Blending creativity and productivity: on the issue of delimiting the boundaries of blends as a type of word formation

Natalia Beliaeva
Blending creativity and productivity: on the issue of delimiting the boundaries of blends as a type of word formation

Natalia Beliaeva

Introduction

1 Setting the limits of blending as a type of word-formation has been a widely discussed question in morphological studies to date. In particular, whether blending is primarily a phenomenon of word creativity, or whether it is a regular and predictable mechanism of word formation, remains an open question. Rather than maintaining an argument in favour of either position, this study aims to investigate factors underlining such distinction, and to develop criteria that could be applied to corpus data to characterize particular examples of blending as points on a cline from word creativity to productive word formation.

2 As has been pointed out by a number of scholars (Algeo [1977], Dressler [2000], Fandrych [2008], Gries [2004], López Rúa [2004], Bauer [2012], to name just a few), blends are remarkably diverse in terms of their formal structure. While the majority of blends are formed out of two constituents by combining the initial segment of one with the final segment of the other, e.g. *dramedy* < drama + comedy, *hangry* < hungry + angry, there are a lot of counterexamples including formations that a) contain three or more constituents, e.g. *Thankshallowistmas* < Thanksgiving + Halloween + Christmas; b) involve intercalation of constituents, e.g. *prowebstinate* < procrastinate + web; c) combine final parts of both words, e.g. *frohawk* < afro + Mohawk; or d) demonstrate other unconventional ways of combining lexical items, e.g. *dublexia* < www + dyslexia. In addition, the constituents of blends display a variety of semantic relationships (see, for example, Algeo [1977] for classification). On the other hand, there is abundant evidence
in the literature showing that the form of a blend can to a large extent be predicted based on phonological and phonotactic properties of their source words (Kubozono [1990], Gries [2012], Arndt-Lappe & Plag [2013], Beliaeva [2014]) and the similarity between them (Kelly [1998], Gries [2006, 2012]). The results of corpus studies such as those in Gries [2006, 2012], as well as experimental data such as those in Arndt-Lappe & Plag [2013] and in Beliaeva [2016] demonstrate that the formation of blends is subject to psycholinguistically relevant constraints such as recognisability of the source words.

In some cases, the formation of blends is not only predictable, but may give rise to productive constructions where part of a word that was once blended is used to form further blends, e.g. -cation (from vacation) in workcation, spa-cation, staycation, etc. As discussed in Lehrer [1998] and as further substantiated by contemporary corpus data in Bauer, Beliaeva & Tarasova [2019], the morphological status of formations containing splinters such as -cation can be compared to that of neoclassical compounds or, to some extent, affixations. At the same time, a lot of blends do not demonstrate such productivity, and factors that influence this are yet to be investigated. An aspect of the use of blends that is essential to consider is the playful character of many blends and their use as expressive means of language (see, for example, Lehrer [2007] and Renner [2015] for the discussion of blends as creative lexical formations). Importantly, factors that enhance the creative and attention-catching properties of blends may at the same time decrease the predictability of their form, as specified by Renner [2015].

This study argues that blending, as a word formation process, can be driven by either factors enhancing predictability of the outcome or factors enhancing playful character of the formation, or both. Corpus-based approach is used to investigate the relationship between well-formedness of the blends (that is, conforming to prosodic and recognisability constraints), and their productivity potential. Data from the Corpus of Contemporary American English (Davies [2008–]), earlier publications on blends, online collections of neologisms such as Word Spy, Urban Dictionary and The Rice University Neologisms Database, and other media sources are used to investigate exemplars of different structures of blends. Data from the NOW Corpus (Davies [2013–]) are further analyzed to construe patterns of recurrent use of blend constituents in various novel formations. The results are used to develop criteria of characterizing individual blends in terms of their productivity and creativity, and to substantiate the identification of blending as a mechanism involving both word creativity and productive word formation at varied degrees.

1. Fuzzy boundaries of blends revisited

A blend word used in advertisement, e.g. Frappuccino® (Urban Dictionary), or in political media, e.g. Brexit (Top Words for the first 15 Years of the 21st century and what they portend) is both attention-catching and thought-provoking. Most blends are attention-catching for language users, and the reasons for this will be considered in Section 2. For linguists, however, this phenomenon has been mind-bogging for decades, because the unusual formal properties of blend words made it difficult to provide an exhaustive description of blends as a word formation category or even define what a blend is. In this section, the much-discussed question of defining the status of blends among other types of word formation will be revisited not with the purpose of putting an end to the

Lexis, 14 | 2019
discussion, but, rather, to accentuate factors that are essential for understanding the connection between blends and other word formation types.

6 Putting together two words to form a compound such as sugar bowl or blackboard can be considered one of the most straightforward ways to form a new lexeme – the universality of compounds is pointed out, for example, in Fromkin, Rodman & Hyams [2012]. A less straightforward and less common way of making a new lexeme using two existing ones is merging two (or sometimes more) words together so that part of the material is lost in the process, which will produce a blend. In fact, in earlier morphological studies such as those by Marchand [1969] and Adams [1973], blends have been classified as a subtype of compounds. Other scholars, such as Devereux [1984] and Cannon [1986], considered truncation a more important prerequisite of blending and therefore classified blending as a type of shortening. Since both shortening and concatenating are essential characteristics of blending, as admitted by many scholars, regarding blending as a word formation type akin to both compounding and shortening should not be a contradiction. For example, López Rúa [2004] categorizes blends as points on a cline between compounds and abbreviations, depending on how much the source words are truncated while forming a particular blend. Similarly, prototypical approach to classification of blends is undertaken by Brdar-Szabó & Brdar [2008] and by Mattiello [2013].

7 Even such a brief glimpse of the vast literature on blending would lead us to a conclusion that the boundaries of the category are fuzzy. What is more remarkable, some researchers claim that the formal diversity that precipitates the fuzzy boundaries is an essential characteristic of blending, while others focus on the regularities that can be observed in linguistic data, and aim to find grounds for a systematic description of blends drawing on those features of blends that are predictable, at least to some extent. The former approach is taken, for example, by López Rúa [2004, 2012] and by Mattiello [2013] whose classifications of blends are based on the results of analyzes involving a plethora of structurally different examples. Taken to extremes, the account of formal diversity of blends can lead to conclusions that they are an “extragrammatical” phenomenon (Dressler [2000]), which is, at best, “minimally predictable” (Mattiello [2013: 96]), and therefore blends can be regarded as instances of word-creation as defined by Ronneberger-Sibold [2010].

8 A factor that is often related to the diversity and unpredictability of blends is the punning nature of many blends if not all of them. Thus, Renner [2015: 130] claims that “[c]oining a new lexical blend is an act of wordplay”, having defined wordplay in the sense that bears some similarity with the concept of word-creation by Ronneberger-Sibold [2010], that is, as follows:

    [...] an intentional and formally ingenious way of associating the semantics of two or more words in a new morphological object. [Renner 2015: 119]

9 Renner substantiates this claim by analyzing examples of lexical blends that involve various degrees of overlapping between the source words, infixation, intercalation, violations of structural well-formedness constraints that are otherwise observable in the language, graphic and semantic play on words. The importance to take into consideration the playful character of blends is reiterated in Beliaeva [2019], where it is also related to the use of blends as expressive means in various domains, e.g. in youth slang (01), journalism (02), political media (03) or as means of creating brand names
or other attention-catching titles such as musical band names (05) – note that parts of the words that are omitted in the process of blending are put in parentheses:

01 brovember < bro + (N)ovember, referring to a month of doing manly things (Urban Dictionary)
02 Brangelina < Br(ad) [Pitt] + Angelina [Jolie] (Urban Dictionary)
03 Merkozy < Mer(kel) + (Sar)kozy, referring to a unified position of France and Germany (Stewart [2011])
04 Teaffee < tea + (co)ffee, the name of a café (http://teaffee.com/index.html)
05 Stratovarius < Strato(caster) + Stra(di)varius, the name of a musical band combining heavy metal and classical music (López Rúa [2012: 31])

As further observed in Beliaeva [2019: n.p.], wordplay cannot be regarded as “the only driving force of blend formation”, because in such cases as creating names of animal hybrids (06), language varieties (07) or hybrid recipes (08) lexical blending may be used in order to reflect the hybrid nature of the objects that are named.

06 zorse < z(ebra) + (h)orse, referring to a hybrid of a zebra and a horse (Urban Dictionary)
07 Spanglish < Span(ish) +(En)glish, used to refer to dialects, pidgins, or creole languages that result from interaction between Spanish and English (Nash [1970])
09 cronut < cro(issant) + (dough)nut, a doughnut-croissant hybrid (Merwin [2013])

Although language play thus cannot be seen as the only factor that influences the formation of blends, it should definitely be taken into consideration in the analysis of blending as a morphological phenomenon.

Scholars arguing for predictability of blends focus on the regularities that may help develop a systematized account of this category, despite its fuzzy boundaries. Such systematizations have been developed in many studies including Kubozono [1990], Gries [2006, 2012], Arndt-Lappe & Plag [2013] and Beliaeva [2014]. Thus, in Gries [2006, 2012] and Beliaeva [2014], inferences about systematic nature of blends are made based on corpus data on the frequency of occurrence of certain types of formations. Some insights into the mechanism of blending are drawn from considering cognitive factors involved in the formation (Arndt-Lappe & Plag [2013], Beliaeva [2014], Gries [2006, 2012]) and processing (Beliaeva [2015, 2016]) of blends.

In sum, two important directions of reasoning can be distinguished in the abundant literature on blends. On the one hand, investigating formal attributes of blends to the full extent of their diversity can provide valuable information about the genesis of blends and their role in language. On the other hand, insights from frequency-based accounts of regularities of blends are no less valuable. However, it appears that focussing on one of these two aspects of the phenomenon may lead to blinkered vision of the complex picture. In Sections 2 and 3, we will consider both the diversity and the predictability of blends, in an attempt to provide a plausible explanation of the fuzzy boundaries of the category and uncover some factors involved in blend formation that may have been understudied to date.
2. Multiformity of blends as interplay between creativity and predictability

2.1. Formal diversity of blends

In most general terms, the mechanism of blending was described in Plag [2003: 123]:

\[ AB + CD = AD \]

Thus blending involves combining parts of two source words so that the beginning of the first one is conjoined with the ending of the second one. However, the linguistic examples that appear to accord with this seemingly simple formula exhibit exceptional formal diversity. As summarized in Beliaeva [2019], following earlier classifications in Algeo [1977], Cannon [1986], Gries [2006] and other studies, the possible outcomes of merging a word AB with a word CD include cases when the beginning of one source words is concatenated with the ending of another one (10), when the same process takes place but the source words overlap where they are merged together (11), when the source words overlap so that the first word is entirely preserved in the blend (12), when the source words overlap so that the second word is entirely preserved in the blend (13), and when the source words overlap so that both of them are preserved in their entirety (14). The overlapping segments in the examples henceforth are in bold type:

(10) tigon < tig(er) + (li)on (COCA)
(11) motel < mot(or) + (h)otel (COCA)
(12) mockbuster < mock + (bl)ockbuster (Urban Dictionary)
(13) jumbrella < jumb(o) + umbrella (Word Spy)
(14) alcoholiday < alcohol + holiday (Urban Dictionary)

This classification is far from exhaustive, as less frequently attested outputs of merging a word AB with a word CD include cases when one word replaces a segment in the middle of another word (15), referred to as infixed blends in Bauer [2012] or as central replacement blends in Beliaeva [2014]. Sometimes the insertion takes place where the two words are phonetically and graphically similar, which thus involves some overlap (16). More rare cases include formations combining final parts of the source words (17–18).

(15) pro-web-stinate < pro(cra)stinate + web (The Rice University Neologisms Database)
(16) li-boob-rian < lib(ra)rian + boob (The Rice University Neologisms Database)
(17) frohawk < (a)fr(o) + (M)o(hawk (Urban Dictionary)
(18) podestrian < (i)Pod + pedestrian (Urban Dictionary)

Occasionally, more than two words are merged into one, such attestations ranging from blends with three elements (19) to more extreme cases of blending six words (20).

(19) Japornimation < Jap(a)n + porn + (a)mination (Mattiello [2013: 122])
(20) Christmahanukwanzadandiwalstice < Christmas + Hanukkah + Kwanza + (R)amadan + Diwali(i) + (Winter sol)stice (Renner [2015: 126])

Although deviating from a more widely attested pattern of coining blends by merging two words, formations such as those in (17–20) bear similarity with more typical blends in other respects. Thus, the source words of blends in (17–20) are merged together in
points where the degree of similarity between them is maximal, and blends are coined in such a way that the source words remain recognisable in the resulting formations. It is worth noting that similarity and recognisability of the source words are factors that have been characterized by some scholars, such as Gries [2006, 2012] and Beliaeva [2014, 2016], as essential properties of typical blends that distinguish them from other types of word formation including clippings. With regard to the degree of similarity between the source words and the relative amount of the source word material preserved in the blend, formations in (17–20) can be considered more conforming to typical properties of blends than coinages combining initial parts of two (21) or more (22–23) source words.

(21) globfrac < glob(alisation) + frag(mentation) (Urban Dictionary)
(22) SoLoMo < so(cial) + lo(cal) + mo(bile) (Word Spy)
(23) COSPAR < Co(mmittee) (on) Spa(ce) R(esearch) (López Rúa [2002: 46])

Formations that are structurally similar to the example in (21) have been described in literature as clipping compounds (Bauer [2012]) or complex clippings (Gries [2006]). The formation in (22) is also similar to (21) but combines initial segments of three words rather than two. This example is also formally similar to acronyms as described, for example, in Quirk, Greenbaum, Leech & Svartvik [1985] and in López Rúa [2002] – also compare (22) to the formation in (23) which was provided as an example of an acronym. An analysis of comparable cases is undertaken in López Rúa [2004], where a prototypical account of blends, compounds and various types of shortenings is used to construct a “categorial continuum” of forms coined by concatenation or merging together parts of existing words. This categorial continuum arranges compounds, neoclassical compounds, blends and acronyms depending primarily on the degree of shortening, that is, on a cline from full preservation of constituents (the case of compounds) to minimal preservation (the case of acronyms preserving one letter from each of their source words such as NATO). A similar conclusion is drawn in Bauer [1998], on the basis of a study of neoclassical compounds. In particular, Bauer [1998: 414] considers blends and clipping compounds as “intermediate stages” between compounds and abbreviations in terms of the degree of shortening involved in their formation. A visualization of such account of blends and related morphological categories is proposed in Figure 1.
As is illustrated in Figure 1, the result of concatenating parts of several existing words can resemble a compound or an acronym to a certain extent, depending on how much material of the source words is preserved. The central area in Figure 1 represents cases that are often referred to as prototypical blends, that is, formations combining the beginning of one word with the ending of another, with some overlap in the middle. Peripheral cases include formations that are similar to acronyms, those similar to compounds, and those which combine word parts in a way that deviates from the more frequently observed pattern exemplified in the middle part of the figure. The more frequently observed pattern, in its turn, is subject to several factors that will be further discussed in Section 2.2.

2.2. Well-formedness and creativity of blends: A seeming paradox

Criteria of well-formedness or prototypicality of blends have been investigated in many studies (Marchand [1969], Kubozono [1990], Kelly [1998], Adams [2001], López Rúa [2004], Gries [2006, 2012], Bauer [2012], to name a few). These criteria of prototypicality concern semantic relations between the source words of the blend, as well as phonotactic and prosodic restrictions the blends are subject to. The semantic considerations of blend formation are beyond the scope of the present study (but see Devereux [1984], Cannon [1986], Renner [2006] for semantic classifications, Gries [2012] and Beliaeva [2014] for relevant quantitative analysis, and Bauer [2012] for a comprehensive summary).

The phonotactic and prosodic factors that are related to well-formedness of blends determine what components of the source words (such as syllable elements and main stress) are preserved in the blend. In particular, blends tend to preserve as much material of both of their source words as possible, while conforming to overall phonotactic constraints of language. Thus, typical blends (as discussed in more detail in Beliaeva [2019], following earlier investigations by Kubozono [1990], Kelly [1998], Arndt-Lappe & Plag [2013] and other scholars) have the same number of syllables as the
longest of their source words (24–25, 27), or are one syllable longer (26). As observed in Beliaeva [2019], in many cases longer blends are formed due to preservation of maximal possible overlapping segment, as is the case in (26). Blends also tend to preserve the main stress of at least one of their source words (the stressed syllables in examples 24–27 are underlined). In some cases (24, 26), the main stress of both source words is kept, or, when this is not possible, most often the stress of the second source word is preserved (25).

(24) bor\textit{dinary} \textless{} boring + ordinary
(25) dram\textit{ality} \textless{} drama + (re)ality
(26) prediction\textit{ary} \textless{} prediction + d\textit{ictionary}
(27) proweb\textit{stinate} \textless{} procrastinate + web (Beliaeva [2019])

If we now reconsider more exotic blend forms exemplified in Section 2.1, we can observe that peripheral cases of blending such as the ones involving intercalation or a combination of more than two elements are highly marked and thus attention-catching. According to Renner [2015], structural complexities such as intercalation in (16) and (19) or the presence of three or more structural elements in (19–20, 22–23) are factors that contribute to the playfulness of blends. Thus, more typical blends which can be regarded as more well-formed, may at the same time fulfill the word play function less successfully. However, note that blends that look not typical in regard to their linear structure, such as a central replacement blend in (27), or blends that do not include word beginnings in (17–18), nevertheless conform to the same phonotactic restrictions, that is, preserve the prosodic contour (the number of syllables and the main stress) of the longest of their source words. This may be related to the recognisability of the source words which, as shown in Gries [2006, 2012] and in Beliaeva [2014], is a key factor affecting blend formation. As gathered in Renner [2015], increased difficulty in recovering the constituents of the blend makes blends more playful. However, for the word play to occur, the constituents of blends have to remain potentially recognisable. Increased complexity of form involves lower recognisability of constituents, and therefore some compensatory formal features increasing recognisability are necessary. The preservation of the prosodic contour of one of the source words (particularly the one which loses its initial part) can be such a compensatory feature, being an essential prerequisite for recognisability, as shown in Gries [2012], Arndt-Lappe & Plag [2013] and Beliaeva [2016]. Therefore, increased playfulness may be associated with atypical form, but not all formal deviations are equally plausible for enhancing playfulness. We will return to this apparent paradox after considering the criteria of well-formedness of blends in relation to the possibility of recurrent use of their constituents.

3. Predictability of blends and its interplay with the productivity of splinters

3.1. Analysis of splinter productivity: Rationale

A distinguishing feature of blends is that, while being blended, words are split at points different from morph boundaries. Occasionally, counterexamples can be found. Thus, in (28–29) a compound is blended with a monomorphemic word so that the latter
replaces either the first (28) or the second (29) compound constituent, and the structure of the resulting blend formally coincides with that of a typical compound.

(28) cookprint < cook + (foot)print (Word Spy)
(29) passthought < pass(word) + thought (Word Spy)

More often, however, one or both source words of blends are cut to what is commonly referred to as ‘splinters’ (Adams [1973], Cannon [1986], Kemmer [2003], López Rúa [2004]). The boundaries of splinters are determined by factors other than morph boundaries. On the one hand, the switch point in blends normally falls on the syllable boundary or the boundary between onset and rime (see Kubozono [1990], Kelly [1998], Plag [2003] and Bauer [2012] for more detailed discussion), which is in agreement with phonotactic restrictions discussed in Section 2.1. On the other hand, the splinter length is subject to recognisability of the source words, that is, splinters tend to be long enough for the words they originate from to remain recognisable. In particular, as shown in a much-cited study by Gries [2006], the switch point in blends is usually close to the ‘recognition point’ of their source words, that is, a point at which the word is distinguished from other words of language that begin or end with the same letter / phoneme string. If we also take into consideration the tendency of blends to maximize overlap between their source words, as mentioned in Section 2.1, it turns out that splinters are the outcome of a complex interaction of phonotactically and cognitively relevant factors. Many splinters are tailored for one particular blend they appear in. However, a notable number of splinters are subsequently used to form further blends. With regard to such cases, Kemmer [2003] concluded that blending can give start to a productive process the result of which will be a lexical family and, eventually, a bound morph.

It is not surprising that splinters can be one-off formations only existing in the blends they appeared in, given the complex interplay of factors involved in their forming to suit a particular environment. The conditions under which splinters are used recurrently to give rise to a family of blends are of particular interest. Building on the investigation in Bauer, Beliaeva & Tarasova [2019], this section will focus on the analysis of recurrent use of splinters in contemporary English language. The term ‘splinter’ will be used throughout the section, although note that some researchers have referred to similar units as combining forms (Lehrer [1998], Warren [1990]). The analysis in Bauer, Beliaeva & Tarasova [2019] included estimating productivity of splinters using different quantitative measures such as type frequency as an estimate of profitability, and the ratio of the number of hapax legomena to the token frequency as an estimate of potential productivity (see Baayen [1992, 2009] for a detailed discussion of the methods and Bauer, Beliaeva & Tarasova [2019] for a discussion of applying the methodology to splinters). Different quantitative measures, however, induced different estimates of splinter productivity, and provide little ground for explaining the observed tendencies in the data:

[...] such models will be of little value without thorough understanding of what the measures actually represent. [Bauer, Beliaeva & Tarasova 2019: 65]

3.2. Data

The present study aims to evolve on findings in Bauer, Beliaeva & Tarasova [2019] regarding factors pertinent to the productivity of splinters, and to set grounds for
developing relevant quantitative analyses. The data for this analysis come from a set of novel formations containing splinters that was collected in 2017–2018 for Bauer, Beliaeva & Tarasova [2019]. The first stage of data collection involved selecting splinters which a) have been used to form novel blends, and b) have been used to form more than one blend. These selection criteria indicated that the splinters had potential for productive use, with regard to criteria of potential productivity discussed, for example, in Baayen [1992] and Arndt-Lappe & Plag [2013]. Therefore, novel blends were selected from a collection of blends attested in COCA or in online dictionaries and neologism databases (Word Spy, Urban Dictionary and The Rice University Neologisms Database) no earlier than in 2000 (as confirmed by a Google search within a specific timeframe, see also Beliaeva [2014] for a detailed discussion). Out of over 500 novel blends in the original collection (Beliaeva [2014]), pairs or sets of blends with identical initial or final splinters were identified. As a result, 22 splinters were selected as having potential for recurrent use. Out of these splinters, 11 were initial, such as *adver*-, originally the initial part of *advertisement* (30), and 11 were final, such as *-cation*, originally the final part of *vacation* (31).

(30) adver(tisement): adverblasting (Urban Dictionary), advergame (Word Spy)
(31) (va)cation: staycation (Urban Dictionary), mancation (Urban Dictionary)

28 The full list of initial splinters, with indication of the words they originate from, is provided in (33), and the full list of final splinters is provided in (34), the parts of the original source words that have been truncated are in parentheses.

(33) adver(tisement), alterna(tive), digi(tal), edu(cation), fabu(lous), fem(ale), fin(ancial), local(l), rob(o)t, virt(ual), yester(day)
(34) (su)urb, (va)cation, (edu)cation, (in)flation, (or)gasm, (ep)isode, (para)noia, (an)orexia, (no)stalgia, (God)zilla, (Mo)zilla

29 Note that some of the final splinters in (34) are homonymous, that is, some of attested novel formations ending in *-cation* are blends using the final part of the word *vacation*, others are blends using the final part of *education*. Likewise, either the word *Godzilla* or the name of an Internet browser *Mozilla* could be a source word of a formation ending in *-zilla*. For each particular case in this data set, the actual source word has been determined by analysing the meaning of the context. The analysis in Bauer, Beliaeva & Tarasova [2019] did not include homonymous splinters and was therefore restricted to 9 final splinters and an equal number of initial ones. The present analysis will include all 11 initial and 11 final splinters from the original data set.

30 In order to investigate the use of splinters in contemporary language and estimate their productivity, words starting or ending with particular splinters were extracted from NOW (Davies [2013–]), a corpus that currently contains about 8.5 billion words from a variety of web-based media from 2010 to the present. After excluding proper names and derivatives starting or ending with the same letter strings, the data set comprised a total of 3007 word types. Examples of formations with different splinters, alongside the number of types of these formations attested in NOW as of the time of collection are provided in Table 1.

Lexis, 14 | 2019
Table 1. Neological formations containing splinters

<table>
<thead>
<tr>
<th>Splinter</th>
<th>Examples</th>
<th>Number of types in NOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>adver-</td>
<td>advercasting, advergame, advertainment</td>
<td>15</td>
</tr>
<tr>
<td>alterna-</td>
<td>alterna-history, alternarock, alternanthem</td>
<td>68</td>
</tr>
<tr>
<td>digi-</td>
<td>digimind, digisafe, digishoppers</td>
<td>605</td>
</tr>
<tr>
<td>edu-</td>
<td>eduloan, edu-speak, educhat</td>
<td>267</td>
</tr>
<tr>
<td>fabu-</td>
<td>fabuhealth, fabulash, faburrito</td>
<td>21</td>
</tr>
<tr>
<td>fem(i)-</td>
<td>femgamer, femicare, femedic</td>
<td>179</td>
</tr>
<tr>
<td>fin-</td>
<td>fincorp, finspy, finvest</td>
<td>36</td>
</tr>
<tr>
<td>loca-</td>
<td>locabrand, locanomics, locavore</td>
<td>28</td>
</tr>
<tr>
<td>robo-</td>
<td>roboracers, robobortion, robob-voting</td>
<td>1180</td>
</tr>
<tr>
<td>virt-</td>
<td>virtnet, virtuheart, virtools</td>
<td>41</td>
</tr>
<tr>
<td>yester-</td>
<td>yestermorth, yester-regime, yesterworld</td>
<td>32</td>
</tr>
<tr>
<td>-burb</td>
<td>boomburb, urbanburb, cluburb</td>
<td>19</td>
</tr>
<tr>
<td>-cation¹</td>
<td>brocation, daycation, farmcation</td>
<td>71</td>
</tr>
<tr>
<td>-(u)cation²</td>
<td>fooducation, musication, NETucation</td>
<td>9</td>
</tr>
<tr>
<td>-flation</td>
<td>foodflation, joyflation, slowflation</td>
<td>38</td>
</tr>
<tr>
<td>-gasm</td>
<td>eargasm, mindgasm, scorgasm</td>
<td>105</td>
</tr>
<tr>
<td>-(i)sode</td>
<td>appisode, operasode, twittersode</td>
<td>14</td>
</tr>
<tr>
<td>-(a)noia</td>
<td>chemonoia, powernoia, juvenoia</td>
<td>18</td>
</tr>
<tr>
<td>-(o)rexia</td>
<td>bleachorexia, brideorexia, manorexia</td>
<td>22</td>
</tr>
<tr>
<td>-(o)stalgia</td>
<td>lustalgia, snowstalgia, technostalgia</td>
<td>23</td>
</tr>
<tr>
<td>-zilla¹</td>
<td>barzilla, sisterzilla, optimizilla</td>
<td>198</td>
</tr>
<tr>
<td>-zilla²</td>
<td>bugzilla, clonezilla, appzilla</td>
<td>18</td>
</tr>
</tbody>
</table>

3.3. Quantitative estimates of productivity and related shortcomings

It is noteworthy that not all items in the data set looked like typical blends, as described in Section 2.2. In fact, in most cases (2541 out of 3007 types, which is 84.5% of
the data), a splinter was attached to a free morph or sometimes a morphologically complex word without any truncation of the latter. In such formations, initial splinters resemble prefixes (35), and final splinters resemble suffixes (36). For subsequent analysis, formations like those in (35–36) have been classified as similar to affixations.

(35) advergame < adver- + game, roboracers < robo- + racers
(36) foodflation < food + -flation, snowstalgia < snow + -stalgia

A much smaller number of items in the data set (7.2%, or 217 out of 3007 types) had properties of well-formed blends. In such formations, initial (37) or final (38) splinters were concatenated with shortened words, which often involved some overlap, or are merged with other words so that they overlap with the splinter – the overlapping segments in (37–38) are in bold type. Note that while determining the length of the overlapping segments, only the recurrently used splinters were considered as constituents of such blends, rather than the original source words, e.g. alternantheme, rather than its source word alternative, was analysed as the constituent of the blend alternantheme in (38). However, if we considered it the result of merging alternative and anthem, this item would still have been classified as blend conforming to properties of typical blends discussed in Section 2.2. In this respect, the distinction between splinters and their original source words does not play a significant role for the present analysis. However, for analyses where it might be relevant (e.g. for determining whether or not the prosodic contour of the source word is preserved), the original source words were taken into consideration.

(37) robstitute < robo- + prostitute, femedic < fem- + medic
(38) alternantheme < altern- + anthem, lustalgia < lust - -stalgia

In some cases (39), splinters were attached not to a free morph but to items that answer the description of neoclassical combining forms as defined in Plag [2003]. Occasionally, combinations of two recurrently used splinters were found (40). Such cases are formally similar to examples in (35–36), the only difference being that splinters are attached to bound morphs, rather than free morphs. To account for this difference, items such as the ones in (39–40) have been classified as similar to combining forms, which comprised 121 types, or 4% of the data.

(39) robocide < robo- + -cide, mega- + -gasm
(40) robogasm < robo + -gasm, locavore < loca- + -vore

In the remaining 128 formations (4.3% of the data set), splinters were attached to clippings (41) or acronyms (42). The items formed in such way can be classified as clipping compounds or in some cases marginal blends, but since some items, e.g. VODzilla in (42), could be placed into either category, such cases were classified as ‘other’.

(41) femfest < fem- + fest(ival), tech(ology) + -gasm
(42) digiTV < digi- + TV, VODzilla < VOD (video on demand) + -zilla

Formations like those in (35–36), where splinters are attached to free morphs or morphologically complex words in an affix-like manner, are the most frequently observed in our data. This is true about all the splinters in the data set, as can be seen in Figure 2, which displays the number of types where each of the splinters under
investigation is attached to the other constituent of a novel formation in one of the following ways: similar to an affix, similar to a blend constituent, similar to a combining form (CF in Figure 2), or other.

Figure 2: Different structural types of blends in relation to additive and subtractive word formation categories

36 Such observation in itself should not be surprising given that affixation is a much more widespread way of forming new words in English than blending or using combining forms. This is not to say that the splinters in our data set have become affixes (which would be a premature conclusion), but this distribution is in accordance with the general observations regarding English word formation, e.g. those in Plag [2003]. Moreover, as observed in Bauer, Beliaeva & Tarasova [2019: 64], splinters with higher type frequency also tend to attach to free morphs in the majority of cases. Using quantitative estimates of productivity, however, did not provide an adequate explanation for the observed tendencies in the data because different measures (type frequency, type/token ratio and the ratio of the number of hapax legomena to the number of tokens) provided different results [Bauer, Beliaeva & Tarasova 2019: 64–65].

3.4. Taking more factors into consideration

37 To further explore the use of splinters and to gain more insights into their potential to participate in word formation, it was decided to take into consideration not only the type of concatenation and type frequency, but also possible factors that have been shown to influence the formation of blends, that is, the degree of overlap and prosodic contour. Therefore, the items in the data set have been coded with regard to the following parameters: the position of the splinter (initial or final), the type of concatenation (affix, blend, combining form, or other), type frequency, overlap (the number of graphemes which overlap where the splinter is attached to the word or other item), and the prosodic contour (that is, whether or not the syllable number and the main stress of the formation coincides with those of its first source word, the second source word, both of them, or none of them).
It is important to note that the degree of overlap and the prosodic contour were estimated with regard to the original source words of the splinters. Thus, the overlap in (43) was estimated as 2 because two last letters of the source word *floor* overlap with the first two letters of *orgasm*, which is the source word of the splinter *-gasm*. Similarly, the formation in (44) was classified as preserving the prosodic contour of its first source word *fabulous*, and the one in (45) as preserving the prosodic contour of its second source word *vacation*, even though the actual coining of these items could have included simple addition of splinters *fabu-* and *-cation*, respectively. Such coding principles were followed in order to avoid ungrounded judgements on the actual origin of the items in the data set, and also not to disregard potentially valuable information. In the process of coding for overlap and prosody, the type of concatenation was therefore ignored. Formations that had the same number of syllables and main stress position as both their source words (46) were classified as preserving the prosodic contour of both source words.

(43) floorgasm < floor + (or)gasm
(44) fabulips < fabu(lous) + lips
(45) braincation < brain + (va)cation
(46) digizen < digit(al) + (ci)zen

In accordance with this procedure, each formation containing splinters with type frequency of up to 100 was manually coded for type of concatenation, overlap, and prosodic contour. For higher type frequencies, a random sample of 100 types has been used for coding all the variables except for the numeric value of type frequency, which remained unchanged. Therefore, the final set amounted to 1419 data points.

A conditional inference tree (decision tree) analysis was used to estimate whether type frequency of splinters is related to other parameters listed above. A decision tree method is used to estimate a regression relationship between the output variable (in this case, type frequency of a splinter) and a set of predictor variables (type of concatenation, overlap, and prosody), as described in Hothorn, Hornik & Zeileis [2006]. If the algorithm rejects the hypothesis that there is no relationship between the dependent variable and any of the predictor variables, the variable which has the strongest association with the dependent variable is selected and the data set is split into two groups (branches of the conditional inference tree), which are significantly different from one another (p<0.05). Such partitioning is repeated until further splits of data are no longer justified. The partitioning can be stopped earlier in order to avoid overfitting the model to the data and to mitigate the interpretation of the results. In the present analysis, the maximal number of levels on which the decision tree can be split was set to three, that is, each group of data could be split no more than three times before reaching the final node. The results of the analysis are displayed in Figure 3. Each node of the decision tree in Figure 3 is titled by the variable which determined the split at that node, the number of data points in each of the resulting nodes is shown on top of the nodes, and the distribution of type frequency in each node is displayed as a boxplot where the black horizontal line marks median type frequency in the respective subset of data.
The results displayed in Figure 3 show that formations in which splinters are concatenated with free morphs like affixes (left branch of the decision tree) differ significantly from other types of concatenation (right branch of the tree) in terms of type frequency. In particular, items which bear formal similarity to affixations are more frequent than blends, combining forms and other types of concatenations. Other factors that appear to have association with type frequency are initial (ini) or final (fin) position of the splinter and, particularly for initial splinters, the preservation of the prosodic contour of the source words. Thus, the largest node in the right part of the tree (n=456) is characterized with the highest type frequency (median=200), affix-like type of concatenation, initial position of the splinter, and the prosodic contour which does not resemble that of any of the source constituents (w0). Such formations bear the most formal resemblance to affixations, as exemplified in (47).

(47) eduportal < edu- + portal, advergame < adver- + game

Interestingly, both for affix-like concatenations, and for the remaining data set, prosody appears to be associated with type frequency of initial splinters, rather than final ones. This seems controversial, given that preserving the prosodic contour of the word that loses its beginning (that is, in the case of the present data set, the source word of the final splinter) was considered an important characteristic of blends (see Gries [2012], Arndt-Lappe & Plag [2013], Beliaeva [2014]). Moreover, the partition of the data in the decision tree suggests that formations using different types of concatenation also differ in terms of prosodic characteristics. This can be illustrated by a mosaic plot in Figure 4.
In a mosaic plot, larger areas correspond to greater number of observations having certain characteristics, and darker areas represent greater deviations of the corresponding numbers from chance. In Figure 4, the type of concatenation is plotted against the horizontal axis, and the vertical axis displays the prosodic characteristics of items in the data set (w0=prosodic contour differs from both source words, w1=the prosodic contour of the first source word preserved, w2=the prosodic contour of the second source word preserved, w1w2=the prosodic contours of both source words preserved). As can be gathered from Figure 4, if the formation conforms to the criteria of blends (type=blend), it is more likely that it will preserve the prosodic contour of its second source word, which is not the case with other types of concatenation. Similar inferences can be deducted from studying the distribution of the data according to concatenation type and the number of overlapping letters, which is displayed in Figure 5. In particular, the majority of items in the data set do not have any overlap between the constituents, but blends differ from the rest of the data in this respect as nearly half of items in this group are characterized with some degree of overlap.
The observed results indicate that the type frequency of formations in the data set is subject to different factors from those associated with well-formedness of blends. The results of the quantitative analysis also show that splinters which have greater type frequency at the same time bear the least resemblance to blend constituents, as the majority of types of neological formations including these splinters do not involve any overlap and do not preserve the prosodic contour of their source words. This may indicate that greater productivity can be associated with losing characteristics of blend splinters and acquiring characteristics of a bound morph, which is also in accordance with observations in Kemmer [2003]. Respectively, novel formations containing splinters that have shown signs of increasing productivity, are less likely to conform to the criteria of well-formedness of blends discussed in Section 2.2.

4. Conclusion: Prototypical blends as points on a cline between creativity and productivity

As has been extensively discussed in the literature, and as reiterated in Section 2.2, the formation of blends is to a large extent predictable. In particular, typical blends consist of two elements, one of which is the beginning or the whole of one word, another is the ending or the whole of another word. In addition, the blend constituents tend to overlap where they are merged together, and the overall prosodic contour of the resulting formation resembles that of one of its source words, typically the second one.

There are many examples of blends and similar formations that deviate from these criteria of well-formedness, but, more importantly, different directions of possible deviation can be distinguished. On the one hand, it is possible to come across less typical formations which, at the same time, can be characterised by increased punning nature and playful character, as discussed in Section 2. On the other hand, if blend
splinters are used recurrently, they may be attached to other language units in such a way that the resulting formations resemble affixations rather than typical blends. From this perspective, the characteristics of typical blends can be regarded as the result of an interplay between factors determining creativity and productivity of the word formation pattern, as visualized in Figure 6.

Figure 6: Deviations from prototypicality in blends in relation to creativity and productivity

In conclusion, the predictability of blend formation and the characteristics of typical blends cannot be ignored in a study of blending. However, the criteria of well-formedness may change for splinters that are used recurrently. What induces recurrent use remains an open question but since any splinter is, at least potentially, subject to recurrent use, it is important to take into consideration factors that may be associated with it. As shown in the present study, recurrent use of splinters may also involve less overlap and greater prosodic flexibility. At the same time, many items containing recurrently used splinters bear characteristics observed in typical blends. Thus, many of the affix-like concatenations discussed in Section 3.4 still preserve the prosodic contour of one of their source words (as was shown in Figure 4, see also discussion in Bauer, Beliaeva & Tarasova [2019]). It may be the case that for certain splinters the preservation of prosodic contour and therefore prosodic restrictions on the items the splinters may be concatenated with are more relevant than for the others, and factors involved in this are worth investigating in future studies.

BIBLIOGRAPHY

ADAMS Valerie, 2001, Complex words in English, Harlow: Longman.


BAUER Laurie, 1998, “Is there a class of neoclassical compounds, and if so is it productive?” Linguistics, 36,3, 403–422.


MERWIN Hugh, 2013, May 9, “Introducing the cronut, a doughnut-croissant hybrid that may very well change your life”, Grub Street, http://www.grubstreet.com/2013/05/dominique-ansel-cronut.html


RENNER Vincent, 2015, “Lexical blending as wordplay”, in ZIRKER Angelika & WINTER-FROEMEL Esme (Eds.), Wordplay and metalinguistic / metadiscursive reflection: Authors, contexts, techniques, and meta-reflection, 119–133.
The formation of blends is predictable to a large extent. Moreover, it may give rise to productive constructions where part of a word that was once blended is used to form further blends such e.g. -cation (from vacation) in workcation, spa-cation, staycation, etc. However, not all blends demonstrate such potential and factors that can influence the productivity of blend splinters are worth investigating. This paper argues that blending, as a word formation process, can be driven by either factors enhancing predictability of the outcome or factors enhancing playful character of the formation, or both. Both qualitative and quantitative data analysis is used to investigate the relationship between well-formedness of the blends and their productivity potential.

La formation de mots-valises est prévisible dans une large mesure. De plus, elle permet de créer des constructions productives dans lesquelles une partie d’un mot qui a déjà été amalgamé est utilisée pour former d’autres amalgames tels que, par exemple -cation (de vacation) dans workcation, spa-cation, staycation, etc. Cependant, tous les mots-valises ne présentent pas un tel potentiel et il s’agit de prendre en considération les facteurs qui peuvent influer sur la productivité des fragments de mots-valises. Cet article avance l’hypothèse que la formation des mots-valises peut être dictée par des facteurs améliorant la prévisibilité du résultat ou par des facteurs renforçant le caractère ludique de la formation, ou les deux. L’analyse des données qualitatives et quantitatives est utilisée pour étudier la relation entre la forme des mots-valises et leur potentiel de productivité.
INDEX

Mots-clés: mots-valises, amalgamation, linguistique de corpus, productivité, création des mots, formation des mots

Keywords: blends, corpus linguistics, productivity, word creation, word formation

AUTHOR

NATALIA BELIAEVA

Victoria University of Wellington, School of Linguistics and Applied Language Studies, New Zealand
natalia.beliaeva@vuw.ac.nz