Te Reo

the Journal of the Linguistic Society of New Zealand

Volume 69
Issue 1

Squib

2026
Pages 102–109

June, 2026

The syntactic source of (some) words

Laurie Bauer
Victoria University of Wellington

This paper is a peer-reviewed contribution from https://www.nzlingsoc.org/journal/current-issue/

Editor: Victoria Chen

The syntactic source of (some) words

LAURIE BAUER
Victoria University of Wellington

Abstract

A problem for word-formation studies is that the notion of a word is not well defined (Dixon & Aikhenvald, 2002), and this has implications for questions of productivity in word-formation. One specific problem is whether words can be created by syntactic rules as well as by rules of word-formation, a question which is important in defining a word. In this paper, two such types are considered: sequences of adjective and noun, as in *blackbird* and also in *red squirrel*, which seem to require distinct analyses, and types like *lady-in-waiting* and *wannabe* where words exploit a wider range of syntactic structures (see also Goldberg & Shirtz, 2025). It is argued that the difference between a word and a syntactic structure is a matter of usage rather than a matter of grammar, largely, but not entirely, determined by how fixed the expression is in common practice.

It is well-known that the term 'word' is extremely difficult to define in linguistics, despite is ubiquity (Dixon & Aikhenvald 2002, and many others). One of problems this raises is in the area of morphological productivity: if we do not know whether something is a word, how can we tell whether morphological productivity applies to it? In particular, even if we consider only grammatical models in which morphology is seen as separate from syntax, is it possible that syntactic structures contribute to morphological productivity? This important question is raised implicitly in a recent article by Körtvélyessy (2026). Körtvélyessy is working with an onomasiological framework, where a word's semantic categories such as 'agent' are generated by a 'cluster' of types of word-formation, such as, for English, *-er* suffixation (*distributer*), conversion (*flirt*), synthetic compound (*spy-buster*), compound (*repairman*), *-ist* suffixation (*shootist*), *-ant* suffixation (*consultant*), *-ent* suffixation (*resident*) and so on. What is central is the semantic category, and cluster of types provides various paths which speaker coining new words can follow. This cluster "guarantees the formation of a new linguistic sign whenever such a need arises, it is 100% productive" (Körtvélyessy, 2026, p. 140). At the same time, Körtvélyessy (2026, p. 151) says that "the onomasiological approach to creativity encompasses all acts of word-formation – both regular and irregular, predictable and unpredictable." I find this problematic, because it is not clear to me how irregular and unpredictable formations can make up part of a cluster. One of the ways in which word-formation can be unpredictable is if words can be formed directly from syntactic structures by way of what Halliday would have called rank shift: in rank shift a unit at one level of analysis operates as part of a structure at a lower level (e.g. a clause operates as a word) (Crystal, 2008, p. 402). Körtvélyessy (2026) does not appear to allow for syntactic processes as part of word-formation, and that is what raises an old question in a new context. The process of changing from a piece of syntax to a word is labelled in various ways in the literature on language change, including univerbation and lexicalization (in one use of the term). What these have in common is that they are sporadic diachronic processes, influenced very largely by the frequency of a specific syntactic expression or collocation. The question that Körtvélyessy's position raises in my mind is

whether things that we might recognise as words can be constructed productively using syntactic processes. This reformulation makes it clear that what is at stake here is not just the old question of whether morphology can arise from syntax, or, indeed, the question from Distributed Morphology as to whether all complex words are syntactic structures, but a much more specific question about processes of word-formation. Körtvélyessy's paper prompts a look at this narrow question. Clearly, this question makes certain presuppositions, most notably that we can recognise words and that there is a distinction to be drawn between morphological processes and syntactic processes. Although both of these presuppositions are widely accepted, both are controversial (see Miller, 2026, for a recent discussion of whether there is an entity – he uses the philosophical term 'kind' – 'word', although the usefulness of the concept in linguistics is nevertheless acknowledged); the discussion here, however, can make sense only if both are accepted.

Bauer (2025a, pp. 131–137, see also Goldberg & Shirtz, 2025, with different presuppositions) discusses the use of syntax in word-formation, and lists various types, listed below:

a. Syntactic bases in affixation: *can-doism*

b. Phrases as modifiers: *it-allows-me-to-express-my-elegant-simplicity sort of black*

c. Phrases as words: *forget-me-not*, *wannabe*

Here, I shall ignore the first two, but consider the third, and one other type where the question of the productive use of syntax in word-formation seems relevant.

The extra type is made up of compounds with the form adjective + noun. Typical examples are *blackbird*, *software*. It is widely accepted that examples like *blackbird*, with initial stress, which the spelling attempts to capture by writing the two elements together, are compounds, while things like *black bird*, with final stress, again captured by the spelling, are not compounds but syntactic structures (Marchand, 1969, p. 21; Carstairs-McCarthy, 2002, pp. 59–60; Aronoff & Fudeman, 2011, p. 40; for a dissenting voice see Bauer, 2021). The question for consideration here is how the initial-stressed forms arise historically.

If we take an example like *blackbird* (which the Oxford English Dictionary, OED, says is created in English by compounding) we find early examples in the OED where it is written as two separate words, but such spellings vanish by the early 17th century. Superficially, this supports an analysis whereby *blackbird* arises through stress shift from an earlier form *black bird*, a syntactic structure. Unfortunately, it is hard to evaluate such evidence, not only because the earliest example in the OED, from ca. 1350, has *blacbrid* [sic] written as a single word, but also because it is not necessarily clear that single-word orthography represents initial stress at the period, or that the pronunciation was settled. Similar comments can be made about examples in Bauer (2025b) where apparently comparable examples are attested from newsreaders. The examples cited there are *BLUE whale*, *FOUL play* and *HIGH cloud*. But it is pointed out that initial stress here may be a result of reading aloud, rather than a result of lexical stress being assigned to the first element. At the same time, the semantics typically associated with forestress in such cases, namely that the adjective is classifying rather descriptive, certainly fit with *blue whale* and *high cloud*, if not so obviously with *foul play*. Bauer (1983, pp. 205–206) says that such compounds arise though a stress shift, but provides no supporting evidence for this opinion, except the observation that some compounds seem to arise after the corresponding phrasal structure, and with the same meaning. According to the OED's citations for *blackboard*, for example, writing the word solid does not become general until the 20th century, *strongman* is not shown as being the default form yet, while *whiteboard* has been written solid or hyphenated since its earliest use, as, apparently, has *bluebird*.

Referees have asked about the relevance of loss of compositionality as a criterion for a compound, since *blue whale*, for example, is a compound (and hence a word) in terms of compositionality. I shall return to this fundamental question below, using a different terminology. But (lack of) compositionality in itself is not a useful criterion because sources disagree about what counts as compositional in precisely the relevant contexts. Some authorities see compounds as being compositional because the meanings of *black* and *bird* are relevant in the interpretation of *blackbird*, and *windmill*, similarly, makes reference to both *wind* and *mill* (see, for example, Aronoff & Fudeman, 2011, p. 110, and less clearly Haspelmath, 2002, and Katamba & Stonham, 2006), while others see compounds as non-compositional because the literal translation of *blackbird* into Sanskrit is understood as the label for a corvid and because *windmill* has a different relationship between the elements from *flourmill* (for some discussion see Bauer, 2017, pp. 11–13). Di Sciullo and Williams (1987) argue that there are words which are not listemes (where lack of compositionality is one factor that leads to lexeme status) and phrases which are listemes, which implies that compositionality is thus not a relevant consideration for wordhood. For one view of this complex problem see Wisniewski and Wu (2012). While the two criteria on which I primarily depend here are not always decisive (and appeal will also be made to other criteria below), at least we know whether they apply or not; with compositionality, we do not seem to know whether it applies or not.

All of this is disappointingly obscure. While a historical development from phrase to compound is plausible in many cases, the emergence of a compound with no phrasal antecedent is also possible. Let us say that the development of an Adjective+Noun phrase to a corresponding compound is likely to be one path in the emergence of Adjective+Noun compounds.

Grimm (1878, p. 638) draws attention to a similar dichotomy in the creation of Adjective+Noun compounds in the history of German. He bases his observations on different criteria, though. He points out that while *Tiefsinn* ‘profundity’ could arise as specialisation (perhaps a metonymic reading. LB) of the phrasal version *tiefer Sinn* ‘deep mind’, the same is not true of *Grossvater* ‘grandfather’, whose meaning does not equate to that of *grosser Vater* ‘large father’. While this might be interpreted in different ways, it supports the idea that not all Adjective+Noun compounds have a common source.

To the extent that a syntactic source is reasonable, lexemes can have as their origin syntactic expressions, and we have yet another instance of ‘today's morphology is yesterday's syntax’ (Givón, 1971, p. 394). But in both types here, we can see a process of word-formation at work as well: prosodic marking on the one hand, and a process of compounding without reference to a syntactic base, on the other. The process of creating such compounds is thus more complex than we might have expected but can be seen to fit with Körtvélyessy’s notion of a cluster of processes being available. This example shows the importance of investigating individual examples carefully.

Another type of name creation involving adjectives and nouns is not so easily explained. This is the type illustrated by *bald eagle*, *blue whale*, *red squirrel*, *small screen*, and so on. These have precisely the same semantic specialisation that the compounds like *blackbird* have, in that the adjective provides a classification of the head noun, but there is no prosodic correlate of the semantic structure. Formally, *blue whale* is like *black bird*; semantically it is like *blackbird* (see Bauer, 2021). Although such expressions are typically treated by grammarians as syntactic structures, because they are names for types (and regularly get listed in dictionaries as a result), they behave syntactically in the ways that words do. If they are words, though, they are words created from syntactic processes rather than through processes of word-formation. It seems to me that these structures cause problems for Körtvélyessy’s point of view. A new

word (lexeme) can arise that has not been created by means of any process that can be in a cluster of word-formation types. Even for those who would argue that such forms are compounds specifically (and not just words of an unspecified kind) because of their lack of compositionality (see above), they differ from the *blackbird* type because there is no prosodic marking of the type which might be seen as word-formation process, so their origin is still more clearly syntactic.

The other type of word-creation via syntax that I wish to consider is the type where a syntactic sequence of a fairly random type is used as a word (Goldberg & Shirtz, 2025, talk of these things being treated *as if* they are words). We can distinguish various syntactic patterns where this has occurred, as in (1), but it is not clear that the syntactic construction used is a relevant consideration for anything but a potential classification.

(1)	by-and-by	coordinated prepositions
	can-do	modal + verb
	cash-and-carry	coordinated words[1]
	drag-and-drop	coordinated verbs
	fly-by-night	verb + prepositional phrase
	forget-me-not	imperative clause[2]
	has-been	auxiliary + past participle
	jack-in-a-box	N + PP (including a determiner)
	lady-in-waiting	N + PP
	lie-abed	verb + adverb
	life-and-death	coordinated nouns
	wannabe	verb + infinitive

It is noticeable that the main piece of evidence we have for the single-word status of these items is the hyphen, which is not a piece of grammatical evidence. This hyphen is a binding hyphen, not a separating hyphen as in *co-operate* and *re-educate*. And when several things we recognise as words from elsewhere in the language are bound in this way, we perceive them as creating a single word. Nevertheless, this hyphen raises some questions.

The first concerns things like *life-and-death*. Following the rules for hyphenation in English, *life-and-death* is hyphenated in attributive position, for instance in (2).

(2) A life-and-death situation

This spelling is not used, however, when the expression is used nominally, as in (3).

(3) This is a matter of life and death

[1] Since coordination in syntax typically demands that the coordinated items should be of the same syntactic type, it is not clear that a coordination of *cash* (a noun) and *carry* (a verb) could be a grammatical piece of syntax.

[2] Although this looks like an imperative in modern English (despite the fact that the syntax is now obsolete, since the flower is not a **don't-forget-me*), its ultimate origin is not necessarily clear, since related forms are found in German (*Vergiß-mein-nich* 'forget-my-not'), Danish (*forglemmigej* 'forget-me-not'), Dutch (*vergeet-mij-nietje* 'forget-me-not-DIMINUTIVE'), with related terms in other European languages (though the French form has now been replaced). The imperative seems to predate English, which makes the OED's comment that the word is created in English slightly odd; it is an English calque of a common European expression.

(In either use, we can have *or* instead of *and.*) Similar comments do not appear to apply to *fly-by-night* and *can-do*. If the hyphenation confers or reflects wordhood, then the phrase is a word when used as an attributive, but a phrase when used after a preposition. The question thus arises as to whether this distinction is justifiable, and if so, what would justify it. The answer is presumably that *life and death* looks like a perfectly normal instance of syntactically coordinated nouns, which could be found in other contexts, such as (4).

(4) Life and death are viewed as incompatible opposites.

Life-and-death, by contrast, requires both nouns in one syntactic phrase, and occurs in places where nouns are supposed to be rare (whether they are or not probably depends on your theoretical stance). However, *life and death* in (3) has precisely the same specialised meaning that *life-and-death* has (namely 'vital importance'), while that meaning is not present in (4). A better conclusion might therefore be that *life and death* has two distinct uses, one as coordinate nouns and one as a single word with a specialised meaning. This, however, loses the link between hyphenation and wordhood, and says that (at least some) multi-word expressions are, in fact, words, something which in our current terminology appears paradoxical, though we should not allow ourselves to be put off by possibly misleading terminology.

The second problem involves things like *water-ski* (the noun is spelt thus in Marr, 2008). The use of hyphens in words of this nature is not fully prescribed, so that Wehmeier (2000) uses the form *waterski* and Allen (2000) uses the form *water ski*. None of these is demonstrably wrong, though house-style and fashion play a role in which is found. If the hyphen confers wordhood in the examples in (1), does it also confer word-hood in *water-ski* (thereby making it the orthographic equivalent of *waterski*), but leave *water ski* as two words? In practice linguists denote things like *waterski* (however spelt) as compounds and determine that status by a set of criteria which are not universally agreed upon. Since compounds are defined as words (that is, as lexemes), it might seem that this is trivial, but since the spelling is not criterial for most linguists' understanding of compounds (except in corpus linguistics on the basis of written texts where no other criterion is available) the number of compounds that are accepted despite their spelling is considerable, and depends on the theoretical stance of the scholars concerned. The same might be argued of items such as *crow's nest*, which might be viewed as compounds or as syntactic phrases.

This leads us to conclude, unsurprisingly and uncontroversially, that orthography is not a good guide to wordhood, although it might be interpreted as indicating something about the perception of something as a word. While it is only one criterion among several other possible criteria, and although orthography is generally assumed to be derived from other linguistic features, hyphenation might thus be said to be an indication that the writer sees the hyphenated sequence as behaving in some ways like a word, without making clear what linguistic features lead to this perception, or how language-users think of a word.

Another thing to notice about the items cited in (1) is that they are all relatively short: two or three underlying items. The same is true of non-hyphenated examples such as *nevertheless* and *albeit*. This would seem to be related to the fact that they are listed in dictionaries: by Zipf's law, short items are likely to be common, and common items are more likely to be listed in dictionaries. Similar factors apply to compounds. While two-element compounds are commonly listed in dictionaries, and while it is generally recognised that compound structures are recursive, so that *university teaching awards committee member* (Plag, 2003, p. 133) is a possible compound, long compounds such as *tweetle beetle bottle puddle paddle battle muddle* (Seuss, 1965) are often considered jokes (as with this example from a children's book) or thought of as headlinese (as with *Rugby World Cup opening night traffic chaos* – Bauer, 2017, p. 43), but in any case not really considered to be words, but thought of as syntactic structures

of some unspecified kind. This leaves us with a problem for examples like that in (5). The hyphenation (and, incidentally, the quotation marks – see Goldberg & Shirtz, 2025) makes it look like a word; its length makes its wordhood seem marginal. Part of that marginality is that we are unlikely to meet the same form elsewhere, though prototypical words are recurrent (Di Sciullo & Williams, 1987).

(5) […] not in this atmosphere of 'we've-all-come-out-here-to-enjoy-ourselves-let's-get-on-with-it'. (Agatha Christie. 1964. *A Caribbean mystery*. Leicester: Ulverscroft (large print), p. 81)

One result is that items which are listed in dictionaries are typically short and thus limited in their structures to expressions which can be short. One widespread pattern, in English, and also in other European languages, is made up of Noun + Preposition + Noun, illustrated in (1) by *lady-in-waiting*. These are often termed compounds, on the basis that they fit the definition of a compound as 'a lexeme which contain two or more stems' (Bauer, 2004, abbreviated here), despite the fact that in English endocentric compounds are right-headed, and these structures are left-headed and that these structures typically fit into syntactic patterns (though see *cash-and-carry* in (1)). Such words are more common in Romance than in Germanic, since compounds which fit a more canonical perception of the category are a more recent development in Romance. French has many examples, such as *chemin de fer* 'way of iron = railway', *fer à repasse*r' 'iron for ironing', *pomme de terre* 'apple of earth = potato', *joie de vivre* 'joy of living', and English may have been influenced by this French pattern. At the same time, some such expressions in French, and in English, are considered to be syntactic structures rather than words: *verre de vin* 'glass of wine'*, homme de génie* 'man of genius'*, royaume de Belgique* 'kingdom of Belgium'; *command of English*, *day of mourning*, *piece of cheese*.

This raises questions of boundaries. When is a structure a syntactic structure and when is it a lexical structure if the elements are apparently parallel? Can individual constructs change from syntactic sequences to lexical items diachronically? Why is *man-of-war* hyphenated (and thus one word?) in several meanings, not all current, while *Man of God* is not (and thus three words?). There are answers to some of these questions. For instance, there are criteria for wordhood that involve matters such as whether the elements in the construct can be independently modified: *rich man-of -war* (where the man is rich) or *man-of-recent war* are not possible versions of *man-of-war*, so *man-of-war* is a word, while *Man of their God* (if we can be sure it is possible) suggests that *Man of God* is a phrase. Unfortunately, intuitions are often insecure with such examples. The example in (5), though is clearly extendable (e.g. by saying …*come out here today*…), and yet its hyphenation says we should think of it as a single word. Bauer (2025c) suggests that what is at stake here is how much variation there is in the input. If speakers hear such a phrase only in one form, they will not use it with modification of individual elements; if they hear it in many forms, they will feel free to add modification. That is, it may be that the structures are not different; what is different is how fixed the expression is in usage (some scholars talk of 'entrenchment' here, though, *entrenchment* refers to usage in the individual not in society). Having said that, for me, *Man of God* is fixed, and although I think I could say *a day of bitter mourning*, *day of mourning* sounds fixed, too. *Piece of Cheddar cheese*, on the other hand, is clearly less fixed. Readers may not agree with these intuitions, but the claim is simply that the borderline between the two types is not necessarily determinable, perhaps, in part, because intuitions do not always correspond with grammatical tests.

Nonetheless, if the apparently restricted syntax in some of these expressions (e.g. the lack of modification) is a matter of usage rather than a matter of genuinely syntactic constraints, we not only have an explanation as to why intuitions and usage might be variable in such instances, we also have an argument for saying that expressions such as *lady-in-waiting* (or more modern

examples, such as *president-for-life* – Barnhart et al., 1990), just like expressions such as *blue whale*, are pieces of syntax rather than forms being created through a process or word-formation. Longer examples, like that in (5) are less fixed because of their length, and are thus more open to variation.

This suggests that, while not all of the patterns that look syntactic but produce new words are necessarily syntactic, there are genuine cases of words being formed by the fixation of syntactic phrases (see also Bauer, 2025c). This has implications for how we model word-formation and confirms the idea that not all words arise though the application of a fixed set of rules of word-formation.

References

Allen, R. (Ed.). (2000). *The new Penguin English dictionary*. Penguin.

Aronoff, M., & Fudeman, K. (2011). *What is morphology?* (2nd ed.). Wiley.

Barnhart, R., Steinmetz, S., & Barnhart, C. (1990). *Third Barnhart dictionary of New English*. Wilson.

Bauer, L. (1983). *English word-formation*. Cambridge University Press.

Bauer, L. (2004). *A glossary of morphology*. Edinburgh University Press.

Bauer, L. (2017). *Compounds and compounding*. Cambridge University Press.

Bauer, L. (2021). Blackbirds and blue whales: Stress in English A+N constructions. *English Language and Linguistics,* 25, 581–600.

Bauer, L. (2025a). *Reflections on English word-formation*. Cambridge University Press.

Bauer, L. (2025b). The paradox of stress in English: Is it vital or unimportant? *English Today,* 1–7. https://doi.org/10.1017/S0266078425100928

Bauer, L. (2025c). Fixity, or why English may not have compounds. In S. Matriciano-Mayerhofer, J. Schnitzer & E. Peters (Eds.), *Patterns, variants and change: Through the prism of morphology. Studies in honour of Franz Rainer* (pp. 39–48). Éditions de Linguistique et de Philologie.

Carstairs-McCarthy, A. (2002). *An introduction to English morphology*. Edinburgh University Press.

Crystal, D. (2008). *A dictionary of linguistics and phonetics* (6th ed.). Blackwell.

Di Sciullo, A., & Williams, E. (1987). *On the definition of word*. MIT Press.

Dixon, R., & Aikhenvald, A. (2002). *Word: A cross-linguistic typology*. Cambridge University Press.

Givón, T. (1971). Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistics Society*, 7, 394–415.

Goldberg, A., & Shirtz, S. (2025). The English phrase-as-lemma construction: When a phrase masquerades as a word, people play along. *Language*,101, 291–320.

Grimm, J. (1878). *Deutsche Grammatik*. Zweiter Theil. Neuer vermehrter Abdruck. Berlin: Harrwitz und Gossmann.

Haspelmath, M. (2002). *Understanding morphology*. Arnold.

Katamba, F., & Stonham, J. (2006). *Morphology* (2nd ed.). Macmillan.

Körtvélyessy, L. (2025). Word-formation at the crossroads of productivity, creativity, and structural complexity. *Linguistics,* 64, 137–163. https://doi.org/10.1515/ling-2024-0038

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

Marr, V. (Ed.). 2008. *The Chambers dictionary* (10th ed.). Chambers Harrap.

Miller, J. (2026). Words. In R. M. Neftd, G. Dupre, & K. H. Stanton (Eds.), *The Oxford handbook of philosophy of linguistics* (pp. 107–122). Oxford University Press.

Oxford University Press. (n.d.). *The Oxford English dictionary* on-line. www.oed.com.
Plag, I. (2003). *Word-formation in English*. Cambridge University Press.
Seuss, T. (1965). *Fox in socks*. Collins.
Wehmeier, S. (Ed.). (2000). *Oxford advanced learner's dictionary* (6th ed.). Oxford University Press.
Wisniewski, E., & Wu, J. (2012). Emergency!!!! Challenges to a compositional understanding of noun–noun combinations. In W. Hinzen, E. Machery, & M. Werning (Eds.), *The Oxford handbook of compositionality (*pp. 403–417). Oxford University Press.