Quantitative corpus data on blend formation: psycho- and cognitive-linguistic perspectives

Stefan Th. Gries

On the basis of a heuristic characterization of intentional blending as a tripartite word-formation process, this paper discusses a variety of case studies concerned with the effects of similarity and recognizability on the formation of blends. More specifically, the case studies focus on (i) the degree of similarity between the two source words that are blended (on different levels of linguistic analysis), (ii) the ordering of source words in blends, and (iii) the ways in which source words are split up and merged into blends. The case studies include comparisons to supposedly related phenomena, viz. speech-error blends and complex clippings, and extend previous work by proposing new, or improving existing, corpus-linguistic operationalizations of relevant concepts and by increasing the sample sizes from previous studies.

Keywords: blends, complex clippings, cut-off point, similarity.

1. Introduction

Blending, the process that underlies the creation of *brunch* or *chunnel* from *breakfast* and *lunch* or *channel* and *tunnel* respectively, is one of the most perplexing word-formation processes, given that

- it is not as rule-governed as derivational processes;
- it is not as productive as most derivational processes;
- it is more creative than most derivational processes;
- it involves conscious effort and word play on the part of the coiner, which often results in “violations” of more rigid morphological rules and includes the “integration” of many kinds of information that are not central to linguistic study (e.g., the interplay between orthography and pronunciation);
- it nevertheless exhibits superficial similarity to other intentional word-formation processes (e.g., compounding, (complex) clipping, abbreviations, acronyms);
- it has an unplanned counterpart in the form of speech-error blends.

Given the interaction of all these characteristics, it comes as no surprise that some have adopted a somewhat pessimistic stance towards blends:

> "in blending, the blender is apparently free to take as much or as little from either base as is felt to be necessary or desirable. […] Exactly what the restrictions are, however, beyond pronounceability and spellability is far from clear." (Bauer 1983: 225)

> "we find no discernible relationship between phonology […] and a viable blend. […] This fact helps to make blends one of the most unpredictable categories of word-formation." (Cannon 1986: 744)

It is true: blends involve a mind-boggling degree of complexity, and the kind of (near-)categorical rules and processes we often find elsewhere in morphology are hard to come by. On the other hand, just because blends do not exhibit many, if any, categorical rules does not mean that blends are unpredictable. In fact, most, if not all, linguistic phenomena are not categorical in nature, but probabilistic and multifactorial – and so are blends. We should therefore adopt a probabilistic approach to the analysis of blends and their structure, but we need larger samples than those studied in some of the classic studies (e.g., 314 in Pound 1914, 132 in Cannon 1986) and statistical methods that can handle probabilistic distributions better than intuition or hunches alone.

For the purposes of this paper, I will adopt the following relatively uncontroversial definition of intentional lexical blends: an intentional fusion of typically two (but potentially more) words where a part of a first source word (sw_1) – usually this part includes the beginning of sw_1 – is combined with a part of a second source word (sw_2) – usually this part includes the end of sw_2 – where at least one source word is shortened and/or the fusion may involve overlap of sw_1 and sw_2. This definition is intended to distinguish such blends from speech-error blends, which are not intentional even though they may be experimentally induced, and complex clippings, which involve the concatenation of the beginnings of two source words. The data to be studied currently include 2329 formations; however, not all of them have already been annotated with regard to all the parameters that will be discussed below.

In what follows, I will present several case studies regarding what one might informally consider the three interrelated "temporal stages" of intentional blending: the selection of the source words to be blended, the (related) decision for a particular order of these source words in the blend, and the (related) decision of how exactly to split up the words and blend them. This division of blending into three different stages is of course somewhat artificial and not intended as a characterization that is isomorphic to the actual psycholinguistic processes, but it is nevertheless a convenient heuristic to approach the phenomenon. The case studies to be discussed here involve (non-standard) elements from many different levels of linguistic analysis:

– graphemes and phonemes;
– graphemic and phonemic *n*-grams;
– syllables and their stress patterns;
– words, their lengths, frequencies (and semantics).

Crucially, I will argue and exemplify that the study of these aspects of intentional blends requires that blends be compared to other intentional word-formation processes as well as (randomly generated) baselines. In addition, I will argue that intentional blend formation involves an interplay of phonemic/graphemic/… similarity on the one hand as well recognizability on the other. Some of the case studies to be discussed are replications of previous work on the basis of a now larger data set, but others will be new or, maybe even more importantly, illustrate that the field is still at a stage where methodological fine-tuning is required by demonstrating that not all previous studies, including my own, have succeeded in operationalizing the relevant parameters optimally.

2. The selection of the source words

When it comes to selecting the words to be blended, a variety of studies have shown that the source words speakers choose to blend are similar to each other. This is true of speech-error blends' phonological characteristics (MacKay 1987; Kubozono 1990), syntactic/POS characteristics (MacKay 1987; Berg 1998), and their semantic characteristics (Levelt 1989; Berg 1998), but it is also true of the intentional blends focused on here most (Kubozono 1990; Kelly 1998; Gries 2004a–c). However, there are many ways in which words can be similar to each other (lengths (of different types of units), frequency/dispersion, phonemes, graphemes, syllables, stress patterns, semantics, etc.) and different ways in which each of these similarities can be operationalized. In addition, when it comes to, say, phonemic/graphemic similarity, source words may be similar to each other in different parts of the words. Finally, similarity measures of all of the above kinds must always be compared against expected/random baselines so as to make sure that whatever similarity value is obtained is not squarely within the range of chance values. In this paper, I will report on several small case studies, each of which is concerned with a different facet of word similarity and supersedes earlier work on this in terms of the amount of data covered and/or in terms of how the data are studied and evaluated.

2.1. The lengths of source words

First, let us explore the lengths of source words of three different kinds of blends: authentic error blends (i.e., unplanned lapses that happened to have been overheard and that were quoted in psycholinguistic studies), induced error blends (i.e., unplanned lapses that were induced in published experimental studies) and intentional word-formation blends (compiled from published studies as well as by the author). Previous studies have argued that, in intentional blends, sw1 is shorter than sw2 (cf. Kelly 1998; Gries 2004c) while, in authentic errors, the reverse tendency was obtained (cf. MacKay 1973 on German error blends). To replicate these findings, I counted for both source words the numbers of syllables, phonemes and graphemes (for intentional blends only) and compared their average lengths (as independent samples); the results are represented in the three panels of Table 1.

Table 1. Medians and interquartile ranges (in parentheses) of lengths of blend types

	authentic error blends (186)		
	syllables	phonemes	letters
sw_1	2 (1, 2)	5 (4, 7)	
sw_2	2 (1, 2)	5 (4, 6)	
$p_{\text{U-test}}$	>0.79	>0.47	
	induced error blends (32)		
	syllables	phonemes	letters
sw_1	3 (2, 4)	8 (6, 9)	
sw_2	3 (2.75, 4)	8 (6.75, 9)	
$p_{\text{U-test}}$	>0.92	>0.97	
	intentional blends (1921)		
	syllables	phonemes	letters
sw_1	2 (1, 3)	5 (4, 7)	6 (4, 8)
sw_2	2 (3, 3)	7 (5, 8)	7 (6, 9)
$p_{\text{U-test}}$	<0.001	<0.001	<0.001

The lengths of the source words of both kinds of error blends are not significantly different from each other (when averaged across blends), which is not compatible with MacKay's above finding. On the other hand, the source words of intentional blends behave differently from the source words of errors and exhibit a significant difference that is compatible with earlier findings: in every comparison, sw_1 is shorter than sw_2.

2.2. The frequencies of source words

A similar picture emerges from the second case study, which is concerned with the frequencies of the source words. Kelly's (1998) data on intentional blends suggested that in intentional blends sw_1 is more frequent than sw_2 whereas MacKay's (1973) data on (German) error blends suggest that the two source words do not differ with regard to their frequencies. The current data yield compatible results in independent-sample comparisons. For each source word of the three blend types, I retrieved its frequency in the Reuters corpus, a corpus of 800,000+ newswire stories (cf. <http://trec.nist.gov/data/reuters/reuters.html> and Lewis et al. 2004). The results are shown in the three panels in Table 2: as expected, the source words of both types of error blends do not differ significantly whereas in intentional blends sw_1 is significantly more frequent than sw_2.

Table 2. Medians and interquartile ranges (in parentheses) of $\log_{10}$ frequencies of blend types

	authentic error blends (186)
sw_1	3.05 (2.2, 4.05)
sw_2	3.11 (1.84, 4.09)
$p_{\text{U-test}}$	>0.7
	induced error blends (32)
sw_1	2.68 (1.57, 3.33)
sw_2	2.8 (1.09, 3.24)
$p_{\text{U-test}}$	>0.88
	intentional blends (1939)
sw_1	2.91 (1.96, 3.68)
sw_2	2.55 (1.54, 3.48)
$p_{\text{U-test}}$	<0.001

2.3. The overall similarity of source words to each other

The third case study is concerned with the phonemic similarities of source words of blends to each other. As mentioned before, previous work (e.g., Gries 2004b-c, 2006) has discovered that source words of intentional and error blends are more similar to each other than expected by chance and that the source words of blends are more similar to each other than those of complex clippings. However, these studies have not compared different kinds of blends to each other or, if they did, only used a bigram-based measure ((X) Dice). In this study, I will use a more sophisticated measure – string-edit distance – and study a larger set of blends to try to replicate the previous findings. For all pairs of source words of the three types of blends already studied above, I computed phonemic Levenshtein string-edit distances, which is a metric based on the number of operations (insertions, deletions, and substitutions) needed to change one word into another. As a control condition, I also included all pairwise similarities of 100 randomly sampled words' pronunciations from the CELEX database. The distributions of the string-edit distances are represented in Figure 1 in the form of cumulative distribution functions.

As one would have hoped for, the similarities of the random word pairs is smallest (because their distances are largest), and Kolmogorov-Smirnov tests show that they differ significantly from the other source word pairs (all these p's<0.03). Also, the source words of authentic error blends exhibit the highest similarity to each other, which is compatible with previous findings. The source words of intentional blends occupy a middle ground, and while nearly all curves differ significantly from each other (all these p's<0.006), the source words of intentional blends do not differ significantly from those of induced errors (p>0.89). In other words, the experimenters' choices of source words of induced errors were very different from the authentic errors they were supposedly intended to represent, but much more like the intentional blends, which most other case studies show are quite different from errors. This is yet another piece of evidence cautioning us to be very careful about generalizations regarding blending that are based on experimentally induced errors.

2.4. The locus of similarity of source words

Another question that arises in this context is where the source words are similar to each other. For example, it is obvious that the similarity of *channel* and *tunnel* is strongest at the end of the words, but it is less obvious whether

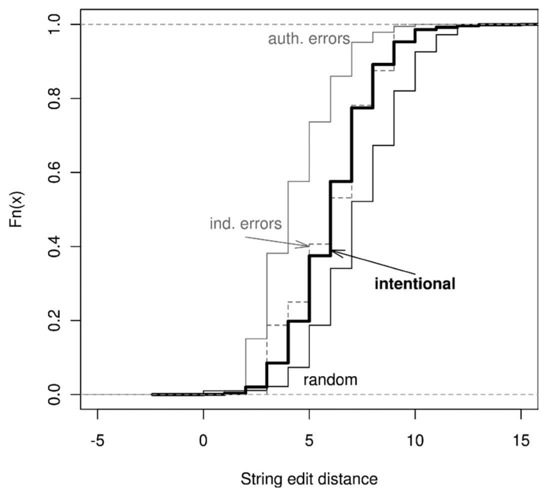

Figure 1. Cumulative distribution functions of similarities of source words of three types of blends as well as 4950 random word pairs

different kinds of blends behave alike in this regard. Gries (2004c) explored this issue, but only in a coarse-grained fashion, namely by checking whether the overlap of source words occurred only/mainly around the breakpoint or not. His results suggested that the similarity of error blends was more global than that of intentional blends. In this paper, I will revisit the issue in more detail.

For all authentic errors, intentional blends and complex clippings, I

- generated all possible substrings of the two source words (i.e., for *tunnel* this yields {*t*, *u*, *n*, *n*, *e*, *l*, *tu*, *un*, *nn*, *ne*, *el*, *tun*, *unn*, *nne*, *nel*, *tunn*, *unne*, *nnel*, *tunne*, *unnel*, *tunnel*};
- computed a position index for each substring that indicates on a scale from 0 to 1 where in the word the substring is located. For example, the trigram substrings *tun*, *unn*, *nne*, and *nel* of *tunnel* received the position indices 0, $^1/_3$, $^2/_3$ and 1 respectively;
- determined for all substrings of a source word whether they occurred in the other source word, too, and for those substrings that did overlap, I computed the mean of their position indices, which therefore indicates where the two words share most of their material.

For the comparison of *channel* to *tunnel*, this resulted in the value of 0.815, i.e., the end part of the word, namely where "nnel" is located. Then, I compared these values for all authentic error blends, all intentional blends and

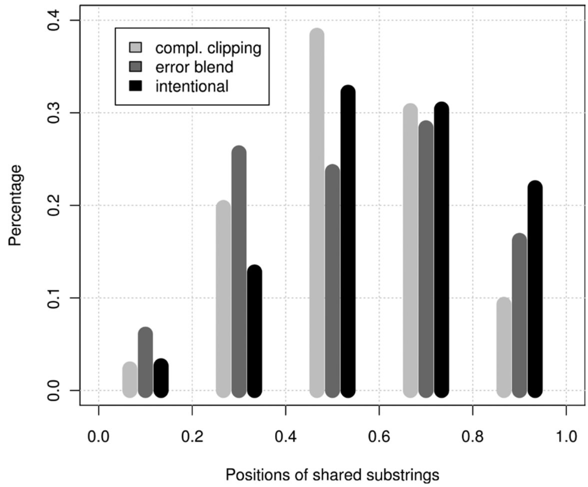

Figure 2. Mean within-word locations of shared substrings (between 0=beginning and 1=end) of complex clippings, authentic error blends and intentional blends

all complex clippings and plotted the results for all source words that are not completely dissimilar into Figure 2. In this graph, results for complex clippings, authentic error blends and intentional blends are indicated in light grey, dark grey and black respectively, such that the position in source words is represented on the *x*-axis (i.e., the beginnings and ends of source words are at $x=0$ and $x=1$ respectively) and the bars indicate on the *y*-axis the percentages of blends that exhibit a particular mean position index.

The results show that, with regard to where their source words are similar, intentional blends differ from error blends (whose similarity of the source words is more widespread across the words while that of the source words of intentional blends is more narrowly concentrated around the middle and end of the word: $D=0.1159$, $p=0.0015$) as well as from complex clippings (whose similarity is located earlier in the words than in intentional blends; $D=0.1293$, $p=0.0087$).

2.5. The stress patterns of source words

As was shown above in Section 2.1, source words of intentional blends are similar to each other in terms of their syllabic lengths. However, the similarity goes even further. On the basis of a smaller data set than in the present

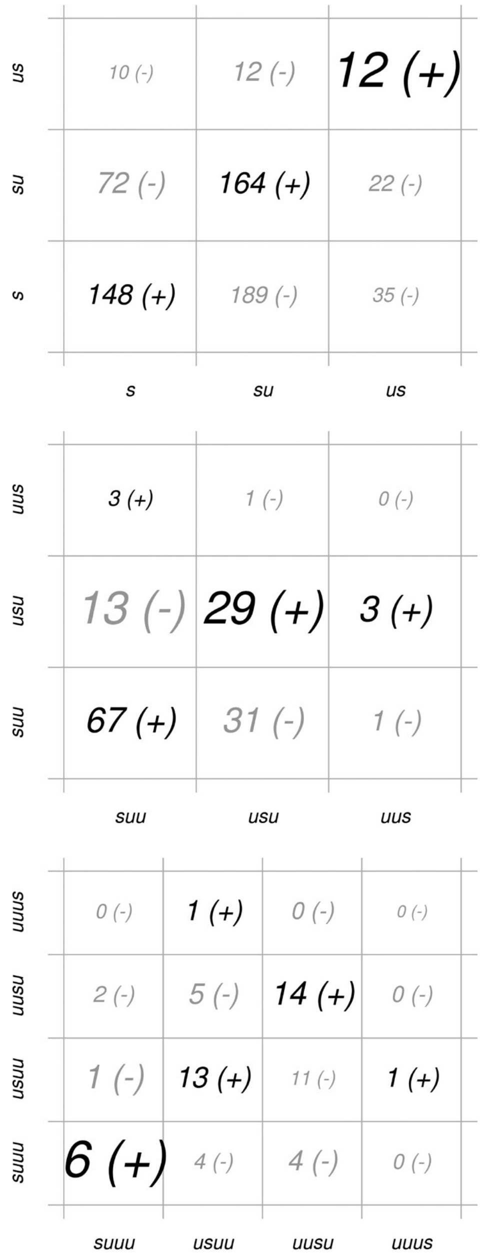

Figure 3. Cross-tabulation plots for source words' stress patterns: 2-syllable source words (top panel, $p_{\text{exact test}} < 0.001$), 3-syllable source words (center panel, $p_{\text{exact test}} < 0.001$), 4-syllable source words (bottom panel, $p_{\text{exact test}} < 0.02$)

article, Gries (2004b) has illustrated how source words of intentional blends that have the same numbers of syllables also tend to have the same stress patterns. To test this result on the basis of the present larger data set, in addition to the syllabic lengths of both source words from above I also identified the stress patterns of the words, i.e., whether each syllable had primary stress ("s") or not ("u"). For example, *webinar* (*web* * *seminar*) was annotated as "s" for *web* and "s-u-u" for *seminar*, *jokelore* (*joke* * *folklore*) as "s" and "s-u", *transponder* (*transmission* * *responder*) as "u-s-u" and "u-s-u", etc. Then, for each set of blends consisting of two equally long source words, I counted which stress patterns of sw_1 were attested with sw_2. The results are represented in the three panels of Figure 3 by means of cross-tabulation plots (cf. Gries 2009). These plots are essentially frequency tables, but frequencies that are larger than expected by chance are printed in black and followed by "(+)", frequencies that are smaller than expected by chance are printed in grey and followed by "(-)", and the physical font size directly represents the size of the deviation of the observed frequencies from the expected ones (based on Pearson residuals).

As is obvious from the overrepresented numbers in the diagonal from the bottom left to the top right corner, when the source words of intentional blends have the same number of syllables, then there is also a significant tendency for them to have the same stress pattern.

2.6. The semantics of source words

In this section, I will briefly (and preliminarily) explore the semantic relationships between the source words of authentic error blends, induced error blends, intentional word-formation blends and complex clippings. In a first analysis of a subset of the whole data set, the semantic relationships between the pairs of source words of 647 forms were classified into the following categories, most of which have been mentioned in previous studies:

- synonymy, as for *deliberal* (*deliberate* ∗ *intentional*), *redupeat* (*reduplicate* ∗ *repeat*), or *tummach* (*tummy* ∗ *stomach*);
- co-hyponymy, as for *magalogue* (publications: *magazine* ∗ *catalogue*), *beefcake* (food items: *beef* ∗ *cheesecake*), or *Frenglish* (languages: *French* ∗ *English*);
- contractive, i.e., when the blend contracts two source words that would have been adjacent as in a compound, as for *carjacking* (*car* ∗ *hijacking*), *skurfing* (*sky* ∗ *surfing*), or *scifi* (*science* ∗ *fiction*);

– frame relation, as for *confrotalk* (*confrontation* ∗ *talk*), *letterzine* (*letter* ∗ *magazine*), or *riverscape* (*river* ∗ *landscape*);
– other (e.g., antonymy, derivation, etc.).

The cross-tabulation of the four types of processes and the five semantic relationships is represented in Figure 4. As is obvious, the data differ highly significantly from chance (χ^2=211.07; *df*=12, *p*<0.001, Cramer's *V*=0.33) and three very clear classes emerge: error blends are characterized very much by their source words being synonyms, and the intentional error blends, where of course researchers chose the source words, unsurprisingly exhibit the same tendency. Intentional blends, on the other hand, involve very many different semantic relationships: while synonyms occur less frequently than expected, they do occur, and all other semantic relationships coded for here are more frequent than expected. Finally, complex clippings have quite a strong preference to involve contractive relations. In other words, intentional blends differ markedly from the other categories, viz. both types of errors and complex clippings.

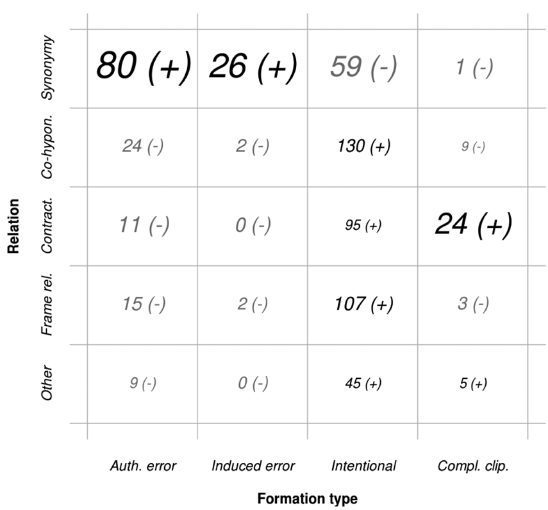

Figure 4. Cross-tabulation plots for source words' semantic relations

2.7. Interim summary

The findings shed some doubt on previous explicit claims and implicit assumptions. Not only do authentic error blends sometimes behave very differently from induced error blends (for instance, with regard to their phonemic similarity), both types of error blends sometimes also differ significantly from intentional error blends (for instance, with regard to their source words' lengths, frequencies, their distribution of similarity over words, and the semantic relation between the source words). Thus, experimental findings from induced error blends may not fully apply to naturalistic errors, and, likewise and contra Kubozono (1990) and Berg (1998), findings from and conclusions based on error blends may not fully apply to intentional errors.

3. The ordering of the source words

Section 2 has shown that the source words that enter into blends are rather similar to each other in a variety of ways. However, especially for intentional blends of course, once the blend coiner has decided on the words to be blended, the question arises of how they are ordered – which source word's contribution comes first and which comes second. Unsurprisingly, one obvious determinant of this process is semantic headedness, as demonstrated by Renner (2010). In this section, I will very briefly explore that question by returning to two variables already explored – lengths and frequencies of source words –, but while the above comparison of sw_1 and sw_2 has lumped all source words in positions 1 and 2 together, here I will perform pairwise – in other words, blendwise – comparisons of both source words.

3.1. Pairwise source word comparisons (in terms of lengths)

In Section 2.1 above, we have seen that the median length of sw_1 of error blends is not significantly different from the median length of sw_2 regardless of how we measure length. However, the non-pairwise computation may mask effects on the level of the individual words and in this section we will therefore compare the lengths of the source words pairwise, i.e., for each error blend. More specifically, for all authentic error blends, I computed the differences length sw_1 minus length sw_2 and sorted them by their size, and I did this for their syllabic and phonemic lengths. Then, these sorted results were plotted for each type of length such that the sorted differences are represented on the *y*-axes; all differences smaller/greater than 0 – i.e., where sw_1 is shorter/longer than sw_2 – were plotted in black/grey respectively, and a paired Wilcoxon test was computed to determine whether or not any differences were significant. The results for the authentic error blends are represented in the panels of Figure 5.

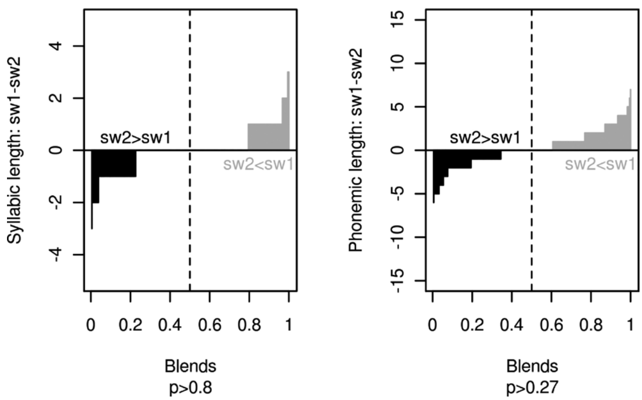

Figure 5. Sorted pairwise length differences for authentic error blends: syllabic lengths (left panel), phonemic lengths (right panel)

It is obvious that the pairwise differences are relatively symmetrically distributed around 0, and the *p*-values of the pairwise Wilcoxon tests show that the source words of error blends do not differ significantly; the same insignificant results were obtained with blends involving synonyms only, where headedness might well play less of a role (thanks to Laurie Bauer for suggesting this test). But what about the intentional blends, where one also needs to include graphemic lengths? The results of the analogous computations are represented in Figure 6. The plots already indicate what the *p*-values then confirm: In all three comparisons, sw_2 turns out to be highly significantly longer than sw_1, namely about half a syllable and one phoneme or grapheme; the same significant results were obtained with blends involving synonyms only.

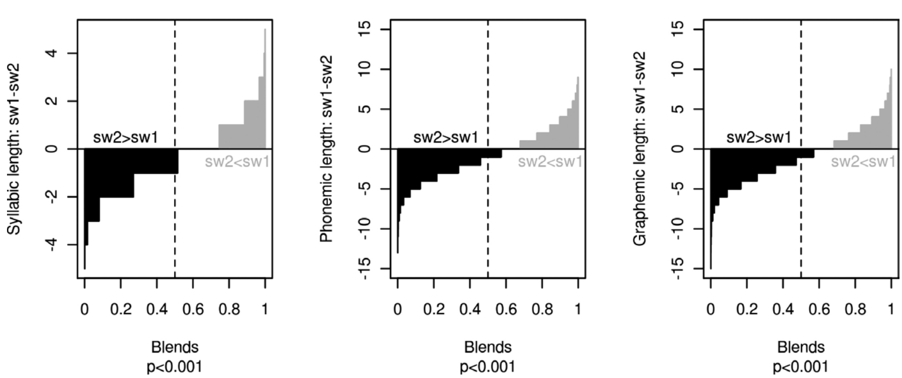

Figure 6. Sorted pairwise length differences for intentional blends: syllabic lengths (left panel), phonemic lengths (center panel), graphemic lengths (right panel)

3.2. Pairwise source word comparisons (in terms of frequencies)

In the same way that Section 3.1 discussed the pairwise test of data from Section 2.1, here, I will very briefly summarize the results of the pairwise comparisons of source word frequencies for authentic error blends, induced error blends and intentional blends. Consider Figure 7.

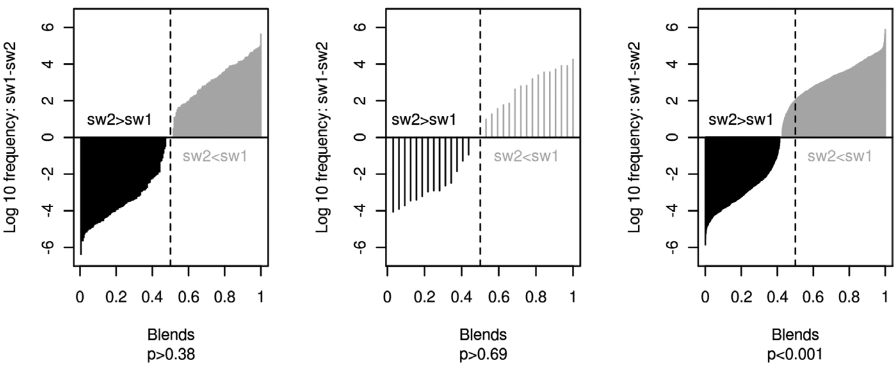

Figure 7. Sorted pairwise frequency differences for authentic error blends (left panel), induced error blends (center panel), intentional blends (right panel)

The results from the pairwise comparison partially support the coarser approach from above: in error blends, the source words are equally frequent, but in intentional blends, sw_2 is significantly less frequent. (Since frequencies are sometimes misleading (cf. Gries 2008), I also tested whether the source words of the three types of blends differed in terms of their dispersion, but the results from the above frequency tests were supported.)

3.3. Interim summary

The results of the pairwise comparisons of source words regarding the ordering of the source words are unambiguous and are compatible with the more coarse-grained results concerned with their selection. For error blends, sw_1 and sw_2 do not differ with regard to their lengths and frequencies while for intentional blends, sw_1 is shorter and more frequent (and more evenly dispersed) than sw_2. As above, this is in contrast to some previous studies and should caution us against prematurely lumping error blends and intentional blends together.

4. The blending of the source words

The final aspect of blending to be considered here is how two source words are blended after they have been chosen (which we dealt with briefly in Section 2) and after their ordering has been decided on (which we dealt with briefly in Section 3). (Again, this division into stages is no more than a convenient heuristic; no psychological/psycholinguistic significance is attached to it.)

4.1. The similarity of source words to the blend

In previous work, I argued that blends are coined under the influence of two opposing factors, similarity and recognizability, and the previous sections have already dealt with similarity of source words on various levels. The question that may arise, however, is whether similarity and recognizability would be opposing rather than correlated factors. The answer to this question is that similarity as I use the term here is concerned with the similarity of the source words to each other while I use recognizability to refer to how recognizable the source words are from the blend. This is of course also a function of the similarity of the blend to the source words, but in a different way. This is because, obviously, the two source words of a blend would be most recognizable if nearly all of their material were present in the blend, as in the examples in (1).

(1) a. *Chevrolet* * *Cadillac* → *Chevrolecadillac*
    b. *Cadillac* * *Chevrolet* → *Cadillachevrolet*

However, such blends are extremely unlikely – especially if they do not involve any overlapping sounds/letters – because they are not fun anymore: while both source words are perfectly recognizable, the blend is not too similar to either source word anymore and the punning/playful character of such blends is largely lost. Ideally, we would have a way to quantify the degree to which a blend strikes a balance between these two forces.

In previous work (e.g., Gries 2004b), I proposed a similarity index *SI* to quantify the similarity of the two source words to the blend. This *SI*, as computed in (2), was intended to be high / close to 1 if both source words were highly similar to the blends and low / close to zero if the source words were not similar to the resulting blend anymore (*el.* stands for *elements*).

(2) $$0.5\left(\frac{\frac{number\ of\ el.of\ sw_1\ into\ blend}{number\ of\ el.of\ sw_1}\times}{\frac{number\ of\ el.\ of\ blend\ out\ of\ sw_1}{number\ of\ el.of\ blend}} + \frac{\frac{number\ of\ el.of\ sw_2\ into\ blend}{number\ of\ el.of\ sw_2}\times}{\frac{number\ of\ el.\ of\ blend\ out\ of\ sw_2}{number\ of\ el.of\ blend}}\right)$$

For example, $^6/_7$ letters of *channel* make up $^6/_7$ letters of *chunnel*, and $^5/_6$ letters of *tunnel* make up $^5/_7$ letters of *chunnel; SI_G chunnel* = 0.665. This seemed like a good idea at the time because, for instance, it made sense intuitively that *chunnel* should score a high value, that *brunch* should score relatively low (SI_G *brunch* = 0.304), and that the hypothetical blend *breakfunch*, in which *breakfast* is much more recognizable than in *brunch*, should score a better value that *brunch* (SI_G *breakfunch* = 0.36). However, while the initial results discussed in Gries (2004b) were encouraging, it turns out there are also more problematic results:

(3) a. *Chevrolet * Cadillac* → *Chevrolac*
(SI_G = 0.414)
b. *Cadillac * Chevrolet* → *Cadillet*
(SI_G = 0.344)

(4) a. *Chevrolet * Cadillac* → *Chevrolecaddillac*
(SI_G = 0.472)
b. *Cadillac * Chevrolet* → *Cadillachevrolet*
(SI_G = 0.531)

It is reasonable to assume that *Chevrolac* and *Cadillet* would be better blends than those in (4), but the *SI*-values do not reflect that sufficiently well. Thus, obviously, the proposed *SI* captures some of what is going on in terms of recognizability, but it does not penalize enough blends which, while boosting recognizability, are too dissimilar to their source words. I would therefore like to propose to instead use the average of the Levenshtein string edit distances (cf. Левенштейн 1966) between both source words and the blend (ASED). This measure provides results that are more compatible with what speakers' intuitions would suggest, as is obvious from (5) to (7).

(5) a. *channel * tunnel* → *chunnel*
(ASED = 1.5)

b. *channel* → *chunnel* 1 (1 substitution)
tunnel → *chunnel* 2 (1 del. + 1 subst.)

(6) a. *Chevrolet* * *Cadillac* → *Chevrolac*
(ASED = 3.5)
b. *Cadillac* * *Chevrolet* → *Cadillet*
(ASED = 3.5)

(7) a. *Chevrolet* * *Cadillac* → *Chevrolecadillac*
(ASED = 8)
b. *Cadillac* * *Chevrolet* → *Cadillachevrolet*
(ASED = 7.5)

That is, as expected, *chunnel* scores a low value (for high similarity, cf. (5)), *Chevrolac* and *Cadillet* score slightly higher values (for less similarity, cf. (6)), and *Chevrolecadillac* and *Cadillachevrolet* score even higher values (for much less similarity, cf. (7)). Note that ASED is also preferable theoretically since we can then use the same type of measure – string-edit distances – for both comparisons: source word to source word (cf. Section 2.3) and source words to blends (here).

To replicate the previous work with the new and arguably more appropriate similarity measure, I computed ASEDs for all error blends in my data set (both authentic and induced), intentional blends and complex clippings, as well as 144 simulated blends, which were created by blending six words in all phonologically possible ways. The results were highly significant according to an ANOVA (adjusted R^2=0.17; $F_{4,\,2394}$=121.6; p<0.001) and are summarized in Figure 8.

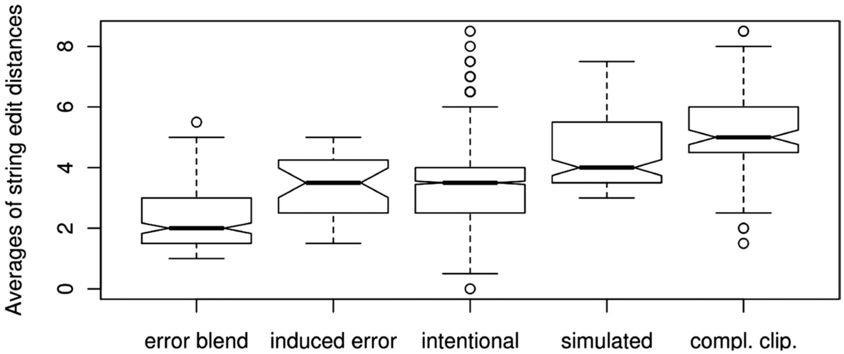

Figure 8. ASEDs as a function of the formation process

The results are very clear and interesting: authentic error blends are most similar to their source words. The two blend types which involve conscious choices and blending of source words – induced errors and intentional blends – involve significantly less similarity, which is another piece of evidence for the fact that, even though researchers choose source words for induced error blends that are comparable to authentic errors in terms of their lengths, frequencies, and semantic relations, the way they are blended still results in somewhat different formations; the results from Figure 8 also lend additional *post hoc* support to the results from Figure 2 in Section 2.4 regarding the location of similarity: if the similarity is distributed more globally in the source words of error blends than in those of intentional blends, then one would expect that error blends can be characterized by a higher degree of similarity to their source words. The intentional blends on the other hand mostly exploit similarity in the middle of the word, where the source words are split up (often around overlapping material), and hence their overall similarity to their source words is lower, but still significantly higher than in the simulated blends, which served as a control group of sorts. The least amount of similarity by far is found between the source words of simulated blends and complex clippings, which supports previous findings testifying to how blends are different from complex clippings (cf., e.g., Gries 2006).

4.2. The point where source words are split up

The previous sections have shown that error blends, intentional blends and complex clippings differ in terms of how similar they are to their source words and how the similarity is distributed across words. What still requires further study is the question of where words are split up into parts. The most detailed study in this regard is perhaps Gries (2006), where I explored how blend coiners have a significant tendency to split up source words near their psycholinguistically defined uniqueness points, which in turn were operationalized corpus-linguistically. More precisely, the uniqueness point of a word *w* was approximated as a so-called selection point *SP*, which is the point after a part of *w* where *w* is the most frequent word with that part (in the British National Corpus). A comparison of intentional blends and complex clippings showed that coiners of intentional blends split up sw_1 nearly exactly at the selection point and sw_2 half a phoneme too early (on average) whereas coiners of complex clippings split up source words much earlier (and thus, less optimally than expected in terms of recognition). This was an interesting finding because it showed how psycholinguistically-motivated determinants can shed light on a complex multifactorial process that has long defied characterization.

Another interesting avenue for exploration, however, is where the splits are made in terms of phonological units. For example, do the different formation processes differ with regard to whether splits occur within or between syllables, at onset-rime or body-code divisions, etc.? In order to explore this question, I cross-tabulated split-point locations of both sw_1 and sw_2 for the annotated sections of error blends, complex clippings and intentional blends in my data set; consider Figure 9 for the results.

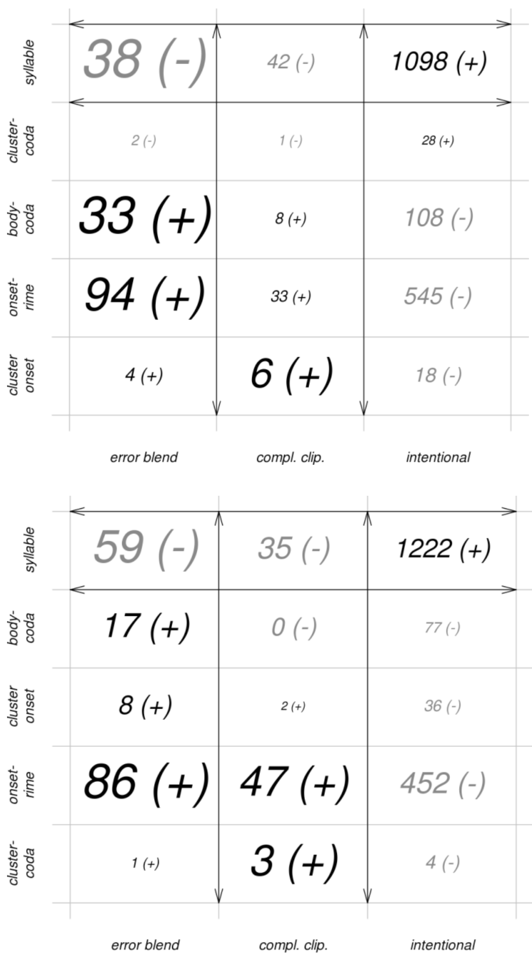

Figure 9. Split point locations for different formation processes: sw_1 (top panel) and sw_2 (bottom panel)

For sw_1, there is a significant difference from chance (χ^2=132.79; *df*=8, *p*<0.001, Cramer's *V*=0.18): error blends "prefer" splits between onset and rime as well as between body and coda, complex clippings involve surprisingly many splits in the onsets that are consonant clusters, and intentional error blends have a strong preference for syllable splits. On the whole, the results are similar for sw_2. Again, there is a significant distribution (χ^2=133; *df*=8, *p*<0.001, Cramer's *V*=0.18) and the only difference to the data for sw_1 is that, for sw_2, complex clippings prefer onset-rime splits much more than for sw_1. Nevertheless, it is very obvious that the three processes differ markedly from each other.

5. Concluding remarks

Given the results of the above case studies as well as those of previous studies, how does the production of intentional blends happen? A not totally serious but still heuristically useful answer would be the following:

- source word selection: a speaker chooses two source words which can communicate what the new formation is supposed to express and are similar to each other in terms of phonemic and/or graphemic length, stress pattern as well as semantics;
- source word ordering: the speaker orders the word so that he either leaves them in the modifier-head order in which they occur anyway or establishes some such structure himself or puts the shorter and more frequent word first;
- source word blending: the speaker cuts the words up at a syllable boundary close to the uniqueness/selection point, fuses them (and uses more of sw_2 in that process) so as to maximize overlap in the middle of the fusion section and maximize phonemic/graphemic similarity elsewhere as much as is still possible, and creates a blend that is more similar to sw_2, which often is the blend's head.

This process will of course be recursive: speakers will consider some words to blend because of their semantics, but will discard them because they are too dissimilar and provide less opportunity for punning overlap than alternative candidate source words. As a purely hypothetical example, the first speaker to ever say *foolosopher* may have discarded *moronosopher* or *moronopher* or *idiosopher* or *idiotosopher* because they result in stress clashes and less humorous exploitation of phonological similarity. This schematic process also indicates how different intentional blends are from both complex clippings and authentic error blends, which in turn are somewhat different from induced error blends. Occasional similarities between the four processes must not detract from the fact that for nearly every parameter that was studied here significant differences were obtained and, thus, make a strong case for treating all four processes in their own right.

While we are beginning to develop a better understanding of how blending functions, many more steps and improvements are necessary. Some of these improvements are quite obvious: we need larger databases of blends and we need more comprehensive descriptions of the patterns they exhibit. Somewhat less trivially, we need a variety of better methods which in turn will improve our descriptions. Given the key role played by similarity and recognizability, most pressing is the need for more flexible and more comprehensive measures of word similarity. More specifically, everything that has been done so far focused on one particular level of resolution: phonemes, graphemes, syllables, and so on. However, this is obviously not how speakers perceive words – naïve speakers have a much more holistic approach, which is why we need measures that allow us to capture and quantify similarity at many different levels at the same time. Consider again *channel* and *tunnel*:

– with regard to articulatory features, the two are quite similar even when the phonemes are not the same (as in their first sounds, /ʧ/ vs. /t/);
– with regard to phonemes and graphemes, they are similar, and they can be blended so that much of that similarity is preserved;
– with regard to CV segments, they are identical – CVCVC –, and they can be blended so that their similarity is preserved;
– with regard to syllable length, they are identical, and they can be blended so that much of their similarity is preserved;
– with regard to stress patterns, they are identical, and they can be blended so that much of that similarity is preserved.
– with regard to their parts of speech, they are identical; etc.

But how do we integrate the information from all these levels? And how do we treat cases where the two source words have different syllabic lengths but both are stressed on the first syllable? And what do we do with unstressed vowels such as the vowel before the /n/ in *impostinator* – is this a /ə/ or a /ɪ/ or, since both will be completely unstressed here, do we even want to consider them separately? These are all tricky and as yet unanswered questions, but once we come closer to answering them, we will be able to make better comparisons between the source words of blends (and similar word-formation processes) as well as the similarities of source words to blends, to determine more precisely what the contributions of source words are (which is especially difficult when we turn to discontinuous contributions of source words, as in *ambidextrous* * *sex* → *ambisextrous* or *carnibbleous* * *carnivorous* → *nibble*), to address more comprehensively the location of cut-off points, etc. One of the very first steps that could be taken is to extend the computation of Levenshtein distances such that they can involve mappings of phonemes/graphemes to handle phoneme/grapheme alternatives, but this is of course only the very first of many steps.

On a more general level, I think we need to leave behind purely descriptive linguistic accounts and turn to psycholinguistic concepts, notions and methods instead. While my previous work and the above have hopefully already constituted first steps in that direction, much more needs to be done. With regard to corpus-based approaches, notions such as neighborhood density – which is related to my approach to recognition/uniqueness points from above – will be useful to explore how source words are chosen from a candidate set and how they are blended. With regard to experimental approaches, it would be interesting to have speakers coin blends of source words while controlling for many of the factors known to influence blending. Theoretically, this may also lead to, or at least underscore the need for, a more flexible approach to the taxonomy/classification of word-formation processes, quite possibly a prototype approach of the type argued for by López Rúa (2002).

In spite of the huge task still ahead of us, one thing is already safe to say: blends are far from unpredictable and their characteristics are not only identifiable using larger databases, reference corpora and statistical techniques (including baseline comparisons), but also firmly grounded in cognitive, but ultimately psycholinguistic and probabilistic mechanisms.

Acknowledgements

I thank the audience at the Conference on Lexical Blending in Lyon, the editors, and especially Laurie Bauer for feedback and comments. The usual disclaimers apply.

References

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

Berg, Thomas. 1998. *Linguistic Structure and Change: An Explanation from Language Processing*. Oxford: Oxford University Press.

Cannon, Garland. 1986. Blends in English word-formation. *Linguistics* 24 (4): 725–753.

Gries, Stefan Th.. 2004a. Shouldn't it be *breakfunch?* A quantitative analysis of the structure of blends. *Linguistics* 42 (3): 639–667.

Gries, Stefan Th. 2004b. Isn't that fantabulous? How similarity motivates intentional morphological blends in English. In Michel Achard and Suzanne Kemmer (eds.), *Language, Culture, and Mind*, 415–428. Stanford, CA: CSLI.

Gries, Stefan Th. 2004c. Some characteristics of English morphological blends. In Mary A. Andronis, Erin Debenport, Anne Pycha and Keiko Yoshimura (eds.), *Papers from the 38th Regional Meeting of the Chicago Linguistics Society: Vol. II. The Panels*, 201–216. Chicago, IL: Chicago Linguistics Society.

Gries, Stefan Th. 2006. Cognitive determinants of subtractive word-formation processes: A corpus-based perspective. *Cognitive Linguistics* 17 (4): 535–558.

Gries, Stefan Th. 2008. Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics* 13 (4): 403–437.

Gries, Stefan Th. 2009. *Statistics for Linguistics with R*. Berlin & New York: Mouton de Gruyter.

Kelly, Michael H. 1998. To *brunch* or to *brench*: Some aspects of blend structure. *Linguistics* 36 (3): 579–590.

Kubozono, Haruo. 1990. Phonological constraints on blending in English as a case for phonology-morphology interface. *Yearbook of Morphology* 3: 1–20.

Levelt, Willem J.M. 1989. *Speaking: From Thinking to Articulation*. Cambridge, MA: The MIT Press.

Левенштейн, В. И. [Levenshtein, V.I.]. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10: 707–710.

Lewis, David D., Yiming Yang, Tony G. Rose and Fan Li. 2004. RCV1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research* 5 (April): 361–397.

López Rúa, Paula. 2002. On the structure of acronyms and neighbouring categories: A prototype-based account. *English Language and Linguistics* 6 (1): 31–60.

MacKay, Donald G. 1973. Complexity in output systems: Evidence from behavioral hybrids. *American Journal of Psychology* 86 (4): 785–806.

MacKay 1987

Pound, Louise. 1914. *Blends: Their Relation to English Word-formation*. Heidelberg: Winter.

Renner, Vincent. 2010. *A contrastive account of blends in the lexical stocks of French and English*. Paper presented at the Sixth International Contrastive Linguistics Conference. Free University of Berlin, 30 Sept.–2 Oct. 2010.