

English Adjective-Noun Compounds and Related Constructions

Paul Bennett

Department of Language and Linguistics
UMIST
P O Box 88
Manchester , M60 1QD, Great Britain
Paul.Bennett@umist.ac.uk

Introduction

The aim of this paper is to examine English constructions consisting of an adjective and a noun, with particular reference to whether this sequence can form a compound as well as a syntactic structure. Emphasis will particularly be placed on two kinds of examples: those where an adjective-noun structure modifies another noun, such as *new-road enquiry*; and those consisting of a relational adjective plus a noun, as in *nervous disease*. We shall touch on issues of productivity and lexicalization, as well as the relations and interactions between compounding and syntax.

The Problem

Let us begin by discussing whether English has a productive process that combines an adjective and a noun to form a compound noun. Selkirk (1982, p. 16) gives some rules for English compounds which include, in addition to the uncontroversial rule (1a), the more contentious (1b):

(1) a. N --> N N

b. N --> A N

However, she provides no explicit justification for (1b), and her examples of AN compounds all seem to be lexicalized and non-compositional in meaning (*high school*, *smallpox*, *sharpshooter*, *well-wisher*).

Comprehensive descriptions of English compounds, such as Marchand (1960, pp. 28-9) and Adams (1973, pp. 83-7), would lend support to the idea that no generative rule is needed for AN compounds, since the examples cited are relatively small in number and non-compositional in meaning. The examples in (2) are from Marchand, those in (3) from Adams:

(2) blackberry, freshman, highschool, hotbed, quicksand, shortbread, greenhouse

(3) dark-room, small talk, high chair, bighead

Since *hotbed*, for instance, does not denote a bed that is hot, it would seem that these examples are lexicalized, i.e. listed as part of the permanent lexicon rather than

generated by a productive rule. The same would apply to *shortlist*, which is non-compositional in that not just any list with a few items on it counts as a shortlist. Even if the lexical entries for these forms show them as having a complex structure (which, for instance, may be needed in order to ensure the plural *freshmen*), there is no basis for a general rule which sanctions this structure. Equally, Bauer (1983, p. 205) briefly discusses AN compounds, using stress as the criterion for their existence, but all his examples are lexicalized.

Jackendoff (1997) has emphasized that the lexicon must contain complex entries, e.g. for idioms such as *kick the bucket*, which specify their meaning and internal structure. He includes lexicalized compounds such as *blueberry* and *redhead* in this. So there can be no problem with including examples such as (2) and (3) in the lexicon, perhaps showing them as having the structure [N A N].

So far we have seen that there is little basis for the rule (1b). In addition, it creates at least two major problems. Firstly, if the rule is used, it would appear that any AN sequence is ambiguous between a syntactic analysis (where an AP is combined with an N') and a morphological one. This would lead to a great deal of ambiguity for which there is simply no justification. Secondly, the rule will massively overgenerate, allowing noun heads in compounds to themselves consist of AN:

- (4) a. *lion young tamer `young tamer of lions'
- b. *house incompetent builder `incompetent house-builder'
- c. *hotel cheap room `cheap room in a hotel'
- d. *shoe big shop `big shop for shoes'

Such examples appear to be uniformly excluded (except for the special cases seen in (6) below), thus adding weight to the view that there should be no such rule as (1b).

Some Further Data

However, an argument for the existence of AN compounds can be found with what are often called relational adjectives (Levi 1978, Beard 1991):

- (5) atomic bomb, nervous disease, sexual abuse, dental technology, solar system, tidal zone

Relational adjectives are related semantically (and often derivationally) to a noun, are not gradable and are not used predicatively.

There are at least three reasons for considering combinations as seen in (5) to be compounds. Firstly, they occur inside noun modifiers:

- (6) a. child sexual abuse
- b. radio religious broadcasting (Bauer 1998, p. 82)
- c. government financial planning

On the grounds that nouns only occur as modifiers in compounds, not in phrases, it would follow that the AN sequences here must form compound nouns. Secondly, in

neutral contexts they have stress on their leftmost member, like most compounds but unlike most phrases (Lieberman & Sproat 1992). Compare the following (where an acute accent marks a stressed syllable):

(7) a. so^ˈlar system

b. do^ˈor handle

c. tall m^ˈan

Thirdly, they can be coordinated with nouns, which again presumably form part of a compound:

(8) a. dental and food technology (Bauer 1998, p. 76)

b. sensory and motor ability (ibid.)

c. architectural and software problems

The view that relational adjective + noun forms a compound noun in English is adopted in Lieberman & Sproat (1992) and Sadler & Arnold (1994). Thus such pairs as *atomic bomb* and *atom bomb* differ only in the category of their left-hand member. However, since relational adjectives clearly differ in many ways from ordinary adjectives, they do not in themselves support the view that AN compounds exist for the general run of adjectives.

A further argument for rule (1b) is that AN sequences can occur in what appears to be the non-head position of a compound. In many cases, this non-head is lexicalized:

(9) high-school dance, smallpox vaccination, blackboard eraser, shortbread recipe

These are not a problem, though: if (say) *smallpox* is listed in the lexicon as a noun (even with an internal structure) it will necessarily be insertable as part of an NN compound. Alternatively, the adjective here maybe a relational one:

10) a. toxic waste campaign

b. solar system diagram

c. tidal zone information

These are equally unproblematic if the AN sequence is a compound noun, as we have argued above.

More challenging are the cases where the complex non-head contains an ordinary adjective but is not lexicalized. Bauer (1998) maintains that the first element in an NN compound is usually not modifiable by an adjective, though he does give *instant noodle salad*. In fact, such examples are easy both to construct and to find in texts:

(11) new-road inquiry, old-clothes shop, old-growth forest, direct-action politics, small-car design, early-morning feeling, small-town life, live-music venue, cold-

weather clothing, fresh-fruit shop, Belgian-beer week, Turkish-language television, mixed-ability group, best-film list, new-books shelf.

(Hyphens have been used here to indicate the intended structure. Actual orthographic practice is very variable.)

Words of this form are also common as newly-coined technical terms (Sager 1990, pp. 71-4):

(12) high-tension wire, modern-style building

Examples with style or type as second element are especially frequent.

The examples in (11) and (12) are primary compounds, with the complex non-head functioning as a modifier of the head. There is, however, less freedom in synthetic compounds, where the non-head is an argument of the head; in these cases, a complex AN non-head is often awkward:

(13) a. ? young-lion tamer `tamer of young lions'

b. ? new-car driver `driver of a new car'

c. ? long-letter writer `writer of long letters'

d. ? extreme-heat resistant `resistant to extreme heat'

Some speakers regard all the above as fine, however, and some generally-acceptable examples can be found:

(14) a. cold-milk drinkers `drinkers of cold milk' (Frazier p. 104n6)

b. warm-beer lovers

c. modern-history teacher

d. new-syllabus proposal

Alegre & Gordon (1996) show that even children as young as three years can correctly interpret examples such as *red-rats eater*. So I will assume here that this structure must be catered for in the grammar.

The examples in (11) and (14) are generable within Selkirk's system, as each would have the following structure, assigned by the two rules in (1):

(15)[_N[_N AN] N]

Such an analysis supports the idea of a rule such as (1b). We seem, therefore, to have an impasse: AN compounds are productive only when part of a larger compound, i.e. where an AN compound modifies another noun (notated from now as AN-N). It should be noted that the following examples fall outside our current concerns.

- (16) a. *long-term* solution
b. *large-scale* model

This is because the italicized forms here are adjectives, as can be seen from the fact that they can be questioned using *how*:

- (17) a. How *long-term* is this solution meant to be?
b. How *large-scale* will the model be?

Such questions are not possible with genuine AN-N structures:

- (18) a. * How *live* music is this venue?
b. * How *small* car is this design?

We should add that adjective + noun + noun sequences are generally interpreted as an adjective modifying a compound noun (Bauer 1998):

- (19) [*comfortable*] [*hotel room*]

No account of this preference is attempted here: it may well be due to processing considerations (Frazier 1990).

Analysis

My proposal is to adopt an analysis which builds on the account of Sadler & Arnold (1994), according to which pre-nominal adjectives combine with a noun to form a 'small' construction where both mother and daughters are zero-level (word) constituents:

- (20) [_N[_A *happy*]] [_N *person*]]

So *happy* *person* is an N, like *person*. This structure is sanctioned in part by rule (1b), which is itself allowed by the principle that an expression can modify another expression of the same level in the grammatical hierarchy: phrases can modify phrases, and words can modify words. A small construction is a weakly lexical one, dominated by a zero-level (word-level) node. A subset of these are strongly lexical, being opaque to syntactic rules. Ordinary NN compounds are strongly lexical, as are relational adjective + N structures. The narrow restrictions on prenominal adjectival constructions follow naturally from the idea that they are weakly lexical, rather than phrasal.

This account can be naturally extended to the claim that all examples of preposed modifiers are zero-level constructions. Thus a *happy person* is unambiguous, with only the structure in (20) available, and the same would apply to *loud music*, *easy exercise*, *new road*, and countless other examples. And if *new road* always forms an N, then it can naturally undergo compounding with another N, to give an example such as *new-road inquiry*), which would be analysed as follows:

(21) [N [N new road] [N inquiry]]

Thus the Sadler & Arnold approach would seem to account for [AN]-N structures rather straightforwardly, with no need to claim that AN sequences are ambiguous between syntactic and compounding structures. The following structure is also unnecessary:

(22) [N [NP new road] [N inquiry]]

This is posited by Alegre & Gordon (1996), and also by Lawrenz (1995) for some comparable structures in German. But it massively overgenerates in that it allows the modifier NP to contain an article or a relative clause, both of which are quite impossible.

However, as have seen above, adjectives may occur as modifiers of non-head nouns in compounds but not in general as modifiers of the head of a compound noun (see (4)). Again, modification of the head noun is possible with leftward-stressed forms, whether lexicalized or not:

- (23) a. home darkroom
b. cocktail party small talk
c. child sexual abuse
d. government financial planning

We have, then, to account for the fact that heads can only be modified by some types of adjective, whereas no such restriction applies to non-heads.

Sadler & Arnold (1994, p. 216) suggest that **shoe big shop* is excluded because it would show a syntactic AN structure inside a compound, which is putatively a syntactic atom. The problem is that this kind of explanation is too strong, as it will exclude all the examples in (11) above, which would also contain a syntactic construct inside a compound.

I want to suggest that the answer is to exploit the Sadler & Arnold distinction between strongly and weakly lexical constructions and to combine this with a distinction based on the head vs. the modifier. Let us set out some of the basic contrasts again:

- (24) a. used car business
b. solar system diagram
c. government financial planning
d. **shoe big shop*

An NN combination is strongly lexical, as are the left-stressed AN combinations, but other AN combinations are weakly lexical (i.e. zero-level, with zero-level daughters, but syntactic, not morphological). Productively-formed AN structures are by default weakly lexical, but are strongly lexical (compounds) if they contain a relational adjective. A single rule combining A and N thus covers two different kinds of structure.

From this viewpoint, (24a) contains a strongly lexical structure whose non-head is formed by a weakly lexical combination. Examples (24b, c) contain a strongly lexical structure whose non-head is also formed by a strongly lexical combination. But (24d) contains a strongly lexical structure with a head which is formed by a weakly lexical combination. I suggest that the following generalization holds:

(25) The head of a strongly lexical (morphological) structure cannot consist of constituents whose combination is weakly lexical (zero-level, but syntactic).

The examples in (23), however, are fine, as the head is either entered in the lexicon as a noun, or consistsof relational adjective + noun. In either case, the head is a strongly lexical combination.

While descriptively adequate, (25) is a somewhat arbitrary statement, and calls out for some further explanation. One might wonder, in particular, why it is the head that is singled out for special treatment in this way, rather than the non-head. I believe that an answer to this can be found in the general claim that a head shares properties with the structure that it is head of. This is reflected in many ways: if a verb phrase is finite, so is its head verb; if a compound noun is plural, so is its head noun; and so on. It seems a natural extension to claim that if a construction is strongly lexical, this fact will be marked on its head, e.g. by a feature such as [+strong]. The idea is that this will constrain the possible combinations of weakly and strongly lexical constructions.

A feature such as this is needed anyway, in order for the stress-assignment rules to apply in the appropriate domains. It is not enough simply to have AN structures without somehow indicating which type they belong to.

I propose that with compound nouns, both mother and head are [+strong]. With AN structures, there is no such requirement built into the rule. Relational adjectives require both their sister and mother N to be [+strong]. Ordinary adjectives require their mother to be [-strong], i.e. to show that there is a weakly lexical construction, but impose no specific value for [strong] on their sister N, which may be a simple noun, a compound noun, or an AN structure.

I now present a formalization of this analysis within the PATR-II framework (Shieber 1986). The following two rules are employed:

(26) N1 --> N2 N3

(N1 strong) = yes

(N3 strong)> = yes

(27) N1 --> A N2

(N2 strong) = (A strongsister)

(N1 strong) = (A strongmother)

(The numbers here are purely to identify which constituent of the rule is being referred to.) For adjectives, the following lexical entries are needed (cat = `category'):

(28) a. *big*

cat = A

strongmother = no

b. *used*

cat = A

strongmother = no

c. *financial*

cat = A

strongmother = yes

strongsister = yes

Essentially, *strongmother=yes* means `mother node must be *strong=yes*', while *strongsister=yes* means `sister node must be *strong=yes*'.

On this analysis, *used car business* will have *strong=yes* on both the N dominating the whole sequence and on its head, the N dominating *business*. It will also have *strong=no* on the N dominating *used car*. In *government financial planning*, the N node dominating *financial planning* is *strong=yes* by dint of being the head of the strongly lexical combination of *government and financial planning*, and also by dint of being mother of the strongly lexical combination of *financial and planning*.

But with **shoe big shop*, there are contradictory features. The N node dominating *big shop* is *strong=yes* because it is head of the NN compound *shoe + big shop*, but is *strong=no* because it is mother of the weakly lexical combination of *big and shop*. On this node, then, there are contradictory features, and on the universal assumption that feature descriptions must be consistent, (24d) will not be well-formed.

No doubt there are other ways of capturing in formal terms the claim made in (25), but the system just sketched does achieve its aim. It may be noted that the idea of a single rule applying in more than one module of the grammar (here, in both morphology and syntax) is not unique to the case discussed here. For instance, Lexical Phonology (Mohanan 1986) recognises that some phonological rules (e.g. Trisyllabic Shortening) may apply both lexically and post-lexically.

Conclusion

I have argued that English does have a productive rule for forming AN compounds, consisting of relational adjective + noun. The same rule is responsible for the AN part of AN-N structures, although with run-of-the-mill adjectives here the AN part is weakly lexical, rather than a compound. The constraint against having weakly lexical structures inside strongly lexical ones needs to be relaxed so that it only prohibits a

weakly lexical structure as the head of a strongly lexical structure: this seems to make the right distinctions between well-formed and ill-formed examples.

A more general conclusion is that our analysis weakens still further the position that lexical and syntactic aspects of grammar can be strictly separated (Di Sciullo and Williams 1987). A single N --> A N rule generates both weakly and strongly lexical constructions, and the question of whether AN forms a compound becomes far less critical.

References

Adams, V. (1973), *An Introduction to Modern English Word-Formation*, Longman: London.

Alegre, M. & Gordon, P. (1996), Red rats eater exposes recursion in children's word formation, *Cognition* 60,65-82.

Bauer, L. (1983), *English Word-Formation*, Cambridge University Press: Cambridge.

Bauer, L. (1998), When is a sequence of two nouns a compound in English?, *English Language and Linguistics* 2, 65-86.

Beard, R. (1991), Decompositional composition: the semantics of scope ambiguities and 'bracketing paradoxes', *Natural Language and Linguistic Theory* 9, 195-229.

Di Sciullo, A. M. & Williams, E. (1987), *On the Definition of Word*, MIT Press: Cambridge, Mass.

Frazier, L. (1990), Identifying structure under X₀. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology*, Vol. 3, Foris, Dordrecht, pp. 87-105.

Jackendoff, R. (1997), *The Architecture of the Language Faculty*, MIT Press Cambridge, Mass.

Lawrenz, B. (1995), Das Graue-Maus-Dasein und das Brave-Mädchen-Image: zur Bildungsweise von A-N-N-Komposita im Deutschen, *Deutsch als Fremdsprache* 32, 39-42.

Levi, J. (1978), *The Syntax and Semantics of Complex Nominals*, Academic Press: New York.

Lieberman, M. & Sproat, R. (1992), The stress and structure of modified noun phrases in English. In I. Sag and A. Szabolcsi (Eds.), *Lexical Matters*, Center for the Study of Language and Information, Stanford, pp. 131-181.

Marchand, H. (1960), *The Categories and Types of Present-Day English Word-Formation*, Otto Harrassowitz: Wiesbaden.

Mohanan, K. (1986), *The Theory of Lexical Phonology*, Reidel: Dordrecht.

Sadler, L. & Arnold, D. (1994), Prenominal adjectives and the phrasal/lexical distinction, *Journal of Linguistics* 30, 187-226.

Sager, J. (1990), *A Practical Course in Terminology Processing*, Benjamins: Amsterdam.

Selkirk, E. (1982), *The Syntax of Words*, MIT Press: Cambridge, Mass.

Shieber, S. (1986), *An Introduction to Unification-based Approaches to Grammar*, Center for the Study of Language and Information: Stanford.

Biodata

Paul Bennett graduated from the School of Oriental and African Studies, University of London, and is now a Senior Lecturer in Linguistics at the University of Manchester Institute of Science and Technology. His main research interests are contrastive linguistics, machine translation and morphosyntax.