Idioms in Morphology

Martin Everaert

Utrecht University, The Netherlands

1   Introduction[1]

Idiomaticity is a pervasive feature of language: 'there is so much of it in every language' (Weinreich 1969, 23). Weinreich also links idiomaticity to morphology when he says, 'Like many facts of word formation (cf. Zimmer 1964), idiomaticity serves to remind us how heavily language is laden with semiproductive patterns'. It is this connection, word formation and idiomaticity, that is the topic of this entry. It will address the following issues: How are we to define 'idiomaticity', and is there reason to assume a difference between syntax and morphology with regard to idiomaticity (Section 2)? The entry will show how two quite different morphological frameworks, Distributed Morphology (DM) and Construction Morphology (CM), deal with idiomaticity (Section 3). Finally, the entry will identify

what I would call 'morphological idioms' in different domains of word formation (Section 4).

Despite Weinreich's quote, it seems as if using the notion idiom in the context of morphology is not straightforward. Searching the word 'idiom' in the index of morphology handbooks, or morphology textbooks, gives very limited results. In the Spencer and Zwicky handbook (1998) the word 'idiom' is not indexed. The same holds for the introductions to morphology by Scalise (1984), Katamba (1994), and Booij (2005). The few entries you will find will refer to passages saying that *kick the bucket* is a listeme, just as *book* is (for instance Katamba and Stonham 2007). However, quite often you do find the entry 'idiosyncrasy', and under this heading the text will give you, to some degree, what I would call 'morphological idioms'. This state of affairs is perhaps not strange. The often-used terms for idioms are 'fixed expressions' or 'multi-word expressions', thus focusing on 'phrasehood' in order to talk about idiomaticity. But that is, perhaps, more a matter of habit than that idiomaticity in the morphological component should be excluded in principle. If we assume that the semantics of linguistic expressions are built regularly on the basis of their constituent parts, this implies that the meaning of polymorphematic words is built on the basis of the meaning of their constituent morphemes (cf. Kooij 1968). If so, idiosyncrasies in meaning are to be expected. However, talking about 'semantic idiosyncrasy' does not necessarily commit oneself to equating that to 'non-compositionality'. Pesetsky (1985, 210) explicitly makes the connection. His 'rules of idiosyncratic interpretation' apply both in syntax (and are then called 'idioms') and at the word level. He, thus, in part follows early work in generative grammar (cf. Katz and Postal 1963) in not differentiating 'lexical idioms', like *baritone*, from 'phrase idioms', like *kick the bucket*. However, Katz and Postal, contrary to Pesetsky, were explicit in how to qualify semantic idiosyncrasy: 'any concatenation of two or more morphemes whose compound meaning is not compositionally derived from the meanings of the concatenated morphemes'.

It is, however, more likely that many morphologists adhere to Aronoff's 'anti-decompositionalism': what happens inside lexemes is qualitatively different from what happens outside them (Aronoff 2007, 804). What Aronoff calls the classical lexicalist hypothesis (Jackendoff 1975; Aronoff 1976; Williams 2007) takes this position not only for derived nominals[2] but also for derivational morphology in general, and for compounding. However, if one wants to address the issue of idiomaticity in morphology as 'non-compositionality', one must assume that there is a 'word syntax', because, without hierarchical structure one cannot talk about compositionality in (formal) semantics.[3]

To illustrate the point, we will briefly discuss so-called 'bracketing paradoxes' (Pesetsky 1985). There is a well-known phonological restriction on the attachment of the nominalizing suffix *-er* in English. It is subject to the constraint that it may attach to monosyllabic adjectives (*green*/*green-er*), and a small class of bisyllabic ones with a light final syllable (*happy*/*happi-er*), while it may not attach to adjectives with two or more heavy syllables: *important* – #*important-er* – *more important*. Now take examples like *unhappier*. If the suffix *-er* obeys the phonological restriction, we

expect that the structure is (1b) and not (1a), because in (1a) the affix would attach to an adjective that is too 'heavy'.

(1) a. $[_A [_A \text{ un } [_A \text{ happy}]] \text{ -er}]$
     b. $[_A \text{ un } [_A [_A \text{ happi}] \text{ -er}]]$

But now look at the meaning side of the structure we have built: the meaning of the word *unhappier* can be paraphrased as 'more not happy', i.e. with 'more' having scope over *un-*, and crucially not as 'not more happy' with reversed scope. Hence the semantically motivated structure is (1a). A potential solution would be to assume *unhappier* has the bracketing in (1b) and call it an idiom, in the sense that the meaning cannot be derived from the way the meaningful elements are structurally arranged. However, there are many more cases of such bracketing paradoxes. Calling them all simply idioms is certainly an option, but we might be missing a point. Pesetsky argues that there is another way of solving these bracketing paradoxes by assuming that the affix is moved as in (2), as a kind of morphological Quantifier Raising (May 1977):

(2) a. $[_A \text{ un } [_A [_A \text{ happi}] \text{ -er}]] \Rightarrow$ b. $[_A [_A \text{ un } [_A [_A \text{ happi}] t]] \text{ -er}]$

In (2a) the *-er* suffix is attached in such a way that the phonological restriction is satisfied while (2b) reflects the meaning of the word in which 'more' takes scope over 'not'.

  Carstairs-McCarthy (1992, 139) points out that such an approach has a surprising by-product – the treatment of idioms at Logical Form (LF). Pesetsky discusses how the noun *rarity* has two meanings: 'the state or quality of being rare' and 'the fact of occurring rarely or in few instances, infrequency'. The former could be called compositional, the latter idiomatic. Now, the derived noun *unrarity* only has the compositional sense, 'not in the state or quality of being rare', Pesetsky claims, and not the perfectly conceivable 'not occurring rarely, being infrequent'. In the idiomatic sense there would be no well-formed bracketing. The categorial requirement for *un-* to be a sister to an adjective at LF imposes the bracketing (3a). However, Pesetsky suggests that there is a general property of idioms such that, for two constituents/morphemes (such as *rar-* and *-ity*) to have an idiomatic interpretation, they must be sisters at LF (cf. 3b).

(3) a. $[_N [_A \text{ un- } [_A \text{ rare}]] \text{ -ity}] \Rightarrow$ b. $[_N \text{ un- } [_N [_A \text{ rare}] \text{ -ity}]]$

These inconsistent requirements would account for the fact that there is no idiomatic reading for *rarity* in *unrarity*. The idiomatic reading is only available for (3b), but the categorial restrictions are not satisfied, therefore blocking the required reading. The notion of semantic idiosyncrasy is, thus, couched in terms of the structural properties that words have, comparable to those of phrases. More specifically the conditions under which idiosyncratic meaning are assigned

to words and phrases are alike. Whether or not this line of argumentation can be upheld (see further discussion in Carstairs-McCarthy 1992), it is a structural approach to meaning at the word level, making no distinction between lexical idioms and phrase idioms, and this issue will be addressed in Section 3.

2   Defining properties of idioms

It is unavoidable to first discuss how we define idiomaticity in order to know what we mean when we say that something is an idiom. Fraser's (1970) definition in (4) reflects most straightforwardly what is taken as an idiom in formal theorizing (generative grammar, formal semantics).

(4)   I shall regard an idiom as a constituent or a series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed.

The notion 'non-compositionality' is made central, giving us phrases such as *to have other fish to fry*, *a feather in one's cap* and *in a nutshell* as idioms. The definition from Mel'čuk (1995) in (5) differs from Fraser's in the sense that it avoids referring to compositionality, but there is more.

(5)   An idiom is a multilexemic expression E whose meaning cannot be deduced by the general rules of the language in question from the meaning of the constituent lexemes of E, their semantically loaded morphological characteristics (if any) and their syntactic configuration. (Mel'čuk 1995, 167)

The central concept in Mel'čuk's approach to idioms is the notion 'phraseme'. Phrasemes are 'set phrases', phrases that are not free, including what Fraser would classify as idioms. However, he also sees examples such as *black coffee*, *French window* and *bacon and eggs* as a proper subset of the class of phrasemes, and such examples would not normally be taken as idioms,[4] but rather as collocations.

Weinreich's (1969) definition is based on a careful lexical semantic analysis of word combinations, and he too doesn't commit himself to (non)compositionality:[5]

(6)   any expression in which at least one constituent is polysemous, and in which a selection of a subsense is determined by the verbal context, is a phraseological unit … A phraseological unit that involves at least two polysemous constituents, and in which there is a reciprocal contextual selection of subsenses, will be called an idiom. (Weinreich 1969, 42)

Weinreich's definition in (6) is twofold. The first part refers to a phraseological unit, a notion that is not fundamentally different from Mel'čuk's phraseme. Subsequently he defines idioms, in the sense of Fraser, as a subset of the phraseological units.

In order to be able to discuss idiomaticity at the word level it is important to focus on some issues in more detail: conventional language use (2.1), collocational restrictions (2.2), reflect more on Weinreich's important insights, and non-compositionality (2.3).

2.1 Convention and language use

When discussing idioms in morphology and syntax we cannot leave aside Hockett's insights on what the defining properties of idioms are. In Hockett (1956) too, an idiom is a grammatical form, the meaning of which is not deducible from its structure, violating, in a sense, a 'rule-based' semantics:

(7) By this definition, every morpheme is an idiom, since a morpheme has no grammatically relevant structure, but so are a vast number of composite forms. (Hockett 1956, 222)

Illustrating this position, Hockett (1958, 171) gives the example of the Chinese form *yóutǔng* having two morphemes *yóu* 'oil, grease' and *tǔng* 'large cylindrical container', the first morpheme modifying the second; 'Given this information, but knowing nothing else about Chinese or the culture of China, we can venture a reasonable guess as to the meaning of *yóutǔng*: "oil container," "oil drum," or the like.' This is indeed the meaning. However, confronted with *mǎshàng* built up from *mǎ* 'horse' and s*hàng* 'space on or above, top, ascend' guessing that the 'meaning of *mǎshàng* would be 'horse's back', 'horseback', or possibly 'on horseback' is wrong: the meaning is 'quickly, right away'. Hockett stresses 'it is one thing to consider a meaning reasonable after we know it, and quite different to deduce the meaning of a form from its structure.'

The following quote from Hockett (1956) adds a crucial dimension to what Hockett considers idiomatic:

(8) In every living language, new idioms are constantly being created, some destined to occur only once or twice and then to be forgotten, others due to survive for a long time … the mere occurrence of a nonce-form [= new/free expression, ME] for the first time does not in itself constitute the creation of a new idiom. An additional requirement is required: something more or less unusual either about the structure of the newly-produced nonce-form, or about the attendant circumstances, or both, which renders the form memorable … Given any such novelty, either of expression or of circumstances or of both, the event installs meaning into the linguistic form which is used, and the latter becomes idiomatic. (Hockett 1956, 222–223)

So, idioms can only be realistically defined in relation to a language community, i.e. only exist in a language community. The same tenet is found in Fillmore et al.:

(9) We think of a locution or manner of speaking as idiomatic if it is assigned an interpretation by the speech community but if somebody who merely knew the grammar and the vocabulary of the language could not, by virtue of that knowledge alone, know (i) how to say it, or (ii) what it means, or (iii) whether it is a conventional thing to say. (Fillmore et al. 1988, 504)

The last aspect, conventionality, is also important in the lexicographic tradition. For instance, Burger et al. (1982, 1) stress that a combination of two or more words is a phrasal lexical item only if the phrase is used in a language community as if it

was a lexical unit ('wenn die Wortverbindung in der Sprachgemeinschaft, ähnlich wie ein Lexem, gebräuchlich ist').[6]

   All these quotes add an interesting perspective, because they focus on the issue 'how do we normally say things' that is generally considered to be separate from meaning, the semantics of a grammatical form. This issue is also raised in Nunberg et al. (1994). They discuss a number of properties that are relevant for an understanding of what idioms are: inflexibility, figuration, proverbiality, informality, affect and conventionality. But Nunberg et al. (1994, 492) note that all idioms only necessarily have the property conventionality: 'Idioms are conventionalized: their meaning or use can't be predicted, or at least entirely predicted, on the basis of a knowledge of the independent conventions that determine the use of their constituents when they appear in isolation from one another.' The fact that *kick the bucket* means 'to die' or that we have to say *spill the beans* and not *spill the peas* does not follow from any underlying grammatical principle or from our knowledge of the world, but simply has to be learned. Their use of the notion 'convention' is based on Lewis (1969, 76), and this means that they adhere to the position that conventionality in language is a relation between a linguistic regularity, a situation of use, and a population that has implicitly agreed to conform to that regularity in that situation out of preference for general uniformity.[7] They, thus, seem to follow the tradition of idiomatic thinking of Hockett and Fillmore et al. (1988).

   Such an approach would include all the examples given above, and many more such as, for example, the 'pragmatic' idioms (10a) from Mel'čuk (1995) or the constructional idioms (10b−c) from Fillmore et al. (1988):

(10)   a.   [on packaging material of perishable food]
             best before … French: *à consommer avant* … German: *mindestens haltbar bis* …
                                      'to consume before'   'at least keepable till'
       b.   … *let alone* X
             'Martin won't drink beer, let alone whiskey.'
       c.   *the* X-*er*, *the* Y-*er*
             'The more, the merrier.'
             'The bigger they are, the harder they fall.'

More importantly, this means that idioms are first and foremost defined in terms of language use in a community. Idioms are, thus, what Chomsky (1986) would call an E-language[8] phenomenon, language as transmitted and shared in a community rather than a phenomenon as perceived and understood in the individual, an I-language phenomenon. In other words, this leaves the question open whether there is any reason, or any way, to define 'idiom' in the sense of I-language. Everyone seems to agree that idioms are part of 'that what needs to be memorized', what Di Sciullo and Williams (1987) have coined *listeme*. But, as they stress, this list is of no interest to the (formal) grammarian (i.e. those interested in I-language). If we interpret Distributed Morphology (DM; Halle and Marantz 1994) correctly, this also seems the perspective DM takes, given how the notion *encyclopedia* is defined. I will come back to this issue in Section 3.1.

2.2 Collocational restrictions

We will follow the line of reasoning of Section 2.1, taking 'conventionality' as a defining feature of idioms. Taking idioms as a regular behavioural pattern in a linguistic community comes close/is identical to saying that all idioms are, whatever their semantic transparency, first (and foremost) collocational restrictions.[9] However, as Weinreich (1960, 338) notes, if we would, for instance, go along with Hockett's position this 'enlarges the domain of idioms rather monstrously'. And he is right, it might result in a position that Mel'čuk (1995, 169) described thus: 'People do not speak in words; they speak in phrasemes.' It adds a crucial language-use dimension to what idioms truly are, 'actual words/phrases'. In fact, Mel'čuk denies I-language as an object of inquiry, just as Gross (1979) rejected the generative enterprise using idioms and collocational restrictions as an argument. Halle's (1973) approach to word formation, identifying the need for a dictionary, the set of all actually occurring words, recognizes the need for both the I-language and E-language dimensions of the lexicon.

Taking this 'liberal' notion of idiom – the phraseological unit of Weinreich, the phraseme of Mel'čuk, an actual phrase/word – idioms are characterized as conventionalized linguistic expressions which can (but need not) be decomposed into potentially meaningful components and exhibit co-occurrence restrictions that cannot be explained in terms of general rule-governed morphosyntactic or semantic restrictions. This would mean that the co-occurrence restrictions, what one could call the collocational dimension, are at the core of what an 'idiom' is.

Collocational dictionaries, like the BBI (Benson et al. 1986), often give a string of choices in a collocation. This also plays a role when discussing 'true' syntactic idioms as is illustrated in (11). Limiting the discussion to *kick the bucket* and *pull the strings* might give the impression that idioms are always completely lexically fixed (apart from open slots as in *know something like the back/palm of one's hand*). That is not the case; idiom dictionaries often give limited choices, as in the following examples:[10]

(11) a. *on/along the lines of* NP
      b. *a dog's breakfast/dinner*
      c. a *white/blue collar job*
      d. to *drop in the bucket/ocean*
      e. to *gather/collect one's wits*
      f. to *count oneself fortunate/lucky*
      g. to *cry/laugh all the way to the bank*
      h. to *dash/shatter* somebody's *hopes*
      i. to *create hell/mayhem/murder*

While *hell* in (11i) can be replaced with *mayhem* or *murder*, we do not see the same lexical choice with the verb *raise* in the case of the idiomatic construction in (12). Likewise, both the verb *collect* and *gather* can be combined with *dust* in (13a) but not in (13b).[11]

(12)   to *raise hell/#mayhem/#murder*

(13) a.   to *gather/collect dust*
      b.   *The dust settles/#collects.*

In some constructional idioms the choice of words is larger, but still limited, and often restricted by a semantic condition on the lexical choice.

(14) *a* day/minute/week/ … *or two*: 'a short time period'
      a centimetre/inch/ ... *or two*: 'a short length'

(15) day/week/month/year/ … *in* day/week/month/year/ … *out*: 'regularly in time'

One could debate whether light verb frames like '*have* a N', '*take* a N' or '*give* a N' in (16) show simply lexical idiosyncratic choices between the light verb and the light verb nominal (Prince 1974), or that such restrictions follow from the lexical semantic properties of the verb and the nominal (Wierzbicka 1982). If we take the latter perspective, the semantic subtleties one needs to account for are illustrated by the Dutch examples in (17), with (17a) giving the light verb variant of the lexical verbs in (17b).[12]

(16) a.   have a drink/*an eat, have a try/*an attempt, etc
      b.   make/*do a proposal versus *een voorstel doen/*maken* (Dutch)

(17) a.   *Ik stelde haar*   *de vraag/#het verzoek hiermee akkoord te gaan.*
           deed   *#de vraag/het verzoek hiermee akkoord te gaan.*
      I   posed her the question/the   agree to go
           request with it
      b.   *Ik vroeg/verzocht haar*   *hiermee akkoord te gaan.*
      I   asked/requested her with it   agree to go
           'I asked her to agree with it.'

It is a well-known fact that the lexical semantic function 'intensity/magnitude' is often collocationally restricted (Mel'čuk and Zolkovsky 1984, among others). To express that there is a lot of rain we would say *heavy rain*; it is not impossible to say #*strong rain* (like *strong wind*) but it is not the unmarked choice. Likewise *stark naked* expresses that one is completely without clothing; #*stark nude* could be used, but it is simply much less frequent. One could take the example in (18a), generally taken as an idiom (see Section 3.2), as another illustration:

(18) a.   It is raining cats and dogs.
      b.   #It is snowing/hailing cats and dogs.

The adverbial complement *cats and dogs* expresses that it is raining heavily. It is a partly compositional idiom in the sense that the verb retains its meaning, but *cats and dogs* in this meaning can only be linked to *rain*, not to other verbs in the same lexical semantic domain (cf. 18b).

   The magnitude lexical function is also found in compounds. To illustrate this with an example from Dutch, the examples in (19) illustrate the category-neutral prefix/affixoid[13] *oer-* 'primal' (19a) and *steen-* (literally 'stone') (19b), all expressing a 'high degree of' when attached to adjectives.

(19) a. *oeroud* 'very old/ancient', *oersterk* 'very strong', *oerlelijk* 'very ugly/repulsive', *oergezond* 'very healthy', *oerdegelijk* 'extremely solid/sound', …

 b. *steendood* 'stone-dead', *steengoed* 'fantastic, great', *steenhard* 'rock hard', *steenkoud* 'freezing cold', *steenrijk* 'immensily rich'

Van der Wouden (2020) takes the examples in (19a) as the result of a productive process with a compositional semantics. Contrary to that, the examples in (19b) are, according to Booij and Hüning (2014), the only AN-compounds beginning with *steen*. They note that, for instance, *steenslecht* 'very bad', the opposite of *steengoed* 'very good/fantastic', is not listed in the dictionary.[14] Assuming that the intensifying element is indeed affix-like we would call (19b) an instantiation of an unproductive word formation rule. But what if we call it a collocational restriction, comparable to the 'unproductiveness' of the adjective *heavy* mentioned above? In itself there is no reason not to do that, but the notion collocation is usually restricted to syntax. The examples in (20), again compounds that feature the magnitude lexical function, might also be called cases of morphological collocational restrictions, albeit of a different type. The Dutch adjective *vol* 'full' has several compounds of the type [$_A$ [x] [$_A$ vol]] all of which have the meaning 'very full', but not all are used in the same context:[15]

(20) a. *De tafel ligt*   *vol, stampvol, tjokvol, #boordevol, propvol boeken*.
          The table lies   x-full x-full x-full x-full   x-full   books
          'The table is full of books.'

 b. *Ik schonk het glas*   *vol, #stampvol, #tjokvol, boordevol, #propvol*.
          I poured the glass   x-full x-full x-full x-full
          'I poured the glass to the brim.'

 c. *Zij zit*   *vol, #stampvol, tjokvol, boordevol, #propvol ideeën*.
          She sits   x-full x-full x-full x-full   ideas
          'She is full of ideas.'

 d. *De kroeg zit*   *vol, stampvol, tjokvol, #boordevol, propvol*.
          The pub sits   x-full x-full x-full x-full
          'The pub is filled with people.'

Judgements may vary between speakers, but it is clear that the compounds are syntagmatically not freely interchangeable. The lefthand side of the compound brings about collocational restrictions of the compound as a whole, collocational restrictions which the head itself doesn't have.

2.3   (Non)compositionality

Compositionality is, in linguistics, generally defined as in (21a) from Partee (1984, 153) or (21b) from Winter (2016, 29), stressing that structure is part and parcel of its definition:

(21) a. The meaning of an expression is a function of the meanings of its parts and of the way they are syntactically combined.

 b. The denotation of a complex expression is determined by the denotations of its immediate parts and the ways they combine with each other.

Compositionality is a central principle of formal semantics concerning the relation between syntax and semantics.[16] But note that semanticists keep compositionality distinct from under-specification of meaning, which we could call vagueness. Formal pragmatic theories consider the way in which sentences are used, and the effects of context on their use, without claiming that their denotations differ. As Winter explains, classic versions of formal semantics did not aim to resolve vagueness, but current semantic theories often interact with pragmatics and describe the way context helps in resolving vagueness in actual language use.

Idioms are always discussed in the context of compositionality because they seem to be a clear example of syntactic expressions that do not obey this principle.[17] But it is important to understand that in formal semantics, 'the principle of compositionality is not an empirical hypothesis. Rather, it must be viewed as a methodological principle, one that represents a choice to do semantics in a particular way' (Groenendijk and Stokhof 2005, 84, based on Janssen 1983). And, to add to that, 'while it makes good sense to ask if a semantics is compositional or not, it makes no sense to ask the same question about a particular phrase' (Westerståhl 2002, 242).[18] The importance of these observations will become clear in the remainder of this section.

Weinreich (1969) contains an important and careful discussion of the semantics of idioms. He illustrates how an idiomatic sense of a complex expression may be the result of a suppression, addition or replacement of some component of meaning. He argues that many morphemes (words) appear in a dictionary with more than one sense. Each sense is contextually specified, and the contextual features may be of several kinds. They may be syntactic; the subsense may vary for transitive (*to walk the dog*) versus intransitive use (*to walk*) of the verb. The contextual feature may be semantic; *blind* has at least two senses, viz. 'unseeing' and 'without exit at opposite end'. The latter is only available in *blind alley*. Finally, the selection of the subsense might be dependent on the contextual presence of a specific morpheme: *blind date* versus #*blind appointment* (Weinreich 1969, 40–41).[19] Weinreich adds that the contextual specialization may work both ways: subsense X on R requires the presence of S, while subsense Y on S requires the presence of R. This is an important theoretical innovation, which we will illustrate below because it allows non-heads to contain contextual restrictions on heads. But before doing so, I will briefly discuss the GPSG/HPSG-approach to idioms because it is the most influential approach to idiomaticity in 'generative grammar', but without really acknowledging and addressing Weinreich's contribution.

In the GPSG(/HPSG) framework, adopted by Nunberg et al. (1994), it is important that there is a one-to-one correspondence between syntactic rules and their semantic consequences. That means that potential restrictions on syntactic flexibility are to be found in the semantics (and pragmatics) of the constructions. In their analysis of idioms, the dependency between the verbs and their objects is considered as semantic in nature. And given this, the fact that the idiom *spill the beans* allows passivization (22a) but *kicking the bucket* does not (22b), must be grammatically accounted for.

(22)   a. to spill the beans / the beans were spilled
       b. to kick the bucket / #the bucket was kicked

Gazdar et al. (1985, 238), following Wasow et al. (1982), suggest an account that makes essential use of the notion of partial function. So, the verb *spill* is assigned two senses (23a); likewise, the object is assigned two senses (23):

(23) a. spill′: literal sense   / spill″: idiomatic sense, roughly 'divulge'
      b. beans′: literal sense   / beans″: idiomatic sense, 'information'

Their compositional semantics will treat the translations of transitive verbs as partial functions from NP-type intensions to VP-type denotations. Such a function may be defined at only one argument, deriving (24):[20]

(24) f (spill′ (the-beans′)) = spill″ (the-beans″)

The fact that (22a) is allowed is accounted for by assuming that it is a syntactically complex lexical item associated with a single undecomposable meaning (Gazdar et al. 1985, 244). The fact that (22b) is not allowed means that the partial function approach in this case simply doesn't apply.

Nunberg et al. (1994) move away from this (and their 1982) analysis. They argue that although their analysis requires the idiomatic NP and the idiomatic verb to co-occur for semantic reasons, it does not exclude the possibility that an idiom chunk could also be semantically compatible with other expressions. Dropping the partial function requirement, they need to assume that the non-literal meaning is 'conventionally and homomorphically associated'. In other words, the partial function approach is replaced by an approach making use of collocational restrictions. Note that it is simply not possible to define the subset of phrasal expressions they call Idiomatic Phrases (22a) or Idiomatically Combining Expressions (22b) on the basis of semantic considerations alone. It is only on the basis of their syntactic flexibility that one could distinguish them.[21]

Note also that the inflectional morphology of the verbal heads in these two classes of idiom does not differ (25). Likewise, the irregular verbal inflectional morphology does not differ in Dutch idioms (26) compared to non-idiomatic phrases (27), and the same holds for the phenomenon of verb-second (26b versus 27b), as observed by Schenk (1995) for Dutch, Bargmann and Sailer (2018) for German:

(25) a. Mary *kicks/kicked/is kicking* the bucket.
      b. Mary *spills/spilled/is spilling* the beans.

(26) a. Ze *heeft* mij *de loef af-gestoken*.
          she has me the luff (weather side) off-cut
      b. Ze *steekt/stak* mij *de loef af*.
          she cuts/cut me the luff off
          'She gets/got the better of/outstrip/outdo someone.'

(27) a. Ze heeft de bocht *af-gestoken*.
          she has the corner off-cut
      b. Ze *steekt/stak* de bocht af.
          she cuts/cut the corner off
          'She cuts/cut the corner.'

More importantly, it is possible to create a compositional derivation of both *to spill the beans* and *to kick the bucket* and, so, still preserve direct compositionality. With some non-trivial differences, Sailer (2003), Richter and Sailer (2004), Everaert (2010), Kay and Sag (2014) and Bargmann and Sailer (2018) sketch such derivations. Crucially, all these analyses make use of Weinreich's insights. Keeping close to what Weinreich suggests we might argue that *to kick* has two subsenses, KICK and DIE, and that *bucket* also has two subsenses, BUCKET and $\emptyset$, and that the *kick*-meaning DIE and the *bucket*-meaning $\emptyset$ have lexical contextual features:[22],[23]

(28) a. $kick_1$ MEANING: 'kick'   CONTEXTUAL FEATURE: [- NP]
          $kick_2$ MEANING: 'die'   CONTEXTUAL FEATURE: [- *the bucket$_2$*]
      b. $bucket_1$ MEANING: 'bucket' CONTEXTUAL FEATURE: [-]
          $bucket_2$ MEANING: $\emptyset$ CONTEXTUAL FEATURE: [$kick_2$ -]

In this way the literal and the idiomatic reading can be computed; the modes of composition in the idiomatic and nonidiomatic reading are distinct, but strictly speaking not violating compositionality.

There are more 'natural' examples of the strategy sketched above. It concerns idioms, reminiscent of cases discussed in Weinreich (1969), that are partially or fully semantically transparent. Example (18a), here repeated as (29a), is only 'partially idiomatic': *cats and dogs* meaning 'heavily' in (29a) might itself be idiomatic, but once combined with the predicate *rain* the derivation is compositional, the verb retaining its normal meaning. In essence it is not very different from bound/unique morphemes in compounds such as (29b), to which I will come back in Section 3:

(29) a. It is *raining cats and dogs*.
      b. *cran*berry, *mul*berry, *cob*web, *luke*warm

What would make such implementation of compositionality unacceptable? Not the fact that some context sensitivity is built into the meaning of lexical items (cf. Kamp and Partee 1995), because using partial functions is exactly doing that. However, one might say that the way in which meanings are defined and combined is not 'natural'. The reason for that would be that the meaning of *bucket* in the idiomatic reading is empty, and the idiomatic subsense of *kick* is quite unrelated to the other subsense.[24] Essentially, what these examples illustrate is the point that Partee (1984, 281) made: 'the principle [of compositionality, ME] can be made precise only in conjunction with an explicit theory of meaning and of syntax, together with a fuller specification of what is required by the relation "is a function of". If the syntax is sufficiently unconstrained and meanings are sufficiently rich, there seems no doubt that natural languages can be described compositionally.'

2.4   Compositionality and productivity at the word level

Above we have shown that any theory of idioms needs to incorporate contextual lexical specifications; in other words, collocational restrictions. To be more precise, one could say (and that certainly holds for all lexicographers) idiomaticity is only

defined by collocational restrictions given the fact that it is extremely difficult to give a proper limitation to what one counts as a compositional analysis. However, for many linguists idiomaticity is defined by compositionality, and as a result a lot of morphological objects are idiomatic because it is very difficult to write a compositional semantics for derivations and compounds. We will illustrate the point below.

Dowty (1979, 295) notes that it is generally agreed that principles of word formation are real enough principles that must be described in any account of a native speaker's knowledge of his language. However, these principles are everywhere subject to exceptions, both in productivity – the level to which potential words are also actual – and in semantics – how the meaning is often not determined by the meanings of the parts.

In the domain of morphology, the distinction between potential and actual is not at issue. Both concepts are addressed in various theories of word formation. For syntacticians focused on I-language, the distinction between potential phrases/sentences and actual phrases/sentences does not exist. However, we have explained above that if we take language use into account, this distinction becomes real. Phrasal collocations are what one could call actual phrases.

Let us concentrate on compositionality in derivational morphology. The question is whether the semantics of morphological rules and the semantics of syntactic rules are different. The semantic interpretation of a phrase/sentence is determined by its structure and the meanings of the words (always relative to a context). The semantic contribution of a syntactic structure is recursively specifiable and, thus, prevents syntactic productivity from creating interpretive problems (Chierchia and McConnell-Ginet 1990, 367). We could also formulate it the other way around: compositionality implies/leads to productivity. So, it is important to understand the link between compositionality and productivity.

Interpretation of a morphologically complex word depends on the interpretations of its constituent expressions and on the morphological structure resulting from the word formation processes involved. Dowty (1979, 296–297) observes that a native speaker can not only distinguish between actual derived words and non-words but can also distinguish an intermediate class of possible but non-occurring words. He observes that English natives readily agree that *beautify* is a word of English whereas *uglify* is not. Still both conform to the same pattern of word formation and *uglify* would clearly mean 'make more ugly' if it were a word that was used.[25] Likewise native speakers of English know precisely what Chandler (a character from the series *Friends*) means when, talking about how 'to mess' with his friends, he says 'the messers become the messees', even though they might not ever have heard the words before.[26] Speakers remember particular derived words that they have heard other speakers use, but they are in general cautious about using a derived word they have not heard in common usage before, even if it conforms to a familiar pattern.[27]

According to Dowty (1979) speakers know idiosyncratic details of the meanings of various derived words which are not predictable by rules. This suggests individual learning. As Anderson (1992, 193) puts it, once something is registered as a word it concomitantly takes on independent lexical existence, it becomes an actual

word. Anderson argues that it is quite natural because, following Aronoff (1976), a word (in the sense of a building block for syntax) is a 'sign', and 'the sign relation is a holistic one between a possibly complex form and its possibly complex meaning'. In other words, language use in a community allows the word to move away from its compositional meaning.

Knowledge of word formation rules and the semantic rules associated with these allows the speaker to know the approximate meaning of a newly derived word upon first hearing it, or to utter a new word which participants in the conversation will understand approximately. But such rules will not prevent them from later deciding that the word has a more specialized meaning than that specified by the general rule. Such meaning details will be inferred by induction on the contexts in which it is used. Let us illustrate this with an example from Dowty (1979). The rule for the suffix *-able* could be as in (30):[28]

(30) a. If α is a member from V, V = transitive in F3 [a fragment of a grammar of English, ME], then [α + *able*] can be added as a member of A.
  b. If $\beta = [_A [\alpha]_V \ able]$, the following constraint is placed on β':
  $\lambda \forall x[\beta'(x) \rightarrow ((\exists y \ \alpha'(y,x))]$.

The semantic component of the rule (30b) constrains the interpretation of intensional predicate calculus translations of words syntactically derived by (30a). The semantics of an adjective triggers an entailment of the form specified in (30b), but leaves open what the meaning of an actual word is in a specific context. Whenever a word is newly coined, like in (31a) (from Katamba and Stonham 2007, 309), the 'true' semantics is clearly visible:

(31) a. Mr Kinnock is not Prime Ministerable.
  b. *ministerable* = has what takes it to be made a Prime Minister.

Summarizing, in derivational morphology the contextually specified meaning of a word often goes beyond what is determined by the interpretation of its constituents and their structural relations. So, these word formation rules typically constrain, but do not determine the interpretation of morphologically complex words relative to an interpretation of the basic morphological units: they are compositional but allowing vagueness.[29]

2.5   Conclusion

The conclusion of this section is that it is not immediately straightforward to use compositionality as a way to make a distinction between expressions (at the word level and at the phrase level) that we perceive as idiomatic and expressions that we perceive as not idiomatic. However, one could (i) allow one's semantics to make use of contextual specification that is defined in terms of phonological forms (cf. the discussion of (28) above), and (ii) put restrictions on what are acceptable subsenses of a basic expression (requiring a theory on how to define polysemy and distinguish it from homonymy).

In other words, if we adopt what one could call Weinreich's approach – adopting lexical contextual features, incorporating collocational restrictions (conventionality) combined with a lexical semantic theory of polysemy, in our grammatical machinery – we will have a principled way of characterizing 'semantic idiosyncrasy', thus achieving descriptive adequacy. Perhaps it makes sense to do that, but it would blur the distinction between semantics (I-language) and pragmatics (E-language) made above. One could also decide not to do so. The grammar machine gives you *The bucket was kicked*, but there are (semantic) reasons for this sentence not to be used. Likewise, grammar gives you both *beautify* and *uglify*, but the latter is not readily used, and therefore not listed in what one could call the 'lexicon'. Section 3 will address such issues.

3   Idioms and the lexicon

Using the word 'lexicon' we must be clear what the notion means, because it is a contentious notion (cf., among others, Scalise 1984, Spencer 1991, Carstairs-McCarthy 1992 and Aronoff and Anshen 2017). If we adopt the Bloomfieldian (1933) sense, it means we refer to the lexicon as a store consisting of 'a list of arbitrary pairings of form and meaning, regardless of the category they belong to' (Aronoff 1994, 21), a list of basic irregularities in the grammar of any language. So, if the sense of a word is unpredictable, irregular, 'lawless', it must be stored in the warehouse called the lexicon. And the same holds for phonological or morphological idiosyncrasy. This is, in essence, the notion 'listeme' that Di Sciullo and Williams (1987, 14) have in mind: the set of idiosyncratic items, 'listed units', primes, primitives of language. Morphemes must be listed because that is the only way to know what the building blocks of words are. But also larger units, composed words or phrases must be listed because they have an idiosyncratic meaning or some other feature that does not follow from its composition. However, it is crucial to observe that such a notion 'lexicon', a list of elements, would have no theoretical meaning (cf. Section 2.1).

As Aronoff notes, in the Bloomfieldian sense of the lexicon we are talking about that which any individual speaker holds in memory, in his/her mental lexicon, a list. The difference between which words exist and which are potential is defined solely in terms of the individual's lexicon and morphological rules. A word that meets all the criteria for being a word of the language but that is not in an individual's mental lexicon does not exist for that person (even though it may exist for another individual).

We could also think of the lexicon as the infinite 'set of potential (regularly derived or compounded) lexemes for any given language' (Aronoff 1994, 10). So, for instance, take a word that is newly coined. If the meaning of the word does not diverge from its predicted meaning, based on its parts and its morphological structure, there is no need for this word itself to be listed. From a lexicalist point of view the generative lexicon incorporates word formation rules (like (30)), but not necessarily the output of those rules.

These two notions of 'lexicon' are unrelated theoretical entities. In both cases the unlisted word is a potential word, and we could argue that morphologically

well-formed complex potential words are provided by the morphology for Bloom-field, but for Aronoff they are the result of the lexicon, although not in the lexicon as such.

This distinction we just discussed reminds us of Halle's distinction between actual and possible words. In Halle's view, the basic units of the lexicon are morphemes, containing unanalysable stems, and affixes. Word formation rules operate in the same way for derivation and inflection.[30] The output of the word formation rules are subject to a filter, basically doing two things. Firstly, the filter adds idiosyncratic features when necessary (both in terms of meaning and sound); that is words whose meanings are not fully determinable from their component morphemes and the word formation rules, and words with phono-logical idiosyncrasies (such as *obesity*).[31] Secondly the filter prevents possible but non-existing words from appearing in sentences by assigning the feature [-lexical insertion] to them, for instance *derival*. Adding this feature creates a dictionary, containing all existing words, with all the inflected forms of every word included. Note, however, that this dictionary is not the mental lexicon of an individual, but something closer to a dictionary in the traditional sense.

In the work of Aronoff and Jackendoff we can distinguish another dichotomy of looking at actual versus possible words. For Jackendoff, the finite set of 'existing' words is listed. It doesn't matter whether their formation is regular or not, whether their meaning is transparent or not. His lexical redundancy rules (Jackendoff 1975) are formulated over the list of words, the words used in a linguistic community. For Aronoff (1976, 43), however, all and only those words which are exceptional, that is arbitrary, idiosyncratic in (some aspect of) their feature constellation, will be part of the lexicon. The Aronovian lexicon, in the sense of listemes, is therefore a subset of the Jackendovian. For instance, the word *think-ing*, being an existing word, will be part of the Jackendovian lexicon, but not the Aronovian one, given the fact that it is regularly (both semantically and phonologically) derived by a morphological rule combining the verb *think* with the affix *-ing*.

There are two frameworks that explicitly address the status of idioms in what one could call morphology, or at least the component of grammar that deals with words in the broad sense of the word: DM (3.1) and CM (3.2).

3.1   Distributed Morphology[32]

For DM the 'syntax of words' is indeed part of syntax, taken as the computational system (Chomsky 1995). We are not going to discuss the full machinery of DM, but limit ourselves to the listing dimension of DM. One can distinguish three distinct lists:

1.   The *Syntactic Terminals*: the list containing the roots and the abstract mor-phemes. Abstract morphemes are the morphosyntactic features which are combined (through syntactic mechanisms) with each other or roots.
2.   The *Vocabulary*: the list of vocabulary items, the exponent list. The vocabulary is the list of the phonological exponents/signals of the morphosyntactic fea-tures of a language, paired with conditions on insertion. Each such pairing

of phonological content with information about the grammatical (i.e. syntactic and morphological) context in which the exponent is inserted is called a vocabulary item. So, vocabulary insertion provides phonological content, a spell out.

3. The *Encyclopedia*: the repository for 'special', partly non-linguistic meanings. The encyclopedia gives the semantic information that must be listed as either a property of a root, or of a syntactically constructed object. To put it differently, it relates (configurations of) vocabulary items, creating compositional meanings, to non-compositional meanings. So, it is the place where, ultimately, sound-meaning pairs are encoded. However, note that the encyclopedia adds 'contextually conditioned non-compositional meaning' which is considered extralinguistic: 'noncompositional meanings of a node may be conditioned by the syntactic environment of the node' (Marantz 1996). So, the encyclopedia lists only the 'special meanings' attached to output of syntax; meaning divergent from what is compositionally derived.

Note that in DM both words, like *cat*, and phrases, like *kick the bucket*, are derived by syntax. If one assumes that the noun *cat*, is derived from the root $\sqrt{\text{CAT}}$ after merger with a categorizing head n, and that what it means – furry domestic feline – is not predictable, it could be called an idiom. Its meaning must be listed, and this is done in the encyclopedia. Following Jackendoff (1997), Marantz stresses that there is no empirical difference between the idiosyncratic meaning of words and phrases.[33]

Thus, the encyclopedia contains the idioms in a language, referring to any expression (even a single word or subpart of a word) whose meaning is not wholly predictable from its morphosyntactic structural description (Marantz 1996, 1997). This formulation does not explicitly commit oneself to compositionality in the formal semantic sense, but it is not excluded.

What are the consequences of this approach? If the encyclopedia encodes 'special meanings', quite a lot could be listed in the encyclopedia. As Siddiqi (2009, 21) points out, words like *scissors* and *terrific* have to be memorized as not being compositional:

(32) a. *terrific* is not compositionally derived from $terror_N$ / $terr\text{-}ify_V$ / $terr\text{-}if\text{-}ic_A$
     b. *scissors* is not compositionally derived from $scissor_{N\text{-}sPL}$

Nevins (2016) gives another example. When we use the word *compárable* as in (33a), the meaning of the word is 'able to be compared'. But there is also *cómparable* in (33b), meaning 'equal, equivalent':

(33) a. *compárable* in 'The contestants are of comparable ages, all in their early twenties.'
     b. *cómparable* in 'The theories are, in essence, comparable.'

Nevins observes that the latter form bears an 'unpredictable' meaning aspect; if you know what *compare* means, and what *-able* means, you won't necessarily know

what *cómparable* means, simply combining the two. The meaning in (33b) will have to be listed in the encyclopedia.[34]

But we can go one step further. If the encyclopedia is the place for encoding non-linguistic knowledge, quite a lot becomes part of the encyclopedia, whether one wants to call it idiomatic or not. Thus, being listed in the encyclopedia comes close to what Hockett calls an idiom (cf. (8)). The encyclopedia seems to be an extensionally defined set of primitives, it encodes knowledge of actual language use, and as such might not be considered part of the object of inquiry in generative grammar: I-language.[35]

Note that there is a consequence to this line of reason that seems to be often ignored. Marantz (1997) makes clear that the compositional meaning of an idiom is not lost. The encyclopedia encodes extra-linguistic knowledge, but the expression is regularly derived and that means that the constituent elements preserve their compositional meanings. This fits in with the well-known observation from Nunberg (1978) that the opposition between #*He kicked the bucket slowly* and *He died slowly* is derived from the fact that the lexical aspect of the verb *to kick* is retained in the idiom.[36]

This approach also gives a perspective on how to deal with bound roots like *-ceive* (Marantz 1996). The root *-ceive* has a non-compositional interpretation entirely dependent on its context – *con-ceive, de-ceive, re-ceive, per-ceive* – but, of course, also a compositional dimension because the semantic relation between the prefixes and the root is preserved.

So, the encyclopedia encodes meaning that is not part of the denotational semantics of an expression. It is a meaning layer on top of a regularly built syntax of meaning, that is also still accessible.

3.2   Construction Morphology

Construction grammar is in essence an idiomatic approach to language. Taking the notion 'construction' as a conventionalized form-meaning pairing (a sign) it uses the core notion of idiomaticity, convention (cf. Nunberg et al. 1994), as the basis for theorizing about the syntax of words and phrases. That is, phrases and complex words are not seen as a hierarchical concatenation of words/morphemes, but as independent meaningful units, Saussurean signs, and there is no necessary connection between a given structure and the meaning it can convey. However, within the sign certain subcomponents, for instance morphemes, may be distinguished based on paradigmatic relations with other words.

CM develops Jackendoff's lexical redundancy rules into an all-encompassing framework. The concept of a lexical redundancy rule/constructional schema requires all words to be listed in the lexicon.[37] That is, lexical redundancy rules are analytic rules that are used to analyse stored forms. Facets of the syntactic, semantic and phonological features of words are expressed by the redundancy rules. They generalize over the stored forms, with all their irregularities and idiosyncrasies, and tease out those properties of complex forms that can be predicted (such as their word class, and many details of the phonology such as the stress pattern). To give an example from Booij and Audring (2018, 59), the set

of de-adjectival nouns in German of the form [A-*heit*]$_N$, words such as *Schönheit* 'beauty' and *Wahrheit* 'truth' can be characterized by the following constructional schema:

(34) $<[A_i\text{-}heit]_{Nj} \leftrightarrow [\text{Property of } SEM_i]_{SEMj}>$

The CM architecture puts schemas/constructions and words at the same level, resulting in a network in which unproductive schemas can be linked to their instances without having to generate them. This allows for the relation to be partial, leaving room for idiosyncrasies such as the non-compositional semantics of a word like *recital* (derived from *recite*), the irregular phonological relation of words like *destroy* and *destruction*, and the absence of a root in *impetuous* (Masini and Audring 2018). In such an approach idiomaticity and productivity do not exclude each other but go hand in hand in a number of constructions. The other way around, productivity does not necessarily imply compositionality, which is the key concept behind the notion of 'constructional idioms' (cf. Jackendoff 1990; Booij 2002, 2010).

A constructional idiom is a fixed syntactic pattern in which some positions are open slots and may be filled by all kinds of words of the right category, whereas other positions are phonologically specified, filled by specific morphemes or words. Booij (2005) argues that English compounds like *Dówn's syndrome* and *Múrphy's law* are instantiations of a constructional idiom 'N's N' that serves to create new lexical expressions. In this case, there is only one morpheme lexically specified, the morpheme *-s*. The two N positions are variable and can be filled by all sorts of nouns. Another example from Masini and Audring (2018) is Dutch NN/NA compounds, of which the left (modifier) constituent may develop a more abstract meaning of 'intensification' (see also Section 2.2). See (35a–b) for examples with the noun *reuze* (derived from *reus* 'giant' plus a linking shwa *-e*) and *dood* 'dead', followed by an adjective.

(35)   a.   *reuze*+A: *reuze-aardig* 'very kind', *reuze-leuk* 'very nice'
    b.   *dood*+A: *dood-gewoon* 'very normal', *dood-eng* 'very scary'
    c. $[dood/reuze]_{Ni} [y]_{Aj}]_{Ak} \leftrightarrow [\text{very}_i \ SEM_j]_k$

The overall semantics is partly compositional, the meaning of the lefthand part of the compound is contextually determined, it deviates from its 'normal' meaning and the construction is productive.[38] Arcodia and Basciano (2018) argue that in Chinese there are many compounds with constituents that do not appear as words by themselves, even though they have a lexical meaning. These roots can be specified by constructional idioms that define the class of compounds with that root, and the corresponding meaning.

4   Morphological idioms

From the preceding section it has become clear that, depending on how you define an idiom, there are quite a lot of 'morphological idioms' (cf. also Jackendoff 1995).

In this section we will discuss some cases. We will not discuss derivation, given the fact that such cases have already been discussed above in Sections 2.4, 3.1 and 3.2.

4.1   Compounds[39]

Compound idioms abound in English, from *mishmash* 'a confused mixture' to *ice cream* 'a soft, sweet, frozen dessert/refreshment, made with milk and cream'. Noun–noun compounds of English allow virtually any type of interpretation (Downing 1977), which means that listing becomes inevitable as it is hard to define any regular semantic rule. Downing (1977, 183) gives the example in (36): the compound identifies a person as described in the gloss:

(36)   *bike girl*
       'a girl who left her bike in the vestibule'

   Downing's description – 'a friend now carries the name the bike-girl … She carries this name not because she habitually parked her bike in the vestibule; indeed, she did so on only one occasion. But this single incident was considered by the coiner of the compound to provide sufficient information for the categorization of this girl …' – exactly fits what Hockett's quote in (8) pointed to: in language use, for a specific context, meanings may randomly be attached to composite forms, and they might, but need not, fossilize. Clark and Clark (1979) describe this process (for noun compounds and denominal verbs) from a diachronic perspective:

(37)   Nearly every denominal verb in the lists above, we assume, was introduced into English as an innovation. This is generally confirmed by the OED – though there may be exceptions. Each verb, then, arrived at its present form by a complex historical process we will call IDIOMATIZATION. Because the process is gradual, each verb passes through several stages on its way from innovation to idiom; and because fresh verbs are being introduced into this process all the time, at any moment there are verbs at each stage. (1979, 804)

   It is tempting to mention here the distinction between decoding and encoding idioms (Makkai 1972), a distinction that is again highlighted in Fillmore et al. (1988). In a decoding idiom one cannot figure out the meaning just by knowing the words and grammar of the language. Encoding idioms are to some extent transparent; based on the meanings of the words one could, in principle, figure out the meaning of the idiom, but one would not know how to use the idiom, given that it is a conventional way of expressing that meaning. The decoding idioms are what one generally considers the 'real' idioms, the non-compositional ones. The *bike girl* example in (36) is, in my view, an encoding idiom given that it is, in the small community described by Downing, the conventional way of referring to a specific girl. One might argue that the same holds for the collocations in (20).
   A more straightforward example of a compound idiom, a decoding idiom, is the word *aardvark*. English speakers knowing the word need to memorize that

this phonological form must be linked to a specific type of animal, the mammal *Orycteropus afer*. However, some will know that this word is a compound originally borrowed from South African Dutch (Afrikaans *erdvark*) and is composed of Dutch *aarde* 'earth' and Afrikaans *vark* 'pig' (Dutch *varken*).[40] Apart from *mishmash*, mentioned above, some more examples of N–N compounds are given in (38): the examples in (38a) are quite opaque, the examples in (38b) less so for some speakers, but all of them for most lexicographers pretty idiomatic:[41]

(38) a. *claptrap* 'cheap showy sentiment; nonsense, rubbish'; *goody-goody* 'someone who behaves in a way intended to please people; an ostentatiously virtuous person'; *nitwit* 'a stupid, silly, or foolish person'; *hogwash* 'any liquid for drinking that is of very poor quality; worthless, ridiculous, or nonsensical ideas, discourse, or writing'; *landlord* 'a man or woman who rents property to others'; etc.
    b. *agony aunt* 'an adviser on personal, psychological, etc., problems'; *bellybutton* 'navel'; *private eye* 'a private detective'; *fairytale* 'any of various short tales having folkloric elements and featuring fantastic or magical events or characters; something resembling a fairy tale in being unreal or incredible, or in having an idealized happy ending'; etc.

Some more encoding type of compounds like *ice cream* are given in (39):[42]

(39) milk man/snowman; railway/motorway; cover girl/cover letter; peanut butter/body butter; etc.

We will now discuss some other types of compound 'subspecies' and their idiomatic behaviour.

A. Exocentric Compounds. One could say that exocentric compounds are 'by definition' idioms. Take AN-compounds as in (40):

(40) *bluebottle* 'a cornflower; a variety of blowfly having an iridescent blue body'; *lazy-bones* 'a lazy person'; *bluenose* 'a person with puritanical ways and attitudes; prude'; *pickpocket* 'a person who steals from pockets or purses, usu. in crowded or public places'; etc.

The constituents do not have a head-modifier semantic relationship; a *bluebottle* is not a blue-coloured bottle but an insect/flower. So, the meaning of an exocentric compound is not 'composed of' the meanings of its constituents. As Katamba (1994, 333) puts it: 'exocentric compounds are opaque like idioms, that is they are not subject to compositionality'. They do seem to have a regular word syntax, the nouns are (in most cases) headed by a noun, but their semantics is not.

B. Copulative Compounds. These are coordinative compounds that express the aggregation of two distinct entities in a collectivity. Take such examples as in (41):

(41) bittersweet, singer-songwriter, child-star, washer-dryer, etc.

Morpho-syntactically (in English/Dutch) the righthand noun is the head,[43] but from a semantic point of view the two elements contribute equally to the meaning of the whole. So, one might argue that these (appositional) coordinative compounds are by definition not semantically opaque. On the contrary, if one takes their semantics as the AND-function, they are compositional. However, note Ralli (2019) who observes that 'the degree of compositionality of coordinative compounds follows a cline from the most transparent to the less transparent ones'.

C. 'Cranberry' Compounds. Compounds containing cranberry morphs seems to be prime candidates for listing, given their semantic opacity. Anderson (1992) argues that these types of compounds are listed in the dictionary, giving the appropriate sense for the whole, which shows that this word contains nominal bases, the first one of which is bound. Similarly, prefix-stem combinations like *re-ceive, per-ceive* are listed in the lexicon in structurally analysed form, but the individual prefixes and stems are semantically opaque, meaning that the meaning of the whole can never be built on the meanings of bound elements. An illustrative example is given in (31), from Katamba (1994, 54):

(42) *sanct-* 'holy'   *vir-* 'man'   *tox-* 'poison'   *loc-* 'place'
      sanct-ify   *vir*-ile   *tox*-ic   *loc*-al
      sanct-uary   *vir*-ago   non-*tox*-ic   dis-*loc*-ate
      sanct-ity   trium-*vir*-ate in-*tox*-ic-ate   *loc*-um

D. 'Univerbation'. We classify the 'phrases as words' under compounding for want of a better place. They are words, synchronically, that started out as consisting of two or more words:

(43) altogether, attorney general, bullseye, dyed-in-the-wool, forget-me-not, wannabe, etc.
(44) to eavesdrop, to manhandle, to boot-lick, etc.

They can still have a transparent meaning, to some extent, but have lost their syntactic structure, are completely fixed and are as such idiomatic.

E. 'Phrasal Compounds'. Under this heading I am mentioning particle verbs (45) and inherent reflexive verbs (46). That is, they are compound-like in consisting of two syntactically independent words, but are at the same time a lexeme.

As Fraser (1970) discusses, there are semi-productive and semantically transparent verb-particle combinations (45a), but there many idiomatic cases, as the examples in (45b) illustrate:

(45) a. gulp down, swallow down, … ; bank away, store away, …
      b. *go off* 'turn sour', *take in* 'deceive', *play down* 'minimize', etc.

I am using the term 'semi-productive' for the examples in (45a) because there are restrictions on the type of verbs that can be combined with *down* or *away*. Like in the case of light verbs (Section 2.2), one can formulate those restrictions as collocational or lexical semantic in nature.

With respect to inherent/intrinsic reflexive verbs, observe the examples in (46–47):[44]

(46) a. *Heidi irrt sich.*
          Heidi mistakes herself
          'Heidi is mistaken.'
      b. *Klaus schämt sich.*
          Klaus shames himself
          'Klaus is ashamed.'

(47) a. *Emma roerde de soep.*   *Emma roerde zich.*
          Emma stirred the soup   Emma objected herself
          'Emma stirs the soup.'   'Emma objected./Emma manifested herself.'
      b. *Dana bedacht iets.*   *Dana bedacht zich.*
          Dana thought.up something   Dana thought herself
          'Dana thought up something.'   'Dana changed her mind./Dana had second thoughts.'

In (46) we have German verbs that only combine with *sich* (for Dutch that would be *zich*, for Norwegian *seg*, etc.) and with no other complement. They are, in a sense, bound verbs, can only be used if combined with a simple reflexive. (47) gives Dutch examples that show that the verb can be used without the simple reflexive, but will then have a (substantially) different meaning.

4.2 Inflection

It is not often discussed, or noted, but also in the inflectional domain one might identify what we would call idioms.

The most straightforward examples seem to be cases like in (48), inflection in the nominal domain:[45]

(48) *waters, grounds, strings, arms*, etc.

      a. *arms* = 'weapons'
      b. *strings* = 'violin, viola, cello, double bass sections of an orchestra'
      c. *waters* = 'the seas and oceans in a particular part of the world; that part of the sea which is regarded as under the jurisdiction of a particular state'
      d. *grounds* = 'an enclosed portion of land of considerable extent surrounding or attached to a dwelling-house or other building'

If you know what *water* is and what plural means it is not enough to get *waters*, in the meaning 'the seas and oceans in a particular part of the world; that part of the sea which is regarded as under the jurisdiction of a particular state'. So, the plural gets a non-compositional meaning, in all these cases. The same holds for Dutch, but with a small idiosyncrasy. Dutch has two productive plural morphemes, *-s* (*nota-s* 'bill(s)') and *-en* (*boek-en* 'book(s)'), and several non-productive morphemes, among which is *-eren* (*kind-eren* 'child(ren)').[46] What we can observe is that some nouns have more than one plural form:[47]

(49) a. *water / water-s* = 'in the sense of (several brands of) mineral water'
          water / water-en = same as above for English: *territoriale wateren / *waters*
          'territorial waters'
      b. *vader / vader-s* = father(s), the regular plural
          vader / vader-en = plural means 'forefathers': *het geloof der vaderen* 'the true
          faith'
      c. *blad / blad-en* = magazin(s)[48]
          blad / blad-eren = leave(s): *gekartelde bladeren* 'crenate leaves'
      d. *kleed / kled-en* = rug(s)
          kleed / kled-eren = robe(s), especially for clergy: *priesterlijke klederen* 'clerical
          robes'

As the translations show, the 'non-regular' plural leads to a special meaning,
which we as such could call idiomatic.

In the nominal domain we might also look at other types of 'inflectional' idioms;
are there idioms in the domain of case? That is, would there be examples where
a word having a specific case would take a totally different meaning, or exclude
polysemy? Not that we are aware of. However, you do find specific inflectional
manifestations in certain cases. Take time expressions in Russian.[49] The instrumen-
tal case marking for seasons is as in (50):

(50) *leto* 'summer', *letom* (instrumental singular) 'in summer'
      zima 'winter', *zimoj* (instrumental singular) 'in winter'

But with months, or days, it works differently:

(51) *mart* 'March' will become: *v marte* 'in + March' (locative singular)
      sreda 'Wednesday' will become: *v sredu*, 'in + Wednesday' (accusative singular)

Moving to the verbal domain, Mittwoch (1992) presents, as far as I can tell, an
early example of an inflectional constructional idiom (see Section 3.2). She adopts
the concept from Fillmore et al. (1988) and shows that the perfect progressive con-
struction in English has a non-compositional use. Mittwoch observes that the pro-
gressive neutralizes the restriction of the perfect to events and their resulting states
(and in that sense is non-compositional). Take a look at the examples in (52):[50]

(52) a. I went to work by bus (in those days).
      b. I have gone to work by bus (?? in those days).
      c. I have been going to work by bus (*these days).

The example (52a) can be interpreted as 'habitually', and such an interpretation
is not possible in (52b), unless we use the progressive, centred around the present,
as in (52c). So, the fact that the perfect progressive is not the sum of its component
parts becomes clear from (53). An achievement verb like *to start* can occur in the
simple progressive and the simple perfect, but not the perfect progressive:

(53) a. The concert is (just) starting.
      b. The concert has (just) started.
      c. #The concert has (just) been starting.

Bonami (2015) discusses periphrasis as an instance of a collocational restriction. Given that collocational restrictions are a necessary ingredient of idiomaticity, it offers another window on idiomaticity in the inflectional domain. The central intuition behind his approach is that periphrasis is the inflectional analogue of a flexible idiom (Idiomatic Combining Expressions in the terminology of Nunberg et al. 1994). Like an idiom, the parts stand in a partially flexible syntactic relation but jointly express morphosyntactic content that is not necessarily deducible from the synthetic morphology of the parts (cf. also Booij 2002). In a nutshell, if we look at the progressive, perfect and passive in English having two auxiliaries and two participles for three constructions, at least one of them has to be semantically idiosyncratic.

(54) a. She is <u>writing</u> a book.
      b. She <u>has written</u> a book.
      c. The book <u>was written</u>.

Bonami (2015) argues that periphrasis is a syntactically flexible idiom and can be formally modelled as such using what he calls 'reverse selection':[51] auxiliaries select a participle, participles select auxiliaries, allowing the three-way distinction in (54), as an example. So, inflectional periphrasis is the phenomenon where a multi-word expression plays the grammatical role normally played by a single word filling a cell in an inflectional paradigm (Bonami 2015, 86): 'while idioms carry semantic content, and are thus multi-word equivalents of lexemes, periphrases carry morphosyntactic content, and are thus multi-word equivalents of inflected words'.

Let me give another example of an inflectional constructional idiom. In Dutch the progressive is not marked by an inflectional affix, such as English *-ing*, but by a constructional idiom (cf. (55b)). It is formed by the verb *zijn* 'to be' followed by a PP of the form *aan het* + verbal infinitive (Booij 2002, 2010):

(55) a. *Hij is aan het werk-en*.
         He is at the work-INF.
         'He is working.'
      b. BE [$_{PP}$ [*aan*]$_P$ [$_{NP}$ [*het*]$_D$ [$_N$ V-en$_{inf}$]]]

So, the verbal inflectional feature 'progressive' is expressed by what we would call a collocational construction, following Bonami.[52]

Mel'čuk and Beck (2006) discuss affixal idioms, a combination of affixes that have a semantic unity. They, for instance, discuss how in German/Dutch, the past participle is formed by a single circumfix *ge- … -t/d* or *ge- … -en* (*ge-frag-t* 'asked' or *ge-fund-en* 'found'). They argue that a circumfix of this type is a sign complex (= a

morphological idiom), essentially similar to such idiomatic expressions as *kick the bucket*. Another example is the Hungarian verbal suffix of the 2PL *-tok*, in which one could distinguish the element *-t* (2PL), which appears only in *-tok* (cf. *-d* (2SG)), and *-k* (PL). They also take the French conditional suffix *-rai/-ri* as an example. This affix combination can be formally (but not semantically!) reduced to *-r-* (FUT) and *-ai/-i*.

5   Summary

In morphology, most literature refers to a list consisting of 'units' which must be memorized because they have particular idiosyncratic properties. The literature often refers to idiosyncrasy of meaning, but listing can also be the consequence of phonological or morphological idiosyncrasy. This fits in with our discussion on how one could best characterize what an idiom is: primarily a collocational restriction in the Weinreichian sense. Summarizing, either (i) something is an idiom because of idiosyncratic meaning, or (ii) something is an idiom because of a very specific lexical realization, often co-occurring with a non-regular/restricted semantics. If we in (ii) replace 'specific lexical realization' with 'non-regular phonology/morphology' we get something that comes close to equating to what morphologists call 'idiosyncrasy' and syntacticians call 'idiomaticity'. Crucial, in my view, is that phrase/lexical idioms, to borrow the terminology of Katz and Postal (1963), are first and foremost manifestations of conventionalized language use, and that means that their bearing on the study of I-language is limited.

SEE ALSO: Coordinative Compounding, Including Dvandva; Cranberry Morphemes; Exocentric Compounds.

Notes

1. I would like to thank Geert Booij, Jan Don and Sabrina Bendjaballah for their comments on an earlier version of this entry. I would like to thank Andrew Spencer and Grev Corbett for sharing their thoughts with me on some issues (not all of which materialized in the entry). A special thanks to the editors of this companion for their encouragement and patience.
2. '[T]he semantic relations between the associated proposition and the derived nominal are quite varied and idiosyncratic' (Chomsky 1970, 188).
3. Sabrina Bendjaballah correctly pointed out to me that one could ask oneself the question how idiomaticity plays a role in non-concatenative processes. Unfortunately, I have not been able to address that issue. I refer the reader to discussion in Kastner and Tucker (2020), and the work they cite there, on the boundaries between idiosyncratic and compositional semantic interpretations in languages with non-concatenative morphology.
4. Masini (2019), for instance, calls them quasi-idioms.
5. In this quote he limits himself to verbal idioms, but his approach is not limited to verbal idioms.
6. Perhaps it is appropriate in this context to refer to Plato's *Cratylus*. Sedley (2020) argues that the topic of debate, the opposition between conventionalism and naturalism is,

in fact, about 'words' in the most generic use. As such the *Cratylus* is Plato's dialogue about language.

7. Millikan (2003) argues against this interpretation of convention.

8. Or 'public language', cf. Millikan (2003), among others.

9. By collocation I mean that in language use the choice of a combination of words (collocation) is restricted even though the language system/grammar would allow other choices.

10. These examples are from Kuiper et al. (2003). See also Nunberg et al. (1994), classifying such examples as Idiomatically Combining Expressions. See the discussion in Section 2.3.

11. In this entry I am using the #-sign to indicate that such word combinations are not found, or hardly found, in corpus searches, but it does not mean that these word combinations could not be used. In (16) I have taken over the *-markings of the source.

12. It is important to understand that collocational restrictions, as used above, are meant to be different from how that notion is used in computational linguistics. That is, it is not simply a frequency phenomenon, but it is, in my mind, a lexical restriction on word combination. Admittedly it is very difficult to distinguish the two interpretations of collocational restrictions.

13. Booij and Hüning (2014) call these 'affixoids' (a notion often used in Germanic morphology): compound constituents with an affix-like behaviour. Such affixoids often have an intensifying meaning.

14. A search in the *Corpus Hedendaags Nederlands* (Corpus of contemporary Dutch) also did not give a result for *steenslecht*. For *steenarm*, the opposite of *steenrijk*, there was one hit: 'voelt de economie als een kermis in een hel: steenrijk versus steenarm' ('the economy feels as a fair in hell: very rich versus very poor'). Moreover, I found several instances of *steenoud* 'very old'.

15. *Tjok-* is a cranberry morph in Dutch (but clearly related to the English *chock* in *chock-full*); *stamp-* and *prop-* are related to the verbs *stampen* 'to stamp' and *proppen* 'to stuff'.

16. I will come back to this issue in Section 2.4.

17. In some literature (non-)compositionality is replaced by formulations such as 'listed syntactic structures which interact with syntax in the way words do and have unpredictable meanings in the way that words do, but which consist of more than one "piece"' (Svenonius 2005, 1), equating idiomaticity with 'idiosyncratic meaning' (Anagnostopoulou and Samioti 2012, 233).

18. An issue succinctly summarized in Dowty (2007).

19. Katz (1966, 312) defines this as a syncategorematic element, i.e. an element whose meaning is a function which operates on other meanings; the meaning of a word x must explicitly refer to a feature of the meaning of a word y, and x and y are in a (close) structural relationship.

20. Gazdar et al. (1985) note that for such an analysis it is necessary to modify the model theory of intensional logic so as to accommodate partial functions.

21. As Weinreich (1969, 76) puts it: 'the relation between idiomatic and literal meanings is so unsystematic as to deserve no place in the theory. It is an essential arbitrary relation, which looks plausible only in retrospect'.

22. Note that Vergnaud (1985, 312–317) describes cases like (28) in terms of a mutual syncategorematic relationship (cf. n. 19).

23. This notation appears to be reminiscent of what Harley and Noyer (2014 [1999], 4) write: 'the encyclopedia entry for "kick", for example, will specify that in the environment of the direct object "the bucket" "kick" may be interpreted as "die"'. However, contrary to (28) it is not immediately clear how the encyclopedia could specify syntactic

constraints on 'the environment' of special meanings (without making use of syntax). Of course, one could introduce reference to the machinery for grammatical computation into the encyclopedia (Marantz 1997), but that seems counterintuitive (cf. also n. 35). If the encyclopedia is encoding 'non-linguistic knowledge', it is not clear that the encyclopedia is itself a grammatical component, contrary to what is intended by (28). See also the discussion in Williams (2007).

24. But cf. Ruhl (1975) for arguments that such an analysis would be justified.

25. As pointed out to me by Geert Booij, nowadays the word *uglify-js* is used, in the sense that it is the name of a Java tool.

26. The OED gives *messer*, but not *messee*.

27. Comedians, comedy-script writers, advertisement copywriters are an exception to this.

28. Following Chierchia and McConnell-Ginet (1990, 369).

29. Cf. Aronoff (1980) on a similar line of argumentation with regard to conversion.

30. The motivation for this choice is that the set of morphophonological operations and the idiosyncratic behaviour typical of derivational processes has parallels in inflection. You can see this approach back in Distributed Morphology.

31. Distributed Morphology, without explicitly saying so, extends what one could call idiomaticity, taken as a notion dealing with idiosyncrasy of meaning, to 'idiosyncrasy of phonology'.

32. Apart from the sources mentioned in the text, I have made use of Embick and Noyer (2012) and Harley and Noyer (2014). I am aware of ongoing discussion in DM on several issues, but I don't think it alters the overall claims I make on the view on idioms in DM.

33. My discussion on what idioms in morphology are does not directly touch upon Borer's 'exo-skeletal' approach (Borer 2013, among others).

34. For a different approach, see Anderson (1992, 193–194).

35. If the encyclopedia is 'the locus of noncompositional meanings in the sense of meanings that can't be structurally represented' (Marantz 1996) it is not clear why it would, for instance, make sense to formulate configurational restrictions on idioms, such as the no-subject idiom constraint (Harley and Stone 2013; Truswell 2016), if correct (Everaert and Kuiper 1997; Williams 2007).

36. Everaert (1996) and McGinnis (2002), among others, show that certain lexical properties of the verb are retained in the idiom. Note that Construction Grammar takes a different position: meanings are introduced as a whole from the lexicon, not allowing us to account for such observations.

37. But note that constructional schemas, if productive, also define non-listed words.

38. But as I discussed in Section 2.2 in some cases of such a construction there are collocational restrictions. In this particular case there are also 'fully idiomatic' cases: *doodleuk* is not 'very nice', but 'coolly', 'casually'.

39. This section relies on Katamba (1994).

40. *erd-* is a variant of *aard-* and is found in compounds in Afrikaans such *erdwurm* '(earth)worm', *erdewerk* 'earth-ware'/'pottery'. See Botha (1968) for an early treatment of idiomatic compounds in Afrikaans, and its relevance for a theory of the lexicon in Generative Grammar.

41. Cross-linguistic analysis reveals many cases where English nominal-based collocational constructs are realized as compounds in West Germanic languages, e.g. English *bunch of keys* = Dutch *sleutelbos*.

42. Although Jackendoff (2011) may be right in saying that the productivity of compounding means that there must also be a set of principles that allows us to interpret new compounds, the question is how to formulate them.

43. Dutch obeys the Righthand Head Rule. For instance, in the Dutch compound *kindster* [$_N$ [$_N$ kind] [$_N$ ster]] 'child-star', it is the righthand element that determines the gender: kind$_{neuter}$ ster$_{common}$ → kindster$_{common}$.
44. All Romance languages and all the Germanic languages having an SE-reflexive in the sense of Reinhart and Reuland (1993) take this 'simplex' reflexive in the case of inherent/intrinsic reflexives.
45. Acquaviva (2008) provides an enlightening discussion of 'lexical plurals'. There is much more semantic idiosyncrasy in this domain, which I am not able to discuss here.
46. Although it is difficult to give a fully general rule, by and large the choice between the two productive plural morphemes is phonologically triggered (at least for standard Dutch spoken in the Netherlands).
47. I agree with Acquaviva (2008, 36) that (i) there is no one-to-one relation between irregular form and irregular meaning and that (ii) the semantic distance between the two plural forms might vary considerably.
48. So, in the case of (49c, 49d) the homonymy disappears in the plural.
49. Some of the facts below were brought to my attention by Grev Corbett and Andrew Spencer. They clearly cannot be held responsible for my use/interpretation of the facts.
50. Based on Mittwoch's examples, but slightly different to avoid interpretational complications.
51. Cf. Sailer (2003), Richter and Sailer (2003), and discussion around example (28) for the gist of what reverse selection is.
52. Depending on what one is prepared to call an idiom.

References

Acquaviva, Paolo. 2008. *Lexical Plurals: A Morphosemantic Approach*. Oxford: Oxford University Press.

Anagnostopoulou, Elena, and Yota Samioti. 2013. 'Allosemy, Idioms and Their Domains: Evidence from Adjectival Participles'. In *Syntax and Its Limits*, edited by Raffaella R. Folli, Christina Sevdali and Robert Truswell, 218–250. Oxford: Oxford University Press.

Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge, MA: Cambridge University Press.

Arcodia, Giorgio, and Bianca Basciano. 2018. 'The Construction Morphology Analysis of Chinese Word Formation'. In *The Construction of Words*, edited by Geert Booij, 219–253. Cham: Springer.

Aronoff, Mark. 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

Aronoff, Mark. 1980. 'Contextuals'. *Language,* 56: 744–758.

Aronoff, Mark. 1994. *Morphology by Itself*. Cambridge, MA: MIT Press.

Aronoff, Mark. 2007. 'In the Beginning Was the Word'. *Language,* 83: 803–830.

Aronoff, Mark, and Frank Anshen. 2017. 'Morphology and the Lexicon: Lexicalization and Productivity'. In *The Handbook of Morphology*, edited by Andrew Spencer and Arnold Zwicky. Wiley Online Library. DOI: 10.1002/9781405166348.ch11.

Bargmann, Sascha, and Manfred Sailer. 2018. 'The Syntactic Flexibility of Semantically Non-Decomposable Idioms'. In *Multiword Expressions: Insights from a Multi-lingual Perspective*, edited by Manfred Sailer and Stella Markontonatou, 1–29. Berlin: Language Science Press.

Benson, Morten, Evelyn Benson and Robert Ilson. 1986. *The BBI Combinatory Dictionary of English: A Guide to Word Combinations*. Amsterdam and Philadelphia: John Benjamins.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Bonami, Olivier. 2015. 'Periphrasis as collocation'. *Morphology*, 25: 63–110.

Booij, Geert. 2002. 'Constructional Idioms, Morphology, and the Dutch Lexicon'. *Journal of Germanic Linguistics*, 14: 301–327.

Booij, Geert. 2005. *The Grammar of Words: An Introduction to Linguistic Morphology*. Oxford: Oxford University Press.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Booij, Geert, and Jenny Audring. 2018. 'Partial Motivation, Multiple Motivation: The Role of Output Schemas in Morphology'. In *Advances in Construction Morphology*, edited by Geert Booij, 59–80. Berlin: Springer.

Booij, Geert, and Matthias Hüning. 2014. 'Affixoids and Constructional Idioms'. In *Extending the Scope of Construction Grammar* edited by Ronny Boogaart, Timothy Colleman and Gijsbert Rutten, 77–105. Berlin: Mouton De Gruyter.

Borer, Hagit. 2013. *Structuring Sense Volume III: Taking Form*. New York: Oxford University Press.

Botha, Rudie. 1968. *The Function of the Lexicon in Transformational Generative Grammar*. The Hague: Mouton.

Burger, Harald, Annelies Buhoffer and Ambros Sialm. 1982. *Handbuch der Phraseologie*. Berlin and The Hague: De Gruyter.

Carstairs-McCarthy, Andrew. 1992. *Current Morphology*. London: Routledge.

Chierchia, Gennaro, and Sally McConnell-Ginet. 2000. *Meaning and Grammar: An Introduction to Semantics,* 2nd ed. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. 'Remarks on Nominalization'. In *Reading in English Transformational Grammar*, edited by Roderick Jacobs and Peter Rosenbaum, 184–221. Waltham: Ginn.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Clark, Eve, and Herbert Clark. 1979. 'When Nouns Surface as Verbs'. *Language,* 55: 767–811.

Di Sciullo, Anne-Maria, and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.

Downing, Pamela. 1977. 'On the Creation and Use of English Compound Nouns'. *Language,* 53: 810–842.

Dowty, David. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Kluwer.

Dowty, David. 2007. 'Compositionality as an Empirical Problem'. In *Direct Compositionality*, edited by Chris Barker and Pauline Jacobson, 23–101. Oxford: Oxford University Press.

Embick, David, and Rolf Noyer. 2012. 'Distributed Morphology and the Syntax–Morphology Interface'. In *The Oxford Handbook of Linguistic Interfaces*, edited by Gillian Ramchand and Charles Reiss. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199247455.013.0010.

Everaert, Martin. 1996. 'The Encoding of the Lexical Semantic Structure of Verbs: The Case of Auxiliary Selection in Idioms'. In *Proceedings of the International Conference on Lexicology and Lexical Semantics Münster,* vol. 2, edited by Edda Weigand and Franz Hundsnurscher, 27–37. Tübingen: Max Niemeyer Verlag.

Everaert, Martin. 2010. 'The Lexical Encoding of Idioms'. In *Syntax, Lexical Semantics, and Event Structure*, edited by Malka Rappaport Hovav, Edith Doron and Yvi Sichel, 76–99. Oxford: Oxford University Press.

Everaert, Martin, and Koenraad Kuiper. 1997. 'Theory and Data in Idiom Research'. In *Parasession on Theory and Data in Linguistics Research of the 32nd Chicago Linguistic Society*, 43–57. Chicago: Chicago Linguistic Society.

Fillmore, Charles J., Paul Kay and Mary O'Connor. 1988. 'Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*'. *Language*, 64 (3): 501–538.

Fraser, Bruce. 1970. 'Idioms within a Transformational Grammar'. *Foundations of Language*, 6: 22–42.

Gazdar, Gerald, Ewan Klein, Geoff Pullum and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*. London: Basil Blackwell.

Groenendijk, Jeroen, and Martin Stokhof. 2005. 'Why Compositionality?' In *Reference and Quantification: The Partee Effect,* edited by Greg Carlson and Jeffrey Pelletier, 83–106. Stanford, CA: CSLI Press.

Gross, Maurice. 1979. 'On the Failure of Generative Grammar'. *Language,* 55 (4): 859–885.

Halle, Morris. 1973. 'Prolegomena to a Theory of Word-Formation'. *Linguistic Inquiry*, 4: 451–464.

Halle, Morris, and Alec Marantz. 1994. 'Some Key Features of Distributed Morphology'. In *MIT Working Papers in Linguistics 21: Papers on Phonology and Morphology*, edited by Andrew Carnie and Heidi Harley, 275–288. Cambridge, MA: MITWPL.

Harley, Heidi, and Rolf Noyer. 2014 [1999]. 'Distributed Morphology'. In *The Second Glot International State-of-the-Article Book: The Latest in Linguistics*, edited by Lisa Cheng and Rint Sybesma, 463–496. Berlin and Boston: De Gruyter Mouton. doi: 10.1515/9783110890952.463.

Harley, Heidi, and Megan Stone. 2013. 'The "No Agent Idioms" Hypothesis'. In *Syntax and Its Limits*, edited by Raffaella Folli, Christina Sevdali and Robert Truswell, 283–311. Oxford: Oxford University Press.

Hockett, Charles. 1956. 'Idiom Formation'. In *For Roman Jakobson: Essays on the Occasion of His Sixtieth Birthday, 11 October 1956*, edited by Morris Halle, Hugh McLean, Horace G. Lunt and Cornelis H. Van Schooneveld, 222–229. The Hague: Mouton.

Hockett, Charles. 1958. *A Course in Modern Linguistics*. New York: Macmillan.

Jackendoff, Ray. 1975. 'Morphological and Semantic Regularities in the Lexicon'. *Language,* 51: 639–671.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1995. 'The Boundaries of the Lexicon'. In *Idioms: Structural and Psychological Perspectives*, edited by Martin Everaert, Erik-Jan Van der Linden, André Schenk and Robert Schreuder, 133–156. Hillsdale, NJ: Lawrence Erlbaum.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2011. 'Compounding in the Parallel Architecture and Conceptual Semantics'. In *The Oxford Handbook of Compounding*, edited by Rochelle Lieber and Pavol Štekauer, 105–129. Oxford: Oxford University Press.

Janssen, Theo. 1983. 'Foundations and Applications of Montague Grammar'. PhD thesis, Department of Mathematics, University of Amsterdam.

Kamp, Hans, and Barbara Partee. 1995. 'Prototype Theory and Compositionality'. *Cognition,* 57: 129–191.

Kastner, Itamar, and Matthew A. Tucker. 2020. 'Non-Concatenative Morphology'. MS, University of Edinburgh, https://blogs.ed.ac.uk/itamar/wp-content/uploads/sites/1083/2020/07/kastner-tucker20.pdf.

Katamba, Francis. 1994. *English Words*. London and New York: Routledge.

Katamba, Francis, and John Stonham. 2007. *Morphology*. Basingstoke: Palgrave.

Katz, Jerrold. 1966. *The Philosophy of Language*. New York: Harper & Row.

Katz, Jerrold, and Paul Postal. 1963. 'Semantic Interpretation of Idioms and Sentences Containing Them'. In *Quarterly Progress Report*, no. 70, 275–282. Cambridge, MA: Research Laboratory of Electronics, Massachusetts Institute of Technology.

Kay, Paul, and Ivan A. Sag. 2014. 'A Lexical Theory of Phrasal Idioms'. http://www1.icsi.berkeley.edu/~kay/idioms-submitted.pdf.

Kooij, Jan. 1968. 'Compounds and Idioms'. *Lingua,* 21: 250–268.

Kuiper, Koenraad, Heather McCann, Heidi Quinn, Therese Aitchison and Kees van der Veer. 2003. *SAID*, linguistic data consortium.

Lewis, David K. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Makkai, Adam. 1972. *Idiom Structure in English*. Mouton: The Hague.

Marantz, Alec. 1996. '"Cat" as a Phrasal Idiom: Consequences of Late Insertion in Distributed Morphology'. MS, MIT.

Marantz, Alec. 1997. 'No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon'. In *Proceedings of the 21st Annual Penn Linguistics Colloquium*, edited by Alexis Dimitriadis et al., 201–225. Penn Working Papers in Linguistics.

Masini, Francesca. 2019. 'Multi-Word Expressions and Morphology'. In *Oxford Research Encyclopedia of Linguistics.* Accessed 1 February 2022. DOI: 10.1093/acrefore/9780199384655.013.611.

Masini, Francesca, and Jenny Audring. 2018. 'Construction Morphology'. In *The Oxford Handbook of Morphological Theory*, edited by Jenny Audring and Francesca Masini, 365–389. Oxford: Oxford University Press.

May, Robert. 1977. 'The Grammar of Quantification'. Doctoral diss., MIT.

McGinnis, Martha. 2002. 'On the Systematic Aspect of Idioms'. *Linguistic Inquiry,* 33: 665–672.

Mel'čuk, Igor. 1995. 'Phrasemes in Language and Phraseology in Linguistics'. In *Idioms: Structural and Psychological Perspectives*, edited by Martin Everaert, Erik-Jan Van der Linden, André Schenk and Robert Schreuder, 167–232. Hillsdale, NJ: Lawrence Erlbaum.

Mel'čuk, Igor, and David Beck. 2006. *Aspects of the Theory of Morphology.* Berlin: De Gruyter Mouton.

Mel'čuk, Igor, and Alexander Zolkovsky. 1984. *Explanatory Combinatorial Dictionary of Modern Russian*. Wiener Slawistischer Almanach Sonderband 14. Vienna: Ges. z. Foerd. Slawist. Studien.

Millikan, Ruth. 2003. 'In Defense of Public Language'. In *Chomsky and His Critics*, edited by Louise M. Antony and Norbert Hornstein, 215–237. Malden, MA: Blackwell.

Mittwoch, Anita. 1992. 'Syntactic Idioms?' In *Proceedings of Idioms,* vol. 1: *International Conference on Idioms*, edited by Martin Everaert, Erik-Jan van der Linden, André Schenk and Robert Schreuder, 75–84. Tilburg: Institute for Language Technology and AI.

Nevins, Andrew. 2016. 'Lectures on Postsyntactic Morphology, Lingbuzz'. https://ling.auf.net/lingbuzz/002587.

Nunberg, Geoffrey. 1978. *The Pragmatics of Reference*. Bloomington: Indiana University Linguistic Club.

Nunberg, Geoffrey, Ivan Sag and Tom Wasow. 1994. 'Idioms'. *Language,* 70 (3): 491–538.

Partee, Barbara. 1984. 'Compositionality'. In *Varieties of Formal Semantics: Proceedings of the 4th Amsterdam Colloquium 1982*, edited by Fred Landman and Frank Veltman, 281–311. Dordrecht: Foris Publications.

Pesetsky, David. 1985. 'Morphology and Logical Form'. *Linguistic Inquiry,* 16 (2): 193–246.

Prince, Ellen. 1974. 'English Aspectual Constructions'. Diss., University of Pennsylvania.

Ralli, Angela. 2019. 'Coordination in Compounds'. In *Oxford Research Encyclopedia of Linguistics*. Accessed 21 May 2022. DOI: 10.1093/acrefore/9780199384655.013.564.

Reinhart, Tanya, and Eric Reuland. 1993. 'Reflexivity'. *Linguistic Inquiry*, 24: 657–720.

Richter, Frank, and Manfred Sailer. 2003. 'Cranberry Words in Formal Grammar'. In *Empirical Issues in Formal Syntax and Semantics*, edited by Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr and Francis Corblin, 155–171. Paris: Presses de l'Université de Paris-Sorbonne.

Richter, Frank, and Manfred Sailer. 2004. 'Basic Concepts of Lexical Resource Semantics'. In *ESSLLI 2003: Course Material I (Collegium Logicum 5)*, edited by Arne Beckmann and Norbert Preining, 87–143. Vienna: Kurt Gödel Society Wien.

Ruhl, Charles. 1975. '*Kick the Bucket* Is Not an Idiom'. *Interfaces: Linguistics and Psychoanalysis*, 4: 2–4.

Sailer, Manfred. 2003. 'Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar'. Phil. diss. (2000), Arbeitspapiere des SFB 340.161, Universität Tübingen.

Scalise, Sergio. 1984. *Generative Morphology*. Dordrecht: Foris.

Schenk, André. 1995. 'The Syntactic Behavior of Idioms'. In *Idioms. Structural and Psychological Perspectives*, edited by Martin Everaert, Erik-Jan van der Linden, André Schenk and Rob Schreuder, 253–271. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sedley, David. 2020. 'Plato's *Cratylus*'. In *The Stanford Encyclopedia of Philosophy*, Winter, edited by Edward N. Zalta. Accessed 15 November 2022. https://plato.stanford.edu/archives/win2020/entries/plato-cratylus/.

Siddiqi, Daniel. 2009. *Syntax within the Word: Economy, Allomorphy, and Argument Selection in Distributed Morphology*. Amsterdam and Philadelphia: John Benjamins.

Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Basil Blackwell.

Spencer, Andrew, and Arnold Zwicky, eds. 1998. *The Handbook of Morphology*. Oxford: Blackwell.

Svenonius, Peter. 2005. 'Extending the Extension Condition to Discontinuous Idioms'. *Linguistic Variation Yearbook*, 5: 227–263.

Truswell, Robert. 2016. 'Verb Meaning and the "No Agent Idioms" Effect'. http://robtruswell.com/assets/pdfs/Potsdam_writeup.pdf.

Vergnaud, Jean-Roger. 1985. *Dépendance et niveaux de représentation en syntaxe*. Amsterdam: John Benjamins.

Wasow, Thomas, Ivan A. Sag and Geoffrey Nunberg. 1982. 'Idioms: An Interim Report'. In *Proceedings of the XIIIth International Congress of Linguistics*, edited by Shirô Hattori and Kazuko Inoue, 102–115. Tokyo: The Proceedings Publishing Committee under the auspices of the CIPL [The Hague].

Weinreich, Uriel. 1960. 'Mid-Century Linguistics: Attainments and Frustrations [Review of *A Course in Modern Linguistics*, by C.F. Hockett].' *Romance Philology*, 13 (3): 320–341.

Weinreich, Uriel. 1969. 'Problems in the Analysis of Idioms'. In *Substance and Structure of Language*, edited by Jaan Puhvel, 23–81. Berkeley: University of California Press.

Westerståhl, Dag. 2002. 'On the Compositionality of Idioms: An Abstract Approach'. In *Words, Proofs, and Diagrams*, edited by Dave Barker-Plummer, David Beaver, Johan van Benthem and Patrick Scotto di Luzio, 241–271. Stanford, CA: CSLI Publications.

Wierzbicka, Anna. 1982. 'Why Can You Have a Drink When You Can't *Have an Eat?' *Language,* 58 (4): 753–799.

Williams, Edwin. 2007. 'Dumping Lexicalism'. In *The Oxford Handbook of Linguistic Interfaces*, edited by Gillian Ramchand and Chris Reiss, 353–382. Oxford: Oxford University Press.

Winter, Yoad. 2016. *Elements of Formal Semantics*. Edinburgh: Edinburgh University Press.

van der Wouden, Ton. 2020. '*oer-*.' *Taalportaal*. Accessed 15 November 2022. https://taalportaal.org/taalportaal/topic/pid/topic-13998813294893391.

Zimmer, Karl. 1964. *Affixal Negation in English and Other Languages: An Investigation of Restricted Productivity*. Supplement to *Word*, 20 (2). London: William Clowes & Sons.