A corpus study of phonological factors in novel English blends[*]

Michael Dow
Université de Montréal

1.   Introduction

A commonly cited limitation on the study of peripheral, extragrammatical and/or ludic processes in morphophonology is the relative difficulty of automated retrieval of their outputs. Whereas machines can be trained to detect and extract products of core word-building processes, thereby facilitating the construction of large corpora, the forms of more informal processes often must be compiled manually. (See, for instance, Gries's (2004a) English corpus of 585 forms touted as "one of the largest blend corpora analyzed so far.") The relative unpredictability of these forms is of direct interest but at the same time complicates large-scale investigations and may thus lead to taxonomies and/or theories based on comparatively few data.

Such processes, moreover, are not of purely empirical interest. In particular, they may both evidence speakers' implicit knowledge of phonological constraints or structures, including those inactive in their native language, and allow us to argue for certain theoretical constructs over others. As an example of the former, Tessier and Becker (2018) find that a nascent class of insult compounds in English, *shitgibbons*, are judged as better formed when the initial vowels, but not consonants, of the two source words are identical. This tendency, especially in its asymmetry, can be established independently in other natural language phenomena. As an example of the latter, Elfner and Kimper (2008) argue from results on *diddly*-infixation (e.g., *wélcome > wel-diddly-élcome*) that phonological reduplication need not be limited to the single segment or subject to rigid locality restrictions.

In this paper, I present another such class of words from a recent language game, *pussy blends* (e.g., *(Margaret) Thatcher + pussy > thatchussy*), and examine their potential contribution to the study of informal word-building processes. As they appear at first glance to be most similar to blends, the behaviour of 1,338 pussy blends in an automatically gathered Twitter corpus are compared against the documented tendencies of blending in the literature. Of particular interest is the amount and shape of material contributed by each source word.

The present article serves as a preliminary report. That is, before we can determine what the theoretical contribution of these forms may be, we need to establish what they

---

are and how they are formed. Ultimately, this study finds that so-called "pussy blends" fit the profile of documented blends fairly poorly. (This term is, nevertheless, retained for the remainder of the paper awaiting further research.) However, they do corroborate what English speakers appear to know implicitly about syllabic structure in English.

The remainder of the paper is structured as follows: §2 summarizes the taxonomy of blends and related forms, as well as the findings of several previous corpus studies on blends. §3 presents the history of "pussy blends" within its popularizing context, that is, the "one thicc bih" meme. The methodology involved in the preparation and treatment of this study's corpus is presented in §4. §5 presents the results, which primarily concern the average grapheme and syllable contribution of these forms' source words. A discussion follows in §6, along with a conclusion and directions for future research.

2.   Blending

Before discussing the history and characteristics of so-called "pussy blends," it is worthwhile to consider the known properties of blends in the literature, especially in English. The types of blends attested within and across the languages of the world are extremely diverse, and as such, the questions and taxonomy arising from their study are complex. Seeing as the main forms of interest in this paper are of a particular type, we shall mainly focus our attention in this section on the behaviour of similar blends.

In its most basic form, blending involves the combination of two base lexemes. One or both of the source words are frequently clipped, though not necessarily, depending on the degree of overlap, e.g., German *Paradies + Diesel > Paradiesel*[1]. Additionally, though circumscribed examples exist (e.g., *hypocritement + critique > hypocritiquement*), blends typically involve a linear juxtaposition. According to Fradin (2015) and Kemmer (2003), this process often chooses as its switch-over point some shared material between the source words, whether graphemic or phonemic or both. This tendency, however, may be language-specific, as blends with overlap in Bertinetto's (2001) English and Italian data amount to only 45% and 24% of cases, respectively, versus those of German and French at 89% and 79%.

Other word-building processes share some of these properties, especially the juxtaposition of lexemes and truncation. Fradin (2015) further distinguishes blending via two other properties: First, within a given language, it is unpredictable which of the two words serves as the semantic head of the resultant blend. This criterion separates blends from compounds, whose semantic interpretation is much more regular within a given language. In addition, compounding rarely, if ever, involves truncation (e.g., Kemmer 2003).

Second, Fradin claims that blending produces "type hapaxes" in a language; that is, any constituent part of a blend cannot freely combine with other lexemes. This distinguishes blending from secreted affixation, as in *chocoholic* or *emailgate*. Tournier (1985) names elements such as *-(a)holic* and *-gate* (from *Watergate*) fractomorphemes. Beyond

---

[1]Examples in this section, if not otherwise stated, come from Fradin (2015).

the increased productivity of secreted affixation, it also involves specific semantic operations not involved in blending (Fradin, 2015). Similarly, though more literal in its interpretation, concealed compounding (e.g., *freeware, shareware*) is excluded by Fradin on the basis of its series-like nature.

Scholars distinguish within blends proper two different types: telescopes and portmanteaux (Piñeros, 2004), or overlap and substitute blends, in Kemmer's terminology. The former function essentially as linear expressions (e.g. Spanish *cuernos* + *nacionales* > *cuernacionales*; 'horns' + 'national' > 'national horns') whose non-converging edges always remain intact, and whose resultant form is always longer than either of the source words in terms of segments (Piñeros 2004, following Algeo 1977). Meanwhile, the contributing words of portmanteaux are selected more for a combination of semantic association and phonological similarity, e.g., *ladrón* + *McDonald's* > *ladrónald's*; 'thief' + 'McDonald's' > 'McDonald's as a rip-off'. As these words tend to mirror closely the phonemic and/or syllabic structure and length of one of their source words, it is argued these are formed by overlapping (Pharies, 1987), or concealment (Piñeros, 2004).

The length of blends in comparison with their source words is a recurrent question in the literature, one which interests us especially here. In terms of syllables, the claims are varied, potentially due to differences in corpora. Kubozono (1990) claims that blends are often identical in syllable count to that of their second source word, while Bat-El (2006) and Cannon (1986) claim it is the longer source word that determines this number. Segmentally speaking, several authors find a strong tendency for the shorter of source words to contribute more material to blends (e.g., Bergström, 1906; Kaunisto, 2000). Gries (2004a,b) confirms this observation, in addition to arguing for a general, if at times conflicting tendency for the second word to contribute more.

The question of which word contributes more is not independent of other factors; in fact, it has been shown to be intertwined with word frequency and prototypicality (Kelly, 1998), order (Kubozono, 1990) and phonological similarity (e.g., Bat-El, 1996; Kemmer, 2003; Gries, 2004a). Given that the order and similarity of the second word are controlled for in the present corpus, however, these issues are set aside here.

The final aspect of blends addressed here concerns syllable shape of blends' constituent parts. Bertinetto (2001) finds a language-specific effect in lexical blends without overlap (the type most representative of pussy blends). Namely, while English shows a preference for "right-branching" types, in which the blend's juncture is situated between onset and rhyme, German, French and Italian prefer "neutral" types (CV-CV juncture). Syllabic structure can also influence order, in that source words with complex onsets may be preferred as the first word and those with complex codas as the second (Bat-El, 2006). This concern is, again, secondary, as order is fixed in our corpus.

3. Background

The following section provides background on pussy blends within and outside of the "one thicc bih" meme that popularized them. First, the origins of the main elements of the

meme's formula are discussed in §3.1. Section 3.2 then examines the content and trajectory of the meme. Section 3.3 concludes with details on pussy blends after the zenith of the meme and offers some speculation on the future of these blends.

3.1   Ancient history (in Internet Years)

All the main elements of the "one thicc bih" formula, i.e., "thicc," "bih," and pussy blends predate the meme itself. The first two can be traced back to AAVE slang, being alternate spellings for "thick" and "bitch," respectively. The former refers to an ample or muscular body which the speaker typically finds attractive. Though the general use of "thick" in this sense is hard to trace, according to KnowYourMeme[2], the earliest known instance on the Internet of the particular spelling "thicc" comes from a 2004 hip-hop fan page. Meanwhile, entries for "bih" date back to 2003 in the Urban Dictionary[3].

Pussy blends are first noted in gay slang, particularly (if not limited to) the formations *boy + pussy > bussy* and *man + pussy > mussy*. These are, again, difficult to pinpoint in time. The website Robert Scott's Gay Slang Dictionary contains both "boy-pussy" and "man-pussy" but not their blended forms. Both were submitted by a London-based user in July 1999.[4] The earliest entries in Urban Dictionary date back to 2004 (though in the earliest entry, defined as "butthole pussy").[5]

The recent popularity in pussy blends can confidently be attributed to an April 2017 Tumblr post which coins *thrussy* from *throat pussy*, garnering over 26,500 "notes" (similar to likes) within 72 hours of publication. Neither Urban Dictionary nor Google Trends shows hits for "thrussy" before April 2017. Before this, it is unclear to what degree pussy blends existing between 2004 and 2017 reflect a freely combinatorial process of word formation (i.e., if novel forms would have been accepted), or whether they belonged to a closed class. For instance, the word "babussy" appears in 2012 in a Ukranian political movement as a combination of "babushka + pussy" (in reference to the band Pussy Riot).[6] Further research may elucidate these questions, but for the moment it is assumed that pussy blending remained marginal and restricted to a few items until the advent of *thrussy* and the "one thicc bih" meme.

3.2   You can't get rid of the Babadook

The "one thicc bih" meme as we know it first appears towards the end of May 2017. The earliest, most popular instances of the meme involve the characters Mr. Krabs from the animated T.V. show Spongebob Squarepants and the Babadook, the eponymous monster

---

[2] https://knowyourmeme.com/memes/thicc

[3] https://www.urbandictionary.com/define.php?term=Bih

[4] Accessible via the Internet Wayback Machine: https://web.archive.org/web/20010410014457/http://hurricane.net:80/~wizard/dict_frame.html

[5] https://www.urbandictionary.com/define.php?term=bussy

[6] https://blogs.wsj.com/emergingeurope/2012/08/17/they-are-calling-it-babussy-riot/

of a 2014 Australian horror movie (blended as *krussy* and *babussy*, respectively). At the time of writing in June 2018, these original posts' videos have garnered over 576,000 and 829,000 views alone (including retweets and shares).[7] On the heels of these notable examples, the meme quickly become popular enough to warrant articles in sources such as New York Magazine, Buzzfeed, and so on.

The essential formula of the meme is as follows: "*x* is one thicc bih, let me see that *y*," where *x* designates a noun (typically a character or famous personality), and *y* represents a blend of that word, or a semantically related word, and the word "pussy". In some instances of the meme, the pussy blend directly references an orifice and/or sexualized part of the referent *x* (akin to "let me see dat ass"), while others, especially novel and rhyming/near-identical forms (e.g., *pharmacy* > *pharmussy*), employ meta-humour to defy expectations and/or provide commentary on the subject or on the meme itself. The semantics of pussy blends are complex and may warrant further research; for the moment, we are most interested in their form.

The "one thicc bih" meme circulated notably, though not exclusively, on Twitter, both in text and accompanied by a video generated by Ditty (Zya Music), a text-to-song app. Text provided by users is dynamically fit to the melody of any of a number of songs and sung by a computer-generated voice. The meme in question uses one of the app's basic songs, "Good Day" by 4Qent100. To illustrate how the algorithm adapts to inputs of various lengths, table 1 breaks down three different examples (containing pussy blends of two, three and four syllables respectively) into simplified groups of 4 beats.

	Mi	ster	Krabs		is	one	thicc bih	
	The	Baba	dook		is	one	thicc bih	
	Ka	ren is	one thicc	bih		lemme		
1	2	3	4	1	2	3	4	

	lemme					see	that kru	ssy
	lemme					see that	babu	ssy
	see	that				compu	tu	ssy
1	2	3	4	1	2	3	4	1

Table 1: Metrical adaptation of 3 memes of varying lengths by Ditty app

3.3 Life after thiccness

As can be extrapolated from the popularity of the phrase "one thicc bih" in Google trends (shown in Fig. 1), worldwide interest in the meme begins to decline in late June and reaches

---

[7]Because of the large number of pussy forms cited in this paper, in the interest of space, the URLs and authors of individual posts are not cited but may be provided upon request. It should be noted, though, that certain posts may have been deleted, or entire accounts may have been deleted or modified in the period following data collection.

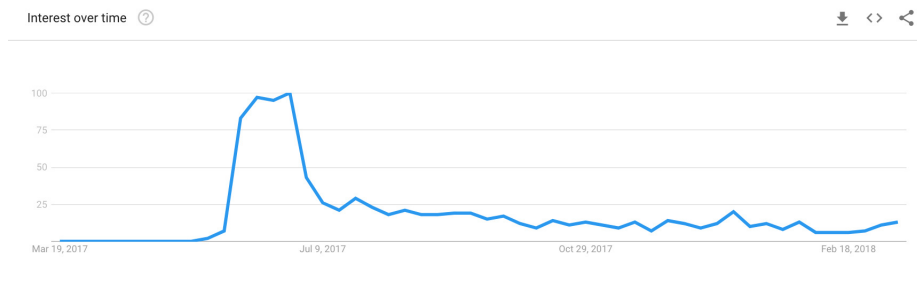its nadir in August 2017, after which time interest remains low but stable.

Figure 1: Google trends for "one thicc bih" (03/2017 - 02/2018)

While the meme's lifespan was not unexpectedly short, it remains to be seen whether pussy blends as an open class of words will follow suit—that is, whether speakers will continue to invent new pussy forms, whether certain forms will persist, but as a closed class, or whether they will all fade into obscurity.

Limitations of data-gathering currently make this question difficult to investigate, as substring searches—that is, all words ending with "ussy"—are not allowed by the Twitter API (or for that matter, most, if not all, API). Data may be gathered *en masse* and post-treated, though initial attempts are showing such forms to be rare. Meanwhile, searches for individual words prove more promising (e.g., a spike in *grussy* following the release of the trailer for the 2018 *The Grinch* film), though the process proves more time-consuming and less exhaustive. The durability of pussy blends may ultimately need to be investigated, along with their interpretability and combinatorial freedom. For the moment, it is assumed that the phenomenon is still fresh in the public memory and therefore worth investigating.

4.   Methodology

In order to examine the factors potentially driving pussy blend formation, a corpus was established along the following lines. The Python script `GetOldTweets`[8] was used to scrape all publicly available tweets containing the phrases "one thicc bih" and "see that" published between June and August of 2017. (The reader is reminded from §3.3 that the reason for using this meme as a framing context is the impossibility of substring searches on this platform.) Though the sources presented earlier in §3.2 suggest that the initial rise in popularity of the meme occurred in May 2017, the script could not gather tweets from this time because of limitations of the free Twitter API. The earliest date represented in the entire corpus is June 17, 2017; however, as only a subset of the data was processed for the present study, and this done by date in descending order, the earliest date represented in this paper is June 23, 2017 (the latest date being August 30).

---

[8]`https://github.com/Jefferson-Henrique/GetOldTweets-python`

After exclusion of retweets and identical tweets, the corpus contains 4,450 tweets, while the subset presented here is based on the 1,500 most recent tweets. The following other information, beyond the text itself, was automatically gathered: date of publication, username, number of retweets and favourites, unique identifier and URL.

The subset of 1,500 tweets (hereafter referred to as the corpus for the sake of brevity) was further processed to extract pussy blends and their intended referents. First, the words immediately following "see that" up to the substring "ssy" (or the end of the line in absence of this ending) was automatically extracted and manually inspected. Forms which could not be conclusively identified as a pussy blend were targeted for elimination. These fall into three types, and though coincidentally the first two are not found in the subset corpus, they are all illustrated here with examples from the larger corpus. First of all, formations unambiguously using other words than "pussy" as the second element of the blend (e.g., *(Don) Draper > Drapenis*) do not qualify. Second, formations such as *Vlad > Vladdy* and *wolf > wolfy* cannot be reliably distinguished from diminutives and thus do not qualify either. Finally, tweets using existing words which cannot be confidently separated into constituent source words were also eliminated. A popular version of this type is illustrated by the tweet "Claire de Lune is one thicc bih, let me see that Debussy", in which the name "Debussy" cannot be separated into "pussy" and a "Deb(u)–"-initial word. Such cases accounted for 38 of all excluded forms. Note that forms appearing to use alternate spellings of "pussy" (e.g., *Mermaid > mermsey*) in the blend were not present in the subset but would not have been excluded. Finally, the pussy blend was standardized to remove effects of orthographic lengthening (e.g., "HarLussYYyyyY" coded as "harlussy").

The referent of the meme was then automatically extracted as the words preceding "is one thicc bih" (variants of the copula, such as "iz," included) and again were manually verified. These referents were then compared with the pussy form to establish both the full first source word and the novelty of the pussy blend. In non-novel cases, the source word was often unambiguous, though certain conventions had to be adopted. For instance, full names were taken as the source word only when all words were included without a space in the pussy blend (e.g., *Ms. Puff > mspussy* coded for "mspuff"); otherwise, only the word explicitly contributing information was tagged (e.g., *Street lego man > legussy* coded for "lego"). Note that any punctuation and spaces in the full referent were removed in the coded source word. In rare cases of potential ambiguity where two or more words of a full name or expression began with the same grapheme(s), the last word was categorically chosen (e.g., "cheese" taken as the base in *Chuck E. Cheese > chussy*).

The novelty of pussy forms was established at this same stage. Forms not taking the explicitly stated reference as its first source word were defined as novel with an important exception. Namely, unexpressed source words which can be interpreted as inherent to the referent, which was limited to first/last name or species—for instance, *Gary* (a snail on Spongebob Squarepants) > *snussy*—were counted as non-novel. These were, however, given a special code and can thus be reclassified later if necessary. The source words of novel forms were finally reconstituted based on the conversation thread, the user's Twitter page, and Google searches. Ultimately, 124 forms had to be discarded for ambiguity. All

in all, 162 forms were discarded for a total of 1,338 pussy blends.

Source word contributions to the blend were then manually coded. Following Gries (2004b), the source word contributions were formatted for a generous interpretation, though strictly locally within each source word. That is, the number of graphemes contributed by *syllabus* and *pussy* in *syllabussy* would be 8 and 4 respectively (i.e., <syllabus> + p<ussy>). A hypothetical example of *Jerusalem > jussy* would not be coded as <j>er<us>alem + p<ussy>, while *jerussy* would be coded as <jerus>alem + p<ussy>. This convention is likely to be complicated when phonemes, rather than graphemes, are considered. The question is left open for the time being.

No single user is overwhelmingly represented in the corpus.[9] The current 1,338 tweets were composed by 1,156 unique users. The max number of forms provided by a single user is 6, while the mean is 1.2 tweets per user. Variation across users is thus assumed in the results rather than directly incorporated.

In order to calculate syllables in source words and blends, vowels were counted in the following way. Contextual digraphs were first processed, e.g., the <i> in <tion> was excluded, and then vocalic digraphs were transformed into single nuclei. Then all single vocalic graphemes were tabulated. The resulting syllable counts were manually verified for all 1,338 forms.

Finally, the Similarity Index of each blend was calculated according to Gries's (2004b) formula. That is, the relative number of graphemes maintained by each word was multiplied by the relative number of graphemes of the blend for that word. These scores were added and divided by two. The formula is provided in table 2, where G = no. graphemes, r = root (i.e., source word), c = contributed material, and b = blend.

$$\frac{\left(\frac{G_{c1}}{G_{r1}} \times \frac{G_{c1}}{G_b}\right) + \left(\frac{G_{c2}}{G_{r2}} \times \frac{G_{c2}}{G_b}\right)}{2}$$

Table 2: Similarity Index equation (Gries, 2004b)

For example, <syllabus> + p<ussy> yields an SI of 0.56, corresponding to well-formed blends in Gries' study.

5.   Results

This section presents some preliminary findings from the study thus far, preceded by some anecdotal remarks from the data. The SI results are then reported along with word contributions by syllable size and shape. Given that, as we will see, the amount of material contributed by the second word is near-invariable, we are interested for the moment in the factors determining the amount of material contributed by W_1.

---

[9]The robot account @Ussybot, which automatically generates pussy blends, was intentionally excluded from the data. It cannot be guaranteed that the corpus contains no robots, but if present, they are assumed to be extremely few.

5.1 Preliminary remarks

The forms discussed in this section are more often than not rare, in comparison with those presented in §5.2, though it remains to be seen whether or not they are truly exceptional. First, several examples of (near-)inclusive blending (Ronneberger-Sibold, 2006) can be noted in the data. This type of blend results in a form whose pronunciation is the same as the longer source word but whose other source word can be recovered from the blend's written form, e.g., *Jew* + *jubilee* > *jewbilee*, an example from the corpus being *democracy* > *democrussy*. Essentially, these amount to cases with a [-əsɪ]-final W_1.

Only a small number of forms take more than two source words, e.g., *octoling chocolate* + *pussy* > *octochussy*. It is unclear how these should be treated, namely whether the derivation should be flat (e.g., <octo>ling <ch>ocolate p<ussy>) or recursive (<ch>ocolate p<ussy> → <octo>ling <chussy> or <octo>ling <ch>ocolate → <octoch> p<ussy>). For the moment, they are treated linearly, the same as 2-word combinations, as combinations of 3+ words are very few in number.

A handful of forms evidenced intrusive consonants, that is, material belonging to neither source word. These largely included <r> (e.g., *Fionn* > *Fiorussy*) and <b> (e.g., *Jake* > *Jabussy*) and can arguably be based in analogy to *thrussy* and *bussy*. A similar but more inexplicable type involves initial-consonant copy (e.g., *Gao* > *gagussy*, *Cameron* > *Cacussy*, *me* > *memussy*).

5.2 Syllables and graphemes

Here, we present a number of calculations in order to paint a portrait of the average blend in the corpus. Forms are generally non-novel (n = 1,187, 88.7%), and only 154 forms (11.5%) have graphemic overlap (only 15 of which are novel). Hiatus is largely avoided, only 114 blends (8.5%) containing a V-V juncture. As for the W_2 contribution, <ussy> is by far the most common (1,238, 92.5%), with <pussy> and <ssy> tied at 50 forms (3.7%) each. The mean SI of non-novel forms is 0.37, while that of novel forms reaches 0.42.

As shown in table 3, there is an overwhelming preference in non-novel forms for a monosyllabic W_1 contribution (that is, considering the syllabification of only the contributed material). Novel forms share equal preference for mono- and disyllabic contributions. Non-novel forms prefer onset-only contributions to disyllabic shapes.

Condition	Onset-only	1σ	2σ	3σ	4σ	5σ
Non-novel	286	719	151	27	1	0
Novel	8	61	63	13	3	2

Table 3: No. forms per W_1 syllable contribution by novelty

Expanding on this, we can see in table 4 that there is a general preference for W_1 contribution to end in a single consonant, regardless of the number of syllables contributed, especially in the non-novel forms. Note that in table 4, only the number of consonants at

Shape	Count (Non-novel)	Count (Novel)
C_1	208	3
C_2	108	6
CV	62	3
CVC_1	466	26
CVC_2	201	32
CVCV	31	6
$CVCVC_1$	63	42
$CVCVC_2$	21	15

Table 4: W_1 contribution syllable shape by novelty

	Condition	Gr (W_1)	σ (W_1)	Gr loss	Count
	$W_1 < W_2$	2.5	0.69	1.2	235
Non-novel	$W_1 = W_2$	2.9	0.66	2.1	318
	$W_1 > W_2$	3.8	1	3.5	634
	$W_1 < W_2$	3.4	0.9	0.5	23
Novel	$W_1 = W_2$	4.1	1.2	0.9	33
	$W_1 > W_2$	6	1.9	1.9	95

Table 5: W_1 contributions and loss by novelty & relative length

the end are considered. For instance, CVC_1 could refer to VC, CVC or CCVC contributions, but CVCC is necessarily distinguished as CVC_2. Given that 3+ consonant sequences are rare, C_2 refers to 2 or more consonants at the end of W_1 contribution. Finally, CV does not distinguish <u>-final W_1 from cases which would lead to true hiatus. Contributions greater than 2 syllables are excluded. As for the sonority profile of C_2-final syllables, there is a marked preference for falling sonority (161 of 272, 59.2%), then rising (84, 30.9%) and even (27, 10%).

So far the results have considered grapheme and syllable contribution regardless of the original length of the source words. Table 5 shows the mean contribution of W_1 (that of W_2 being constant at 4 graphemes) according to the relative lengths of the two source words, separated by novelty. Not unexpectedly, within the same condition, W_1 contribution is greater in novel blends. Within non-novel and novel blends, W_1 contribution increases across conditions; that is, the first word contributes the most material when longer than the second. These are the majority of cases, regardless of novelty, though note that in the case of non-novel blends, the result is effectively equal contribution as W_2 (3.8 vs. 4).
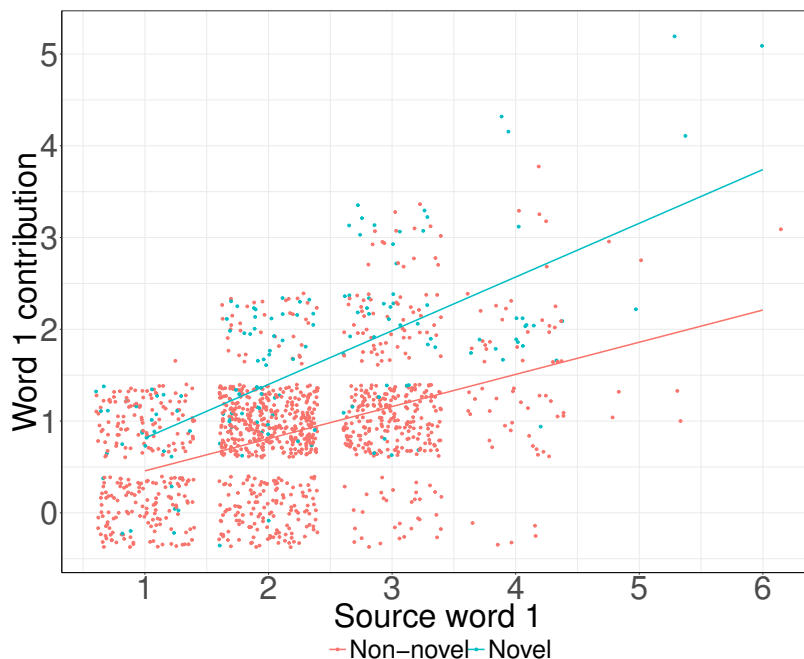
Figure 2: No. syllables, source word vs. contribution (W_1)

Finally, we consider the interaction between the number of syllables provided by W_1 against the source word's original length, as shown in figure 2. In the non-novel condition, contribution clusters around 1 syllable, while contribution in novel blends starts on average higher and appears to increase proportionately with source word length.

6.   Discussion and Conclusion

The questions asked in this report are twofold: First, what factors, presumably phonological, influence the amount and shape of material contributed in so-called "pussy blends"? Second, can they rightfully be considered blends?

To remind the reader, we know the factors determining the resultant blend of any two words to be multifaceted, among which are phonological similarity, word order, relative length, language-specific parameters, and so on. The very nature of the present language game allows us to isolate just a few. Namely, word order is fixed, as is (largely) W_2 contribution, and phonological similarity is not a prerequisite. We can thus focus on what attributes of a given word, in relation with the word "pussy," may determine the material contributed to their combination.

In non-novel conditions, W_2 contributed more on average, regardless of relative length. This is slightly in keeping with one of the observations in Gries (2004b), where it is claimed:

"...x segments of the beginning of a word increase its chance of being recog-

nised more than *x* segments of its end (cf. Nooteboom 1981), given that the normal way we encounter words is from beginning to end rather than vice versa. Therefore, it makes sense that, if both source words are equally long, the second word contributes more because this would enhance its recognisability by compensating for the fact that it is not processed in the normal way."

In another study, though, using tests of statistical significance, Gries (2004a) finds source words contribute equal material when equally long. This he attributes to phonological similarity and overlap counteracting the above mentioned findings. Similarity and overlap are largely absent from this corpus and thus cannot be invoked.

Gries's other observation, a competing constraint that the shorter word tends to contribute more material (see also Kaunisto 2000), takes precedence outside of this condition; indeed, the strongest factors in Gries (2004a) are that the shorter words contribute more, regardless of position. Again, this is not borne out in the present data in the non-novel condition.

Meanwhile, in the novel condition, W_2 contributes more when it is longer, contributes equally when the two are equal, and W_1 contributes more when it is longer. The equal length data match Gries (2004a), but otherwise the results skew in the opposite direction; that is, the longer word contributes more.

Relative length matters not only in the literature on blends as a whole, but also in the subcategorization of blends, i.e., telescopes vs. portmanteaux. In some respects, pussy blends are clearly closer to telescopes than portmanteaux. The relative lack of overlap, especially in novel forms, suggests that the interaction of phonological overlap and semantic association (effectively amounting to the wit or felicity more characteristic of portmanteaux blends) plays a minor role, if at all, in pussy blends. Rather, their parts stand in a syntagmatic relation, "*x* + pussy" essentially meaning a sexualizable part of *x*. However, the non-negligible presence of C+*ussy* forms (15.7% of the corpus) goes against Piñeros's (2004) claim that telescopes are always longer than their source words. In fact, more than half the C+*ussy* forms (112 of 211) are equal to or shorter than W_1 in addition to W_2.

Moreover, beyond these shorter forms, as seen in tables 3, 4 and 5, there may be evidence in the non-novel condition for a CVC(C) templatic contribution from W_1. We see this both in the number of closed monosyllabic contributions and more subtly suggested when we consider the relative lengths of each word. (The effects of stress, though, must be sorted out, especially in these longer words.) This runs counter to Bat-El's statement that "the number of syllables in a blend is often identical to the number of syllables in the longer base word" (Bat-El, 2006, 67). Pussy blends may, in fact, be maximally, if not ideally trisyllabic. (Ideally in the sense that, especially context of the "one thicc bih" meme, few disyllabic blends are likely to be distinct enough.)

Speaking of syllables, let us briefly turn our attention to the relative shapes of "pussy blend" source word contributions, in lieu of Bertinetto (2001). As argued earlier, W_1 contributions more often than not end in at least one consonant, regardless of the number of syllables preceding. Contribution-final CC sequences also tended to fall in sonority. Meanwhile, W_2 is almost categorically a vowel plus an extra CV syllable (i.e., -[ʊsi]). These

correspond in Bertinetto's (2001) system to Onset + Rhyme and Closed syllable + Rhyme types (the latter in the case of falling CC sonority sequences). These right-branching types were the most frequent type in English (though not cross-linguistically).

To summarize, what have been called "pussy blends" largely fail to convene to most aspects of blends. In terms of relative length and contribution, they rarely match the trends observed in previous corpus studies, at times reversing them. Support that these forms can be categorized as telescopes is questionable; their formation does not seem to follow that of true telescopes. Finally, their series-like nature constitutes a technical but immediate disqualification for Fradin (2015).

In sum, "pussy blends" are more likely to fall into the category of secreted affixation, in which case *-ussy* would have become, however ephemerally, a fractomorpheme (cf. *Watergate* > *-gate*). Further supporting this is the seemingly regular semantic operation transforming the meaning of "pussy," in the same way that the meaning of "Watergate" is transformed into "scandal connected with *x* involving high-ranking people" (Fradin, 2000).

To my knowledge, no large-scale investigation focussing on secreted affixation has been conducted. As such, the forms investigated here (and to be investigated in the larger corpus) may reveal more nuanced structural differences between these and blends, beyond the well-established semantic differences. In the future, judgement tasks and surveys from meme creators and/or consumers may also allow us to further elucidate open questions, in particular the following: whether stress can influence the switch-over point, whether W_1 contribution is indeed ideally a monosyllabic template and whether this word-formation process, or even certain of its forms, is expected to remain in the public eye or disappear into the annals of Internet history.

References

Algeo, John. 1977. Blends, a structural and systemic view. *American Speech* 52: 47–64.

Bat-El, Outi. 1996. Selecting the best of the worst: The grammar of Hebrew blends. *Phonology* 13: 283–328.

Bat-El, Outi. 2006. Blend. In *Encyclopedia of Language & Linguistics*, ed. Keith Brown. Elsevier, 66–70.

Bergström, Gustav A. 1906. *On blendings of synonymous or cognate expressions in English: a contribution to the study of contamination*. H. Ohlsson.

Bertinetto, Pier Marco. 2001. Blends and syllabic structure: A four-fold comparison. *La gramàtica i la semàntica en l'estudi de la variació* : 59–112.

Cannon, Garland. 1986. Blends in English word formation. *Linguistics* 24(4): 725–754.

Elfner, Emily and Wendell Kimper. 2008. Reduplication without RED: Evidence from diddly-infixation. In *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Cascadilla Press Somerville, MA, 150–158.

Fradin, Bernard. 2000. Combining forms, blends and related phenomena. *Extragrammatical and Marginal Morphology* 12: 11.

Fradin, Bernard. 2015. Blending. In *Word-formation. An International Handbook of the Languages*

of Europe, ed. Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer. Berlin: De Gruyter Mouton, 386–413.

Gries, Stefan Th. 2004a. Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics* : 639–668.

Gries, Stefan Th. 2004b. Some characteristics of English morphological blends. In *Papers from the 38th Regional Meeting of the Chicago Linguistics Society, Vol. 2*. 201–216.

Kaunisto, Mark. 2000. Relations and proportions in English blend words. Fourth Conference of the International Quantitative Linguistics Association.

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

Kemmer, Suzanne. 2003. Schemas and lexical blends. In *Motivation in Language: Studies in Honor of Günter Radden*, ed. Hubert Cuyckens, Thomas Berg, René Dirven, and Klaus-Uwe Panther, vol. 243. John Benjamins Publishing, 69–97.

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

Nooteboom, Sieb G. 1981. Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics* 9(4): 407–424.

Pharies, David A. 1987. Blending in Spanish word-formation. *Romanistisches Jahrbuch* 38: 271–289.

Piñeros, Carlos-Eduardo. 2004. The creation of portmanteaus in the extragrammatical morphology of Spanish. *Probus* 16(2): 203–240.

Ronneberger-Sibold, Elke. 2006. Lexical blends: Functionally tuning the transparency of complex words. *Folia Linguistica* 40(1-2): 155–181.

Tessier, Anne-Michelle and Michael Becker. 2018. Vowel but not consonant identity and the very informal English lexicon. In *Proceedings of the 2017 Annual Meeting on Phonology*. 1–12.

Tournier, Jean. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Champion Books.