

# Phrases inside compounds: a puzzle for lexicon-free morphology

Andrew Carstairs-McCarthy

*Is the wellformedness of a complex word ever dependent on whether some constituent of it is lexically listed? The fact that many words are not lexically listed and many lexically listed items are not words encourages us to think that the answer should be 'no'. The behaviour of some compounds suggests otherwise, however. Spencer's (1988) approach to bracketing paradoxes seems at first to support the 'no' conclusion, but on closer inspection does not. These facts may support Sadock's (1998) suggestion that in some languages, including English, compounds illustrate a third pattern of grammatical organization, neither syntactic nor morphological.*

## 1. Introduction: the Lexicon-Free Morphology Hypothesis<sup>1</sup>

Why is there morphology alongside syntax? That is, why in language are there two types of grammatical structure, not just one? There is no obvious reason for there to be two types of structure, and this fact has contributed to a widespread view among linguists that there really is only one. Perhaps morphology is really just syntax below the word level, a view whose classic formulations within generative linguistics are due to Selkirk (1982) and Lieber (1992). Alternatively, perhaps phenomena that one thinks of as morphological can all be 'distributed' among phonology, syntax and the lexicon, a view that gives its name to the Distributed Morphology approach originally propounded by Halle and Marantz (1993).

Despite such arguments, most grammatical theorists probably still agree that not all of morphology can be tidied away in this fashion. Too much of it seems irreducibly *sui generis*. A popular alternative view, then, is that morphology is needed because of language's need for lexical items, complex as well as simple. It is not an accident, according to this view, that dictionaries function not only as lists of items that are idiosyncratic but also as lists of words. Complex words have a compactness that makes them intrinsically better suited than phrases for lexical listing. This view underlies the approach to phonology and morphology known as Lexical Phonology (Kiparsky 1982).

The trouble with this view is that the correspondence between being a word and being idiosyncratic is so loose. There are many words and wordforms that are not listed in any dictionary because their formation is regular and their meaning and grammatical function are predictable: not only many inflected wordforms, but also many derived lexemes. That is, there are many words that are not 'lexical items' in my sense of this term. Indeed, the possibility of using a complex word that has never been used before presupposes that this should be the case. At the same time, by contrast, there are many linguistic items whose meaning requires them to be included in any complete list of a language's idiosyncrasies, but whose internal structure is syntactic rather than morphological. These are what we call idioms. Thus there are many lexical items that are not words. Hence a language may in principle possess complex lexical items (in my sense) without having any recourse at all to morphology, or 'word-structure', in the formation of them.

A third view, then, is that word structure has no more to do with lexical item status than syntax has. Di Sciullo and Williams (1987) argue that the only reason why words are more likely than phrases to be idiosyncratic and hence to be 'listemes' is that words are in

general shorter than phrases. But this third view leaves unanswered the question with which we started: why does morphology exist?

Developing an answer to that question is part of a project that I am just now embarking on (Carstairs-McCarthy, in preparation). I will say no more about that project here, except to remark that it derives part of its justification from the fact just mentioned: the looseness of the connection between being a complex word and being a lexical item. Let us formulate a proposal that I will call ‘the Lexicon-Free Morphology Hypothesis’ as follows:

- (1) **Lexicon-Free Morphology Hypothesis (LFMH):** The question of whether a given complex word or wordform is well-formed or not never depends on whether or not one of its components is a lexical item.

A linguist, in stating whether one element can be combined morphologically with another to create a well-formed complex word or wordform, may make crucial reference to many aspects of it, such as its phonological shape, its syntactic category, its internal complexity, its meaning, or its inflection class—but never to whether or not it is lexically listed.

The LFMH is not essentially a new claim. Rather, it is a corollary of DiSciullo and Williams’s view of the relationship, or lack of it, between wordhood and listedness. But is the LFMH correct? In this paper I will present evidence of a novel kind to suggest that it is not correct, despite the fact that there are ways of handling superficial counterexamples that seem plausible at first sight. If so, then the need for complex lexical items may after all provide sufficient motivation for the existence of complex words, as Lexical Phonology assumes, and the need for some alternative answer to the ‘Why morphology?’ question diminishes. On the other hand, the crucial evidence is of a rather limited kind, restricted to compounding. It may well be relevant, then, that some linguists (e.g. Sadock 1998) have expressed doubts about whether compounding (at least in a language such as English, where it involves free forms rather than bound forms) really belongs to morphology at all. Perhaps compounding deserves to be regarded as a third pattern of grammatical organisation, distinct from both syntax and morphology. If so, the LFMH may be correct after all, and the need for an answer to the ‘Why morphology?’ question remains undiminished.

This conclusion is somewhat convoluted as well as tentative. However, the issue that provoked my original concern (‘Why morphology?’) is important. I hope that the thinness of my conclusion may stimulate other readers of SKASE to investigate the matter further.

## 2. Spencer’s (1988) approach to morphosemantic mismatches

Let us consider an issue that at first sight seems far removed from the LFMH. ‘Morphosemantic mismatch’ is the name given by Stump (1991) to the phenomenon of ‘bracketing paradoxes’, illustrated by items such as *nuclear physicist*, *transformational grammarian*, *Little Englander*, *Big Endians* (the name of a social group in Jonathan Swift’s *Gulliver’s Travels*, meaning not ‘endians (?) who are big’ but ‘people who cut open their boiled eggs at the big end rather than the little end’). In all these the individual components seem semantically appropriate to the meaning of the whole, but the way in which they are put together grammatically does not reflect how the meaning of the whole is structured:

- (2) Grammar: [nuclear<sub>A</sub> [physic-ist]<sub>N</sub>]<sub>N'</sub>, i.e. [ A [ B C ] ]  
 Meaning: [[NUCLEAR PHYSICS] -IST], i.e. [ [ A B ] C ]

In the 1980s, various solutions were suggested for these paradoxes, involving different bracketings at different levels of grammatical and semantic analysis. For example, Pesetsky (1985) suggested that the suffix *-ist* was ‘raised’ into the appropriate position in Logical Form (a level of semantic and syntactic structure posited in the then-current version of Chomskyan syntactic theory), while Marantz (1988) posited an operation of Morphological Merger, which can be thought of as a kind of ‘lowering’ at the level of word structure. But Spencer (1988) showed that these solutions are inadequate because there are expressions in which similar mismatches occur, but in which no similar rebracketing is possible. Examples are *chemical engineer*, *southern Dane*, and Spencer’s coining *aerobic gymnast* ‘practitioner of aerobic gymnastics (whatever that may be)’. The item *chemical engineer* is semantically paradoxical in just the same way as *nuclear physicist* is, yet it contains no counterpart to the suffix *-ist* that might be available for ‘raising’ or ‘merger’. (The suffix *-eer* looks superficially as if it might be such a suffix, but it is not: a chemical engineer is an expert in chemical engineering, not in ‘chemical engines’.) Spencer therefore proposes a better solution: when a semantic slot (e.g. EXPERT IN X, INHABITANT OF X) needs to be filled, a language may use to fill it any more-or-less appropriate item that the grammar (syntax or morphology) renders conveniently available, even if its grammatical structure does not compositionally reflect its meaning. Hence the meanings of both *nuclear physicist* and *chemical engineer* impose no need for any grammatical bracketing other than the obvious ones: [nuclear<sub>A</sub> [physic-ist]<sub>N</sub>]<sub>N'</sub>, [chemical<sub>A</sub> [engineer]<sub>N</sub>]<sub>N'</sub>

The words ‘when a semantic slot needs to be filled’ deserve commentary. The need in question implies some degree of lexicalisation or at least institutionalisation: in this instance, an expectation that there should be an institutionalised expression with the meaning EXPERT IN X, where the domain of X includes academic disciplines such as chemical engineering and nuclear physics. Institutionalisation and the lack of it can lead to a contrast in ambiguity between two superficially parallel expressions. Compare, for example:

- (3) *rural historian* (i) ‘historian living in the country’  
 (ii) ‘expert in the history of the countryside (as opposed to towns)’

and:

- (4) *suburban historian* (i) ‘historian living in a suburb’  
 (ii) ?? ‘expert in the history of suburbia (as opposed to town centres)’

The contrast between (3ii), which is clearly available as one interpretation of (3), and (4ii), which is not nearly so obviously available as an interpretation of (4), is due to the fact that the history of the countryside is an institutionalised domain of inquiry (a specialism within history), whereas the history of suburbia is not.

Compare also:

- (5) *agricultural economist* ‘expert on the economics of agriculture’  
 (6) ? *horticultural economist* ? ‘expert on the economics of horticulture’

- (7) ?? *botanical economist* ?? ‘expert on the economics of botany’

The diminishing acceptability of these reflects the fact that the economics of agriculture is an established specialism within economics, whereas the economics of horticulture is not, so far as I know (though it could conceivably become one), and the economics of botany is not and never could be, because it is hard to see what such a specialism would cover.

### 3. A compounding problem for the Lexicon-Free Morphology Hypothesis?

Now let us turn to some phenomena that are superficially more relevant to the LFMH. There are some ambiguous collocations of  $X^0$  items whose ambiguity seems to point to a need to permit phrases as constituents of compound words. An example is *American history teacher*, whose two meanings seem to point towards two distinct grammatical structures:

- (8) [American<sub>A</sub> [hístory<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub>]’ ‘American teacher of history’

- (9) [[American<sub>A</sub> hístory<sub>N</sub>]<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub> ‘teacher of American history’

The stress pattern of *history teacher* in (8) (stress on the nonhead element, as in *greenhouse* rather than *green house*) confirms uncontroversially that it should be classified as a compound rather than a phrase. The same goes, it seems, for the whole expression *American history teacher* in (9), where the main stress is on *history*, part of the nonhead element. What is striking, then, is that this nonhead element, *American history*, looks like a phrase. We thus appear to have a phrase inside a compound word, that is a syntactic unit inside a morphological one. This violates what has been called the No Phrase Constraint (Botha 1981: 18) – a whimsical term, in that it is a violation of itself. So is the No Phrase Constraint simply wrong?

Lieber (1992) lists a considerable literature on this issue. One thing that seems clear is that not just any phrase can appear inside a compound:

- (10) \*[[glorious<sub>A</sub> hístory<sub>N</sub>]<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub> ‘teacher of glorious history’

- (11) \*[[dull<sub>A</sub> hístory<sub>N</sub>]<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub> ‘teacher of dull history’  
 (*contrast* [dull<sub>A</sub> [hístory<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub>]’ ‘dull teacher of history’)

In fact, it seems as if only lexicalised or institutionalised phrases (clichés) can appear freely inside compounds:

- (12) [[defective compóntent] problem]  
 (13) ?[[expensive compóntent] problem]  
 (14) \*[[Norwegian compóntent] problem]

The phrase *defective component* is a cliché, whereas *expensive component* and *Norwegian component* are not. It is not that *defective component* necessarily occurs more commonly than the other two (in fact, *expensive component* yielded far more Google hits than *defective component* on 27 October 2005), but *defective component* differs from the other two in that it belongs to the institutionalised, cliché-filled jargon of the technical manual and the manufacturer's guarantee ('Any defective component will be replaced without charge ...').

Two similar sets of examples to compare are (15)-(17) and (18)-(20):

- (15) [[broken gláss] injuries]
- (16) ?[[broken pláte] injuries]
- (17) ??[[broken wíng] injuries]
- (18) [[capital cíties] lesson]
- (19) ?[[British cíties] lesson]
- (20) ??[[dangerous cíties] lesson]

The phrase *broken glass* is a cliché, whereas *broken plate* and *broken wing* are not. Notice that it is not that *broken pláte injuries* and *broken wíng injuries* are uninterpretable, nor that occasions for their use are unimaginable. It is easy to visualise 'broken plate injuries' occurring in the dining room of a ferry during a rough crossing of the Cook Strait (the turbulent stretch of water between the North and South Islands of New Zealand). 'Broken wing injuries' could arise at a microlite aircraft show if a mishap causes pieces of an aircraft to land among spectators. Similarly, *capital cities* is a cliché (the name of a possible topic for a primary school geography lesson) whereas *British cities* and *dangerous cities* are not, even though one can easily visualise a lesson on these topics. (Senior executives who travel the world extensively, for example, may benefit from instruction about those cities where special safety precautions must be taken.) What makes the compounds (15) and (18) more acceptable than the others is not that one cannot visualise circumstances where the others might be useful, but that the non-head phrase in (15) and (18) is institutionalised.

If this is correct, it seems bad news for LFMH. What determines the relative well-formedness of (15) and (17), or of (18) and (20), is precisely whether or not one of its elements is a lexical item, in the sense of being institutionalised and stored as a whole (whether or not its meaning is unpredictable). In the next section, however, we will explore whether Spencer's approach to morphosemantic mismatches suggests a mode of analysis that avoids this conclusion.

#### 4. A Spencer-inspired solution to the problem

The essence of Spencer's analysis of *nuclear physicist* and the like is that bracketing can be dissociated from interpretation. If so, do we really need the two distinct bracketings of *American history teacher* at (7) and (8)? Perhaps we can make do with only one, with two possible interpretations (Carstairs-McCarthy 2002:79-81):

- (21) [American<sub>A</sub> [hístory<sub>N</sub> teacher<sub>N</sub>]<sub>N</sub>]<sub>N</sub>' 'American teacher of history' or  
'teacher of American history'

This opens up radical alternatives to the bracketings suggested at (10)-(20). What if *defective compónt problem* and *broken gláss injuríes*, even with the pragmatically expected readings 'problem due to defective components' and 'injuries from broken glass', are structured grammatically not as in (12) and (15) but as in (22) and (23)?

- (22) [defective<sub>A</sub> [compónt<sub>N</sub> problem<sub>N</sub>]<sub>N</sub>]<sub>N</sub>'

- (23) [broken<sub>A</sub> [gláss<sub>N</sub> injuríes<sub>N</sub>]<sub>N</sub>]<sub>N</sub>'

In these instances, for pragmatic reasons, the only available reading is one that conflicts with the bracketing. However, so far as the grammar is concerned, this need matter no more than the apparently anomalous bracketing of *nuclear physicist*. Likewise, for an example such as (24) only one bracketing is possible:

- (24) [Norwegian [compónt problem]]

The reason why this can mean only 'component problem in Norway', not 'problem due to Norwegian components', is simply that *Norwegian component* is not a cliché, unlike *defective component*.

Under this analysis, items such as *defective compónt problem* and *broken gláss injuríes* are not compound words but phrases. So we no longer, perhaps, need to recognise a class of compounds that contain phrases. *A fortiori*, the problem that such 'compounds' seemed to pose for the LFMH disappears, because, being after all phrases, with an internal structure that is syntactic rather than morphological, the LFMH has nothing to say about them.

## 5. Why the Spencer-inspired solution will not work

For the solution just proposed to work, it must work for all the problematic data, not just some of it. Unfortunately, it does not. Consider the expression *capital cities lesson*. This seems well-formed because *capital cities* is a cliché, unlike *British cities* or *dangerous cities*. So the Spencer-inspired solution requires us to amend the bracketing of (18) as follows:

- (25) [capital<sub>A</sub> [cíties<sub>N</sub> lesson<sub>N</sub>]<sub>N</sub>]<sub>N</sub>'

However, this is unattractive because *cities lesson* is not a well-formed compound. The first element in an English compound cannot be plural unless it is a *plurale tantum* (e.g. *alms-giving*, *arms race*) or possibly an irregular plural: *mice-infested*, *teeth-marks* versus *\*rats-infested*, *\*claws-marks* (Kiparsky 1982) (though an irregular plural is not by itself sufficient to guarantee acceptability: *\*mice-hole*, *\*teethbrush*).

This is not an isolated example. Consider the following:

- (26) a days of the wéek lesson
- (27) a months of the yéar lesson
- (28) \*a problems of the wéek discussion
- (29) \*a day of the mónth decision

None of these expressions is pragmatically strange. Lessons about days of the week and months of the year are no doubt held in primary schools throughout the world. In a company or office, one can well envisage a meeting to discuss the problems of the week. One can also envisage a decision about the day of each month on which to hold a regularly scheduled meeting (say, the first Wednesday). Why, then, are (26) and (27) well-formed while (28) and (29) are not, at least in my judgement? The answer is that *days of the week* and *months of the year* are clichés whereas *problems of the week* and *day of the month* are not.

Can we then analyse (26) and (27) in terms of a Spencer-inspired bracketing that avoids incorporating a phrase inside a compound? Such an analysis will attribute to *a days of the wéek lesson* and *a months of the yéar lesson* the same grammatical structure as, for example, the phrase *the days of the whéat harvest*, even though the relationship between their grammatical structure and their meaning is not parallel:

- (30) the days [of the [whéat<sub>N</sub> harvest<sub>N</sub>]<sub>N</sub>]<sub>PP</sub>
- (31) a days [of the [wéek<sub>N</sub> lesson<sub>N</sub>]<sub>N</sub>]<sub>PP</sub>
- (32) a months [of the [yéar<sub>N</sub> lesson<sub>N</sub>]<sub>N</sub>]<sub>PP</sub>

But while the grammatical analysis indicated by the labelled bracketing at (30) is plausible and indeed standard, the parallel analyses at (31) and (32) are ludicrous. The noun that the singular article *a* specifies grammatically in (31) and (32) cannot be a plural noun such as *days* or *months*. It must instead be the singular noun *lesson*. A number mismatch between the indefinite article and its noun is totally unacceptable in all varieties of English, so far as I know. So (31) and (32) must be structured on the following lines instead:

- (31) a [[days of the wéek]<sub>N'</sub> lesson<sub>N</sub>]<sub>N</sub>
- (32) a [[months of the yéar]<sub>N'</sub> lesson<sub>N</sub>]<sub>N</sub>

The trouble is that this structure is in relevant respects exactly parallel to the problematic structure proposed at (9) for *American history teacher* ‘teacher of American history’, which our Spencer-inspired analysis was designed to avoid:

- (33) [[American<sub>A</sub> hístory<sub>N</sub>]<sub>N'</sub> teacher<sub>N</sub>]<sub>N</sub>

Our attempt to rescue the LFMH by means of an analysis inspired by Spencer's treatment of morphosemantic mismatches has thus failed, seemingly.

## 6. A (not necessarily desperate) alternative

'Compounding' in English is unlike compounding in, for example, German and perhaps all other Germanic languages (including Afrikaans, despite its morphological simplification) in an important respect. In English, compounding involves only juxtaposition and prosody, never any segmental linking elements or stem alternations of the kind seen in German *Straußenfeder* 'ostrich feather', *Arbeitsmarkt* 'labour market'. This opens for compounding in English, and in other languages where it involves pure juxtaposition, an intriguing possibility. Could it be that this sort of compounding is distinct from both syntax and morphology?

At first sight, this proposal may seem like a desperate effort to rescue the LFMH at all costs. Yet the suggestion that compounding belongs outside both morphology and syntax has a respectable pedigree (Sadock 1998). From a language-evolutionary point of view, English-style compounding as a grammatical phenomenon may constitute a relic of 'proto-linguistic' stringing-together of lexicalised units (Jackendoff 2002: 249-50). If so, the problem posed for the LFMH by examples such as *broken glass injuries* again dissolves, but for a quite different reason from that suggested in section 4. Again, if so, it would seem necessary to find an evolutionary origin for morphology proper that is distinct from that of compounding. But there are at least preliminary indications that such an explanation may be available (Carstairs-McCarthy 2005).

I said earlier that my conclusion would be tentative and might seem thin. If it does seem so, I apologise. I hope other linguists will be stimulated to do better with this data. But one aspect of my conclusion I do not apologise for: its invocation of language evolution. Many aspects of the way language is will not be properly understood until we know more about how the human biological capacity for language came to be the way it is (Carstairs-McCarthy 1999).

## Notes

<sup>1</sup> I am grateful to Pavol Stekauer for the encouragement to contribute this article. I am also grateful to a seminar audience at the University of Canterbury for comments on an earlier version.

## References

- Botha, Rudolf P. 1981. "A base rule theory of Afrikaans synthetic compounding." In: M. Moortgat, H. van der Hulst and T. Hoekstra (eds.), *The Scope of Lexical Rules*. Dordrecht: Foris, 1-77.
- Carstairs-McCarthy, Andrew. 1999. *The Origins of Complex Language: An Inquiry into the Evolutionary Beginnings of Sentences, Syllables and Truth*. Oxford: Oxford University Press.



- Carstairs-McCarthy, Andrew. 2002. *An Introduction to English Morphology*. Edinburgh: Edinburgh University Press.
- Carstairs-McCarthy, Andrew. 2005. "The evolutionary origin of morphology." In: M. Tallerman (ed.), *Language Origins: Perspectives on Evolution*. Oxford: Oxford University Press, 166-184.
- Carstairs-McCarthy, Andrew. In preparation. *Morphology and Language Design*. Oxford: Oxford University Press.
- Di Sciullo, Anna-Maria and Williams, Edwin. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.
- Halle, Morris and Marantz, Alec. 1993. "Distributed Morphology and the pieces of inflection." In: K. Hale and S. J. Keyser (eds.), *The View from Building 20*. Cambridge, MA: MIT Press, 111-76.
- Jackendoff, Ray. 2002. *The Foundations of Language*. Oxford: Oxford University Press.
- Kiparsky, Paul. 1982. "From Cyclic Phonology to Lexical Phonology." In: H. van der Hulst and N. Smith (eds.), *The Structure of Phonological Representations (Part I)*. Dordrecht: Foris, 131-175.
- Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory*. Chicago: University of Chicago Press.
- Marantz, Alec. 1988. "Clitics, morphological merger, and the mapping to phonological structure." In: M. Hammond and M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. San Diego: Academic Press, 253-270.
- Pesetsky, David. 1985. "Morphology and Logical Form." *Linguistic Inquiry* 16, 193-246.
- Sadock, Jerrold M. 1998. "On the autonomy of compounding morphology." In: S. G. Lapointe, D. K. Brentari and P. M. Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI, 161-87.
- Selkirk, Elisabeth O. 1982. *The Syntax of Words*. Cambridge, MA: MIT Press.
- Spencer, Andrew. 1988. "Bracketing paradoxes and the English lexicon." *Language* 64, 663-682.
- Stump, Gregory T. 1991. "A paradigm-based theory of morphosyntactic mismatches." *Language* 67, 675-725.

*Andrew Carstairs-McCarthy*  
*Linguistics Department*  
*School of Classics and Linguistics*  
*University of Canterbury*  
*Christchurch*  
*New Zealand*

*andrew.carstairs-mccarthy@canterbury.ac.nz*