

Laurie Bauer, Natalia Beliaeva & Elizaveta Tarasova

Recalibrating Productivity: Factors Involved

Abstract: This paper approaches productivity by considering three case studies: compounds, blends and phrasal verbs. The aim of the paper is to encourage a discussion about the factors involved in the notion of productivity, and to show why so many of the established measures are not completely satisfactory or are interpreted in a way that is not.

Keywords: productivity, blend, compound, phrasal verb, availability

1. Introduction

There is an extensive literature on morphological productivity, and yet the nature of productivity remains obscure (for a summary of much of the material, see Bauer 2001). In this paper,* we attempt to examine some data on productivity which does not fit easily into current models, and to suggest new ways of looking at the phenomena involved.

Productivity is generally defined as the extent to which some morphological process is exploited by speakers. Following Corbin (1987), it has become standard to distinguish between availability (whether a given morphological process can be used at all) and profitability (the degree to which an available process can be used). This implies that there are degrees of productivity, although precisely what influences such degrees remains controversial. One view, for example, is that productivity is the result of constraints on the morphological process: the more constraints there are on a process, the less productive (profitable) it is, until eventually the constraints prohibit any use at all (for discussion see Bauer 2001: § 2.3). Whether or not this is the case, there have been various attempts to provide a quantitative measure of productivity (Aronoff 1976, Baayen 1992, Baayen and Lieber 1991, Plag 1999), none of which has been entirely acceptable, although many of these measures have been used with some success. More recently, Hay (2003, Hay and Plag 2004) has offered a processing view of productivity, by which the most productive processes are semantically transparent and phonologically distinct. This has made linguists consider productivity as an outcome of cognitive principles associated with coding and processing, rather than as a formal feature associated with the processes themselves. We view this as progress, even if we think there is more to be said.

* We should like to thank the two anonymous referees who gave us very detailed feedback on an earlier draft.

Productivity has mainly been discussed in relation to derivational morphology. Some scholars (notably Dressler 2003) have extended this discussion to inflectional morphology as well. Following Yang (2016), we would like to see productivity as a phenomenon which does not solely affect morphological processes, but also affects syntactic and phonological processes. Yang cites such phenomena as dative alternation (linking *She gave the teacher an apple* to *She gave an apple to the teacher*), and English stress rules as areas which are plausibly treated as being subject to an analysis in terms of productivity. We do not wish to recapitulate Yang's work here, but one of the areas we consider, the area of phrasal verbs, has traditionally been viewed as syntactic in English, and we contend that whatever is affecting the use of verb + particle combinations in English is comparable to what affects the use of base + affix combinations, even if one is a matter of syntax and the other is a matter of morphology.

In this paper we consider some of the weaknesses with current theories of productivity by looking at the influences on various processes which, we would claim, are not easily taken into account in current theory. We are less concerned with constructing a theory of productivity than we are with pointing out factors which we feel any such theory will need to consider.

The structure of this paper is as follows. After some preliminary discussion in Sections 2 and 3, we examine three particular cases: the productivity of N+N compounds in English, the productivity of blends in English, and the productivity of phrasal verbs in English. In each case we use different methodologies to make the point that general assumptions about productivity are not sufficient. In the final section (Section 7), we draw together the various strands and provide some suggestions on the subject of what a theory of productivity must be able to deal with.

2. What does productivity measure?

The various measures of productivity in the literature (often called indexes of productivity, Baayen and Lieber 1991: 803) estimate the overall vitality of an individual affix or process. Frequently, this is done in a comparative manner: the productivity of *-ness* is compared with that of *-ity*, and so on. Whether the measures are statements of the situation in a particular corpus, or whether they are intended as a statement of what happens in “the language”, they locate productivity in the individual morphological process (typically the affix).

Yet we know this to be misleading. As is pointed out by Aronoff (1976), the productivity of individual affixes varies enormously according to domain. This is illustrated in Table 1, which shows the number of hits in COCA (Davies 2008) for various combinations of suffixes, where the last suffix is either *-ness* or *-ity*. What Table 1 shows is that any result for the relative

productivity of these two suffixes is merely an average over widely diverse patterns of productivity. Note that this variability is not simply a matter of learned versus native affixes preferring to collocate with matching affixes. Note further that the examples below the gap in Table 1 fail to show any particular preference in either direction.

Tab. 1: Types in COCA illustrating different combinations of final suffixes with -ness and -ity.

<i>-ity</i> preferred		<i>-ness</i> preferred	
-ality	>1000	-antness	6
-alness	91	-antity	only <i>quantity</i>
-arity	264	-esqueness	3
-ariness	46	-esquity	0
-bility	>1000	-fulness	154
-bleness	81	-fulty	0
-icity	410	-ishness	138
-icness	35	-ishity	0
-inity	118	-ousness	435
-ineness	9	-osity	133
-ority	226	-someness	32
-oriness	15	-sometry	0
-anity	86	-oidness	0
-anness	57	-oidity	0

Most of the types that give rise to the patterns illustrated in Table 1 will be well-established words of English, and the result of past productivity. They may not reflect present or past productivity in all topic areas or in all genres. Nevertheless, they indicate that usage is dependent upon morphological context or domain, and thus that simple statements about individual affixes are insufficient. Such a conclusion is not new: for previous statements along similar lines see, for example, Hay and Plag (2004), Baayen (2009).

An alternative view of how productivity should be measured might be prompted by an onomasiological approach to word-formation (see, e.g. Štekauer 2000). In an onomasiological approach, the speaker creating a new form has to choose between a number of possible patterns to capture the appropriate meaning. A coiner seeking a new word for a person who captures protected birds in the wild in order to band them for scientific research could, in principle, coin a *bird-catcher*, a *parrot-bander*, a *catch-person*, a *catcher*, a

snarist, *an entrapmentant*, and so on. Accordingly, one view of productivity could be that productivity is a measure of the likelihood of any given morphological process (say compounding or *-er* affixation) being used in such a semantically-determined environment. This might or might not be further constrained by the presence of particular final suffixes in the potential bases, as illustrated in Table 1.

If we accept the view proposed in Baayen (2009) that there are three kinds of productivity – realised productivity (Bauer 2001 talks about ‘generalisation’ here), expanding productivity and potential productivity, it is clear that measuring ‘productivity’ is difficult, and clear why different measure focus on different aspects of the topic. Different approaches highlight different facets of productivity, many of which may be simultaneously valid, although some may contradict each other (as illustrated in Section 5.3). What this shows is that it is not necessarily clear just what the ideal index of productivity (assuming such a thing exists) should be a measure of. While we are not necessarily committed to any particular view of precisely what should be measured, we feel that it is potentially problematic to assume any of them without discussion.

3. Some preliminaries

In order to put our examples and discussion into a framework, we need to define some of the terms we will be using. We have already mentioned availability and profitability as branches of productivity, and we will use these two labels in our discussion.

Many complex words are, in the terms of Meys (1975), item-familiar. That is, whether they are created by morphological processes or not, individual speakers recognise them as words of which they have experience. This is trivially true of simplex words like *cat*, *elephant*, *house*, *sky*, but is also true of words such as *blackmail*, *capitalise*, *loneliness*, *parental*, *truth*. These are all morphologically complex, but their form-meaning relationship is set in the language. We refer to such words as established words. Words can be established by being lexicalised (that is, they were once created by rules or forms which are no longer available in current English), such as *blackmail* and *truth*, or because they are institutionalised (their form and/or meaning is conventionally fixed, even if they could be derived in the present state of the language system). For discussion of this division see Bauer (1983: 48–50). Because the distinction between institutionalisation and lexicalisation can be hard to determine in individual cases, and because the distinction is not of importance to us here, we avoid the issue through the use of the term *established*.

Frequency, especially relative frequency, is an important concept in dealing with productivity. Like others before us, we distinguish between type frequency and token frequency. If we assume that we did a search of a small corpus

and found the following forms including *-dom*: *bachelordom*, *Christendom*, *earldom*, *fairydom*, *kingdom*, *kingdom*, *kingdoms*, *martyrdom*, *officialdom*, *officialdom*, we could say that *-dom* in this corpus has a type frequency of 7, that *kingdom* has a token frequency of 3, and that *-dom* has a token frequency of 10. That is, the type frequency is the number of different words that contain the affix, and the token frequency is the total number of times that affix is found. Problems are raised by items such as *subkingdom* which could be seen as tokens of *kingdom* or as new types, but such matters are not of concern to us here.

Finally, we use the term construction as it is used in Construction Grammar, or, more specifically, in construction morphology (Booij 2010). A construction is a pairing of form and meaning. It is often, but not necessarily, the case that some of the meaning arises from the pairing. A single instantiation of a construction containing all relevant lexical material is a construct (Booij 2010: 12). So *fairydom* is one construct in the construction [Ni] +[dom] ↔ [SEM; ‘place where Ni holds sway, collection of Nis, state of being an Ni’].

4. Compounds: Availability of the pattern or availability of a noun?

Endocentric root N+N compounds, i.e. compounds that are hyponyms of their rightmost element (Bauer 2017: 37) with the latter not being derived from a verb (henceforth – N+N compounds), like *house mouse*, *riot police*, *table tennis*, provide an interesting set of questions that challenge some of the grounding claims of productivity theories. The discussion on the issues of productivity of N+N compounds is based on the PhD research of one of the authors (Tarasova 2013), which looks at the peculiarities of the semantics of N+N sequences by analysing the database comprised of 98 compound word families.

It is commonly assumed that compounding in English, especially N+N root compounding is a highly productive morphological process with practically any noun being available for use either as a head or as a modifier in an N+N sequence, e.g. *police car*, *police dog*, *police uniform* vs. *road police*, *riot police*, *peace police*. At the same time, saying that compounding is productive is equivalent to the statement that derivation is productive. Measures like those provided in Hay and Baayen (2002, 2004, 2005) look at a far deeper level of productivity of individual affixes, and we believe that a similar approach should be taken to the measures of productivity in compounding.

Morphological productivity studies assume that there is a relation between the productivity of a word-formation process and the frequency of the outcomes of this process (Bauer 2001; Hay and Baayen 2002; Hay 2003; Fernández-Domínguez, Díaz-Negrillo, Štekauer 2007). Three main frequency

models include the type frequency, token frequency and relative frequency models. In the type frequency model productivity is calculated by means of counting types containing a particular process, i.e. the higher the number of types, the higher the productivity. The token frequency model assumes that there is a correlation between the number of tokens and productivity of the process. The relative frequency model considers the frequency of the base lexeme and the derived word, with a word-formation process being viewed as productive if the lexical bases are more frequent than their derived counterparts (Fernández-Domínguez et al. 2007: 35–37). As shown by Fernández-Domínguez et al. (2007), the problems with all these models is that they provide mere quantifications, which are not always informative in terms of information on the availability of the word-formation process analysed. In other words, the main challenge of productivity studies has been related to finding a way to estimate present productivity rather than past profitability (see, for example, Bauer 2001: 146–157 for discussion).

A different view on productivity is presented in probabilistic models, which assume that a productive word-formation process should be characterised by a large number of low frequency items as compared to high frequency items that employ the same word-formation process (Baayen and Lieber 1991). The notion of hapax legomena, i.e. the item that occurs in a given corpus only once, is central in probabilistic models, since they assume a correlation between the number of hapaxes and the number of novel formations. A large number of new lexemes with low frequency is believed to be indicative of the productivity of a word-formation process. These methods of measuring productivity that are based on the number of hapax legomena in a given corpus, further developed in Baayen (1992) and Baayen and Renouf (1996), have been proposed as solutions for the challenges in productivity studies mentioned above and are aimed at estimating expanding productivity or potential productivity of a process (Baayen 2009). However, it is essential to bear in mind that the number of hapaxes is only an indirect indicator of the rate of expansion of a morphological category, since the number of hapaxes is not a direct reflection of the number of neologisms coined by a given morphological process, as observed in Baayen (2009: 906). Another problem with this approach, especially when applied to compounds, lies in the notion of hapax legomena itself, since when we deal with compounds, the number of hapaxes may be limited, especially if a large-scale corpus (like the BNC) is used. If N2 in a compound is represented by a countable noun, the chances are that it will occur in the corpus both in its singular and plural forms. We assume that *loan risk* and *loan risks* should be lemmatized as a single type.

What is more important for our discussion is that studying the factors that may influence the productivity of individual elements can be much more informative than trying to find a unified measure of productivity of N+N

compounds. In order to account for such factors, the productivity of individual elements that comprise a compound also needs to be considered. In this regard, the first question that we can ask is what features of nouns allow for their being used in the formation of N+N sequences to satisfy immediate demands of communication.

In this section, we focus on the question of how the availability of the N+N compounding pattern coincides with the way speakers apply it to individual nouns, because there seem to be certain limitations on the position that nouns can take in a given compound. The switch of the position of a noun in an N+N sequence results in the change in meaning, e.g. a *garden flower* is a flower that grows in a garden, whereas a *flower garden* is a place where flowers grow. Sometimes the change in the position of nouns in an N+N results in a nonce-formation (or at least we cannot immediately think of a denotatum for such formations), e.g. *crocodile skin* vs. **skin crocodile*, *poker trophy* vs. **trophy poker*, *friend zone* vs. **zone friend*.

Observations of the behaviour of individual nouns when they are used as constituents of N+N compounds demonstrate that the number of types of compounds using one and the same noun is not always equally distributed between the head and the modifier, and that some nouns are more likely to occur as modifiers, whereas others demonstrate the preference for being modified. Baayen (2010) claims that in the majority of cases, the constituents of lexicalised compounds are position-bound, that is they demonstrate a strong preference to being used either as a head or as a modifier (but not both). As shown in Tarasova's (2013) corpus-based study of compound word families, this claim holds true for semantically transparent, non-lexicalised items too. For example, such nouns as *police*, *tourist*, and *water* have a much higher type frequency when used as modifiers in an N+N sequence, whereas nouns like *problem*, *success*, and *story* are mainly used as heads.

It is important to note though that in case of non-lexicalised items, position-boundedness should be viewed as a scalar phenomenon, since different nouns demonstrate a different degree of preference for being used in one position over the other. These observations raise the issue of the potential distinction between the productivity of compounding as a generalised construction and the productivity of compounding using particular constituent elements. Once we view the productivity of compounding in terms of individual constituents, identifying the factors that determine an element's preference for one position or the other becomes a matter of importance.

4.1 Morphological properties

One of the possible factors for predicting the noun's potential for being used as a modifier may be explained by the morphological properties of nouns. In

Tarasova's (2013) dataset, nouns that demonstrate a strong preference for the modifier position, like *chocolate*, *lemon*, *future*, *animal*, etc., are grammatically ambiguous, and it is often difficult to say whether we are dealing with a noun or an adjective, since they can be categorised (and listed in dictionaries) as both. However, there does not seem to be any ambiguity in terms of such words' grammar when they occur in other constructions (as shown in (1) below).

- (1) You can't buy happiness but you can buy *chocolate*.
 The juice of one *lemon*.
 Think about your *future*.
 A domestic *animal*.

Because these elements tend to arise in modifier position more often than in head position in N+N compounds, they have some of the features of attributive adjectives. Analyses of such words (e.g. that in Bell 2012) have raised questions about the criteria that might be used to distinguish nouns from adjectives in such a position but have been largely inconclusive. Therefore, we believe that the ability of such words to be used for modification of another noun (a function which is generally associated with adjectives) is not sufficient for categorising them as adjectives. In our view, words like these should be analysed as nouns since they pass the typical tests for nounhood in English, i.e. allowing a plural form (for countable nouns), forming an NP with a range of determiners, being modified by adjectives, allowing substitution by pronouns, etc.

4.2 Constituent families and schemas

Another factor that may promote positional preference of nouns when used in N+N sequences is the extant paradigm of compounds containing a given noun used in one of the positions, sometimes called the constituent family. Psycholinguistic studies provide evidence that the number of types in a constituent family (i.e. noun-word + N or N + noun-word) influences the ease of processing of a compound's meaning (Baayen 2003, Schreuder and Baayen 1997). If this is the case for understanding a compound, we would suggest that this factor could be important for the formation of new compounds too. Our knowledge of the way a noun is used in compounds is expected to be based on our previous experience with this noun as an element of a compound, and this should influence the productivity of compounds containing this noun. According to the Construction Grammar approach, we follow certain abstract schemas in the process of forming a new compound, and

[...] compound words beginning with or ending in the same constituent may form word families that can be characterized in terms of schemas for complex words in which one of the constituents is lexically specified (Booij 2010: 96).

This means that schemas should determine the use of a noun as a head or modifier, at least on the level of structural representation.

Booij (2008) explains that a limited number of lexicalised compounds may serve as prototypes within a paradigm of compounds containing one and the same word. These lexicalised compounds then promote further formation of structurally similar items by means of analogy. As we come across superficially similar constructions that consistently demonstrate correlation between form and meaning, we inevitably make generalisations about the nature of the template used as the basis for such constructions. We then employ these templates to form new items, in the hope that a listener/reader will be able to access the meaning we intend by applying the principles of analogy (Booij 2008).

A good example of this is the word *head* used attributively in the meaning ‘chief, principal’ in English N+N compounds, e.g. *head office*, *head teacher*, *head chef*, etc. The search in different corpora demonstrates that the use of the word *head* in this meaning is largely limited to the modifier position serving as an attribute of the right-hand element of the nominal construction. This allows for a speculation that to express the idea of ‘main N’, one of the available patterns is *head + N* compound, which forms a schema $[[\text{head}]N]_N$ (based on Booij 2010). In this case, we can speak about the productivity of the schema defined by the presence of a particular noun in a particular position rather than just the productivity of compounding as an overall pattern, and the availability of the schema promotes the extension of the paradigm in which a compound is used in either the head or the modifier position.

The fact that the noun *head* still occurs independently but in slightly different meanings (cf. the meaning ‘a person in charge of something’ in *head of the department*) prompts the suggestion that the word’s meaning may have been altered through its use in compounds, which also resulted in the change of its grammatical characteristics and partial loss of lexical content. Following the discussion in (Booij 2010), one might speak of *head* acting as an affixoid or semi-affix rather than a noun when it is used attributively in this meaning in N+N compounds. The implications of this are that we need to consider applying the principles of analysis suitable for this grammatical category (affixoids) and need to speak about the productivity of the element of a compound rather than the pattern.

Such changes in the lexical content of a noun in N+N compounds are not limited to the modifier position. If we look at cases with the head noun *fish*, we come across examples like *starfish* and *shellfish*, whose denotata no longer have anything to do with the main meaning of the word.

The loss of lexical characteristics can also be seen in compounds like *fireman*, *policeman*, *freshman*, *chairman*, *layman*, *Englishman*, *Frenchman*, etc. especially before the second half of the twentieth century, when the word *man* was used to refer not only to adult males but also to human beings in general, regardless of sex. The productivity of formations of this kind and the lack of direct connection to the lexical meaning of the word led to the change in the lexical and grammatical characteristics, where *man* as the head element of such compounds is reported to be a combining form rather than a noun (OED).

The compounds containing the element *man* are also interesting because they demonstrate how extra-linguistic factors may influence productivity of the noun in compound formation. According to the OED, the word *man* was traditionally combined with other words to denote an occupation, origin or role of the referent of the whole sequence. The use of *man* in the head position reflected the dominance of men in the society and *N + man* pattern was productive. With the changes in the role of women, the use of more gender-neutral terms like *firefighter*, *police officer*, *chairperson*, *freshener* became preferable. The use of *man* in compounds is limited to cases in which the reference to a specific male is obvious/required by a situation, which means that even though it is still available for use in compounds, its productivity is declining.

The examples above support the claim about nouns used as constituents of compounds being bound to one position and forming productive schemas that serve as the basis for the formation of new items. It is noteworthy that the degree of productivity of schemas in compound formation varies. Even though we can still talk about the overall tendency for a given noun to be used only in one of the roles (Tarasova 2013), there is no guarantee that the situation will not change with time, as is the case with *N + man* compounds.

4.3 Frequency of prototypes

If we extend Booij's (2008) claim about the limited number of prototypes which promote the formation of similar items, the next question is how to determine a prototype for non-lexicalised compounds. We cannot use etymological dictionaries or corpora, since they do not tell us what contrasting compounds were in current use at the time of the original formation of the compound in which we are interested. Nor can we rely on corpora if we want to look at how long a given compound (and a concept it denotes) has been in use. For example, in the consideration of the compound family that contains the noun *farm*, the first mention of the compound *farm boy* in COHA refers to the 1860s, which does not mean that farm boys did not occur in the language before that time. Since we have no etymological data, we need to look for other reliable sources in order to identify the units that could be

considered prototypes. One obvious piece of evidence is a constituent family size. We assume that the more often speakers are exposed to a given item (or items that demonstrate similar formal properties), the more likely they are to use this item in an analogical formation (i.e., that employs the same schema).

If this is the case, then we should expect that compounds with higher frequency of occurrence need to be considered possible prototypes for the formation of similar sequences. Therefore, the frequency of use of such prototypes could be a strong predictor for the dominance of one constituent paradigm over the other. A good example of this is the compound word family for the noun *service*, in which the number of types in the *N + service* paradigm is four times the size of the *service + N* paradigm. A closer look at the head and modifier paradigms of the noun *service* reveals compounds that may be viewed as good candidates for being prototypes in each of the cases. This means that in both paradigms we have compounds which are so frequent that it may be plausible to view them as prototypes on the basis of this criterion alone. For example, the dominance of *N + service* over *service + N* paradigm can be attributed to the presence of such a well-established and frequently used compound as *health service* (with 27796 hits in the NOW Corpus). The highest frequency compound in the *service + N* paradigm, i.e. *service sector*, has 5861 hits (NOW Corpus; Davies 2013).

The next question to consider is whether the frequency of use of the prototype(s) in the paradigm can be applied to a wide range of compounds. If so, then we should expect all dominant constituent families to behave in a similar way to compounds with the noun *service*; however, this does not always seem to be the case. The analysis of the *N + holiday* and *holiday + N* paradigms (as outlined in Tarasova 2013) demonstrates the selectivity of this criterion. The highest frequency token in the dominant constituent word family of *holiday + N* paradigm is the compound *holiday home* (143 hits in the BNC). This is opposed to the highest frequency compound *bank holiday* (455 hits in the BNC) in the non-dominant constituent family. Following the assumption above, we should expect the preference for the noun *holiday* to be used as a head rather than a modifier, which is not the case. Tarasova (2013) reports that *holiday* as a head only occurs in 19 compound types, as opposed to 52 sequences in which it is used as a modifier. Overall, *N + holiday* compounds are better established, and high frequency items like *bank holiday*, *school holiday*, *family holiday*, *package holiday*, *skiing holiday*, etc., account for the fact that the cumulative frequency for *N + holiday* paradigm in the BNC is comparable with the *holiday + N paradigm* (1367 and 1398 respectively).

Another factor that needs to be taken into account in regard to frequency and potential prototypes is the range of contexts in which a prototypical compound can be found. For example, the compound *language family* would

probably have a very high frequency in a corpus based on linguistic literature (should one exist), but not in general English use (only 4 hits in the BNC and 291 in the NOW Corpus).

Thus, we can see that the frequency of the items that may be considered prototypes promoting the extension of a constituent family cannot be used as an unambiguous criterion that would account for productivity of the noun as an element of a compound. And even though the importance of the influence of high-frequency items on the extension of the paradigm should not be excluded, pinpointing just one case that may have served as a prototype may neither be viewed as a reliable predictor nor as being possible. It is also the fact that compounds with a high frequency of occurrence are quite rare. Several large-scale corpus studies that utilise different corpora, e.g. BNC, WordNet, CELEX (Andrews, Miller and Rayner 2004, Plag 2006, Plag et al. 2007, etc.), demonstrate that compound words generally tend to be very low-frequency (with a considerable number of hapax legomena within a given compound word family). It is also the presence of a large number of low-frequency items that is commonly believed to be a marker for productivity (Baayen and Lieber 1991, Baayen and Renouf, 1996).

This suggests that it is not the high frequency item (or a limited number of such items) that drives the formation of new compounds but rather the abundance of low-frequency N+Ns that matters. Taking this into consideration, we can assume that speakers rely less on their experience with one and the same compound, but rather on the knowledge of how a noun may behave in the paradigm of structurally similar formations – N+Ns in our case – to (re-)create the meaning of low frequency compounds when we encounter them (rather than access it automatically).

4.4 The role of established compounds

Looking at individual paradigms of compounds that contain one and the same word provides some interesting insights regarding the influence of established compounds on the extension of the paradigm. It is important to note though that categorising a compound as an established one or a non-established one is not as straightforward as it may seem. One and the same unit may be established (or ‘item familiar’, in the terminology of Meys 1975) for one speech community but not another. For example, an expression like *origami poker* (the style of playing without taking risks), overheard by one of the authors, will not make much sense to people who do not belong to a particular poker community. Even though the use of dictionaries as reference sources for determining listedness has its limitations, in this study, we used the *OED* as a proxy for identifying whether an N+N compound is linked to an established concept.

We checked the compound data in Tarasova (2013) to see if the presence of established compounds can be considered a possible predictor for the size of the paradigm where a given noun is used either as a head or as a modifier. We looked at compound families in which nouns demonstrate different degrees of preference for being used in one of the positions in an N+N sequence in order to understand if there is a connection between the productivity of a noun as a compound constituent and the number of established compounds in the constituent families. The results of our analysis demonstrate that this criterion also works selectively and cannot be used as a benchmark for predicting a noun's productivity when it is used as an element of an N+N sequence. For example, the modifier paradigm of compounds containing the word *family* (which demonstrates a strong preference for modifier position, preferring modifier position to head by a ratio of 10:1 in Tarasova's data), contains a considerable number of established *family + N* compounds, including *family member*, *family business*, *family car*, *family court*, *family album*, *family life*, etc. (33 out of 72 *family + N* compounds in Tarasova's (2013) data are listed in the *OED*); and only one in *N + family* paradigm (*language family*) is attested in the corpus. A similar trend holds true for some other nouns that demonstrate a preference for modifier position, e.g. *health*, *art*, *air*, but not others. For example, in the analysed dataset, the modifier paradigm of the noun *court*, which is twice as likely to be used as a modifier than as a head, only contains two compounds that are listed in the *OED* (*court case* and *court costs*). At the same time the (non-dominant) head constituent family for the same noun contains seven compounds that are listed in the *OED* (*police court*, *law court*, *county court*, *district court*, *centre court*, *youth court*, *family court*).

Referring back to the issue of categorising a compound as being established by using a dictionary as a reference point, the obvious question that comes into mind is whether the listedness of an item really provides an objective measure for that. For example, the *OED* lists a compound *health fascism*, which to some speakers might seem less established than, say, *health risk* or *health care*, which are not listed.

4.5 Summary

With such lack of consistency in compound formation, we believe that analogy could be considered one of the influential factors for an element's positional preference. However, as the corpus data imply, it is difficult to pinpoint a single factor as determining whether a given noun will be preferred as part of a head paradigm or of a modifier paradigm. A better explanation of why some noun concepts are head-oriented and others are modifier-oriented is probably grounded in the size of the constituent family. In other words, the

larger the constituent family is, the more likely it is to increase further. It is hard to explain, however, what makes one constituent family large in the first place and we can only make suggestions rather than offer relevant solutions at this stage.

One of such suggestions can be offered by research in applied linguistics and second language acquisition. According to Yang (2016), if the learner is exposed to a particular pattern sufficiently but not exposed to the contrary pattern, then the dominant pattern is taken to be productive and the minority pattern is taken to be exceptional. Changes in exposure can change the conclusion. If we apply this idea to our discussion, then we can say that the speaker's experience of how a noun is used in an N+N compound may be considered responsible for the extension of the constituent family paradigm, which becomes a dominant pattern for this noun. The amount of exposure can be influenced by any or all of the following: high frequency of use of a limited number of units, the number of such units, the use of units in a wide range of contexts and genres, the importance of the concept denoted by a language unit for the life of the language community, etc. It may also be the case that the more factors are involved for a given noun, the more likely it is that its productivity as a compound constituent will be higher. The absence of one or more of such factors could explain the differences in the degree of productivity.

While there is no question that compounding is a productive process, just what evidence should be used to support this conclusion is not necessarily clear. Multiple factors that might contribute to the productivity of compounding have been considered here, but we do not currently have a model which can take proper account of the relevant ones. Little attention has been paid to the availability of the individual elements involved in compounding, for instance, and many of the factors discussed above are difficult to quantify.

One of the fundamental problems that faces every study of productivity – whether corpus- or dictionary-based – is that we are forced to make the assumption that the data available to the linguist is in some way representative of the data available to speakers. This is almost certainly wrong, and we do not know how it is wrong. In the absence of better data, we have made the assumptions that all scholars working in this field are forced to make, but we acknowledge the potential problems associated with this.

5. From splinters to affixes: A study of the productivity of novel combining forms

In the light of a widely accepted assumption that 'productive mechanisms are best investigated in rare formations rather than in frequent formations' (Arndt-Lappe and Plag 2013: 539; see also Baayen 1992; Plag 1999 for discussion),

this section will focus on lexical blending as a source of potentially productive constructions. Blending is ‘a type of word formation in which two or more words are merged into one so that the blended constituents are either clipped, or partially overlap’ (Beliaeva to appear), e.g. *tigon* (*tiger* + *lion*) or *glassphalt* (*glass* + *asphalt*). Although blends are much less common than derivations, they are notably widespread in contemporary language, especially in particular domains such as hybrid names, media and youth slang (Beliaeva to appear; López Rúa 2004; Renner 2015a). Although some scholars have discussed blends as ‘extragrammatical’ creative coinages (Dressler 2000; Mattiello 2013), there is extensive evidence in literature (Arndt-Lappe and Plag 2013; Beliaeva 2014, 2016; Gries 2006, 2012; Lehrer 2007 and other studies) that the formation of blends can be described in terms of predictable constraints on their phonological, semantic and structural properties. Moreover, as observed in Lehrer (1998, 2007), Mattiello (2017) and other literature, the existence of one blend can give rise to the formation of further blends containing the same *splinter*, that is, part of a base word which is retained in a blend. For example, the initial splinter *fem-* can be found in such blends as *femstronaut* (*female* + *astronaut*), *feminister* (*female* + *minister*) and *femocracy* (*female* + *democracy*); the final splinter *-cation* can be found in *staycation* (*stay* + *vacation*), *workcation* (*work* + *vacation*), and other blends. In line with the argument developed in Section 4.5, this part of the discussion considers factors influencing the productivity of individual splinters that are used recurrently. Furthermore, the analysis here, involving different quantitative measures of productivity, is used to substantiate the claim that such qualitative investigation is important.

5.1 Splinters in blends and beyond

The term ‘splinter’ is widely used in literature on blends to refer to a portion of a word that has become part of a blend (Adams 1973; Bertinetto 2001; Fandrych 2008; Gries 2012; López Rúa 2004 and other studies). A splinter may be part of a morph (e.g. *fem-* is part of a root) or may include more than one morph (e.g. *-cation* in *staycation* includes a suffix plus a portion of the root), but is not equal to an established morph. In many studies, e.g. in Bauer (2006) and Bauer et al. (2013), this term is used to refer to a portion of a word which has been attached recurrently to new bases, as is the case for *fem-* and *-cation*. In contrast, Lehrer (1998) claims that once a splinter has demonstrated some productivity, it should be classified as ‘combining form’. In other literature, e.g. in Adams (1973), the term ‘splinter’ is used to name any portion of a base word that has become part of a blend or blends, regardless of its potential to be used recurrently. The distinction between a splinter that has been used recurrently and one that appears in one blend only is somewhat arbitrary, given that new words (including blends) appear

every day and may not be registered in dictionaries or even large corpora (see also Beliaeva 2014 for a discussion). This section focusses on splinters that have been found in more than one new word, but the observations and conclusions herein can be extrapolated to splinters that have not (to date) been used recurrently.

The distinction between a splinter and an established morph requires some clarification. The blend *staycation* originates from two source words: *stay* and *vacation*. The first source word preserves its full form in the blend, the second is only partially preserved, i.e. its final splinter *-cation* is cut off to be retained in the blend. In some cases, full words overlap to form blends such as *stoption* (*stop* + *option*) or *predictionary* (*prediction* + *dictionary*) so that no part of any of the source words is cut off. In such blends, it is not possible to define the borders of splinters. There are also cases when one of the source words is a compound or another morphologically complex word, as in the blend *cookprint* (*cook* + *footprint*). The portion of the second source word that is retained in the blend (*-print*) is a free morph and therefore cannot be classified as a splinter. Such cases are outside the scope of the present study (though it is clear that, like any other blends, such blends have the potential to give rise to productive constructions). It is also worth noting that some splinters that have been used to form a high number of blends are now listed in dictionaries as affixes and therefore can no longer be classified as splinters. For example, *-(a)holic* which is listed in dictionaries as a suffix meaning ‘a person who appears to be addicted’ to a thing or activity (*OED*) was originally a splinter in the blend *workaholic* [1947] ‘a person addicted to working’ (*OED*) and was then used productively in such formations as *newsaholic*, *spendaholic*, *computerholic* and others (all examples from the *OED*). In addition to relatively high type frequency, there are other factors that may be used to distinguish affixes (including novel affixes such as *-(a)holic*) from splinters. These factors will be discussed in Sections 5.3–5.4, after some preliminary remarks on the data investigated for this study. For the purposes of the present investigation the term ‘splinter’ will be used to refer to any parts of words which are used to form blends, provided that they are not listed in dictionaries as morphs (including cases when established morph is a part of a splinter, as in *-cation*).

5.2 The productivity of splinters: Rationale and data

As pointed out in Section 1, productivity has been discussed in terms of availability and profitability. One of the challenges of productivity studies to date has been related to finding a way to estimate present productivity rather than past profitability (see, for example, Bauer 2001: 146–157 for discussion). Quantitative estimates of the productivity using hapax legomena have their

shortcomings, primarily the fact that they do not directly account for the availability of the morphological processes, as discussed in Section 4. The approach suggested in this section is to focus on admittedly recent formations (which has to be controlled for at the stage of data selection). Thus, instead of using the number of hapax legomena as approximation to the number of neologisms, we propose to investigate coinages which have not been established in the available media sources for a long time and which, therefore, would all be likely neologisms. The number of hapax legomena in this case would be a closer approximation to expanding productivity in the sense of Baayen (2009) than the number of hapax legomena in a data set collected without prior selection considering the date of occurring in sources.

The data for this study were taken from a corpus of novel English blends that were attested in media and / or corpora no earlier than January 1, 2000 (a date selected arbitrarily to control for relative novelty of formations). The corpus was collected in 2012–2014 and discussed in Beliaeva (2014, 2016). Splinters that appeared in more than one novel blend (in particular, 9 initial and 9 final splinters) were selected for the present investigation. Further examples of novel formations containing these splinters were collected from NOW (Davies 2013), which is composed of over 3 billion words from web-based media from 2010 to the present, and is the largest corpus of English to date. All word tokens beginning or ending with a particular letter string (e.g. beginning in *loca-* or ending in *-flation*) were extracted from NOW. Derivatives of the original source words (e.g. *localization* and *locate*), as well as proper names (e.g. *Locarno*) were excluded from the dataset. Furthermore, the same letter strings were searched in COCA, but the items that were extracted from NOW did not show in the COCA results, which is an additional indication that they are neological formations. The final dataset included 2931 types, as exemplified in Tables 2 and 3.

Tab. 2: *Items containing initial splinters.*

Splinter	Source word	Examples	Total number of types in NOW
loca-	local	localife, locanomics	28
edu-	education	eduloan, edusoft, educise	266
robo-	robot	robocraft, robo-signing, robocide	1180
virt-	virtual	virtnet, virtuonica, virt-screen	42
digi-	digital	digibank, digitour, digipedia	605
fabu-	fabulous	fabulips, fabutan, faburrito	21
adver-	advertise(ment)	advergame, advertorial	15
fem-	female	femplay, fem-care, femocracy	179
alterna-	alternative	alterna-pop, alterna-teen	68

Tab. 3: *Items containing final splinters.*

Splinter	Source word	Examples	Total number of types in NOW
-burb	suburb	ethnoburb, exburb, joburb	20
-noia	paranoia	powernoia, ebolanoia	18
-cation	vacation	brocation, runcation, petcation	71
-stalgia	nostalgia	motostalgia, snowstalgia	23
-(o)rexia	anorexia	drunkorexia, tanorexia	22
-flation	inflation	trumpflation, foodflation	38
-zilla	godzilla	barzilla, boss-zilla, mumzilla	202
-(i)sode	episode	minisode, operasode, clip-isode	14
-gasm	orgasm	eargasm, scoregasm, statsgasm	105

The fact that each of the splinters was used in a number of formations attested in NOW can be regarded as confirming that these units are available for word formation in current English. As shown in Tables 2 and 3, there are notable differences between the splinters in terms of type frequency. This may indicate that these splinters differ in profitability, but this gives no information about possible reasons for it. Some factors that may be relevant for analyzing the productivity of splinters will be considered in Sections 5.3–5.5.

5.3 Morphological properties and profitability

The data set is not uniform in terms of morphological structure. In particular, there are cases when splinters attach to free morphs (2a), to combining forms (2b), to other splinters while forming blends (2c), or to other formatives such as clippings or acronyms (2d). Depending on the formatives that the splinters are attached to, and also on whether or not there is overlap involved, neological formations in our data set have been classified into groups, as exemplified in (2).

- (2) a Affixations:
 – edu-business, edukids
 – digikiller, digigirl
 – food-flation, slowflation
 – speedzilla, gigzilla
- b Combining forms:
 – locanomics, loca-paedia
 – robocide, robonomics
 – alternaverse
 – robogasm, megagasm

- c Blends (merging two items together so that at least one is shortened, often involving overlap):
 - edupreneur: edu- + (entre)preneur
 - localytics: loca- + (ana)lytics
 - cluburb: club + -burb
 - robocalypse: robo- + (ap)ocalypse
- d Complex clippings (attaching to a clipping or acronym, with no overlap):
 - robo-tech: robo- + tech(nology)
 - digiTV: digi- + TV
 - statsgasm: stat(istic)s+ -gasm

In (2a), the splinters are attached to free morphs, e.g. the splinter *edu-* is attached to *business* as a prefix, and the splinter *-flation* is attached to *food* as a suffix. In such formations splinters can be compared to affixes in the sense that they have a fixed phonological and graphical representation and a stable meaning, as discussed, for example, in Fradin (2000). Like affixes, splinters can attach to morphologically complex words, e.g. in *robolawnmowers*, the splinter *robo-* is attached to a compound. Occasionally, splinters can attach to a phrasal unit, as in *drunkbridezilla*.

In (2b), the splinters are attached to items that are not usually used as free morphs. This includes neoclassical combining forms discussed, for example, in Bauer (1998) and Plag (2003), and also bound morphs that originated as splinters and are classified as combining forms in Lehrer (1998) and Fradin (2000). This category is fuzzy, as discussed in Section 5.1, and exact classification is outside the scope of this study. The examples in (2b) differ from those in (2a) in that splinters attach to bound rather than free morphs, which includes cases when two splinters are concatenated, as in *robogasm*.

In (2c), splinters are attached to shortened versions of words which aren't established morphs, and this may also involve some overlap. In other words, the units in (2c) can be classified as blends, according to the definition provided in this section. Note that the formations in (2c) match the criteria of blends even though splinters, rather than the corresponding words, are accepted as source forms in each case. Thus, *edupreneur* can be regarded as a blend of *edu-* and *entrepreneur* in which part of the source word *entrepreneur* is cut off. It is possible to analyse this formation as a blend of *education* and *entrepreneur*, formed by analogy to numerous other blends containing the splinter *edu-* (see Mattiello 2017 for a detailed discussion of recurrently used splinters in analogical formations). Such analysis, however, would not change the status of formations like *edupreneur* in the morphological classification that is developed here. Note also that, like many blends, examples in (2c) retain the prosodic contour of one of their constituents, e.g. the word

localytics has the same number of syllables and preserves the main stress of its source word *analytics*.

Finally, in (2d) splinters are attached to clippings or acronyms, e.g. the initial splinter *robo-* and the final splinter *-gasm* are attached to institutionalised clippings (*tech* from *technology* and *stats* from *statistics*, respectively), and the splinter *digi-* is attached to the acronym *TV*. The formations in (2d) are similar to the examples in (2a) as the splinters in (2d) also attach to free morphs, albeit shortenings rather than full words.

The overall distribution of morphological types in the data set is summarised in Table 4.

Tab. 4: Morphological classification of items containing splinters. The most frequent morphological types are highlighted in shades of grey (light shade of grey denotes 50% types or more, medium shade – 75% or more, dark shade – 90% or more).

Splinter	Affixation types	Blend types	Combining form types	Complex clipping types	Tokens	Hapax legomena
loca-	14	7	7	0	1038	9
edu-	210	25	14	17	4539	32
robo-	1101	30	33	16	12894	529
virt-	27	4	5	6	261	6
digi-	493	27	36	49	5995	112
fabu-	17	2	2	0	64	11
adver-	11	0	3	1	2283	2
fem-	84	16	54	25	2736	47
alterna-	65	1	1	1	107	39
-burb	15	0	4	1	39	10
-noia	10	0	7	1	208	7
-cation	71	0	0	0	1511	30
-stalgalia	16	1	6	0	74	8
-(o)rexia	12	2	7	1	640	7
-flation	32	0	3	3	2588	18
-zilla	194	0	7	15	1070	69
-(i)sode	13	0	0	1	223	6
-gasm	91	4	7	3	414	40
Total	2495	196	119	142	45155	999

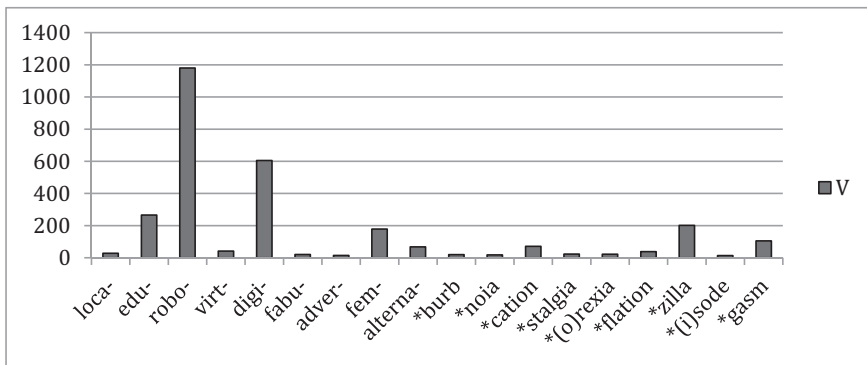
As shown in Table 4, in the abundant majority of types in our data set splinters are attached to free morphs in an affix-like manner: 2495 out of 2931 types, which comprises 85.1% of the data set, are classified as affixations. This type of concatenation is significantly more frequent in the data than all the other

types as confirmed by a contingency table analysis of the number of types in the corpus over the type of word formation ($\chi^2 = 468.14$, $df = 57$, $p < 0.001$, Contingency Coeff. = 0.406, Cramer's $V = 0.256$). It is also noteworthy that the splinters with the highest type frequency (*robo-*, *digi-*, *edu-* and *-zilla*) also demonstrate a strong tendency to attach to free morphs (over 75% of types containing these splinters).

Despite the fact that all the splinters, by definition, originated from blends, the observed pattern of repetitive use in an affix-like manner is associated with greater profitability. In this respect, the initial splinter *fem-* stands out as it is used in blends more than other splinters (1/3 of all types containing *fem-* are blends, and less than half are affixations). This may be explained by the semantics of the splinter that is primarily used in analogical formations with a meaning 'female variant of something', e.g. *femocracy* (*female democracy*), *femoir* (*female memoir*). As pointed out in Mattiello (2017), analogical formations often have playful nature and are used as attention-catching means. Similar functions are observed in blends (see Renner 2015b for a discussion), and therefore the tendency of particular splinter to be used in blends may be explained by its predominant use in jocular formations.

In regard to the productivity of the splinters, each of them is attested in a number of (a minimum of 14 and a maximum of 1180) different types of neological formations, as shown in Figure 1.

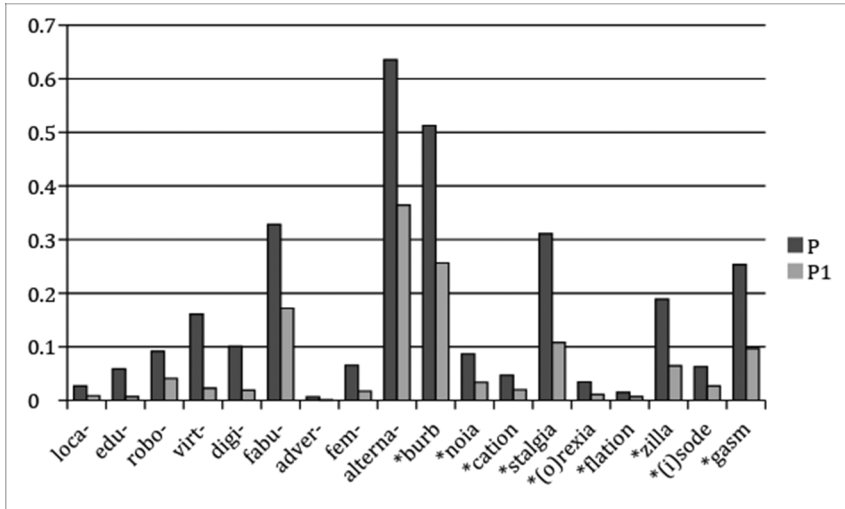
Fig. 1: Type frequency V of different splinters in NOW.



The splinters *robo-* and *digi-* appear to have greater realised productivity, as can be estimated with regard to higher type frequency. This analysis was complemented by using measures that are believed to reflect present, or potential, productivity. Two of such measures were used for the present data set: the type-token ratio V/N (Aronoff 1976) which is indicative of the lexical diversity of a morphological process, and the ratio of the number of hapax legomena n_1 to the token frequency N (Baayen 1992), as an estimate

of potential productivity (see also Baayen 2009 for the discussion of the method). The results are displayed in Figure 2.

Fig. 2: Estimated productivity of the splinters: $P=V/N$; $P_1=n_i/N$.



As can be seen from Figures 1 and 2, different productivity measures induce dissimilar results for the given data set. The splinters that have higher type-token ratio are also characterised by higher P_1 , however, though, the splinters that appear to have higher productivity according to Figure 2 (*alterna-*, *-burb*, *fabu-* and *-stalgia*), are characterised by relatively low type and token frequency. On the contrary, the splinters *robo-*, *digi-*, *edu-* and *-zilla* which have the highest type frequencies, as per Table 4 and Figure 1, do not appear prominent in regard to productivity measures P and P_1 . A possible explanation of this apparent contradiction is that profitability measures may produce false predictions for marginally productive processes, or that the profitability measures are overly sensitive to the corpus size, as pointed out in Bauer (2001: 148). It is also plausible that different measures are indicative of different aspects of productivity which is, as noted, for example, in Plag (2003) and Baayen (2009), a multifaceted phenomenon. It is potentially beneficial to study the relationships between various measures of productivity, as is proposed in Baayen (2009), but such models will be of little value without thorough understanding of what the measures actually represent. For example, it can be inferred from Figures 1 and 2 that splinters *alterna-* and *-burb* have lower realized productivity, compared to *robo-* and *digi-*, while at the same time demonstrating greater potential productivity. However, it is not entirely clear what this result implies, e.g. whether it can be predicted that splinters like *alterna-* and *-burb*

can become productive in future, or whether they have been used for a number of times to coin occasionalisms but have failed to gain wider use in language for some reasons. What is clear from the observed data is that, in addition to quantitative measures, a qualitative analysis of specific examples is required to understand what factors may contribute to the productivity of the splinters. Some of these factors will be considered in Sections 5.4–5.5.

5.4 Phonological and prosodic factors affecting profitability

As observed in Section 5.3, some neological formations containing splinters retain the prosodic contour of one of their source words. Predominantly, the prosodic contour of the second source word is preserved, as in *localytics* and *robocalypse*. This refers not only to items that are classified as blends, but also to some of the combining forms such as *locanomics* and *loca-paedia*. Moreover, some of the words in which a splinter attaches to a free morph like an affix, retain the prosodic contour of the word the splinter originated from, as illustrated by examples in (3).

- (3) a staycation, daycation, mancation
 b foodflation, joyflation, trumpflation
 c chemonoia, powernoia
 d fabutan, fabulips

In each of the examples in (3a), the splinter *-cation* is attached to a monosyllabic word, and as a result, the prosodic contour of the original source words *vacation* is retained. In a similar way, the splinter *-flation* tends to attach to monosyllables, so that the words in (3b) retain the prosodic contour of *inflation*. The splinter *-noia* in (3c) is attached to bisyllabic words, thus retaining the prosodic contour of *paranoia* in the resultant formations. This tendency conforms to what has been observed in literature on blends which tend to retain the prosodic contour of at least one of their source words to enhance their recognisability (Arndt-Lappe and Plag 2013, Beliaeva 2014, Gries 2012). The studies on blends show that, predominantly, the prosodic contour of the second source word is preserved. This is also the case for examples in (3a–c) which preserve the prosodic contour of the source word of their final splinter. On the other hand, words in (3d) preserve the prosodic contour of *fabulous* which is the source word of the initial splinter *fabu-*. The tendency to preserve a particular prosodic contour in case of recurrently used splinters may bring forth specific constraints on the words the splinters can attach to – that is, monosyllables in (3a–b, d) or disyllables in (3c).

As shown in (3), the original source of the splinter can be made more salient by the preservation of its prosodic contour, that is, main stress and syllable number. However, counterexamples to this can be found, e.g. *solocation*,

Colorado-cation, *ebolanoia*, etc. It is worth noting that the splinters that have comparatively high type frequency (initial splinters *digi-* and *edu-*, as well as final splinters *-zilla* and *-gasm*, see Table 4) also demonstrate flexibility in terms of prosodic contour, as exemplified in (4).

- (4) a digicable, digimedia, digipower, digisphere, digilockers, digi-discipline
 b educhild, edu-establishment, edufun
 c bridesmaidzilla, gossipzilla
 d audiogasm, careergasm, moviegasm

The splinter *digi-* in (4a) originates from the word *digital*, and in order to preserve the prosodic contour of its source words it would be selectively attached to monosyllables only. This is not what is observed, however. Among the words that the splinter can attach to, monosyllables are clearly in the minority (e.g. in *digiart* and *digi-age*). Similarly, despite the fact that such words as *edubusiness* and *educentre* preserve the prosodic contour of *education*, the source word of the splinter *edu-*, there are numerous cases where *edu-* attaches to words of various syllabic lengths, such as the ones in (4b). Final splinters *-zilla* and *-gasm* can also attach to words of various lengths, as shown in (4c–d). However, it is worth noting that these two splinters tend to attach to monosyllables and therefore preserve the prosodic contour of their source words (*Godzilla* and *orgasm*, respectively) more than the initial splinters in (4a–b). This is in line with findings of the research on blends (e.g. Arndt-Lappe and Plag 2013) and confirms that the preservation of the prosodic contour is more important for recognition of the source word if the word beginning is not retained. Moreover, some neological formations that are formed by blending an initial splinter and another word (e.g. *femvertising*, *digipendent*, *robocopter*), the prosodic contour of the second source word (*advertising*, *independent* and *helicopter*, respectively) is retained, which conforms to the overall tendency in blends.

The data in this section demonstrate that coining new words using splinters from institutionalised blends is, in many cases, constrained by such factors as the prosodic contour of the source words of the splinters. If splinters demonstrate flexibility in terms of prosodic contour, this may signify that they have become productive enough to start losing connection to their source words. This, in its turn, further supports the idea that the profitability of a morphological process is related to the number of constraints on it, and also shows that specific methods of estimating productivity have to be developed for different morphological processes. Firstly, different factors are at work for initial and final splinters, in parallel to different factors contributing to the productivity of prefixes and suffixes. Moreover, although our data show a tendency for splinters to form new words in an affix-like way, the observed

differences between splinters and affixes indicate that the productivity of splinters should be analysed using specific methods, different from those estimating the productivity of affixes. Finally, the productivity of individual splinters may be subject to specific constraints. Thus, some factors may reinforce the need to maintain the association with the model for analogy, as in the case of words containing the splinter *fem-* which bears the meaning ‘a female version of something’. Semantic considerations that may influence productivity will be addressed in more detail in Section 5.5.

5.5 Semantic factors affecting productivity

As observed in Lehrer (1998), splinters that demonstrate a certain degree of productivity often undergo semantic resegmentation, that is, specific components of the meaning of the word a splinter is secreted from, becomes associated with the splinter in further use. This is the case for some of the splinters in the present study. For example, the formations containing *-zilla* (5a), bear the meaning ‘someone or something giant or monstrous’, thus secreting a certain component of the meaning of the source word *Godzilla*. Likewise, each of the words containing the splinter *-gasm* in (5b) denotes a pleasant or ecstatic feeling of emotional, rather than physical, nature.

- (5) a *-zilla*: gossipzilla, avozilla
 b *-gasm*: bookgasm, musicgasm, statsgasm

Another factor that may influence the productive use of splinters is semantic analogy. That is, by analogy with an existing formation in which a splinter attaches to a word, the same splinter can be attached to other words of the same semantic domain. Thus, *groomzilla*, *wedzilla* and *bridesmaidzilla* may have been formed by analogy with *bridezilla*. Similarly, semantic analogy can underlie the formation *eyegasm*, *eargasm* and *foodgasm*. It is worth noting that recurrent use of splinters to form new words is in itself driven by analogy – this type of word formation is classified in Mattiello (2017) as ‘analogy via schema’. However, examples such as *bridesmaidzilla* and *wedzilla* demonstrate that analogy may also influence the selection of elements a splinter can attach to.

As a consequence of the analogical mechanism of formation, words containing a particular splinter may belong to specific semantic fields. The meaning of the splinter may also determine a particular semantic field or contexts in which such words are used. For example, terms in (6a) belong to the domain of media and the ones in (6b) are related to technological innovations.

- (6) a *-(i)sode*: webisode, appisode, twittersode, videoblogisode
 b *virt(u)-*: virtunet, virt-screens, virtusoft, virtuwatt

Specific contexts of use may be a restriction on productivity. For example, the productivity of the splinter *-isode* denoting a type of video episode is restricted by the number of types of broadcast media. On the other hand, this doesn't seem to be the case for the splinter *digi-* which demonstrates relatively high productivity despite its confinement to the domain of informational technology.

5.6 Summary

The examples in this section show that existing quantitative profitability measures are not sufficient for estimating the productivity of word formation units such as splinters. As discussed in Lehrer (2007) and Mattiello (2017), once a splinter is used in a blend, it may appear in more blends by analogy, gain productivity over time and eventually acquire a status of an affix. The data in this section illustrate how this may happen. For a splinter to become more productive and thus to be used in a greater number of novel formations, certain phonotactic, prosodic and semantic requirements have to be satisfied. These requirements concern, for example, the number of syllables in a word the splinter may attach to, or the semantic domain the new formation is to be used in. Increased productivity may also be associated with semantic resegmentation. Thus, the analysis in this section aimed at outlining phonological, prosodic and semantic considerations that have to be taken into account when analysing the productivity of splinters and their potential to become attested as affixes. These considerations may not be exhaustive, and further qualitative and quantitative research is needed to develop models predicting the productivity of splinters. On the other hand, analysing the factors that trigger the productivity of splinters and cause their attestation as affixes is beneficial for deeper understanding of the productivity of affixes, and for the development of a more general theory of morphological productivity.

6. 'Phrasal verbs'

In this section, we consider verb + particle combinations in English. The terminology in this area is confused, as is the number of sub-types that need to be distinguished (Bolinger 1971; Quirk et al. 1985; Huddleston 2002). Although we recognise it as a problem, we use the label 'phrasal verb' here as a cover term for any such verbs, independent of what tests they meet, since our claim will be that multiple types may simultaneously affect outcomes, since our claim will be that multiple types may simultaneously affect outcomes.

Phrasal verbs are generally treated as syntactic constructions in English. This contrasts with most of the other Germanic languages, where corresponding verbs are usually treated as lexical. We suspect that this has to do with the infinitive form. German *aussterben* corresponds directly to English

die out (it is an element-by-element translation and the same meaning), but *aussterben* appears to be a single orthographic word (and so a morphological construction) while *die out* is written as two words, and thus looks like a syntactic construction. Whether or not this is the real reason for the distinction in treatment, it illustrates the point that such verbs have something of a dual nature, part syntactic, part lexical: they are often lexicalised or idiomatised, they act as bases for further word-formation, yet they allow free interruption by other elements in certain circumstances. We take the view here that even if phrasal verbs are purely syntactic (an extreme view, which we would not espouse), they provide a suitable input to a discussion on productivity, since which combinations are used depends on factors other than the simple availability of the overall patterns involved.

As a further preliminary comment, we should overtly recognise that many – perhaps most – phrasal verbs are polysemous (in many instances, homonymous): you can work out a maths puzzle or work out in a gym; you can make out in the back of a car or make out a shape in the mist. When we use examples of phrasal verbs to illustrate particular meanings or usages, therefore, we do not necessarily imply that all usages of that phrasal verb fit the same pattern, only that at least one does. Many of the examples of phrasal verbs we present, and much of the data on usage with established examples, comes from Courtney (1983).

In this paper, we will be concerned only with verbs using the particle *out*, and we will be mainly concerned with two distinct patterns with *out*, which will be introduced later. First, though, we will consider verbs with *out* in more general terms. In this part of the discussion, qualitative rather than quantitative analysis is used to demonstrate the factors involved in the availability of patterns.

6.1 Phrasal verbs with the particle *out* in general

There are many phrasal verbs with *out* (see (7) for some examples). An earlier corpus study by Beliaeva and Tarasova (2012) based on the BNC found nearly 350 such verbs in some 90,000 tokens. This must be an underestimate, given that it is clear that some of these phrasal verbs must count as homonyms and also that (as claimed below) some patterns with *out* are productive. Given this *embarras de richesses*, it would be open to the linguist simply to say that any new verb is an extension of the general pattern with verb + *out*. Not only does this seem, to us, to involve rather too much hand-waving and too little analysis, it seems a dereliction of duty not to consider what subpatterns there may be and how they may affect new formations. A list such as that in (7) is not, of itself, very enlightening.

- (7) act out, back out, bail out, black out, block out, blot out, bottle out, break out, carry out, cash out, check out, chill out, clock out, come out, cop out, count out, contract out, cross out, deal out, die out, dine out, dish out, dole out, drop out, edge out, eke out, face out, find out, fit out, fork out, get out, hold out, iron out, lay out, let out, make out, muck out, nose out, opt out, point out, pull out, root out, rule out, seek out, send out, single out, spell out, spread out, strike out, take out, time out, try out, whip out, work out, yell out

Our first task, therefore, is to show that a list such as that in (7) can, at least to some extent, be systematised, and that there are recurrent patterns, both formal and semantic, within this morass of data. We do not attempt to be exhaustive here, only to illustrate some patterns to which we will refer later. The patterns we illustrate are not ordered in terms of priority, but we deal with formal patterns before semantic patterns.

The first pattern we choose to comment on is one where the non-particle element (let us call it the ‘head’) is clearly verbal and **active**. This is the major, expected pattern. Some examples are given in (8).

- (8) Active: ask out, bear out, broaden out, carry out, come out, die out, pull out, take out

The second pattern is one where the head is, or may be, **denominal**. The hedge here is that because of the amount of conversion in English, it is not always clear whether something is denominal or merely correlates with a noun of the same form. We do not draw a distinction here between established instances of conversion, and instances of conversion which seem to arise purely in the phrasal verb. Examples are given in (9).

- (9) Denominal: average out, bug out, clock out, comb out, cop out, crap out, freak out, hammer out, hose out, map out, parcel out, puzzle out, ration out, relay out, sack out, space out, veg out

Equivalently, there is a set of phrasal verbs where the head is **deadjectival**. Examples are given in (10).

- (10) Deadjectival: black out, brazen out, dim out, dry out, even out, grey out, open out, plump out, red out, single out, tough out

Next there is a set whose members are predominately used in the **passive** (according to Courtney 1983). Some examples are given in (11).

- (11) Passive: flood out, kit out, knock out (of a competition), pad out (a manuscript), salt out, scorch out

When it comes to semantic categories, we start with the default reading of *out*, which involves movement from within an enclosed space to beyond that space. We will term this the ‘**from within**’ reading, and we illustrate it in (12).

- (12) From within: back out, bring out, dig out, force out, let out, march out, scrape out, smuggle out, throw out

Almost all other readings are figurative in some way. The figure is not always the same in all cases, and there are innumerable different semantic readings of the figurative extensions. We do not attempt exhaustiveness in this section, but draw attention to some of these readings to which we will want to make reference later. In some instances, it is hard to determine the borders of the categories. Consider, for example, the set which we will call ‘**thoroughness**’, illustrated in (13). The phrasal verbs here indicate that the action is carried out in a thorough manner. It is not clear to what extent this set is distinct from the set which we will call the ‘**extinction**’ set, where the action is carried out in such a way that the affected object no longer exists in the same form. Examples of this set are given in (14).

- (13) Thoroughness: clean out, clear out, muck out, sweep out

- (14) Extinction: burn out, buy out, count out, die out, drive out, drown out, edge out, scratch out, sell out

Another class is the one which indicates **lack of awareness** or mental control. This is illustrated in (15).

- (15) Lack of awareness: black out, block out, knock out, pass out

There are, of course, many other patterns of phrasal verbs with *out*, but we ignore them here, not because they are unimportant, but because they are not directly relevant for our purposes.

We now turn to the two specific patterns that we wish to focus on in this paper. It will be shown that, although they are distinct patterns of phrasal verb with *out*, they share features with earlier patterns that have already been discussed.

6.2 Construction 1: Noun-*ed out*

Examples of this construction are given in (16)–(20) below. The head is always a noun with an *-ed* suffix. The pattern is extremely productive in current usage: individual examples are rarely listed, but the pattern recurs. The use of *all* before the head is common, but not necessary. We have not managed to find a starting date for this construction, but have been aware of it since the early 1990s. Although the BNC has examples which seem to fit the *all *ed out* frame, none of them seem to be this particular construction. COCA presents some, listed as (21)–(26).

- (16) You get all cathedraled out and museumed out (Overheard, 2010).
- (17) By the time Cap returned to the hotel at nine, Lou was conferenced out. (Palmer, Michael 2014. *Resistant*. Farmington Hills, Michigan: Wheeler, p. 83)
- (18) When I was moused out, I blagged a cleaning job. (Christer, Sam 2011. *The Rome prophecy*. New York: Overlook, p. 11)
- (19) I'm all Madonna'd out. (Hilary, Sarah 2015. *No other darkness*. London: Headline, p 152)
- (20) I'm all caffeined out. (Hilary, Sarah 2015. *No other darkness*. London: Headline, p 203)
- (21) She was all drugged out when I went to see her the other day. (COCA 1991)
- (22) McGregor's all junked out, his head goes back and he's got a cigarette in his mouth. (COCA 1998)
- (23) We're all Seinfeld-ed out by now, right? I don't need a "Seinfeld" blooper tape (COCA 1999)
- (24) You all partied out from the balls last night? (COCA 2001)
- (25) "The lead singer was all coked out." # Pearl turned down the radio. "Sounds dangerous." # (COCA 2005)
- (26) if you are all burgered out already and its [sic] the beginning of the summer, have no fear (COCA 2009)

The pattern illustrated here has some peculiarities of its own. The nouns which make up the root of the head are often polysyllabic, while most phrasal verbs (not all) have monosyllabic heads. To this root is added an *-ed* suffix, which although it is not in origin a verbal suffix (it creates adjectives from nouns) is homophonous with a verbal suffix, and is used as the past/passive participle.

Although there are pre-existing patterns which might look as though they are parallel to this (illustrated, for instance, in *We have got it all worked out*), they are not. The pattern here is new, very productive – at least in colloquial usage, and apparently influenced by the denominal pattern mentioned earlier, the passive pattern mentioned earlier, the thoroughness and extinction patterns mentioned earlier. The pattern cannot be subsumed under any of these earlier patterns, but must be the product of multiple influences, including the patterns discussed here. We have here a new construction with its own meaning, ‘to have dealt with N to such an extent that one does not wish to deal with it any more’. The choice of *out* as the particle in this construction is not random, but is encouraged by various pre-existing patterns, none of which, on its own, is sufficient to explain the semantics and form of the new pattern. The high profitability of the new pattern must be explained by some onomasiological need, but the creation of the new pattern, that is its availability, demands that we consider multiple sources simultaneously.

6.3 Construction 2: Adjective *out*

The second construction to be discussed here is much rarer, and much less obviously productive. It is illustrated with the example of *weird out*, which seems to have arrived in the 1970s. This looks as though it is part of the series of deadjectival verbs mentioned in Section 6.1, but semantically it is distinct. The verbs in this series are all verbs of affected psychological state. A person who is weirded out is made to feel weird by some phenomenon and loses some mental control as a result (see ‘lack of awareness’ in Section 6.1).

The *OED* does seem to have a suitable gloss for *out* in such expressions. It is ‘From one’s normal, equable, or amicable state of mind, or ordinary course of action; into confusion, anger, or disturbance of feeling.’ The meaning is attributed solely to the *out* element rather than to the whole construction because the *OED* tends to treat phrasal verbs as syntactic constructions, and this to treat any element of meaning which cannot be attributed to the verb as being a meaning of the particle. The *OED* examples, however, focus of the phrasal verb *throw out* (*Seeing her there acting the part of a governess... threw him out*). The construction here is different because of the adjective in the first element.

There are a few apparently parallel constructions (though the degree of parallelism is rather variable). Some of these have nouns rather than adjectives in the first element. *Creep out*, for example, must be assumed to have a noun in the first element because its past participle is *creeped*, while if it were the verb *creep*, we would expect *crept*. *Psych out* may belong to the set, though the precise nature of the *psych* element is obscure; the *OED* sees it as a verb, as in *psych up*. *Psych out* again arises ca. 1970. *Freak out* seems to be parallel

in meaning to *weird out*, but *freak* is, again, presumably a noun – it is not related in meaning to an earlier verb *freak* ‘variegate’. *Gross out* (thanks to Ross Clark for this example) seems to be the closest parallel, with an adjective in the first element. According to the *OED*, it first arises ca. 1965, about the same time that *freak out* is first registered.

While there are clear influences of this set from deadjectival and denominal sets, and perhaps the meaning of *out* shown in *throw out* and a few other verbs, in this particular case there seems to be influence on the pattern from individual verbs, perhaps most clearly from *gross out*. To the extent that we can talk of the productivity of any pattern here, it seems to have lasted no more than ten years and have been only weakly productive. Nonetheless, it illustrates the point that the bourgeoning of individual patterns and their becoming available are influenced by multiple factors, not just one or not just a rule format.

6.4 Summary

What these examples show is that the availability of phrasal verbs is influenced by multiple factors all at once. It is not enough to say that there is a pattern of verb + *out*, several patterns coexist and may support each other in the productivity of individual verbs. We can rephrase this by saying that productivity in such instances is not a matter of having a rule allowing the existence of a wide range of semantic patterns, but a matter of multiple paradigms influencing the possibility or the probability of a new form. Although we have not illustrated this here, we would assume that multiple supporting paradigms are likely to increase the productivity of a given pattern. As with the other examples discussed above, we see a large range of low-frequency items forming a basis for productivity.

7. Conclusions

There are certain generalities which emerge from the kind of data presented in this paper.

1. The productivity of a morphological process or of an affix is an average measured over several different patterns and levels of productivity, and is, to this extent, potentially misleading. Given the number of meanings of ‘productivity’ distinguished by Baayen (2009) and the fact that averages must apply to all of them, any statement of the productivity of an individual process must be of very low value.
2. Productivity is local: the individual lexemes and morphemes involved may influence productivity.

3. Productivity is influenced by multiple factors simultaneously. These factors may be linguistic (different constructions) or non-linguistic (the context, the community in which the word is first used).
4. Productivity is shown by, and potentially also encouraged by, a number of low-frequency items in a single schema, more often than by single exemplars of high frequency. One of the best guides a speaker has to productivity may be low-frequency items.
5. Productivity has a qualitative side, not simply a quantitative side. Measures have to be placed in a theoretical space, but often that theoretical space has not been fully described or fully understood.
6. The notion of analogy seems preferable to the notion of rule in dealing with productivity. Although we have not espoused a particular model of analogy, there are several in the literature (e.g. Skousen 1989, Mattiello 2017) and we are not here concerned with the detailed operation of such models, but with the fundamental principle.

All of these factors point to an approach to productivity which is different from that which has generally been taken in the recent past. Partly the difference is a matter of taking a viewpoint which has been influenced by various types of Cognitive Linguistics, rather than a rule-governed model. We believe that this difference of approach will help focus future approaches to productivity.

References

- Adams, Valerie. 1973. *An introduction to modern English word formation*. London: Longman.
- Andrews, Sally, Brett Miller & Keith Rayner. 2004. Eye fixation measures of morphological segmentation of compound words: There is a mouse in the mousetrap. *European Journal of Cognitive Psychology* 16(2). 285–311.
- Arndt-Lappe, Sabine & Ingo Plag. 2013. The role of prosodic structure in the formation of English blends. *English Language and Linguistics* 17(3). 357–563.
- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge MA: MIT Press.
- Baayen, R. Harald. 1992. Quantitative aspects of morphological productivity. *Yearbook of Morphology* 1991. 109–149.
- Baayen, R. Harald. 2003. Probabilistic approaches to morphology. In Rens Bod, Jennifer Hay & Stefanie Jannedy (eds.), *Probabilistic Linguistics*, 229–287. Cambridge, MA: MIT Press.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 900–919. Berlin: Mouton de Gruyter.
- Baayen, R. Harald. 2010. The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In Susan Olson (ed.),

- New Impulses in Word-Formation (Linguistische Berichte Sonderheft 17)*, 383–402. Hamburg: Buske.
- Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling *The Times*: Productive Lexical Innovations in an English Newspaper. *Language* 72. 69–96.
- Baayen, R. Harald, & Rochelle Lieber. 1991. Productivity and English derivation: A corpus based study. *Linguistics* 29(5). 801–843.
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 1998. Is there a class of neoclassical compounds, and if so is it productive? *Linguistics* 36(3). 403–422.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2006. Compounds and minor word-formation types. In Bas Aarts & April McMahon (eds.), *The handbook of English linguistics*, 483–506. Oxford: Blackwell.
- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Beliaeva, Natalia. To appear. Blending in morphology. In Rochelle Lieber (ed.), *The Oxford Encyclopedia of Morphology*. Oxford University Press.
- Beliaeva, Natalia. 2014. A study of English blends: From structure to meaning and back again. *Word Structure* 7(1). 29–54.
- Beliaeva, Natalia. 2016. Blends at the intersection of addition and subtraction: Evidence from processing. *SKASE Journal of Theoretical Linguistics* 13(2). 23–45.
- Beliaeva, Natalia & Elizaveta Tarasova. 2012. Going off with *off* and figuring out *out*: A case study of English particle verbs. Seminar presented at Victoria University of Wellington.
- Bell, Melanie J. 2012. The English noun-noun construct: a morphological and syntactic object. In Angela Ralli, Geert Booij & Sergio Scalise (eds.), *On-line Proceedings of the Eighth Mediterranean Morphology Meeting (MMM8)* Cagliari, 14–17 September 2011, 59–91.
- Bertinetto, Pier Marco. 2001. Blends and syllabic structure: A four-fold comparison. In Mercè Lorente, Núria Alturo, Emili Boix, Maria-Rosa Lloret & Lluís Payrató (eds.), *La gramàtica i la semàntica per a l'estudi de la variació*, 59–112. Barcelona: Promociones y Publicaciones Universitarias.
- Bolinger, Dwight. 1971. *The phrasal verb in English*. Cambridge, MA: Harvard University Press.
- Booij, Geert. 2008. Construction morphology and the lexicon. In Fabio Montermini, Gilles Boyé, & Nabil Hathout (eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, 34–44. Somerville, MA: Cascadilla Proceedings Project.

- Booij, Geert. 2010. Compound construction: Schemas or analogy? A Construction Morphology perspective. In Sergio Scalise, & Irene Vogel (eds.), *Cross-disciplinary issues in compounding*, 93–108. Amsterdam & Philadelphia: Benjamins.
- Booij, Geert. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Niemeyer.
- Courtney, Rosemary. 1983. *The Longman dictionary of phrasal verbs*. Harlow: Longman.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA): 400+ million words, 1990–present*. <https://corpus.byu.edu/coca/>.
- Davies, Mark. 2013. *Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day*. <http://corpus.byu.edu/now/>
- Dressler, Wolfgang U. 2000. Extragrammatical vs. marginal morphology. In Ursula Doleschal & Anna M. Thornton (eds.), *Extragrammatical and marginal morphology*, 1–10. München: Lincom Europa.
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Rivista di Linguistica* 15. 31–62.
- Fandrych, Ingrid. 2008. *Pagad, chillax and jozi: A multi-level approach to acronyms, blends, and clippings*. *Nawa Journal of Language and Communication* 2(2). 71–88.
- Fernández-Domínguez, Jesús, Ana Díaz-Negrillo & Pavol Štekauer. 2007. How is low morphological productivity measured? *ATLANTIS* 29(1). 29–54.
- Fradin, Bernard. 2000. Combining forms, blends and related phenomena. In Ursula Doleschal & Anna M. Thornton (eds.), *Extragrammatical and marginal morphology*, 11–59. München: Lincom Europa.
- Gries, Stefan Th. 2006. Cognitive determinants of subtractive word-formation processes: A corpus-based perspective. *Linguistics* 17(4). 535–558.
- Gries, Stefan Th. 2012. Quantitative corpus data on blend formation: Psycho- and cognitive-linguistic perspectives. In Vincent Renner, François Maniez & Pierre J. L. Arnaud (eds), *Cross-disciplinary perspectives on lexical blending*, 145–167. Berlin: Mouton de Gruyter.
- Hay, Jennifer. 2003. *Causes and consequences of word structure*. London: Routledge.
- Hay, Jennifer & R. Harald Baayen. 2002. Parsing and productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 2001*, 203–235. Dordrecht: Kluwer Academic Publishers.
- Hay, Jennifer & R. Harald Baayen. 2004. Phonotactics, parsing and productivity. *Italian Journal of Linguistics* 15(1). 99–130.
- Hay, Jennifer & R. Harald Baayen. 2005. Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Science* 9(7). 342–348.
- Hay, Jennifer & Ingo Plag. 2004. What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22(3). 565–596.

- Huddleston, Rodney. 2002. The clause: complements. In Rodney Huddleston & Geoffrey K. Pullum (eds.), *The Cambridge grammar of the English language*, 213–321. Cambridge: Cambridge University Press.
- Lehrer, Adrienne. 1998. Scapes, holics, and thons: The semantics of combining forms. *American Speech* 73. 3–28.
- Lehrer, Adrienne. 2007. Blendalicious. In Judith Munat (ed.), *Lexical creativity, texts and contexts: The morphology/stylistic interface*, 115–133. Amsterdam & Philadelphia: Benjamins.
- López Rúa, Paula. 2004. The categorial continuum of English blends. *English Studies* 85(1). 63–76.
- Mattiello, Elisa. 2013. *Extra-grammatical morphology in English: Abbreviations, blends, reduplicatives, and related phenomena*. Berlin: Mouton de Gruyter.
- Mattiello, Elisa. 2017. *Analogy in word-formation*. Berlin & Boston: Mouton de Gruyter.
- Meys, Willem J. 1975. *Compound adjectives in English and the ideal speaker-listener*. Amsterdam: North Holland.
- OED. The Oxford English Dictionary online. www.oed.com.
- Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin & New York: Mouton de Gruyter.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.
- Plag, Ingo. 2006. The variability of compound stress in English: Structural, semantic and analogical factors. *English Language & Linguistics* 10(1). 143–172.
- Plag, Ingo, Gero Kunter, & Sabine Lappe. 2007. Testing hypotheses about compound stress assignment in English: A corpus-based investigation. *Corpus Linguistics & Linguistic Theory* 3(2). 199–233.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Renner, Vincent. 2015a. Lexical blending as wordplay. In Angelika Zirker & Esme Winter-Froemel (eds.), *Wordplay and metalinguistic / metadiscursive reflection: Authors, contexts, techniques, and meta-reflection*, 119–133. Berlin: Mouton de Gruyter.
- Renner, Vincent. 2015b. Panorama rétro-prospectif des études amalgamatives. *Neologica: Revue Internationale de La Néologie* 9. 97–112.
- Schreuder, Robert, & R. Harald Baayen. 1997. How simplex complex words can be. *Journal of Memory and Language* 37. 118–139.
- Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.
- Štekauer, Pavol. 2000. *English word-formation: A history of research (1960–1995)*. Tübingen: Narr.
- Tarasova, Elizaveta. 2013. *Some new insights into the semantics of English N+N compounds*. Unpublished PhD Thesis. Victoria University of Wellington.
- Yang, Charles. 2016. *The price of linguistic productivity. How children learn to break the rules of language*. Cambridge MA: MIT Press.

Laurie Bauer
School of Linguistics and Applied Language Studies
Victoria University of Wellington
PO Box 600
Wellington 6140
New Zealand
laurie.bauer@vuw.ac.nz

Natalia Beliaeva
School of Linguistics and Applied Language Studies
Victoria University of Wellington
PO Box 600
Wellington 6140
New Zealand
belayanaveta@gmail.com

Elizaveta Tarasova
IPU New Zealand
57 Aokautere Drive
Fitzherbert
Palmerston North 4472
New Zealand
etarasova@ipu.ac.nz