DOI: 10.26346/1120-2726-223

A syntactic and morphological account of English nonaffixal deverbal compounds

Jesús Fernández-Domínguez

Department of English and German Philologies, University of Granada, Spain
<jesusferdom@ugr.es>

The exploration of English compounds like *heartbeat*, *mouse click* or *gunfight* reveals complexities because, despite an apparently straightforward noun + noun structure, their internal constituency is less obvious. Even if these formations have been sometimes treated as regular primary noun + noun compounds, today there is agreement that their chain of word-formation involves deverbal conversion ($click_V > click_N$) followed by noun + noun compounding ($mouse_N + click_N$). Precisely because the head noun has previously undergone verb-to-noun conversion, such compounds may be seen sometimes as primary and sometimes as synthetic. This article explores a selection of ca. 800 non-affixal (de)verbal compounds (NDVCs), inspected in the light of their BNC frequencies, (non-)argumental nature, type of orientation and productivity degree. The results point to: (i) a significant role of the extralinguistic context for meaning interpretation, (ii) a connection between the argumental specifications of the verbal base and the converted head noun (i.e. *click* in *mouse click*), and (iii) the possibility to understand a given NDVC at the same time as primary and as synthetic.

KEYWORDS: argument structure, compounding, conversion, corpus-based, lexicon, productivity.

1. Introduction

The nature of formations such as *snake bite*, *haircut* or *moon walk* has for some time passed unnoticed within the category of English nominal compounding. The initial obstacle may lie in Marchand's principle underlying compounding: "to see a thing identical with another one already existing and at the same time different from it" (1969: 11). So-called non-affixal (de)verbal compounds (**NDVCs**, Lieber 2010) were initially perceived as a subclass of plain root/primary compounding resulting from the concatenation of two simple nouns, analogue to units like *hospital staff* or *summer dress*, but have been found out to arise from the interaction between compounding and conversion.

The prevailing view on NDVCs is that verb-to-noun conversion applies first (1a) and that compounding between head and modifier follows (1b):

(1) a. $bite_V + \emptyset_N > bite_N$
b. $snake_N + bite_N = [snake\ bite]_N$

A consequence of this deverbal nature is that the left-hand member of an NDVC functions as an argument of the action expressed by the head. NDVCs are in need of closer attention in view of their uncertain status, even if an array of morphologists has noted features that set these formations apart from other classes of compounds (Marchand 1969; Allen 1978; Bauer 1983; 2017: 80; Adams 2001; Bauer & Renouf 2001; Bauer & Huddleston 2002).

This paper sets out to further study NDVCs by use of a set of units larger than previous attempts, complemented with corpus information. This makes it possible to explore compounds in the light of a number of variables, for example their subject *vs* object orientation, range of semantic readings or degree of productivity. The specific objectives of the investigation are:

(i) To explore the (extra)linguistic conditions which may favour or hinder the creation of NDVCs in the context English nominal compounding,
(ii) To propose an analysis for NDVC heads as either simple or converted nouns, which in turn affects the semantic relationship between the two constituents.

The investigation is structured as follows: after this introduction, section 2 describes the methodological decisions as well as the tools and materials employed, while section 3 delves into the study of NDVCs by focusing on their twofold synthetic/primary nature (3.1), on their functional and structural ambivalence (3.2), and on their productivity rates (3.3). Section 4 offers some conclusions.

2. Data management

This paper is interested in NN subordinative compounds whose head is a deverbally converted noun, e.g. *snakebite* (13)[1] or *boat ride* (13), that is, formations with the underlying structure $N + [V]_N$. One preliminary terminological note seems pertinent at this point. As regards the ever-present demarcation compound-phrase, I count myself among the "lumpers" in Bauer's (1998) terminology, that is, in the absence of further evidence, constructions like *oilspill* (33) or *rainfall* (409) are here treated as belonging to the same class regardless of their different spelling (see Bauer 1998; Schäfer & Bell 2020).

For collection of the data, the first step was to enlarge the sample, given that the lists of NDVCs available in the literature are scarce (see Appendix). With the purposes of this article in mind, the *British National Corpus* (BNC, Davies 2004-) was selected for its suitability in terms of time range, genre representativeness and balance. Besides the shortage

of NDVCs available, one added problem is their corpus retrieval (see Lieber 2010: 129), given that POS-tagging in the BNC is derivation-blind, and searching for NN sequences will yield sequences of any two nouns, the majority of which are primary compounds and hence irrelevant for these purposes, e.g. *world war*, *county council* or *health service*.

The alternative was an initial selection of NDVCs based on the following references: Adams (2001), Bauer (2020), Bauer & Renouf (2001), Lieber (2009; 2010; 2016a), Jackendoff (2010) and Cetnarowska (2020) (see Appendix). This 84-item sample was enlarged through BNC queries, using as input the heads of the units in the initial list under the assumption that the retrieved units would also be cases of deverbal conversion, and hence of NDVCs. For example, based on *peace talk* (341; from Bauer 2020), 20 other formations with *talk* as head were extracted: *table talk* (35), *pillow-talk* (3), *streettalk* (8), *shoptalk* (6), etc. In other words, because in *peace talk* the head $talk_{N}$ is converted from $talk_{V}$, according to the OED and to Bauer (2020), I propose that the same analysis applies to other *talk*-compounds. This procedure was followed with the heads of the rest of units in the Appendix.

The following specifications were set during retrieval:

i) The right-hand constituent is POS-tagged as a noun and the same form is attested also as a verb, which ensures a relationship of conversion as much as possible. For example, *car crash* (181) is obtained at this stage because it consists of two nouns, and this step checks that the right-hand one ($crash_{N}$) exists in the BNC as a verb ($crash_{V}$). The directionality of conversion is certainly not disclosed by this procedure, but the verbal and nominal forms point towards a relevant relationship with each other. In assessing conversion, right-hand constituents were checked for possible cases of polysemy and homonymy, of which there were none.

ii) The compound is a non-lexicalized common noun, a condition which leads to meaning compositionality, i.e. the meaning of the compound stems from the meaning of its constituents, as in *water rise* (2; 'the water rises'). NN combinations with a non-compositional meaning were ruled out, since their analysis cannot be suitably carried out by means of the three-level orientation used here (see below for details).

iii) The BNC does not attest any lexeme that could involve a relationship of back-formation, which would rule out conversion, e.g. **shoptalking* > *shoptalk* (6).

iv) The spelling of the unit may be open, hyphenated or solid. This implied using different BNC searches for each spelling variant and then adding up their corpus frequencies (see Sanchez-Stockhammer 2018).

After preparing the initial list of NDVCs, frequency values were tracked down from the BNC by hand via POS-based searches and by allowing for inflectional variations in the compound's head (i.e. singular and plural). Frequency values of singulars and plurals, and of different spellings were totalled and the most commonly used spelling was kept for the entry in the database: *oil spill* (20), *oil spills* (2), *oilspill* (7), *oil-spills* (1), *oil-spill* (3) = *oil spill* (33). This stage also retrieved the frequency values required for the productivity models for all corpus entries: type frequency, token frequency, hapax legomena and frequencies of right-hand constituents (see section 3.3.1).

Besides the above, two filters were set towards irrelevant entries:

i) Units which appear in the input list (Appendix) but do not occur in the BNC, as lack of their frequency would prevent their quantitative analysis: *flea bite*, *brain bleed*, *government collapse*, *eyewink*, *ballkick*, *fee hike*, *traffic alert*, *robot repair*, *clam bake*, *onion smell*, *wing support*, *surface drag*.
ii) Units POS-tagged as nouns but which are apparently verbs in their BNC context, e.g. *mouse squeak*:

(2) *Better to hear the lark sing in the woods and fields than the* mouse squeak *in the corridors* (HU0 1374)

The last stage implied the analysis of the NDVCs in terms of their syntactic orientation: Subject-, Object- or Prepositional-Object-. Every entry was individually inspected and allocated to one of these categories. Whenever the BNC context suggested that more than one orientation is possible for the same compound, this was annotated, with the result that some entries have two or three possible orientations (see section 3.1).

The online *Cambridge Dictionary* was checked in order to assess the usage of every verb as transitive or intransitive. This dictionary is selected for its focus on British English, which is also the variety of the BNC, and thus ensures language compatibility between both resources. A verb's having a transitive use does not necessarily prevent it from having an intransitive one. For example, *drift* is first listed in the dictionary as intransitive and then two transitive senses are given, so *drift* was analysed as potentially transitive, thus allowing for the possibility of it having an Object (<dictionary.cambridge.org/dictionary/english/drift>).

For the semantic analysis of primary compounds, Levi's (1978) Recoverably Deletable Predicates (**RDPs**) are employed given their standing and prevalence: CAUSE, HAVE, MAKE, USE, BE, IN, FOR, FROM and ABOUT (see ten Hacken 2009; 2016 for an assessment of Levi's work).

The above steps result in 803 formations representative of non-affixal synthetic compounding, each accompanied by the following information: unit's token frequency, number of hapax legomena, frequency of the head, and syntactic orientation.

3. The interaction between compounding and conversion

This section reviews notions of compounding relevant to the present study and stresses the limits between argumental and non-argumental compounds. To do so, it first focuses on the morphological nature of NDVCs (3.1), then it turns to their (non)argumental interpretation (3.2), and finally considers their productivity measurement (3.3).

3.1. Argumental and non-argumental NN compounds

Synthetic compounds (also called *deverbal* or *verbal*) can be generally defined as those which feature a deverbal element as head and an argument of that verb as the non-head.[2] The understanding of what counts as a synthetic compound is far from undisputed and has progressively evolved from an early view (Marchand 1969; Roeper & Siegel 1982; Selkirk 1982; Lieber 1983) where *-er*, *-ing* and *-ed* were the only legitimate suffixes for this type of compounds, as in (3), to more open views (Allen 1978: 157; Bauer & Renouf 2001: 117-120; Bauer & Huddleston 2002: 1652-1654; Lieber 2009; 2010; Jackendoff 2016; Mattiello & Dressler 2022), which contemplate also *-age*, *-al*, *-ion* or *-ure*, as in (4).[3] In all cases, the left-hand constituent of the compound can be read as an argument of the right-hand constituent (for overviews of synthetic compounding, see Spencer 1991: 324-343; Olsen 2017; Melloni 2020; ten Hacken 2023).

(3) *truck-driver, drug dealing*

(4) *food spoilage, budget approval, muscle relaxation, heart failure*

Such suffixes are frequent candidates for synthetic compounding, but conversion has been obliquely omitted from the discussion until recently. Although an array of morphologists has noted features that set compounds with converted heads apart from similar classes, the label *non-affixal (de)verbal compounds* is coined by Lieber (2010; 2016a; 2016b: 169-174), and their close inspection has been occasional thereafter (Jackendoff 2010; Bauer, Lieber & Plag 2013: 470-471; Cetnarowska 2020). Probably because of their NN makeup, NDVCs have become blended in the amalgam of compound-related studies over-

whelming the research community since the early 2000s and which concern European and non-European languages (Lieber & Štekauer 2009; Brunner, Engelberg & Hein 2021), the semantics of compounding (Lieber 2010; Maguire, Wisniewski & Storms 2010; ten Hacken 2016; Schäfer & Bell 2020), spelling and stress assignment (Giegerich 2004; Plag *et al*. 2008; Rakić 2009; Sanchez-Stockhammer 2018), headedness (Nóbrega & Panagiotidis 2021) and works on compounding in general (Bauer & Renouf 2001; Olsen 2015; Bauer 2017; 2020; Melloni 2020).

Alongside compounding, the second process involved in the creation of NDVCs is conversion. This paper rests on an understanding of conversion as the creation of a lexical item without any modification on the shape of the input base (Bauer 1983: 32; 2005; Katamba 1994: 70; Kastovsky 2005; Lieber 2016b: 112; Andreou & Lieber 2020: 335-338; Lieber & Plag 2021: 3-6). As explained, the deverbal origin of their head technically sets NDVCs as synthetic compounds and, as a consequence, the meaning scope is constrained because the deverbal head preserves the argument structure of the base verb (Olsen 2017: 21). Three orientations are possible in NDVCs based on the link between the head and the left-hand constituent: subject- (*snake bite*, 13), object- (*haircut*, 305) or prepositional object-orientation (*moon walk*, 3). Subject-orientation has been reported as the most frequent in NDVCs in a number of works (Lieber 2009; 2010; Bierwisch 2015: 32; see Haspelmath 2014: 202; Härtl 2015: 886), while results are more heterogenous elsewhere (Adams 2001: 78-79; Bauer, Lieber & Plag 2013: 471; Lieber 2016a; 2016b).

(5) a. *snake bite* 'the snake bites'
b. *haircut* 'someone cuts the hair'
c. *moon walk* 'a walk on the moon'

This threefold taxonomy keeps things quite simple concerning meaning interpretation thanks to the occurrence of an argument and its action. This straightforwardness is especially evident in comparison to primary compounds, which lack a verbal element and therefore enjoy a much higher number of potential meaning interpretations (Bauer 2020: 270). It has been largely argued and illustrated that, thanks to their ease of creation and to the lack of a verbal element, the possible meanings of primary NN compounds are numerous, often restricted only by the context of occurrence and encyclopaedic knowledge. The semantic underspecification and potential interpretations of primary NN compounds have become a recurrent topic in modern English linguistics which, interestingly, cannot be regarded as settled yet (see Allen 1978; Bauer 1983: 202-204; 2020: 269-272; Bauer & Renouf 2001: 117; Bauer &

Huddleston 2002: 1646-1650; Giegerich 2004: 10-19; Lieber 2009: 359-362; Maguire, Wisniewski & Storms 2010; Jackendoff 2016; Brunner, Engelberg & Hein 2021). Thus, various potential meanings can emerge from the primary NN compound *museum pamphlet*:

(6) a. 'a pamphlet which is produced in the museum'
b. 'a pamphlet which deals with museum-related topics'
c. 'a pamphlet whose shape is museum-like'

These interpretations of *museum pamphlet* all seem possible, even simultaneously, but others could be equally valid within the appropriate context, i.e. any multiple meanings stand "in cooperation rather than competition [and we] might call such a word *promiscuous*" (Jackendoff 2016: 20; see also Marchand 1969: 56; Allen 1978; Bierwisch 2015: 33; Schäfer & Bell 2020). As seems obvious, this loose semantic character means a marked contrast with NDVCs like those in (5). The magnitude of the above is considerable because it dissects NN compounds into two kinds: argumental (synthetic) and non-argumental (primary). The delicate side of the matter lies in the fact that both types are formally identical, which may explain their joint treatment in the past, as originally noted in Selkirk's (1982: 32-33; see Lieber 2004: 59-60) paradigmatic discussion of *tree eater*.[4] A misleading feature is the shared surface structure in (7), and one crucial attribute is the fact that the word-class of the compound's head is ambiguous between verb and noun (Bauer 1983: 205; Lieber 2004: 60). An analysis of such units should further specify such constituent so as to explicitly state its morphological complexity, i.e. whether it is a simplex noun (7) or a deverbal noun (8).

(7) N + N

(8) $N + [V]_N$

The choice between (7) or (8) matters because the underlying structure of a compound will favour its argumental or non-argumental nature, and in turn its most likely reading (see ten Hacken 2016; Brunner, Engelberg & Hein 2021). The non-argument and the argument hypotheses of *snake bite* are presented in (9) and (10):[5]

(9) $*snake_N + bite_N$

(10) $snake_N + [bite_V]_N$

In (10), *bite* is a deverbal noun and hence its reading is an argumental one, in this case most probably subject-oriented: 'the snake bites X'. The expression in (9) is unviable for *snake bite*, since a non-argumental reading would yield less coherent non-verbal interpretations: 'the

bite is caused by the snake' or 'the bite comes from the snake'. Following the exposition of these two subtypes of NN compounds, the next subsection tackles how they interact within NDVCs.

3.2. (Non-)argumental collaboration in NDVCs

A distinction has been made between synthetic compounds whose first element is an adjunct and those where it is an argument (Adams 2001: 79; Bauer, Lieber & Plag 2013: 482-483; Melloni 2020; see Bierwisch 2015: 8-9; Härtl 2015: 881). Among the former occur meanings like locative IN (*airport parking*), temporal IN (*afternoon trainer*) or ABOUT (*AIDS meeting*), which means a difference with argumental compounds because the context can admit virtually any meaning.[6] Although compounds with converted heads have not been discussed exhaustively, the examples in 3.1 evidence that this twofold behaviour is found beyond compounds with affixed heads.

The argument here is that what has been conventionally presented as two different categories of compounding (synthetic *vs* primary) actually overlap in the case of NDVCs: the same compound can be read both argumentally and non-argumentally.[7] This underspecified (non-)argumental behaviour and structural correspondence has been acknowledged in relation to affixal synthetic compounds (Spencer 1991: 325-326; Adams 2001: 79; Bauer & Renouf 2001: 118; Olsen 2015; 2017; Andreou & Lieber 2020; Melloni 2020), but only tangentially regarding non-affixal ones (Bauer, Lieber & Plag 2013: 482-483). In fact, in their discussion on synthetic compounds, Bauer (2020) uses the title "A single type or two types?", and Olsen (2017) speaks of "A single class of compounds?" (see also Härtl 2015: 882). This attitude seems indicative of changing perceptions of the question. Despite their verbal origin, I claim NDVCs to display a hybrid nature because they allow the two treatments indicated above. Further examples are (11-13), *baby care*, *cancer fight* or *leaf fall*, which can be read doubly: argumentally (a) and non-argumentally (b):

(11) a. 'X takes care of the baby' (Prepositional-Object orientation)
b. 'the scope of the care is the baby' (RDP ABOUT)

(12) a. 'X fights cancer' (Object orientation)
b. 'the purpose of the fight is cancer' (RDP FOR)

(13) a. 'leaves fall' (Subject orientation)
b. 'the fall involves leaves' (RDP ABOUT)

Section 2 carried out an analysis of the corpus units in terms of their possible syntactic orientation: Subject, Object, or Prepositional Object. It has also been shown that the same NDVC may have more than one simultaneous interpretation, depending on its context of occurrence. More

precisely, three groups of corpus units can be discerned based on their number of possible orientations: one, two or three. Out of the 803 corpus units, 597 have a univocal interpretation and 206 can be read variously:

ONE ORIENTATION		TWO ORIENTATIONS		THREE ORIENTATIONS	
O	333	S/O	117	S/O/PO	28
PO	145	S/PO	7		
S	119	O/PO	54		
Subtotal	597	Subtotal	178	Subtotal	28
Total				803	

Table 1. Number of NDVCs per number of possible orientations.

As expected from the syntactic binding posed by the verb, the majority of corpus units (598; 74.37%) enjoy one possible orientation. This corresponds to the features of synthetic compounding outlined in 3.1, namely that meaning interpretation is rather straightforward thanks to the presence of the verbal action (Štekauer 2005: 79-81). Out of the one-orientation units, the largest group (333) corresponds to NDVCs where the left-hand constituent can be only Object because the verb involved has a transitive use (*design*, *feed*, *reform*, *shake*, *worship*) in formations like *cancer fight* (2), *divorce reform* (4), *job design* (29), *milkshake* (6) and *sun worship* (8). In these cases, the action is carried out by an underlying implicit agent. Next, the 145 compounds with a Prepositional-Object orientation are syntactically and semantically similar to the Object-oriented ones, with the difference that a preposition is required for the expression of the non-head: *hand stand* (25), *gunfight* (18), *negligence claim* (5), *trumpet-call* (2) or *moon walk* (3). The last set of units with one possible reading are Subject-oriented, 119 compounds participated by both transitive and intransitive verbs. In particular, some of the transitive verbs used for Object-oriented NDVCs are exploited for Subject-orientation too, e.g. *shake* (*moonshake*, 16) or *design* (*author design*, 1). Other verbs appear exclusively with a Subject reading: *earthquake* (684), *flame-burst* (1), *groundswell* (56), *heartbeat* (290), *snakebite* (13), *sand slide* (2) or *snowfall* (47). Interestingly, with a few exceptions (*population drift* 4, *user design* 1), the Subjects in this group are semantically inanimate or non-human, with referents often in the body or in natural phenomena.

Regarding NDVCs with two possible orientations, the weight of transitive verbs is felt in the high number of units that can have both Subject- and Object-orientation. This depends essentially, first, on the verb's semantics and, second, on the context of occurrence (Lieber 2004:

60), but sometimes even checking the BNC concordances does not make it possible to opt for a definitive interpretation. For instance, *kite crash* has a frequency of 1, and the corpus occurrence in example (14) does not disambiguate a Subject ('the kite crashes') or an Object ('someone crashes the kite') reading. A similar argument can be deployed for *monster fight* (1), with the further remark that the Subject and Object interpretations are possible at the same time (see example 15), if we can imagine a context where *monster* can simultaneously execute and receive the action (e.g. a fight between monsters).

(14) *So Alan must have had well in mind the fact that any Rokkaku bout is going to offer short odds on a* kite crash*!* (CA1 1366)

(15) *Muzzily overlaying this film onto recorded shots of the Daleks firing into the jungle – complete with 'negative' effect-gave an image similar in impact to the Id* monster fight *featured in the classic sf film Forbidden Planet.* (F9Y 1682)

Verbs occurring for Subject and Object orientation are, among others, *burst*, *worship*, *crash*, *fight* or *increase*, in units like *balloon burst* (2), *fan worship* (1), *helicopter crash* (10), *computer design* (20) or *bear hunt* (3). After Subject/Object ambiguity, a smallest set is that of Subject/Prepositional-Object-orientation, comprised by formations with the verbs *talk* and *claim*: *baby-talk* (3), *womentalk* (1), *god-talk* (1), *government claim* (9) or *court claim* (2). In all of them a Subject reading ('the baby talks', 'the government claims') as well as a Prepositional-Object reading are possible ('to talk about the baby', 'a claim about/against/concerning the government').[8]

Finally, 54 units stand between an Object and a Prepositional-Object orientation. These are usually NDVCs where the left-hand constituent is a name of place or time which can be seen as the Object but also as the location of the action: *factory reform* (6), *studio design* (3), *morning worship* (7), *street design* (2) or *tree worship* (3). Often, either the Object or the location reading seems more probable, but nothing in the compound's structure prevents the alternative interpretation. Units with a more probable Object reading are *church design* (7), *jail reform* (1), *station design* (14) or *tree worship* (3), and units with a more likely location reading are *morning worship* (7) and *evening worship* (1).

The last group is NDVCs which may be potentially read under three orientations: Subject-, Object, or Prepositional-Object-. Only two verbs are in fact involved in these 28 units, *design* and *reform*, which have in common a left-hand constituent that is a locative: *court reform* (1), *education reform* (195), *cabinet reform* (1), *company design* (2), *ministry design* (2), *organization design* (2). A name place favours the Prepositional-Object reading ('a reform taking place in court', 'a design taking place at the

organization'), but Subject and Object readings are perfectly possible as well ('the court reforms X' and 'the organization designs X' for the former, 'X reforms the court' and 'X designs the organization' for the latter). As in previous cases, disambiguation by context checking proves difficult:

(16) *What turned out to be the most far-reaching of all the post-emancipation measures of Alexander II derived, ironically, from the government's concern for the gentry. The* court reform *is sometimes thought to have originated in concern for the peasantry.* (HY7 1588)

(17) *Cohen, March and Olsen (1972) provided a stark contrast to the planning-oriented literature on* organization design *and decision making with their 'garbage-can model', which in essence claims that people and systems in organizations were in possession of solutions to problems and predispositions to take certain actions when problems or specific situations arise.* (GUC 1587)

The inspection of orientation types can be supplemented by dissecting the individual impact of syntactic orientations (Subject, Object, Prepositional-Object) on the corpus data globally. In Figure 1, each column indicates how many NDVCs carry one orientation by including one, two and three possible interpretations. The column for Subject-orientation counts formations where Subject is the only possible reading (e.g. *nosebleed* 18), formations where we may have Subject/Object (e.g. *computer design* 17), Subject/Prepositional-Object (e.g. *baby talk* 3) and formations with potentially three readings Subject/Object/Prepositional-Object (e.g. *university reform* 1). The same method was applied to the Object and Prepositional-Object categories.

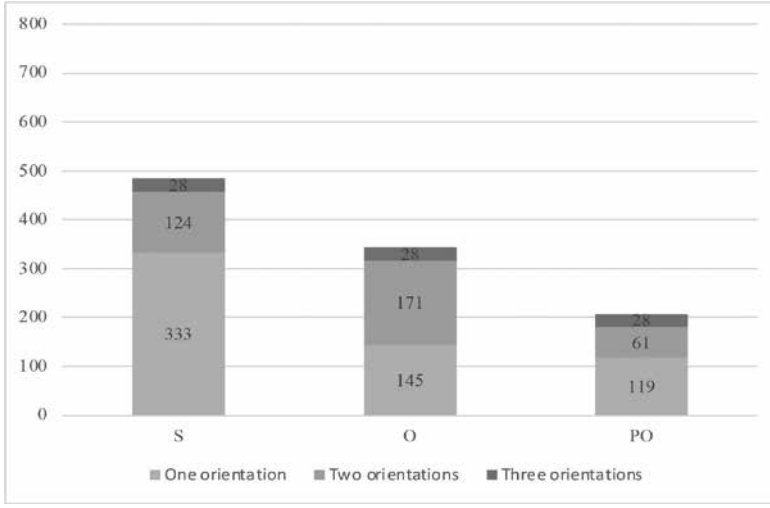

Figure 1. Number of NDVCs per orientation type.

Out of the 803 corpus units, 485 (60.39%) can have a Subject reading, 344 (42.83%) can have an Object reading and 208 (26.01%) a Prepositional-Object one. Note that each column is to be assessed independently and therefore NDVC values do not amount to 803, and percentages do not add up to 100%. Figure 1 exposes a dominance of Subject-orientation, notably in the category of one orientation (333 units), while Prepositional-Objects are less frequent in all cases. The marked presence of Subjects must be directly linked to the nature of English verb patterns: any verb, be it transitive or an intransitive, includes a Subject in its underlying argument structure, which means that all 803 NDVCs could potentially include a Subject as its left-hand constituent. This is different for Objects and Prepositional Objects, which occur exclusively in bivalent clause patterns, and are hence much less common overall.

The exploration of the possible syntactic orientations and their distribution in the sample indicates that the lack of an argument-bearing element in NDVCs leads to a wide range of possible interpretations, which happens much less frequently in affixal formations (e.g. *bus-driver*). In this case particularly, Subjects are predominant due to the argumental nature of the verbs in question, but it must be stressed that these compounds out of context may express various possible orientations. Section 3.3, next, delves into the productivity rates of NDVCs.

3.3. Morphological productivity

NDVCs have been shown to display the capacity for a wide range of meanings, some of them closer to argumental compounds, some closer to non-argumental compounds. This shared NN structure, contrary to what could be expected, entails not a disadvantage but a benefit that facilitates a twofold interpretation (see 3.2), thus implying that ease of production outweighs ease of interpretation. In order to discover the productivity potential of NDVCs, a pending matter according to some (Bauer, Beliaeva & Tarasova 2019: 48), mainstream productivity models are introduced in section 3.3.1 and the units' corpus values are exploited variously in section 3.3.2. This will allow checking whether Object- and Prepositional Object-oriented units are regarded as more productive than Subject-oriented ones, as seems to be the case.

If not habitually, remarks have been made on the unforeseen high productivity degree of NDVCs (Bauer, Lieber & Plag 2013: 470-471; Lieber 2016b: 111-112). Morphological productivity has been described as the language faculty whereby a word-formation process is used for the creation of morphologically complex lexemes (Aronoff 1976; 1983; Bauer 1983; 2001; Spencer 1991; Plag 1999; Bauer, Beliaeva & Tarasova 2019). Based on Corbin (1987), it is customary for productivity studies to distinguish two

sides to it: profitability (the degree of attested usage of a process, a quantitative notion), as opposed to availability (whether a process can be used or not, a qualitative notion). In the following, the synchronic availability of NDVCs is taken for granted and their profitability is brought into focus, i.e. we concentrate on the degree of usage of NDVCs under the assumption that this word-formation rule is available in Contemporary English.

NDVCs are versatile and apparently more fertile than it would seem, probably because

> form is generally minimized when it is not required for comprehension, and this may happen for a variety of reasons: A meaning may be inferable from the pragmatic context […], it may be predictable from the grammatical context […], or it may be predictable from the fact that the meaning occurs significantly more often than a contrasting meaning. (Haspelmath 2014: 198)

The frequency distributions of the study sample (Table 2), first, faithfully reflect Zipf's Law, namely, that in a given corpus, the individual frequency of a word is inversely proportional to its rank in the frequency table (Zipf 1949; see Altmann 2002). As can be noticed, high-frequency units in the corpus amount to a very small percentage, while the number of units per range goes up as we go down the frequency value. As expected, units with frequency 1 (hapax legomena) represent a sizeable proportion (more than one third of the total).

FREQUENCY RANGE	N° OF UNITS	%
1,666-1,000	3	0.37
999-500	6	0.74
499-200	10	1.24
199-100	13	1.61
99-50	16	1.99
49-20	46	5.72
19-10	70	8.71
9-5	94	11.70
4	42	5.23
3	94	11.70
2	116	14.44
1	293	36.48
Total	803	100

Table 2. Frequency distributions in the study corpus.

3.3.1. Models for productivity measurement

In modern productivity studies, hapaxes of a given word-formation process have been taken to be indicative of a profitable status, so Table 2 tentatively points to a high productivity value for NDVCs in general terms (Baayen 1994; 2009; Baayen & Lieber 1991; Plag 1999; Bauer 2001; Bauer, Beliaeva & Tarasova 2019). This initial perception, however, needs to be qualified through a comprehensive approach to the subtypes of NDVCs. A number of proposals exist for productivity measurement, from Aronoff's (1976; 1983) early estimates through Baayen's (1994; 2009) statistical models to dictionary-based or mixed calculations (Bolozky 1999; Plag 1999). In all cases, the assumption is that the morphological productivity of a given process can be measured, and corpus or dictionary data is employed for a numerical estimation thereof.

To evaluate the productive behaviour of NDVCs, various productivity formulae have been selected: type frequency (*V*), token frequency (*N*), type-token ratio (*V/N*), productivity in the strict sense (*P*), and relative frequency (*RF*) (see Bauer 2001: 125-162; Bauer, Beliaeva & Tarasova 2019). The *V* and *N* frequencies indicate, respectively, the number of different units occurring for a process, and the number of total occurrences (including repetitions) of a process. For example, the 39 NDVCs with the meaning location are a subtype of the study sample, and 39 is its *V* frequency because there are 39 different forms. One of these types is *streettalk*, which appears 8 times in the BNC, so 8 is its *N* frequency because those 8 occurrences of *streettalk* are repeated forms of the same type. One of the earliest productivity-related formulae is precisely the *V/N* ratio (Aronoff 1976; 1983), aimed at measuring the lexical diversity of a process and taken as an indirect indication of high productivity. Such correlation, however, has been questioned on different grounds and is not normally taken as significant in itself, but when combined with other procedures (Baayen & Lieber 1991; Bauer 2001: 145-147; Baayen 2005).

A more sophisticated option is Baayen's (1994; 2009; Baayen & Lieber 1991) *productivity in the strict sense* (*P*), whereby the productivity potential of a given process is gauged based on its likeliness to create new words. *P* is calculated through the formula below, which divides a process's hapaxes (n_1) by its token frequency (*N*), and where the higher the result, the more probable it is to come across new lexemes created by that process.

$$P = n_1 / N$$

In Baayen's view, hapaxes are indirect representations of neologisms because highly productive processes create a great number of one-off formations which, during corpus sampling, will be captured only

once. This entails that, given a sufficiently large corpus (Baayen employs an 18-million-word one), each hapax can be taken as a measure of one-off formations which were created by a process but did not make it to the corpus. This is also why hapaxes are not necessarily neologisms, but they are indications of a process's productive potential. Baayen's model is originally devised for affixational processes, but it has been here applied to NDVCs under the assumption that its rationale will still hold, namely, that more hapax legomena should indicate a higher degree of productivity. Let us look at NDVCs meaning location for illustration. In order to obtain *P* for this semantic category, a division is required between its n_1 value (19) and its *N* value (191), the result of which is 0.0994. This figure is not perhaps telling in itself, but once compared with the *P* values of other processes (see section 3.3.2).

Finally, *relative frequency* (*RF*) is a proposal by Hay (2001; 2003) which considers not only the token frequency of the derivative but also that of its lexical base, e.g. the *N* value of *recyclable* and also of *recycle*. Following psycholinguistic postulations, the rationale of Hay's model is that derivatives which are less frequent than their base will reveal a higher future productivity than derivatives which are more frequent that their base. In adapting Hay's proposal to compounding, it was decided to use the *N* of the NDVC and that of its head (since in compounding there is no lexical input analogous to that of derivation). The widely agreed right-headedness of English NN compounds seems reason enough to bring this constituent into play here (Allen 1978; Selkirk 1982; Giegerich 2004). The *RF* formula therefore divides the token frequency of the whole NDVC (*N*) by the token frequency of its head (N_h), e.g. *shoe polish* (17) and *polish* (411). In all formulae above a higher result means a higher productivity degree, with the exception of the *RF* model, where lower values point to higher productivity:

$$RF = N / N_h$$

3.3.2. Productivity degrees of NDVCs

Table 3 and Table 4 summarize, in turn, the values and the ranking of each orientation type for the five productivity models introduced in section 3.3.1. Results have been split by number of possible orientations in order to facilitate separate inspection. Table 4 omits the three-orientation section, as it consists of just one category (S/O/PO) and no ranking is possible:[9]

	V	n_1	*N*	*V/N*	*P*	*RF*
	One orientation					
S	119	34	7,649	0.0155	0.0044	0.0127
O	333	140	5,127	0.0649	0.0273	0.0014
PO	145	53	2,926	0.0495	0.0181	0.0029
	Two orientations					
S/O	117	41	1,519	0.0770	0.0269	0.0015
S/PO	7	3	21	0.3333	0.1428	0.0002
O/PO	54	14	383	0.1409	0.0365	0.0006
	Three orientations					
S/O/PO	28	8	666	0.0420	0.0120	0.0026
Total	803	293	1,087.221	–	–	–

Table 3. Results of productivity models.

Table 3 displays interesting aspects. There is, for instance, a dominance of the one-orientation categories for the *V* and *N* values, as could be expected from the structural predisposition of NDVCs towards meaning interpretation (see 3.2). In particular, formations with one interpretation amount to 597, while those with two possible interpretations are 178, and 28 units have three possible interpretations. We next find a total of n_1 293, a very high proportion considering the overall *V* value 803. This should be interpreted as an indicator of extreme productivity, although it must be conceded that the size of the study corpus is small by Baayen's (1994; 2005; 2009) standards, with corpora of normally no less than millions of tokens (see Berg 2020: 1120-1121). The n_1 value of 293 has a strong impact on the results of other models, as shown below. Hapaxes rank especially high in the category Object (140) and are rather low for Subject-orientation (34), among others. To this must be added a comparatively high *V* value in Subjects and a lower one in Objects, with the consequence that the productivity estimates of *P* and *RF* clearly favours an Object reading. The reasoning would be that NDVCs with Object-orientation are created frequently and therefore their *N* value is the minimum possible (i.e. 1); in contrast, NDVCs with a Subject-orientation do not arise so regularly, hence the same units are repeatedly used, so their *N* value rises (Aronoff 1983; Anshen & Aronoff 1988; Baayen 1994; 2009). Indeed, Subject-oriented units are those with higher frequencies in the corpus: *sunshine* (1,244), *headache* (1,105), *earthquake* (684), *sunset* (598)

or *waterfall* (588) (see Baayen & Lieber 1991). This means a marked divergence with the 140 hapaxes in the Object category.

If we turn to the three formula-based models, anomalous figures are found for the categories of S/PO and S/O/PO, since extremely low *V* values are regarded as not representative (Baayen 1994), and therefore these two groups should be taken with a pinch of salt. Moreover, it has been claimed that Baayen's models tend to disfavour processes whose units have a high *N*, and they are hence not adequate to compare processes with very different *N* values (Berg 2020: 1121). Other than that, the *V/N* ratio confirms a high value of Object-orientation categories and the opposite for Subject-orientation, which is explained in terms of a higher lexical richness in Object-oriented NDVCs, i.e. that these units resort to more diverse lexicon and are thus more varied in their composition. As for Baayen's *P*, Object-oriented stands as the most profitable set (0.0273) in the group of one orientation, and O/PO (0.0365) is the most profitable one among two-orientation NDVCs. The least productive categories involve Subjects, both under one possible orientation (0.0044) and in the aforesaid unproductive cases of S/PO and S/O/PO. Finally, the results of RF appear in the same vein with a dominance of Object-orientation both under one possible orientation (0.0014) and under O/PO (0.0006), in fact the most productive category for *RF*.

	V	n_1	*N*	*V/N*	*P*	*RF*
	ONE ORIENTATION					
1st	O	O	S	O	O	O
2nd	PO	PO	O	PO	PO	PO
3rd	S	S	PO	S	S	S
	TWO ORIENTATIONS					
1st	S/O	S/O	S/O	S/PO	S/PO	S/PO
2nd	O/PO	O/PO	O/PO	O/PO	O/PO	O/PO
3rd	S/PO	S/PO	S/PO	S/O	S/O	S/O

Table 4. Ranking of orientation types per productivity model.

Table 4 displays the evidence differently, but again validates the common assessment of Object-oriented categories as more productive than Subject-oriented ones. The reasons are various, for example the adaptability of NDVCs to various contexts, which would here favour an

Object-oriented reading (see 3.1), or the fact that left-hand constituents in the corpus entries are largely inanimate entities where a Subject-oriented reading would seem forced, e.g. *escalator design* (1), *calendar reform* (2), *camera-shake* (4), or *ice pick* (5). From a semantic-cognitive angle, Table 3 and Table 4 confirm that NDVCs are semantically adaptable to different contexts as well as interpretively-friendly. These results are also in line with descriptions that portray a lower dominance of Subject-orientation and a salient role of context (Bauer, Lieber & Plag 2013: 471; Lieber 2016a; 2016b).

Notwithstanding these findings, recent concerns have been raised about frequency-based formulae. Among others, Bauer, Beliaeva & Tarasova (2019: 44) rightly point out that a number of productivity models have been proposed, "none of which has been entirely acceptable", and Berg (2020) goes through the pitfalls of some productivity methods in word-formation. Along these lines, Dressler, Libben & Korecky-Kröll (2014) raise the *tooth-brush objection* against the excessive weight given to corpus frequencies: the fact that some lexemes may not have a high token frequency in corpora, but even so have a high pragmatic relevance. Dressler, Libben & Korecky-Kröll illustrate their point with the compound *tooth-brush* which, despite its low frequency (*N* 191 in the BNC), comes to the mind of the average speaker "with regular frequency and such thoughts are generally verbally shaped in silent inner language". They generalize:

> [T]he text frequency of either written or oral texts cannot be the source of frequency effects: the real source is the frequency of verbal thinking. And such frequency cannot be established by counting occurrences in corpora but only by judicious rating experiments. (Dressler, Libben & Korecky-Kröll 2014: 188)

This point, already raised by Bauer (2001: 36-37) when weighing up the notion of *item-familiarity*, is certainly reasonable in acknowledging an evident importance of pragmatics and the real world, but it poses the risk of boycotting any frequency-based results. In our case, with all the due caution, it seems evident that the findings in Table 3 and Table 4 provide significant insights into NDVCs and prove useful in confirming our intuitions, namely that Subject-orientation is less widespread than it could initially seem, and that Objects and Prepositional-Objects play a prominent role.

With the above in mind, it may be sensible to perceive a collaborative more than competitive relationship between the two facets of NDVCs: argumental and non-argumental. It has been shown that formations like *drumbeat* (11) can be analysed doubly as primary ('the beat comes FROM

the drum') and as synthetic ('X beats the drum', Object-oriented), while in other cases a primary (*motorway crash* under RDP IN) or a synthetic reading (*ancestor worship* as Object-oriented) is more likely. Crucially, the head of NDVCs is in all cases a deverbal converted noun, and this entails a verbal nature that can apparently remain latent and be activated when needed. This collaborative nature between the primary and synthetic sides of NDVCs is reminiscent of Jackendoff's (2016: 20) notion of promiscuity, originally restricted to the meaning potential of primary NN compounds (see Bierwisch 2015: 33; Schäfer & Bell 2020).

4. Conclusion

The literature has largely described NDVCs as a subtype of synthetic compounding, but this hyponymic relationship is less straightforward than it would seem. Parallel to the descriptions of the affixal types of synthetic compounding, NDVCs allow for both argumental and non-argumental readings. It has been shown that contextual meaning is often decisive for (non-)argumental disambiguation, but also, as in (18), that context does not always resolve interpretation.

(18) *The polymer treatment might be used straight away as a daily* mouthwash *for mentally handicapped people to stop plaque sticking to teeth.*

In that respect, it may be concluded that an argumental reading (triggered by a deverbal head noun) reduces semantic indeterminacy; in the case of a non-argumental reading, the meaning of the NDVCs will correspond to that of a primary NN compound and the traditional semantic vagueness of NNs will hold. Certainly, it is important to contextualize the findings of productivity models, as their outcome is largely influenced by the weight of some specific units in the sample, whose high *N* value crucially affects the overall figures of processes (see Table 3).

Regarding deverbal heads, the valency of the base verb which derives the head noun is fundamental to its range of meanings. In such cases, valency is retained from the lexical input (e.g. $design_V > design_N$), which is why as many readings are possible as potential arguments occur in the verb, in the case of *design*, a Subject or Object interpretation, e.g. *computer design* as 'a design done by a computer' or 'someone designs the computer'. This is in agreement with statements that have stressed a noun-verb dependence in conversion nouns, and confirms that the tendency found in regular V>N conversion applies also to V>N conversion within NDVCs (Bauer 2005: 22; Lieber 2016b: 111-112;

Iordăchioaia *et al.* 2020). The "morphosemantic and morphotactic condensation" in these compound type (Dressler, Libben & Korecky-Kröll 2014: 186) makes them doubly effective thanks to conversion, although this "may be unsatisfactory from the point of view of descriptive economy" (Kastovsky 2005: 32).

Finally, a natural conclusion is that there are sound reasons to perceive argumental and non-argumental NDVCs as belonging to "a single class", as has been done for synthetic compounds with overt suffixes (Bauer 2020; Olsen 2017; see Bierwisch 2015: 35-36; Härtl 2015: 882). Bearing in mind the structural and semantic similarities between argumental and non-argumental readings, and the fact that the analysis of a unit greatly depends on extralinguistic factors, it would seem that there are more pros than cons for a unified view, although the term and characterization of such category deserves an investigation *per se*. As Bauer (2020: 271) puts it, the main problem for attempts at compound categorization is "that there is invariably some remainder of compounds which fit uneasily into any such classification". The one-category option would at least provisionally allow a coherent analysis of NDVCs and would avoid an analysis *a la* Procrustean bed.

A number of questions remain unanswered beyond the above findings, for example whether, under parallel circumstances, the argumental or the non-argumental reading is preferential during interpretation, or which specific meanings are transferred from the verb to the converted nominal.

Acknowledgements

I am grateful to the two anonymous reviewers for insightful feedback on the form and content of the article, and to Pius ten Hacken for assistance with the bibliography. This publication has been supported by the project PID2020-119851GB-I00 MCIN/AEI/10.13039/501100011033, funded by the Spanish State Research Agency (AEI), the Ministry of Science and Innovation (MCIN) and European Regional Development Fund (FEDER).

Notes

[1] Corpus units are accompanied by their bracketed lemmatized BNC token frequency. No frequency value means that the unit does not come from the study corpus.

[2] As pointed out by one of reviewers, the term *synthetic compound* has been used in at least another different sense, e.g. by Neef (2015), where the condition for a [N + V + *-er*] structure to be 'synthetic' is the non-existence of both $[N + V]_V$ and $[V + er]_N$. In this perspective, *truck driver* is not synthetic, but there are many synthetic compounds that do not have a V as a second element, as in German *blauäugig* 'blue-eyed' [A + N + *-ig*].

[3] The earliest observation on this matter comes from Marchand's (1967: 14) definition of *expansion*, of which compounds are a subtype: "An expansion is a combination AB which is analysable on the basis 'B determined by A', with AB belonging to the same word class and lexical class to which B belongs. […] Semantically speaking, the determinatum represents the element whose range of applicability is limited through the determinant" (see Kastovsky 1999; 2005: 33).

[4] One lesser-known approach to synthetic compounding is Levi's (1978), which contemplates derivation by nominalization, where four types are possible: ACT, PRODUCT, AGENT and PATIENT. Levi, however, does not explicitly address compounds involving nominalization by conversion.

[5] A different formalization is found in Jackendoff (2010: 427-448; 2016), developed only succinctly here for reasons of space. Framed in the Parallel Architecture framework, two schemas are proposed, one for argumental and one for non-argumental compounds. The Argument Schema corresponds to synthetic compounds, while the Modifier Schema corresponds to primary compounds (see also Bierwisch 2015: 31-36).

Argument schema
$[N_1\ N_2] = [Y_2 (\ldots, X_1, \ldots)]$
'a N_2 by/of/ … N_1'
Modifier Schema
$[N_1\ N_2] = [Y_2^{\alpha}; [F (\ldots, X_1, \ldots, \alpha \ldots)]]$
'an N_2 such that F is true of N_1 and N_2'

[6] The kind of orientation in *baby care* (35), *gunfight* (18) or *pub-talk* (1) has been called *adjunct* (Spencer 1991: 325; Lieber 2009) and *prepositional object* (Lieber 2010; Bauer, Lieber & Plag 2013). Certainly, a subject- or object-oriented reading does not seem the most likely ('care FOR a baby', 'a fight BY MEANS OF guns', 'talk AT the pub'), although they would be possible (e.g. *baby care* as subject-oriented in a fictional context). Given the syntactic implications of each term, the former is kept here (see Huddleston 2002; Mittwoch, Huddleston & Collins 2002; Duffley 2020 for complements *vs* adjuncts).

[7] In nominalization studies, this dual nature has been referred to as *referential/result/R* reading, in contrast to an *eventive/E* reading (see Grimshaw 1990; Lieber 2016b: 111-112; Andreou & Lieber 2020: 338-339; Iordăchioaia *et al.* 2020: 121-123; Lieber & Plag 2021). In *cancer fight* (2), the R reading using Levi (1978) would be 'the fight is ABOUT cancer', while its E reading would be 'X fights cancer' (Object-orientation).

[8] Note that a second Prepositional-Object reading ('to talk to the baby') is possible here, disregarded due to the fuzzy nature of the distinction argumental *vs* non-argumental compound.

[9] Orientation types are abbreviated are as follows: S (Subject), O (Object), PO (Prepositional-Object). Productivity models appear as follows: *V* (type frequency), n_1 (hapaxes), *N* (token frequency), *P* (productivity in the strict sense), *RF* (relative frequency).

Bibliographical References

Adams, Valerie 2001. *Complex Words in English*. Harlow: Longman.

Allen, Margaret R. 1978. *Morphological investigations*. PhD dissertation. University of Connecticut, Storrs.

Altmann, Gabriel 2002. Zipfian linguistics. *Glottometrics* 3. 19-26.

Andreou, Marios & Lieber, Rochelle 2019. Aspectual and quantificational properties of deverbal conversion and -ing nominalizations: The power of context. *English Language and Linguistics* 24,2. 333-363. <DOI: 10.1017/s1360674319000108>.

Anshen, Frank & Aronoff, Mark 1988. Producing morphologically complex words. *Linguistics* 26,4. 641-655. <DOI: 10.1515/ling.1988.26.4.641>.

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

Aronoff, Mark 1983. Potential words, actual words, productivity and frequency. In *Proceedings of the XIII International Congress of Linguists, August 29–September 4, 1982, Tokyo*. Tokyo: Permanent International Committee on Linguistics. 163-171.

Baayen, Harald R. 1994. Productivity in language production. *Language and Cognitive Processes* 9,3. 447-469. <DOI: 10.1080/01690969408402127>.

Baayen, Harald R. 2005. Morphological productivity. In Köhler, Reinhard; Altmann, Gabriel & Piotrowski, Rajmund G. (eds.), *Quantitative Linguistics. An International Handbook*. Berlin: Walter de Gruyter. 243-256.

Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (eds.), *Corpus linguistics*. Berlin / New York: Mouton de Gruyter. 899-919. <DOI: 10.1515/9783110213881.2.899>.

Baayen, Harald R. & Lieber, Rochelle 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29,5. 801-844. <DOI: 10.1515/ling.1991.29.5.801>.

Bauer, Laurie 1983. *English Word-Formation*. Cambridge: Cambridge University Press.

Bauer, Laurie 1998. When is a sequence of two nouns a compound in English? *English Language and Linguistics* 2,1. 65-86. <DOI: 10.1017/s1360674300000691>.

Bauer, Laurie 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie 2005. Conversion and the notion of lexical category. In Bauer, Laurie & Valera, Salvador (eds.), *Approaches to Conversion/Zero-Derivation*. Münster: Waxmann. 19-30.

Bauer, Laurie 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.

Bauer, Laurie 2020. Compounds. In Aarts, Bas; Bowie, Jill & Popova, Gergana (eds.), *The Oxford Handbook of English Grammar*. Oxford: Oxford University Press. 262-280.

Bauer, Laurie & Renouf, Antoinette 2001. A corpus-based study of compounding in English. *Journal of English Linguistics* 29,2. 101-123. <DOI:

10.1177/00754240122005251>.
Bauer, Laurie & Huddleston, Rodney 2002. Lexical word-formation. In Huddleston, Rodney & Pullum, Geoffrey K. (eds.), *The Cambridge Grammar of the English Language*. Cambridge University Press. 1621-1722. <DOI: 10.1017/9781316423530.020>.
Bauer, Laurie; Lieber, Rochelle & Plag, Ingo 2013. *Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
Bauer, Laurie; Beliaeva, Natalia & Tarasova, Elizaveta 2019. Recalibrating productivity: Factors involved. *Zeitschrift für Wortbildung / Journal of Word Formation* 3,1. 44-80. <DOI: 10.3726/zwjw.2019.01.02>.
Berg, Kristian 2020. Changes in the productivity of word-formation patterns: Some methodological remarks. *Linguistics* 58,4. 1117-1150. <DOI: 10.1515/ling-2020-0148>.
Bierwisch, Manfred 2015. Word-formation and argument structure. In Müller, Peter O.; Ohnheiser, Ingeborg; Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*. Berlin: de Gruyter Mouton. 1056-1099. <DOI: 10.1515/9783110246278-016>.
Bolozky, Shmuel 1999. *Measuring Productivity in Word Formation: The Case of Israeli Hebrew*. Leiden: Brill.
Brunner, Annelen; Engelberg, Stefan & Hein, Katrin 2021. The distribution of constituent words in nominal compounds and its impact on semantic interpretation: An empirical study. *Zeitschrift für Wortbildung / Journal of Word Formation* 5,1. 6-36. <DOI: 10.3726/zwjw.2021.01.01>.
Cetnarowska, Bożena 2020. Interaction of conversion and compounding in construction morphology. Paper Presented at the Workshop *Derivational Zero Affixes*, at the *53rd Annual Meeting of the Societas Linguistica Europaea*, 26 August–1 September 2020.
Corbin, Danielle 1987. *Morphologie dérivationnelle et structuration du lexique*. Berlin / New York: de Gruyter. <DOI: 10.1515/9783111358390>.
Davies, Mark 2004-. British National Corpus (from Oxford University Press). <www.english-corpora.org/bnc>.
Dressler, Wolfgang U.; Libben, Gary & Korecky-Kröll, Katharina 2014. Conflicting vs. convergent vs. interdependent motivations in morphology. In MacWhinney, Brian; Malchukov, Andrej & Moravcsik, Edith (eds.), *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press. 180-196. <DOI: 10.1093/acprof:oso/9780198709848.003.0011>.
Duffley, Patrick 2020. Clause structure, complements, and adjuncts. In Aarts, Bas; Bowie, Jill & Popova, Gergana (eds.), *The Oxford Handbook of English Grammar*. Oxford: Oxford University Press. 358-374.
Giegerich, Heinz J. 2004. Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics* 8,1. 1-24. <DOI: 10.1017/s1360674304001224>.
Grimshaw, Jane B. 1990. *Argument Structure*. Cambridge, MA: MIT Press.
Härtl, Holden 2015. Argument-structural restrictions on word-formation patterns. In Müller, Peter O.; Ohnheiser, Ingeborg; Olsen, Susan & Rainer, Franz (eds.), *Word-Formation. An International Handbook of the Languages of Europe*. Berlin: de Gruyter Mouton. 876-894. <DOI:

10.1515/9783110246278-005>.
Haspelmath, Martin 2014. On system pressure competing with economic motivation. In MacWhinney, Brian; Malchukov, Andrej & Moravcsik, Edith (eds.), *Competing Motivations in Grammar and Usage*. Oxford: Oxford University Press. 197-208. <DOI: 10.1093/acprof:oso/9780198709848.003.0012>.
Hay, Jennifer 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39,6. 1041-1070. <DOI: 10.1515/ling.2001.041>.
Hay, Jennifer 2003. *Causes and Consequences of Word Structure*. New York: Routledge.
Huddleston, Rodney 2002. The clause: Complements. In Huddleston, Rodney & Pullum, Geoffrey K. (eds.), *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press. 213-322. <DOI: 10.1017/9781316423530.005>.
Iordăchioaia, Gianina; Schweitzer, Susanne; Svyryda, Yaryna & Cabrera, María Camila Buitrago 2020. Deverbal zero-nominalization and verb classes: Insights from a database. *Zeitschrift für Wortbildung / Journal of Word Formation* 4,2. 120-142. <DOI: 10.3726/zwjw.2020.02.07>.
Jackendoff, Ray 2010. *Meaning and the Lexicon: The Parallel Architecture, 1975-2010*. Oxford: Oxford University Press.
Jackendoff, Ray 2016. English noun-noun compounds in Conceptual Semantics. In ten Hacken, Pius (ed.), *The Semantics of Compounding*. Cambridge: Cambridge University Press. 15-37. <DOI: 10.1017/cbo9781316163122.002>.
Kastovsky, Dieter 1999. Hans Marchand's theory of word-formation: Genesis and development. In Carls, Uwe & Lucko, Peter (eds.), *Form, Function and Variation in English. Studies in Honour of Klaus Hansen*. Frankfurt am Main: Peter Lang. 19-39.
Kastovsky, Dieter 2005. Conversion and/or zero: Word-formation theory, historical linguistics, and typology. In Bauer, Laurie & Valera, Salvador (eds.), *Approaches to Conversion/Zero-Derivation*. Münster: Waxmann. 31-49.
Katamba, Francis 1994. *English Words*. London: Routledge. <DOI: 10.4324/9780203495971>.
Levi, Judith N. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
Lieber, Rochelle 1983. Argument linking and compounds in English. *Linguistic Enquiry* 14,2. 251-285.
Lieber, Rochelle 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
Lieber, Rochelle 2009. IE, Germanic: English. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 357-369.
Lieber, Rochelle 2010. On the lexical semantics of compounds. In Scalise, Sergio & Vogel, Irene (eds.), *Cross-Disciplinary Issues in Compounding*. Amsterdam: John Benjamins. 127-144. <DOI: 10.1075/cilt.311.11lie>.
Lieber, Rochelle 2016a. Compounding in the lexical semantic framework. In ten Hacken, Pius (ed.), *The Semantics of Compounding*. Cambridge: Cambridge University Press. 38-53. <DOI: 10.1017/cbo9781316163122.003>.
Lieber, Rochelle 2016b. *English Nouns: The Ecology of Nominalization*. Cambridge:

Cambridge University Press.
Lieber, Rochelle & Štekauer, Pavol 2009. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.
Lieber, Rochelle & Plag, Ingo 2021. The semantics of conversion nouns and *ing* nominalizations: A quantitative and theoretical perspective. *Journal of Linguistics* 58,2. 307-343. <DOI: 10.1017/s0022226721000311>.
Maguire, Phil; Wisniewski, Edward J. & Storms, Gert 2010. A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory* 6,1. 49-73. <DOI: 10.1515/cllt.2010.003>.
Marchand, Hans 1967. Expansion, transposition and derivation. *La Linguistique* 3,1. 13-26.
Marchand, Hans 1969. *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach*. 2nd edition. München: C.H. Beck.
Mattiello, Elisa & Dressler, Wolfgang U. 2022. Dualism and superposition in the analysis of English synthetic compounds ending in *-er*. *Linguistics* 60,2. 395-461. <DOI: 10.1515/ling-2021-0235>.
Melloni, Chiara 2020. Subordinate and synthetic compounds in morphology. In *Oxford Research Encyclopedia of Linguistics*. <DOI: 10.1093/acrefore/9780199384655.013.562>.
Merriam-Webster Dictionary. <www.merriam-webster.com>.
Mittwoch, Anita; Huddleston, Rodney & Collins, Peter 2002. The clause: Adjuncts. In Huddleston, Rodney & Pullum, Geoffrey K. (eds.), *The Cambridge Grammar of the English Language*. Cambridge University Press. 663-784. <DOI: 10.1017/9781316423530.009>.
Neef, Martin 2015. Synthetic compounds in German. In Müller, Peter O.; Ohnheiser, Ingeborg; Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*. Berlin: de Gruyter Mouton. 582-593. <DOI: 10.1515/9783110246254-035>.
Nóbrega, Vitor A. & Panagiotidis, Phoevos 2020. Headedness and exocentric compounding. *Word Structure* 13,2. 211-249. <DOI: 10.3366/word.2020.0168>.
Olsen, Susan 2015. Composition. In Müller, Peter O.; Ohnheiser, Ingeborg; Olsen, Susan & Rainer, Franz (eds.), *Word-Formation. An International Handbook of the Languages of Europe*. Berlin: Mouton de Gruyter. 364-386.
Olsen, Susan 2017. Synthetic compounds from a lexicalist perspective. *Zeitschrift für Wortbildung / Journal of Word Formation* 1,1. 17-45. <DOI: 10.3726/zwjw.2017.01.03>.
Plag, Ingo 1999. *Morphological Productivity: Structural Constraints in English Derivation*. Berlin: Mouton de Gruyter.
Plag, Ingo; Kunter, Gero; Lappe, Sabine & Braun, Maria 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84,4. 760-794. <DOI: 10.1353/lan.0.0072>.
Rakić, Stanimir 2009. Some observations on the structure, type frequencies and spelling of English compounds. *SKASE Journal of Theoretical Linguistics* 6,1. 59-83. <www.skase.sk/Volumes/JTL13/pdf_doc/04.pdf>.
Roeper, Thomas & Siegel, Dorothy 1978. A lexical transformation for verbal compounds. *Linguistic Inquiry* 9. 199-260.

Sanchez-Stockhammer, Christina 2018. *English Compounds and Their Spelling*. Cambridge: Cambridge University Press.
Schäfer, Martin & Bell, Melanie J. 2020. Constituent polysemy and interpretational diversity in attested English novel compounds. *Semantics and Psychology of Complex Words* 15,1. 42-61. <DOI: 10.1075/ml.00013.sch>.
Selkirk, Elisabeth O. 1982. *The Syntax of Words*. Cambridge, MA: MIT Press.
Spencer, Andrew 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Blackwell.
Štekauer, Pavol 2005. *Meaning Predictability in Word Formation: Novel, Context-Free Naming Units*. Amsterdam: John Benjamins.
ten Hacken, Pius 2009. Early generative approaches. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Compounding*. Oxford: Oxford University Press. 54-77.
ten Hacken, Pius 2016. *Semantics of Compounding*. Cambridge: Cambridge University Press.
ten Hacken, Pius 2023. Synthetic compounding. In Ackema, Peter; Bendjaballah, Sabrina; Bonet, Eulàlia & Fábregas, Antonio (eds.), *The Wiley Blackwell Companion to Morphology*. Malden, MA: Wiley-Blackwell.
Zipf, George Kingsley 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley.

Appendix. Input NDVCs by source

- Adams (2001): ice fall, flea bite, telephone call
- Bauer (2020): peace talks
- Bauer & Renouf (2001): mouse squeak
- Cetnarowska (2020): oilspill, bear hunt, mouseclick, skin tear, hotel offer, ice pick, prison guard
- Lieber (2010): dog attack, bee sting, earthquake, heartbeat, rainfall, mouse squeak, brain bleed, government collapse, heartburn, eyewink, ballkick, cost control, court reform, energy audit, fare increase, fee hike, sun worship, manslaughter, bodyguard, blood test, boat ride, homework, day dream, moon walk, pub crawl, gunfight, table talk, hand stand, age limit, traffic alert, spending cuts, plane crash, robot repair, baby care, computer design
- Lieber (2016a): landslide, slide rule, snowdrift, haircut, bloodshed, clam bake
- Jackendoff (2010): tooth decay, speed limit, onion smell, birth order, power supply, doorstop, wine press, hair dye, noise filter, bookmark, stomach pump, wing support, eyewash, mouthwash, toothbrush, nailbrush, paperclip, mouse trap, shoe polish, neck brace, ear plug, chicken feed, nail file, toothpick, dog whistle, hearing aid, morning swim, sunburn, knife wound, surface drag, safety lock