound nouns being used as premodifiers.

Mr Curtis-Jenkins retorts: 'You do a *patient-satisfaction* study, and you'll find that it's the patients themselves who say they don't want to be sent off to a psychiatrist and stigmatised.'

... was taken in the light of the damage the coup was doing to the future of the archipelago, *Frenchforeign-office* officials said yesterday.

*Patient-satisfaction* and *foreign-office* are words that could stand alone as nouns, and it seems to us to make more sense to see the constructions they form a part of as multi word units (MWUs) of some type rather than as adjective + noun constructions. We term these constructions MWUs rather than compounds here because a compound is defined for us in this study as being written as a single word (possibly

including a hyphen). There are problems with our stance, but the problems seem to be more practical than theoretical. Consider, for instance, the following example:

Anyone with a B-cup or bigger should consider wearing two sports bras, a regular bra shape with a *cropped-top* style over the top. There are four bra styles to choose from in this high-performance range: three bra shapes and one *cropped top* covering three 'impact' levels.

Here we see the listener *cropped top* occurring as two words when it is not used as a premodifier but as one when it is used as a premodifier, and this is a well-known orthographic principle (if one rather inconsistently applied in our data). Strictly, on the orthographic definition of compound, these items are compounds only when used as premodifiers. We assume that this orthographic decision is one that has no real linguistic consequences.

However, this principle of grammatical hyphenation leads to the inclusion in our lists of a number of items that superficially appear to be acting as compound adjectives (or at least premodifiers) that may not only be spurious adjectives but also spurious compounds.

From Old Headington, Elizabeth Leyland sends me the stupendously useful Oxford University Pocket Diary, which runs on *academic-year* lines from September onwards, and features, for fed-up students, a useful map of the London Underground.

But deciding whether the company is in breach of US *false-advertising* laws could be tricky.

In the past, ITV has broadcast *extended-length* episodes of hit series such as Prime Suspect on successive evenings or over the weekend.

The Welsh trio led 10-3 after nine ends but the Scots rallied and forced their share of the spoils with a *final-end* single.

*Heavily-overcoated* crime reporters were the monarchs of all they surveyed.

In the cupboard, *heavier-gauge* steel-based enamel ware from British manufacturers includes an excellent 25-year-old cream and green oval pot-roaster and an indigo blue stockpot.

The seizure is the latest in a series of large-scale cannabis imports in recent years, taking advantage of Ireland's *largely-unpatrolled* 2,000 mile coastline.

The constructions here seem to be phrasal structures made to look like lexical structures by our methodology, but with no real claim to that status. Indeed, some of the adjectives that occur in these constructions (*academic, final, regular comparatives*) cannot normally occur in the modifying position in compounds, which provides a formal argument to support the intuition. Accordingly, we conclude that these items, although apparently fitting the orthographic definition of compounds, are not real compounds but syntactic structures.

There are innumerable other instances in our data where the hyphenation is unhelpful or distinctly counterintuitive. Consider, for example, the following:

to fill AB social *class-type* jobs

*ex-vice* queen of Hollywood

While *AB-social-class-type* would be sensible for the former, the latter, contrary to the evident intention, appears to refer to the queen of former vice rather than to a former queen of vice. Examples like these are the best evidence that hyphenation should not of itself be taken to imply compound status in English.

Among the remaining potential types is the translative compound (for the terminology, see Bauer and Huddleston forthcoming). These are made up, in the examples given here, of two nouns in a coordinate relationship, but the same construction cannot appear as a noun with the same meaning. These can only be used attributively.

This unconscious *angel-beast* division was not unique to psychoanalysis.

Even though my parents were not there, it was a typical *grandmother grandchild* relationship in that I could do no wrong.

The *love-pain* equation packs a devastating punch.

The construction is to be distinguished from the superficially similar one found in

A combination of ripe pears, apples and crisp white cabbage tossed with grated dates and a *honey-yoghurt* dressing usually proves popular.

Here the use of the coordinated construction as a noun is at least conceivable (even if an attributive usage is more likely). It is not clear to us whether exclusive attributive usage is sufficient to allow something to be classified as an adjective, but there is at least an argument to be made in this direction. If the argument is accepted, these compounds join the class of exocentric compounds since they have no adjectival head.

A similar argument may hold for a category not dealt with by Marchand (1969) but mentioned by Bauer (1983,212): the category made up of a superficial adjective (possibly functioning adverbially) and an element that is probably a verb (though in some cases it could formally be a noun). This type has already been mentioned as a type of headless compound. Some examples from our data are as follows:

Thank goodness for copious doses of chung, the *quick-brew* Tibetan wine, which washed down the sticky mass.

And from there by analogy to the compelling thought that C&B also manifests in earthy form as Waistline salad cream, 3-minute noodles or *quick-mix* Hollandaise.

The *low-build*, skimpy tyres-with a Corsa suspension that's tuned to match make the Tigra hunker down close to the tarmac.

The turkey doesn't need to roast forever, either. Current culinary creed favours cooking it on a higher heat for a shorter time rather than the 10-hour *slow-cook* marathon that usually means some poor soul-invariably Mumhas to stagger out of bed at 5 a.m.

Haspelmath (1996) argues that adverbial-ly in English is inflectional rather than derivational. This would explain the adjectival form in many cases since inflectional affixes are generally not permitted compound-internally.<sup>2</sup> *Quick* rather than *quickly* in *quick-brew* would thus arise from the application of a wider principle and not require specific explanation. If this argument is accepted, examples such as *largely-unpatrolled* cited above prove themselves to be something other than compounds, and there is another formal argument for the claim made above that they are syntactic.

The real point in all of this is, of course, that it appears to be possible to do much more in attributive position than in most other positions. Items occurring in this position are by definition premodifiers, but it is less clear that they are also adjectives indeed, we claim above that they are not all adjectives. The patterns are not uncommon, though individual examples may be, and it is often difficult to parse individual examples because of the common homophony of nouns and verbs (and even adjectives) in English. We list some examples here, without any implication that these should all be treated alike in a grammar but with an implication that a satisfactory descriptive grammar needs to consider these as possible types.

'It was,' he says, in the ringing clichés that come so easily to his lips, 'the *get-ahead*, buoyant Eighties, and I was on my way to making a lot of money.'

The Biro hit the shops of Britain 50 years ago. In the run-up to Christmas 1945, this was no el cheapo *chuckaway* item costing a few pence but a luxury purchase. At 55 shillings (pounds 2.75), it cost the weekly wage of a secretary.

This part-anthem, part-abrasive, all-disco album was brimming with sex, fags and bad clothes. The Big Town Players Energetic *jive-jump* band The Grahamophones.

With both sides adopting a *run-everything* policy the result, not that it mattered, was a victory to Evans's British Isles XV.

This was a '*cost-nothing*' play.

"He has a very *kick-arse* attitude," one says. "He'll say, 'right, here's the strategy. Go do it.' "

The *lynch-happy* citizens of Parliament. . . .

Bauer (1983,211) notes that the V + N pattern is not used productively with compound nouns but only with elements used attributively. Again, there are problems in deciding what category such items should have, though their function is clear. These things cannot be ignored as part of the panoply of English word-formation patterns.

### Compounds with Plural Inflection on First Element

The standard view of English compounding is that it does not allow internal inflections, and yet there is a fairly extensive literature (e.g., Mutt 1967; Dierick 1970; Johansson 1980) commenting on the use of plural modifiers and usually claiming that the phenomenon of plural modifiers is increasing in contemporary English.

To our surprise, we did not find particularly many of these in our data sample, and most of those we found fit into one of a small number of patterns.

1. Those where the modifier has a different meaning in the plural than in the (apparently corresponding) singular form or where the noun is found only in the plural: *arms-issue*, *arts-patronising*, *Beatles-watcher*, *blues-type*, *newsbusiness*, *roots-quest*, and, with a derivative rather than a compound, *specs-ist*.
2. Those where the modifier has an irregular plural (these were not common in our data): *media-comprehensible*, *people-smuggling*, *women-friends*.
3. Those where the modifier is in the plural because this is the unmarked form for that modifier: *spuds-wise*, *relations-wise*, *yachts-woman* (note the paral



lel with *yachtsman*). To this type we could add the derivational type illustrated by *gourdsful*, where the internal inflection shows the original syntactic construction.

4. Those like the words in (I) above but where the use of the plural is only clarificatory rather than necessary: *drugs-induced*, *forms-compatible*, *savings rate*, *singles-only*. *Drugs in drugs-induced* seems to be used to distinguish illegal drugs from legal ones; "a drug-induced sleep" would be something ordered by the doctor, and the drugs-induced teenage rampage that we are dealing *with in this* text is clearly related to drug abuse. The plural *in forms-compatible* seems to be used to clarify that we are dealing *with* the kind of form that has to be filled in rather than form just meaning 'shape.'

Precisely how this type works and what the plural *is* used to signal *is* not clear to us; *it* may be that *it is* misleading to label such cases "instances with a plural modifier" and that the semantics *is* more complex, although it does look superficially as if the *-s is* being used to prove countability in some cases.

Despite the fact that category (4) *is* relatively broad and may encompass examples like *fireworks-Jest*, *graphics-laden*, and *sports-vision*, there still seem to be some instances that remain outside this classification. Consider the following examples:

Most of the Cantona repertoire was on *view*, including a *studs-high* booking and even a feet-first foray *into* the crowd.

. . . the next year saw that handsome colt Track Spare *win* the first *stalls-started* event *in* England, at Newmarket.

Total mergers and *acquisitions-related* fees for the year could easily top pounds *Ibn*.

The only explanation for such instances we can see *is* a genuine wish to stress the plurality of the first element-something that ought not to be possible according to standard descriptions (although see note 3).

Although, as noted above, we found relatively few compounds with *this* kind of construction, another look at our data suggests reasons why *this* may be. Because our criterion for a compound was orthographical, we failed to note the cases where the plural modifier was written as a separate word from the head. Even *in* our examples of compound forms, several are being used as plural modifiers, including the

following:

Twenty years ago, writers Dick Vosburgh, Barry Cryer, Peter Vincent and I were writing the *all-impersonations* television series *Who Do You Do?*

From next year, a new module will be able to access the Worldspan *computer reservations* system directly.

He multiplied the houses, increased their size and developed a 3,000 tons per-week *feeding-stuffs* mill at Belle Eau Park, thought to be the largest in Britain at the time.

this has meant re-evaluating the benefits of its previous *free-calls* offers against the use and value placed on them by various types of customer.

The series culminates with the finals on 23-25 February in Paris, where the *top-points* earners will compete for dollars 700,000 in prize-money. [Note that, as so often, the hyphenation in this example is unhelpful.]

Speculation about names centres on Kenneth Boey, a previous director of the club's *membership-services* department.

It is the normality of such examples that suggests that the use of plural modifiers, not usual in lexicalized items, may indeed be increasing in current English and that descriptions that ignore the use of plural modifiers are missing one important facet of the construction of vocabulary today.

## Synthetic Compounding

Synthetic compounds are those like *tax-payer*, as opposed to primary compounds like *taxman*. The synthetic compound is interpreted as having a verb in the second element and an argument of that verb in the left-hand element. *Box spanner* would be a synthetic compound if it meant 'person or thing *which* spans boxes' but is a primary compound with the meaning 'spanner shaped like a box.' The area of synthetic compounding in English is a descriptive and terminological nightmare. In terminology, there are those who distinguish between synthetic and verbal compounds (Botha 1984) and those who do not (Selkirk 1982), and the terminologies of "primary" and "root" compound for the complementary set of compounds are also totally unhelpful. Where description of the structures is concerned, the whole problem is tightly tied up with a matter of definition: for some scholars, synthetic compounds are more tightly defined than for others so that some include structures under the rubric that others exclude by definition. This leaves subsequent scholars in a state of perpetual confusion in trying to interpret the competing claims made about synthetic compounding. For example, Roeper and Siegel (1978, 199,206) discuss only compounds whose second element ends in one of the three suffixes -*er*, -*ing*, and -*ed* as being verbal compounds; others (Botha 1984; Marchand 1969; Selkirk 1982) specifically allow for other suffixes. So for Roeper and Siegel, *slum clear*

*ance*, *grain storage*, and *consumer protection* are not verbal compounds, while for the others they are.

Roeper and Siegel (1978,207) allow compounds where the first element is interpreted as locative with respect to the verb in the head element. For them, both *cave dweller* and *sea-going* are verbal compounds. For Katamba (1993, 309-10), on the other hand, *Sunday closing* and *low-flying* are not verbal compounds because locatives do not "function as an argument of the verb." Botha (1984, 62) discusses this problem, noting that for Selkirk (1982), locatives are apparently not arguments, though for the scholars who are cited by Selkirk as the origin of her notions of what is an argument, locatives do count as arguments.

For Selkirk (1982,34), "The SUBJ [Subject] argument of a lexical item may not be satisfied in compound structure," while Alien (1978, 173) specifically includes compounds such *asfood spoilage* and *insect flight*, where the left-hand element acts as the subject of the intransitive verb. Selkirk is clearly aware of a problem here, for in a footnote (1982, 128, n.13) she makes the startling suggestion that *consumer spending* might not be a compound at all but that in such a construction, *consumer* might be an adjective. This piece of sophistry would not be necessary if the major claim about subjects had not been made first.

It is not our purpose here to become embroiled in the discussion of where the boundaries of synthetic compounding should run; in particular, discussions based on contrasting definitions are fruitless since what holds true under one definition may not hold true under another. Rather, we can illustrate some points that appear to have been masked in the theoretical discussion.

The major point to be made is that our data sample reveals many cases where there appears to be a subject-verb relationship between the two parts of the compound. Whether these things are synthetic compounds is-as just discussed-a matter of definition.

Some of the examples may be ambiguous between a subject-verb reading and some other reading. For example,

Caught in the Act also shows couples in supposedly secret sexual embraces.

Barrie Goulding is no stranger to *video-shockers*. Earlier this year he gained notoriety with Executions.

It may not be clear whether we are talking about a video that is a shocker (i.e., awful) or a video that shocks people, but other examples are clearer. We find examples with inanimate subjects and intransitive verbs:

(The make-up man is surely up for some technical prizes for the quality of his lesions and *blood-pooling*.)

After a *calf-aching* climb up from Cromford station. . . .

'Directed with *shrapnel-flying* immediacy,' declared the Financial Times.

We also find other examples, perhaps even clearer examples of the type, where the subject is animate and sometimes human:

but those in Saudi Arabia have not elicited a *mouse-squeak* of complaint by the US and British governments.

Is it supposed to help us make up our minds in a tricky *consumer-choice* situation ('What do you fancy from the tap tonight honey? Thames? Welsh?')

Vodafone was the best-performing blue chip. It was helped on Tuesday by *director-buying* and gained a further 12.5p.

Note that although many of the clearest examples in this set involve a head noun that is derived from its verb by conversion, the typical *-ing* suffix of the most productive pattern of synthetic compounding is also found. Note also that the verbs used are not canonical intransitives, and transitivity is possible with an absolute usage.

One type is frequent enough to be worthy of separate comment. It is made up of compounds ending in the element *-speak*, something of a vogue formation in the 1980s and 1990s (with examples recorded by Algeo 1991 from 1984 onwards). One pattern with this element (though by no means the only pattern) is for the modifying element in the compound to be the speaker, which again produces a subject + verb pattern:

Though this appears to be a slice of worthy *controllerspeak*, it is a revealing remark.

Yesterday Nato formally authorised the deployment- 'G-Day' in *armyspeak*.

This is *waiter-speak* for all the tasty bits: livers, kidneys, pigs' cheeks, sweet breads, brains.

These examples echo Algeo's (1991, 218, 231) *Haigspeak* and *Valley Girlspeak* and *childspeak*. In the light of such examples, it is clear that compounds that have to be interpreted as involving a subject + verb relationship are productive in current English.

It has been suggested to us that interpretations of compounds containing nominalizations where the first element is the subject of the nominalized verb are

permitted and easy where the nominalization is not productive and not permitted where the nominalization is productive. *Consumer-choice* in the examples above is easy to coin because *choice* cannot productively be derived from *choose*, while *director-buying* is (the suggestion runs) actually ungrammatical despite being attested because the derivation of *buying* from *buy* is productive. This suggestion has a great deal of merit and explains why some scholars limit themselves to the suffixes *-er*, *-ing*, and *-ed* when discussing this type of construction since these are among the most productive suffixes in English. There are two problems with it in this simplistic statement of the case. The first is that conversion (zero-derivation) is productive in Modern English, and instances like *mouse-squeak* cited above do not appear to be too problematical. The second is that the examples with inanimate subjects and the productive suffix *-ing* above all seem to indicate that a more sophisticated statement of what is actually going on is required. Even if the ultimate solution to the problem turns out to be based on this insight, we are not yet convinced that the exceptions to the general rule are a motivated group, and we still believe that sets of data of the kind we are presenting here are valuable in showing precisely what needs to be described.

### Conclusion

In this article, we have shown how the detailed analysis of corpus material can expose some unexpected trends even in a well-explored area such as compounding in English. The gaps in the descriptions of compound verbs and compound adjectives are matters that, in one sense, are relatively trivial. However, unless attention is drawn to such phenomena, descriptions cannot improve, and without a thorough descriptive base, theoreticians are likely to build inadequate theoretical constructs. The questions of headedness seem to us to be more fundamental in nature, raising questions that lie at the very heart of word-formation studies. These include not only the whole question of headedness and how it should be defined but questions of diachronic change and whether and how diachronic change can affect headedness and why. The data we have presented may lead to a great deal of speculation: for us, it certainly raises more questions than it provides answers. But in the end, if these patterns are found in English, the grammar of English needs to be able to explain them and build them in. Ignoring them and hoping that no one will notice is not likely to help us elaborate an explanatory picture of what is happening in English.

### Notes

1. One result of the proliferation of types of a premodifier is that some adjectives that are usually seen as being excluded from appearing in compound construc

tions nevertheless can occur in compounds. Levi (1978) argues that non predicate adjectives and nominal premodifiers are fundamentally in complementary distribution: either there is an adjective + noun syntactic construction, or there is a noun + noun lexical construction. In a few cases such as *atom bomb* versus *atomic bomb*, both constructions coexist. But non predicate adjectives can occur in compounded premodifiers.

We have macho *multiple-divorce* Clive, hiding his tenderer feelings under a show of hard drinking bluster and public urination.

Pre-nip and tuck, clubbers would dance in a dark, cold *industrial-look* basement complete with steel girders.

He also revealed that Mr Yeltsin has the suitcase containing the so-called *nuclear-launch* button with him.

2. Compound-internal plural marking provides a well-known exception to the general pattern, though not a particularly frequent one in our data (see below). These exceptions can be explained either in semantic terms (*bowls-shop* in our data is not synonymous with a putative *bowl-shop* and so requires the *-s* to guarantee the meaning) or by the argument that plural-marking is actually derivational in English (Beard 1982).

### References

- Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. London: Longman.
- Algeo, John. 1971. The Voguish Uses of 'non.' *American Speech* 41:87-105.
- . 1991. *Fifty Years among the New Words*. Cambridge, UK: Cambridge University Press.
- Allen, Margaret R. 1978. *Morphological Investigations*. Ph.D. dissertation, University of Connecticut.
- Baayen, R. Harald, and Antoinette Renouf. Forthcoming. On Mock-Scholarly, Faux-Casual and Cod Philosophy: Patterns of Diachronic Change in a British Newspaper. Paper presented at the ACH-ALLC Conference, Glasgow, UK.
- Bauer, Laurie. 1978. *The Grammar of Nominal Compounding*. Odense: Odense University Press.
- . 1983. *English Word-Formation*. Cambridge, UK: Cambridge University Press.
- . 1994. Head and Modifier. In *The Encyclopedia of Language and Linguistics*, edited by R. E. Asher, 1529-32. Oxford, UK: Pergamon.

- Bauer, Laurie, and Rodney Huddleston. Forthcoming. Lexical Word-Formation. In *The Cambridge Grammar of the English Language*. Cambridge, UK: Cambridge University Press.
- Beard, Robert. 1982. The Plural as Lexical Derivation. *Glossa* 16:133-48.
- Botha, Rudolf P. 1984. *Morphological Mechanisms*. Oxford, UK: Pergamon.
- Collier, Alex, and Mike Pacey. 1997. A Large-Scale Corpus System for Identifying Thesaural Relations. In *Corpus-Based Studies in English: Papers from the 17th International Conference on English Language Research on Computerized Corpora*, edited by M. Ljung, 87-100. Amsterdam: Rodopi.
- Dierick, Jean. 1970. Why Are Plural Attributives Becoming More Frequent? In *Linguistique Contemporaine: Hommage a Eric Buyssens*, edited by Jean Dierick and Y van Lebrun, 39-46. Bruxelles: Editions de l'Inst. de Sociologie de l'Universite Libre.
- Di Sciullo, Anna-Maria, and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.
- Haspelmath, Martin. 1996. Word-Class-Changing Inflection and Morphological Theory. In *Yearbook of Morphology 1995*, edited by Geert Booij and Jaap van Marle, 43-66. Dordrecht, the Netherlands: Kluwer.
- Higginbotham, James. 1985. On Semantics. *Linguistic Inquiry* 16:547-93.
- Johansson, Stig. 1980. *Plural Attributive Nouns in Present-Day English*. Lund: Glerup.
- Katamba, Francis. 1993. *Morphology*. Basingstoke, UK: Macmillan.
- Levi, Judith. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- Lieber, Rochelle. 1983. Argument Linking and Compounds in English. *Linguistic Inquiry* 14:251-85.
- \_\_\_\_\_. 1992. *Deconstructing Morphology*. Chicago: Chicago University Press.
- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word Formation*. 2d ed. Munich: Beck.
- Mutt, Otto. 1967. Some Recent Developments in the Use of Nouns as Premodifiers in English. *Zeitschrift für Anglistik und Amerikanistik* 15:401-8.
- Renouf, Antoinette. 1993. A Word in Time: First Findings from Dynamic Corpus Investigation. In *English Language Corpora: Design, Analysis and Exploitation*, edited by Jan Aarts, Pieterde Haan, and Nelleke Oostdijk, 279-88. Amsterdam: Rodopi.
- \_\_\_\_\_. 1996. The ACRONYM Project: Discovering the Textual Thesaurus. In *Papers from English Language Research on Computerized Corpora (ICAME 16)*, edited by Ian Lancashire, Charles Meyer, and Carol Percy, 171-87. Amsterdam: Rodopi.
- Renouf, Antoinette, and R. Harald Baayen. 1998. Aviating among the Hapax Legomena: Morphological Grammaticalisation in Current British Newspaper English. In *Explorations in Corpus Linguistics*, edited by Antoinette Renouf, 181-89. Amsterdam: Rodopi.