A corpus study of phonological factors in novel English blends

Annual conference of the CLA (University of Regina)

Michael Dow
Université de Montréal

May 30, 2018

Introduction

Introduction

- "Pussy blends" are becoming more and more commonplace online (popularized by "one thicc bih" meme), for instance *thick pussy* > *thussy*:

  Jean shorts have to be tight and then i [sic] *get a thussy and it's annoying* (`step2myworld`, 2018, May 22)

- Irregularity of what a single blend can mean, e.g. *thussy* < `Thor`, `Thanos`, `Margaret Thatcher`, etc.
- Irregularity of what form a single subject will yield, e.g. *Thanos* also > `thanussy`; *Margaret Thatcher* also > `thatchussy`

Previous study

A pilot study (Dow 2018) of ~100 forms found:

- Contribution of material from first word (W_1) increases if longer than W_2 (i.e., "pussy") and if not directly expressed in the meme or thread (= "novel")
- Equal preference between onset-only and CVC_1 contribution of W_1 in non-novel forms
- Greater W_1 contribution when containing low sonority-high sonority CC juncture (e.g. *Manray* > *manrussy*) and internal sC clusters (e.g. *toadstool* > *toadstussy*)
- Internal fricatives and <r> encourage loss of <u> from W_2 (e.g. *Ms. Puff* > *puffsy*, *Barney* > *barsy*)

Today

Preliminary findings of an expanded corpus study.

Q: What implicit factors are at play, i.e., what makes a good (or bad) "pussy blend"?

A: W_1 contribution increases as a function of source word length (up to final rhyme); W_2 contribution remains constant.

Q: Can we consider these words as true blends? If not, what else could they be?

A: These words go against a number of properties of blends. They might instead be considered as mid-clipped compounds.

Outline

1. Introduction
2. Meme
3. Blending
4. Methodology
5. Results
6. Discussion

Meme me up, Scotty

-*(u)ssy* blends & the "one thicc bih" meme

- Appearance & explosion of "one thicc bih" meme (text and Ditty videos, e.g. Fig. 1) in May 2017
- Format: "x is one thicc bih, let me see that y"; x = character or famous personality; y = blend of x (or related word) and *pussy*

Fig. 1: Babadook > babussy

Source: `dcparkers`, 06/2017

What's in a meme?

- Documented -*(u)ssy* forms date back to `early 2010s` in gay slang, re-popularized by an April 2017 `Tumblr post`, in particular:
  - Boy, man > bussy, mussy
  - Throat > thrussy
- *thicc* & *bih* AAVE slang (together ≈ "sexy individual"), each documented back as far back as early 2000s

Life cycle of a meme

- Widespread media recognition (e.g. `New York Magazine`, `Buzzfeed`) ⟶ Ditty app #1 on iTunes store (May 2017)
- Decline around July 2017 (Fig. 2)

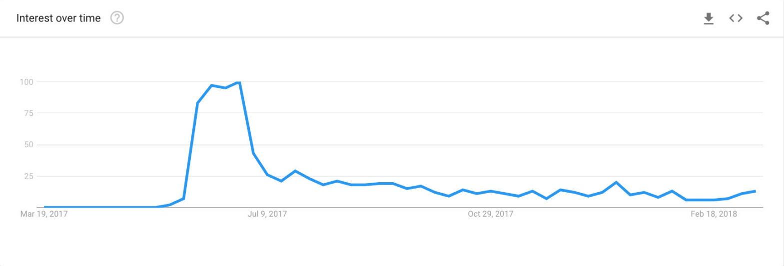

Fig. 2: “one thicc bih” in Google trends

Life after thiccness

- Popularity hard to quantify outside of meme: no substring searches on Twitter
- Several individual searches turn up recent results, though not as common as in summer 2017
- Remains to be seen if the process and/or certain forms survive, but “not dead yet” for the moment

Why study it?

- Memetic nature of "ussification" may resolve some empirical problems in study of blends:
  - Difficulty of automatic collection/recognition (Fradin 2015) $\longrightarrow$ ease of collecting large corpus of meme
  - High degree of variation within and across languages (different "species") $\longrightarrow$ controlled setting (W_2 remains constant) allows for isolation of factors in W_1
  - Differing degrees of felicity $\longrightarrow$ several metrics (e.g., meme-user judgments, retweets & likes) can make sense of variation

Will it blend?

Properties of blends

- Definition: "[I]ntentional coinage of a new word by fusing parts of at least two source words of which either one is shortened in the fusion and/or where there is some form of phonemic or graphemic overlap of the source words" (Gries 2004)
- Three salient properties (Fradin 2015)
  - No preservation of lexical integrity: stems are rarely maintained intact & their alteration is variable
  - No fixed pattern of compositionality: head member is unpredictable
  - "Type hapaxes": blends cannot form series (e.g. *élevage* 'breeding' + *vache* 'cow' ⟶ *élevache* 'cow breeding' but **élechien* 'dog breeding')

Overlap

- What determines the respective contribution and order of words in a blend?
- Semantic motivation (*brunch*) vs. phonological selection *glitterati*, cf. Fradin's (2015) criterion of overlap

	A. Trunc. = both	B. Trunc. = 1	C. Trunc. = 2	D. Trunc. = 0
+OV +LIN	*daxpór* *daxáf* × *laxpór*	*knáuros* *knástos* × *áuros*	*Müllionärin* *Müll* × *Millionärin*	*Paradiesel* *Paradies* × *Diesel*
+OV −LIN	*dialügisch* *dialogisch* × *Lüge*	*carnibbleous* *carniverous* × *nibble*	—	*hypocritiquement* *hypocritement* × *critique*
−OV +LIN	*brunch* *breakfast* × *lunch*	*klafúda* *klára* × *fúda*	*smothercate* *smother* × *suffocate*	*sálkal* *sál* × *kál*
−OV −LIN	*agitprop* *agitation* × *propag.*	—	—	*rajolivissant* *ravissant* × *joli*

Fig. 3: Typology of blends (Fradin 2015)

Chunnel vs. brunch

- Gries' (2004) Similarity Index (SI), proportionate amount of material contributed by each word:
  - *<ch>a<nnel> + t<unnel>* = 0.67
  - *
eakfast + l<unch>* = 0.3
- Average SI of intentional & error-driven blends ≈ 0.5, vs. random word pairings ≈ 0.3

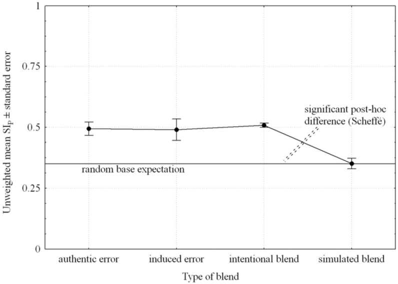

Fig. 4: SI by blend type

Extragrammaticality $\neq$ irregularity

- Debate over blending as morphological (e.g., Bat-El 1996, Plag 2003) vs. extragrammatical (e.g., Bauer 1988, Dressler 2000) process
- Extragrammaticality does not exclude influence of regular/universal linguistic forces, especially phonological for blends (Fradin, Montermini & Plénat 2009)
- Other peripheral (informal) processes evidence knowledge of grammar-external structures or forces, e.g., expletive infixation (McCarthy 1982), *shitgibbons* (Tessier & Becker 2018)

Blend-trends

- The shorter source word of a blend more likely to contribute more information for intelligibility (Kaunisto 2000)
- Gries' (2004) results confirm this + a (competing?) tendency for W_2 to contribute more:

which word is larger?	which word contributes more to the blend? =	source word$_1$	source word$_2$	row totals
=	47 (++)	22 (--)	53	122
source word$_1$	58	45 (---)	211 (+++)	314
source word$_2$	125	316 (+++)	111 (---)	552
column totals	230	383	375	988

Fig. 5: Contribution by length, phonemes

Methodology

Corpus

- 4450 Tweets scraped using `GetOldTweets`
- Criteria:
  - June 2017 - August 2017
  - Contains phrases “one thicc bih” & “see that”
- Information automatically gathered:
  - Username
  - Date
  - Retweets & favourites
  - Link, ID
  - Text

Processing

- First 1500 manually annotated for:
  - Full referent
  - Deduced W_1 "base" & contribution
  - W_2 contribution
  - Standardized blend word (e.g. reduction of *yyyyy* to *y*)
  - Novelty of blend (if $W_1 \neq$ referent)
- Unclear forms (e.g. inside jokes), tweets using existing words (e.g. Claire de Lune > Debussy), and retweets/identical tweets excluded $\longrightarrow$ 1334 forms
- Some educated guesses on bases (e.g., Chuck E. *Cheese* > chussy)

Users & lifespan

- 1156 unique users
- Max no. tweets/user = 6, mean 1.2
- Peak popularity in corpus at end June, declines early in July (Fig. 6, cf. Fig 2)

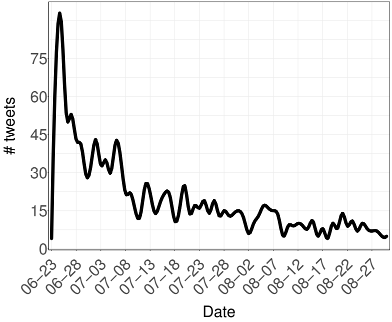

Fig. 6: No. tweets over time in annotated corpus

More processing

- Generous count of shared graphemes between words, e.g., graphemes in *Bass* > *bassy*: $W_1 = 4$, $W_2 = 3$
- Number of syllables identified as number of vowels (digraphs & graphemes such as <i> in <-tion> pre-processed), manually verified

SI

- Proportion of material each word contributes to the blend (Gries 2004)
- Calculation performed on graphemes (G = no. graphemes, r = root, c = contribution to blend, b = blend), generous interpretation:

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

- For example, *syllabus* > *syllabussy*:

$$\frac{\left(\frac{8}{8} \times \frac{8}{10}\right) + \left(\frac{4}{5} \times \frac{4}{10}\right)}{2} = 0.56$$

Results

Anecdotes from the data

- /-əsi/ final words rare in corpus but highly felicitous; not specially quantified by current measures (e.g. *democracy* > *democrussy*)
- Non-contiguous blends also rare but may be of further interest, for instance:
  - *octoling chocolate* > *octochussy*
  - *hentai Trump* > *hentrumpussy*
- Presence of "intrusive" letters (belonging to neither word) in a handful of forms:
  1. <r> (e.g. *Fionn* > *fiorussy*)
  2. <b> (e.g. *Jake* > *jabussy*)
  3. <p> (e.g. *Kirsten* > *kirpussy*)
  4. C_1 (e.g. *Gao* > *gagussy*, *Cameron* > *cacussy*, *me* > *memussy*)

Preliminary numbers

- Only 130 forms (9.7%) have motivated graphemic overlap (i.e. <u-u> or <p-p> overlap)
- 1187 forms (88.7%) are non-novel
- <ussy> by far the most comment W_2 contribution (1238), vs. <pussy> (50) and <ssy> (50)
- Only 114 forms (8.5%) have true hiatus at juncture (i.e., <u-u> overlap ignored, as in *communism* > *commussy*)
- Mean SI = 0.37 (non-novel), 0.42 (novel)
- Sonority profile of last 2 consonants of CC(C)-final W_1 contribution: even (27), rising (84), falling (161)

Syllable size

Preference for 1-syllable W_1 contribution over onset-only regardless of novelty:

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

Table 1: No. forms per W_1 syllable contribution by novelty

Expanded syllables, W_1

Non-base-final consonant sequences & <u-u> junctures simplified. Max 2 syllables.

Shape	Count (Non-novel)	Count (Novel)
C	208	3
CC	106	6
CCC	2	0
CV	62	3
CVC	466	26
CVCC	183	28
CVCCC	18	4
CVCV	31	6
CVCVC	63	42
CVCVCC	19	14
CVCVCCC	2	1

Table 2: W_1 contribution syllable shape by novelty

Results (graphemes)

	Condition	Gr (W_1)	Gr (W_2)	Loss	Count
Non-novel	$W_1 < W_2$	2.5	4	1.2	235
	$W_1 = W_2$	2.9	4	2.1	318
	$W_1 > W_2$	3.8	4	3.5	634
Novel	$W_1 < W_2$	3.4	4	0.5	23
	$W_1 = W_2$	4.1	4	0.9	33
	$W_1 > W_2$	6	4	1.9	95

Table 3: Mean grapheme (Gr) contribution & loss by novelty and relative length

Results (syllables)

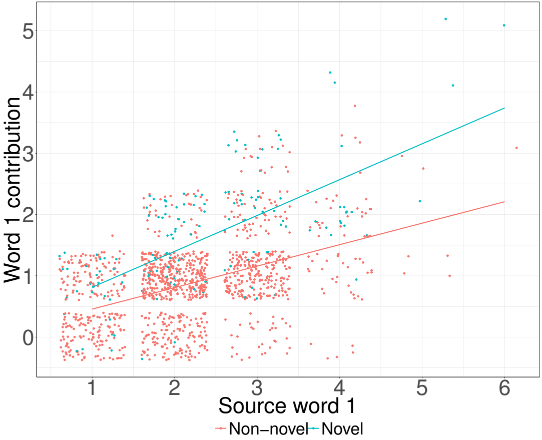

Fig. 7: No. syllables, source word vs. contribution (W_1)

Discussion & future work

Findings

- Grapheme contribution of W_1 increases with source word length regardless of novelty, while W_2 contribution stays the same.
- Novel forms on average lose less information.
- Competing preference between C(VC) and "all but final rhyme" templates.
- Loss of <u> from W_2 is rare and currently not predictable by any factor.

Are they blends?

- *-(u)ssy* forms meet some criteria, but not series-like nature & tendency for W_1 to contribute more as it gets longer, esp. in novel forms
- Many forms are dependent, not transparency of form, for meaning
- Forced combination regardless of overlap: *<chick>en + p<ussy>* ⟶ *chickussy* (more common) vs. *<Bloss>om + pu<ssy>* ⟶ *blossy*
- Low SI, especially in non-novel forms
- May in fact be median-clipped compounds (Tournier 1985, Jamet 2009), e.g. *smoke fog* > *smog*, though not always distinguished from blends

Future work

- Finalization of corpus annotation
- Integration of stress & phoneme counts (size, contribution, better sonority profiles) into results
- Comparison with fandom pairing names (DiGirolamo 2012), especially stress match (location of juncture) & onset conservation (e.g. *Clyde* + *Rani* > *Clani*)
- Judgment task (variation & strength of factors)

Thank you!

Acknowledgments

- The makers of the Ditty app
- Darius Kazemi (maker of the `UssyBot`)
- The audience members of the MOT 2018 (including Jeffrey Lamontagne & Daniel Curie Hall)

Works Cited I

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

Bauer, L. (1988). *Introducing Linguistic Morphology*. Edinburgh University Press, Edinburgh.

Becker, M. and Tessier, A.-M. (2018). Vowel but not consonant identity and the very informal English lexicon. In *Proceedings of the 2017 Annual Meeting on Phonology*, pages 1–12.

Bernard, F., Montermini, F., and Plénat, M. (2009). Morphologie grammaticale et extragrammaticale. In Fradin, B., Kerleroux, F., and Plénat, M., editors, *Aperçus de morphologie du français*, pages 21–45. Presses Universitaires de Vincennes, Saint-Denis.

DiGirolamo, C. M. (2012). The fandom pairing name: Blends and the phonology-orthography interface. *Names*, 60(4):231–243.

Dow, M. (2018). Let me see that truncussy: Elucidating patterns in a novel blending meme. Montreal-Ottawa-Toronto Phonology Workshop.

Dressler, W. U. (2000). Extragrammatical vs. marginal morphology. In Doleschal, U. and Thornton, A., editors, *Marginal and extragrammatical morphology*, pages 386–413. Lincom Europa, München.

Works Cited II

Fradin, B. (2015). Blending. In Müller, P. O., Ohnheiser, I., Olsen, S., and Rainer, F., editors, *Word-formation. An international handbook of the languages of Europe*, pages 386–413. De Gruyter Mouton, Berlin.

Gries, S. T. (2004). Some characteristics of English morphological blends. In *Papers from the 38th Regional Meeting of the Chicago Linguistics Society, Vol. 2*, pages 201–216.

Jamet, D. (2009). A morphophonological approach to clipping in English: Can the study of clipping be formalized? *Lexis. Journal in English Lexicology*, (HS 1).

Kaunisto, M. (2000). elations and proportions in English blend words. Fourth Conference of the International Quantitative Linguistics Association.

McCarthy, J. J. (1982). Prosodic structure and expletive infixation. *Language*, pages 574–590.

Plag, I. (2003). *Word-Formation in English*. Cambridge University Press, Cambridge.

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