



**Lexis**  
Journal in English Lexicology  
**14 | 2019**  
**Blending in English**

---

## Variable base-word positioning in English blends

Aviv Schoenfeld, Evan Gary Cohen and Outi Bat-El

---



### Electronic version

URL: <http://journals.openedition.org/lexis/3781>

DOI: 10.4000/lexis.3781

ISSN: 1951-6215

### Publisher

Université Jean Moulin - Lyon 3

### Electronic reference

Aviv Schoenfeld, Evan Gary Cohen and Outi Bat-El, « Variable base-word positioning in English blends », *Lexis* [Online], 14 | 2019, Online since 16 December 2019, connection on 21 April 2020. URL : <http://journals.openedition.org/lexis/3781> ; DOI : <https://doi.org/10.4000/lexis.3781>

---

This text was automatically generated on 21 April 2020.



Lexis is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

---

# Variable base-word positioning in English blends

Aviv Schoenfeld, Evan Gary Cohen and Outi Bat-El

---

## Introduction

- 1 In the vast majority of English blends [Adams 1973, Bryant 1974, Algeo 1977, Bat-El 2006], the order of the base words is invariable. For example, *smoke* × *fog* yield *smog* rather than *foke*, and *lábrador* × *póodle* yield *lábradoodle* rather than *póodrador* (the accented vowel letter marks the nucleus of the primarily stressed syllable; we ignore secondary stress). However, there is a small but non-negligible number of cases where the order of the base words varies, yielding two (or more) blends, i.e. blend doublets. Cases noted in the literature include *absotívely* ~ *posolútely* (< *absolútely* × *positívely*) [Algeo 1977: 60], *tígon* ~ *líger* (< *tíger* × *lío*n) [Bat-El 2006], and *chévroloc* ~ *cádillet* (< *chévrolet* × *cádillac*) [Gries 2012: 160]. In some doublets, the blends are synonymous, e.g. *plúmpricot* ~ *ápriplum* ‘plum-apricot hybrid’ (< *plum* × *ápri*cot), while in others they are not, e.g. (*egg* × *prégnant* ⇒) *égnnant* ‘pregnant with egg’ ~ *prégegg* ‘egg that counts down pregnancy’.
- 2 In this paper, we explore the conditions that result in variable base-word positioning in English blends. We argue for the relevance of the difference between synonymous vs. non-synonymous blend doublets, and endocentric vs. non-endocentric blends. In non-synonymous doublets, the main factor is that in endocentric blends, the base word that contributes the semantic head is right-aligned. For example, *égnnant* denotes a way of being pregnant, whereas *prégegg* denotes a kind of egg. By contrast, in synonymous blend doublets (where the vast majority of blends are non-endocentric), variable base-word positioning results from at least four factors interacting: (i) segmental faithfulness – maximizing segmental similarity; (ii) Pāṇini’s law – positioning the short base word before the long one; (iii) monosyllabic integrity – keeping the monosyllabic base word in the same syllable; and (iv) syntagmatic faithfulness – matching the linear order of the base words to their order in a would-be syntactic constituent.

- 3 In the § 1., we introduce the corpus of the present study. We then turn to the analysis (§ 2.), where we distinguish between non-synonymous (§ 2.1.) and synonymous (§ 2.2.) blend doublets.

## 1. Corpus: Monosyllabic Base Blends

- 4 The goal of the corpus is to determine what influences variable base-word positioning in English blends. For this, we use blends with exactly one monosyllabic base word. The rationale behind this is that the fewer segments a base word has, the more flexible its positioning potentially is. To illustrate, consider blends with the base words *entreprenéur* and either *mom* or *móther*. While *montreprenéur*, *entrepremóm* and *motherprenéur* occur in webpages as common nouns in full sentences in non-metalinguistic contexts, shown in (1), *entremother* (stress pattern unknown) does not. Thus, the monosyllabic *mom* has more flexible positioning than the disyllabic *móther*.

(1) Flexible positioning of monosyllables (*mom*)

- a. Becoming a **montreprenneur** can enable you to have truly fulfilling family and professional lives.
  - b. Ever since she could remember, **entrepremom** Rachele Chua-Villaceran's biggest dream was to have her own ice cream parlor.
  - c. For many, many years I was the most goal oriented **motherprenneur** you can think of.
- 5 Besides exactly one base word being monosyllabic, we limit our corpus in two additional ways: (i) the blend has the same number of syllables and stress pattern as the polysyllabic base word, and (ii) the monosyllabic base word is not truncated segmentally or orthographically. Blends that violate (i) include *cómpucar* (< *compúter* × *car*), with a different stress pattern from *compúter*, and *plúmcot* (< *plum* × *ápricot*), with a different number of syllables from *ápricot*. Blends that violate (ii) include *áprium* (< *ápricot* × *plum*), where *plum* is truncated segmentally and orthographically, and *bláxicen* (< *black* × *Méxican*), where *black* is truncated orthographically. Admitting such blends into the corpus would have made it more heterogenous, which might have gotten in the way of the object of study.
- 6 Our corpus consists of 407 pairs of base words, 366 of which (~90%) yield one blend, and 41 (~10%) yield two blends with variable base-word positioning, bringing the blend total to 448; 101 are from the corpus of Bat-El and Cohen [2012], and the rest (n=347) were collected specifically for this study. 306 were found by browsing the Wiktionary list of English blends from <a> to <e>, and 41 were found by searching Google for blends that would form doublets with those from the two aforementioned sources.
- 7 The blends in the sub-corpus of variable base-word positioning (n=82) have at least one webpage occurrence that meets three criteria: (i) the blend is not a proper noun with two common nouns as base words, and it occurs (ii) in a full sentence (iii) in a non-metalinguistic context (e.g. not a dictionary or question about whether the blend is a real word). These prevent the criterion for what counts as an attested blend from being too liberal: many blends occur as proper nouns (e.g. usernames, website or company names), not in full sentences (e.g. hashtags) or in user-submitted online dictionaries, but less meet the aforementioned three criteria.

- 8 From the aforementioned webpage occurrences, the meaning of the blends in the sub-corpus of variable base-word positioning is ascertained. This turns out to be relevant for base-word positioning, as shown in §2. The blends in the sub-corpus of invariable base-word positioning (n=366) are not coded for meaning.
- 9 To illustrate how the corpus was compiled, consider three blends from the corpus of Bat-El and Cohen [2012]: *assmósis* (< *ass* × *osmósis*), *guésstimate* (< *guess* × *éstitute*), and *spórtlon* (< *sport* × *nýlon*). We found no webpage occurrence of *osmósass* that meets the aforementioned three criteria, so *assmósis* is in the sub-corpus of invariable base-word positioning. Next, we found such webpage occurrences of *guésstimate* and *éstitute*, where both mean ‘guess & estimate’, so they are in the sub-corpus of synonymous blend doublets (§ 2.2.). Finally, we found such webpage occurrences of *spórtlon* and *nýlsport*, where the former means ‘athletic sock’ and the latter ‘snowmobile suit’, so they are in the sub-corpus of non-synonymous blend doublets (§ 2.1.).
- 10 The forms in our corpus are analyzable in the morphological representation of native speakers as what linguists would classify as blends, but some are morphologically ambiguous. Unlike earlier studies, which draw categorial distinctions between blends and affixed forms or clipped compounds, we recognize morphological ambiguity. For example, Bat-El and Cohen [2012: 67] write that their corpus does not include affixed forms, what they call COMBINING FORMS, but their corpus includes *irritáinment* and *digitéria*, which are analyzable as including the suffixes *-táinment* and *-téria*; the former has an entry in *Wiktionary*, and the latter has an entry in the *Oxford English Dictionary*, with citations as early as 1923. Morphological ambiguity ‘in action’ is reported in Danks [2003: 174], who recounts a scenario where her sister perceived *slugger* as a blend of *wormery* and *slug*, whereas Danks herself perceived it as an affixed form with *slug* and *-ery*. Perceiving *slugger* as a blend or affixed form is reducible to different morphological and mental representations. The affixed representation /sɫɒg-ɪ/ goes along with a mental activation of *slug* and *-ery*, whereas the blend representation /sɫɒgɪ/ goes along with a mental activation of *slug* and *wormery*. Our corpus includes such morphologically ambiguous items, and in § 2.2. we explain several cases of variable base-word positioning by appealing to word formation processes other than blending.

## 2. Data and Analysis

- 11 In this section, we present and analyze the 41 blend doublets with variable base-word positioning. 15 include non-synonymous blends, such as *éggnant* ‘pregnant with egg (bird)’ ~ *préegg* ‘egg that counts down pregnancy’ (< *egg* × *prégnant*), and 26 include synonymous blends, such as *plúmpricot* ~ *ápriplum* ‘plum-apricot hybrid’ (< *plum* × *ápricot*). In the next two subsections, we treat each sub-corpus in turn.

### 2.1. Non-synonymous blend doublets (n=15)

- 12 To discuss non-synonymous blend doublets, we introduce the notion of a semantic head (see Williams [1981] for a notion of a syntactic head in compounds and affixed forms). We begin with the doublets where both blends have a semantic head (n=7), then continue to those where one or neither has a semantic head (n=8).

### 2.1.1. Both blends are endocentric

- 13 In 7 of the 15 non-synonymous blend doublets (Table 1), both blends are such that their semantic head is contributed by exactly one base word (henceforth ENDOCENTRIC BLENDS), and this semantic head is aligned to the right edge of the blend. MonoL and MonoR in the column headers indicate blends where the monosyllabic base word is aligned to left and right edge of the blend respectively (apologies for offensive language in Table 1 and throughout).

Table 1: Non-synonymous blend doublets – Semantic head *right*-aligned

	MonoL	MonoR
a.	blogébrity    celebrity blogger	celébriblog    blog about celebrities
b.	blogógraphy    photography service for blogs	photógrablog    photography blog
c.	cúntada    Canada, derogatory	cánacunt    Canadian cunt
d.	cuntécticut    Connecticut, derogatory	connécticunt    Connecticuter cunt
e.	dateívérsary    anniversary of first date	annivérsadate    date in honor of anniversary
f.	déskercise    exercise done at desk	éxerdesk    desk for exercising
g.	éggnant    pregnant with egg (bird)	préegg    egg, counts down pregnancy

The 7 blend doublets in Table 1 indicate the activity of the constraint formulated in (2), which captures the language-specific fact about English that *soap dish* denotes a kind of dish, whereas *dish soap* denotes a kind of soap. This constraint is formulated generally enough to apply to both blend and compounds.

(2) HEADR: every base word that corresponds to a semantic head is aligned to the right edge of the form (compound or blend).

- 14 We conceive of (2) as a violable constraint (in the spirit of Optimality Theory [Prince and Smolensky 1994/2004]), rather than principle, because it is violable in blends [Shaw 2013: 58]. For example, a so-called LEFT-HEADED BLEND in Shaw's corpus is *entremaníre* (< *entreprenéur* × *manúre*), a type of entrepreneur rather than manure. Our corpus includes four left-headed blends (Table 2).

Table 2: Semantic head *left*-aligned

a.	cánacunt	Canada, derogatory	(semantic head <i>Canada</i> , not <i>cunt</i> )
b.	connécticunt	Connecticut, derogatory	(semantic head <i>Connecticut</i> , not <i>cunt</i> )
c.	obámbush	Obama, derogatory	(semantic head <i>Obama</i> , not <i>Bush</i> )
d.	tótstitute	heavily made up toddler	(type of <i>tot</i> , not <i>prostitute</i> )

Two notes are in order on the blends in Table 2. First, *cánacunt* and *connécticunt* from Table 1 reoccur in Table 2, but with different meanings. *Cánacunt* occurs (in webpages) as meaning either 'Canadian cunt' or 'Canada, derogatory', and only the former obeys HEADR. Second, of the four blends, three have at least one proper noun base word: *cánacunt* and *connécticunt* have one, and *obámbush* has two. It is unclear whether this is indicative of a larger pattern; we leave this for future research (see Moreton *et al.* [2017] for proper nouns in blends). At any rate, *tótstitute* is the odd one out, because neither base word is a proper noun.

### 2.1.1. One or neither blend is endocentric

- 15 Because HEADR is formulated as universal quantification over base words, it does not influence base-word positioning in non-endocentric blends, where both or neither base words contribute a semantic head. When both base words contribute a semantic head, as in COORDINATIVE blends like *jewfugée* ~ *refujév* ‘Jew & refugee’ (< *jew* × *refugée*), the initial base word in both blend alternatives makes the blend violate HEADR once (see Renner [2008] for coordinative compounds). Likewise, when neither base word contributes a semantic head, as in EXOCENTRIC blends like *átpersand* ~ *ámpersand* ‘the @ symbol’ (< *at* × *ámpersand*), both blend alternatives vacuously obey HEADR. Given the identical effect of HEADR in coordinative and exocentric blends, the distinction between them is irrelevant for present purposes, so we combine them under the title NON-ENDOCENTRIC blends. This relieves us from needing to determine for each non-endocentric blend whether it is coordinative or exocentric, which is not always straightforward. For example, conceiving of *bárstaurant* (< *bar* × *restaurant*) as both a bar and restaurant is a coordinative conception, but conceiving of it as a third separate category is an exocentric conception.
- 16 In the 7 doublets in Table 1, both blends are endocentric. However, in the 8 other doublets in the sub-corpus of non-synonymous doublets (Table 3), one (a-f) or neither (g-h) blend is endocentric, and those that are endocentric (shaded) obey HEADR.

Table 3: One or neither blend in doublet is endocentric

	MonoL		MonoR	
a.	<i>airvéyor</i>	pneumatic conveyor	<i>convéyair</i> <sup>TM</sup>	grain vac
b.	<i>chátiquette</i>	chatroom etiquette	<i>étichat</i>	website to chat about etiquette
c.	<i>trashmérica</i>	America, derogatory	<i>améritrash</i>	boardgame genre
d.	<i>bóokazine</i>	book-magazine hybrid	<i>mágabook</i>	magazines bounded to book
e.	<i>cártsket</i>	cart-basket hybrid	<i>báscart</i>	shopping cart
f.	<i>dúckodile</i>	reptile with duck head	<i>crócoduck</i>	duck with crocodile head
g.	<i>spórtlon</i>	athletic sock	<i>nýlsport</i>	snowmobile suit
h.	<i>artreprenéur</i>	artist entrepreneur	<i>entreprenárt</i>	program that gives artistic youth entrepreneurial skills

To reiterate, in the non-endocentric blends (not shaded) in Table 3, HEADR does not influence base-word positioning. Consider a blend that denotes a brand of athletic socks with the base words *nylon* and *sport*. An athletic sock is not a kind of nylon nor sport, so *spórtlon* and *nýlsport* both satisfy HEADR vacuously. In practice, *spórtlon* was chosen, for reasons that are irrelevant for the present study. However, what is relevant is the reason that blend doublets like *spórtlon* and *nýlsport* are attested. In cases where HEADR has no effect, it is better to examine synonymous blend doublets, because something about the meanings of ‘athletic sock’ and ‘snowmobile suit’ might support *spórtlon* and *nýlsport* respectively. For this, we turn to synonymous blend doublets in the next subsection.

## 2.2. Synonymous blend doublets (n=26)

- 17 In this subsection, the synonymous blend doublets are analyzed in three chunks, distinguished by the degree of segmental similarity between the monosyllabic and polysyllabic base words, where similarity is in terms of a proportional string edit

distance. First, we treat the cases where the monosyllabic base word is most segmentally similar to a non-initial part of the polysyllabic base word (Table 5). Next, we treat those where the monosyllabic base word is more or less equally segmentally (dis)similar to two parts of the polysyllabic base word (Table 6). Finally, we treat those where the monosyllabic base word is most segmentally similar to the initial part of the polysyllabic base word (Table 7).

- 18 In the sub-corpus of synonymous blend doublets ( $n=26$ ), the vast majority of blends are non-endocentric (all but *tóstitute* ~ *próstitot*), meaning HEADR does not affect base-word positioning. In what follows, we introduce two constraints that are known to be active in blends – segmental faithfulness and Pāṇini’s law (short before long) – and show the circumstances under which their interaction yields variable base-word positioning. Afterwards, we introduce two constraints that are apparently strong enough to override both these constraints: monosyllabic integrity and syntagmatic faithfulness.
- 19 Beginning with segmental faithfulness, recall from § 1. that our corpus is constrained as follows: (i) exactly one base word is monosyllabic, (ii) the blend has the same number of syllables and stress pattern as the polysyllabic base word, and (iii) the monosyllabic base word is not truncated segmentally or orthographically. This means that not all segments of the polysyllabic base word are preserved in the blend, with the exception of blends that are homophonous with their polysyllabic base word, such as *arkeólogist* (< *archeólogist* × *ark*). Put differently, the vast majority of these blends are segmentally unfaithful to their polysyllabic base word. There are three such ways to be unfaithful: segment-zero mismatches, segment-segment mismatches, and linearity mismatches. In what follows, we illustrate these mismatches, and explain how they enter into calculating a PROPORTIONAL STRING EDIT DISTANCE (PSED), which quantifies dissimilarity between strings [Levenshtein 1966].
- 20 To illustrate segment-zero mismatches (Figure 1), consider two blends from the sub-corpus of invariable base-word positioning: *barcáde* (< *arcáde* × *bar*) and *eyéborg* (< *cyborg* × *eye*); each has a segment-zero mismatch between it and its polysyllabic base word. In *barcáde*, the zero is in the polysyllabic base word, whereas in *eyéborg* it is in the blend. In the mismatch illustrations, the first line is the polysyllabic base word, the second line is the blend, and the monosyllabic base word is underlined.

Figure 1: Two segment-zero mismatches

<i>arcáde</i>	∅	a	ɹ	k	e	j	d	<i>cyborg</i>	s	a	j	b	ɔ	ɹ	g
<i>barcáde</i>	b	a	ɹ	k	e	j	d	<i>eyéborg</i>	∅	a	j	b	ɔ	ɹ	g

In the PSED (recounted fully in the appendix), segment-zero mismatches cost 1, and the proportional edit distance between two strings is the cost divided by the number of segments in the longer string. Thus, the PSED between *arcáde* and *barcáde* is  $1/7=0.14$ , and the same goes for *cyborg* and *eyéborg*. The PSED runs between 0 and 1, where 0 is identity and 1 is complete dissimilarity. The PSED is non-directional, so it does not distinguish between zero-segment (*arcáde* – *barcáde*) and segment-zero (*cyborg* – *eyéborg*) mismatches. See Bailey and Hahn [2005: 346] for a discussion on symmetric vs. asymmetric similarity measures.

- 21 Next, to illustrate segment-segment mismatches (Figure 2), consider *bluenette* (< *brunette* × *blue*) and *búycott* (< *bóycott* × *buy*); each has a segment-segment mismatch between the blend and its polysyllabic base word – consonantal in *bluenette*, and vocalic in *boycott*.

Figure 2: Two segment-segment mismatches

<i>brunette</i>	b	ɹ	u	n	ɛ	t	<i>bóycott</i>	b	ɔ	j	k	a	t
<i>bluenette</i>	b	ɹ	u	n	ɛ	t	<i>búycott</i>	b	a	j	k	a	t

In the PSED, segment-segment mismatches cost between 0-1, depending on the number of differing distinctive features between the segments. We use Bailey and Hahn's [2005] four-feature system for consonants (Appendix Tables 10-11), and a five-feature system for vowels (Appendix Tables 12-13). Per these systems, [ɹ-l] mismatches cost 0.2, meaning the PSED between *brunette* and *bluenette* is  $0.2/6=0.03$ , and [ɔ-a] mismatches cost 0.5, meaning the PSED between *bóycott* and *búycott* is  $0.5/6=0.08$ .

- 22 Lastly, *entrepornéur* (< *entrepreneur* × *porn*) has a rare case of a linearity mismatch (Figure 3), as well as a base word embedded in the middle of a blend. Our corpus of 448 blends has 7 with an embedded base word, the others being *margabéerta*, *adórable*, *edjewcátion*, *ambiséxtrous*, *entertóynment* and *induhvídual* (the monosyllabic base word is underlined).

Figure 3: Two linearity mismatches, one segment-segment mismatch

<i>entrepreneur</i>	a	n	f	ɹ	ɔ	p	ɹ	ɔ	n	ɹ
<i>entrepornéur</i>	a	n	f	ɹ	ɔ	p	ɔ	ɹ	n	ɹ

In the PSED, a linearity mismatch costs 1. The above blend also has the segment-segment mismatch [ɔ-ɔ], which costs 0.17, so the PSED between *entrepreneur* and *entrepornéur* is  $1.17/10=0.12$ .

- 23 Segmental similarity affects base-word positioning by favoring blends that maximize segmental similarity between them and their polysyllabic base word, as stated in (3) (for other versions, see Bat-El [1996: § 5.1.], Piñeros [2004] and Gries [2004: § 3., 2012: § 4.1.]):

(3) FAITHPOLY: maximize segmental similarity between blend and polysyllabic base word (measured by proportional string edit distance).

- 24 FAITHPOLY is a combination of several constraints, which together have the effect of similarity maximization. There is the requirement for input-output identity, which means segmental identity between the base words and blend. It is, however, impossible to achieve total identity while preserving all segments from the two base words and restricting the size of the blend to that of the polysyllabic base word. Therefore, identity is maximized but rarely achieved (except cases like the aforementioned *arkeólogist* < *archeólogist* × *ark*).
- 25 FAITHPOLY, at least partially, explains why the alternatives to the four blends in Figures 1-2 are unattested and intuitively sound bad: *\*arbár* (instead of *barcáde*), *\*cýeye*



(instead of *eyeborg*), \**brublúe* (instead of *bluenétte*) and \**bóybuy* (instead of *búycott*). FAITHPOLY might just be a partial explanation, because something additional might be wrong with these unattested blends, e.g. something phonotactically wrong with *cýeye* [sajaj]. At any rate, Table 4 shows that the four blends in Figures 1-2 are more segmentally similar to their polysyllabic base word than their unattested alternative (green indicates closer to 0, i.e. identity, and red indicates closer to 1, i.e. complete dissimilarity):

Table 4: FAITHPOLY illustration

	Poly		MonoL		MonoR (unattested)			
a.	<i>arcáde</i>	arkejd	<i>barcáde</i>	barkejd	0.14	<i>arbár</i>	arbar	0.30
b.	<i>cýborg</i>	sajborg	<i>eyeborg</i>	ajborg	0.14	<i>cýeye</i>	sajaj	0.41
c.	<i>brunétte</i>	b.runet	<i>bluenétte</i>	blunet	0.03	<i>brublúe</i>	b.rublu	0.54
d.	<i>bóycott</i>	bójkat	<i>búycott</i>	bajkat	0.08	<i>bóybuy</i>	bójbaj	0.23

Just as segmental similarity between a blend and its polysyllabic base word is a factor, so might be similarity between a blend and its monosyllabic base word. In our corpus, the PSEDs in this regard differ when the blend alternatives have a different number of segments. For example, the PSED between *plúmpricot* and *plum* is  $\frac{6}{10}=0.6$ , whereas that between *ápriplum* and *plum* is  $\frac{4}{8}=0.5$ . Thus, a FAITHMONO constraint can be formulated, parallel to FAITHPOLY in , but we have not found it necessary to appeal to FAITHMONO to explain variable base-word positioning in our corpus.

- 26 The second constraint that is known to be active in blends is Pāṇini’s law, developed originally for coordinative compounds by the ancient Sanskrit grammarian Pāṇini (circa 350 B.C.). Cooper & Ross [1975: 78] state the law as follows: “other factors being nearly equal, place 1 elements contain fewer syllables than place 2 elements.” In their study of conjoined elements (aka binomials) in English (e.g. *kit and caboodle*, \**caboodle and kit*), Pāṇini’s law has been found to be the strongest phonological principle to determine ordering, overriding six other principles [Cooper & Ross 1975: 79]. Pāṇini’s law has also been found to be active in English blends by Kelly [1998: 582], Gries [2004: 421] and Renner [2014: § 3.2.1.]. This law manifested in our corpus as the constraint stated in (4), a specific case of Pāṇini’s law.

(4) MONOL: the monosyllabic base word is aligned to the left edge of the blend.

- 27 In our corpus, there are two manifestations of MONOL. First, in the sub-corpus of invariable base-word positioning (n=366), the monosyllabic base word is left-aligned in 288 (79%) of the blends. Note that this sub-corpus includes endocentric blends like *átrogene* ‘gene that influences muscle atrophy’ (< *átrophy* × *gene*), where the monosyllabic base word *gene* is right-aligned for semantic reasons, namely to respect HEADR. If such blends were not considered, the percent of blends that respect MONOL might be higher (recall that the blends in this sub-corpus are not coded for meaning).
- 28 The second manifestation of MONOL in our corpus comes from the seven doublets in the sub-corpus of synonymous blend doublets (n=26), where the monosyllabic base word is most segmentally similar to a non-initial part of the polysyllabic base word (Table 5); in (f) a middle part, and a final part in the rest. In such cases, MONOL directly conflicts with FAITHPOLY: the former pressures the monosyllabic base word to be initial, while the latter pressures it to be non-initial. The variable base-word positioning in Table 5 is

attributed to this conflict. The number to the right of each blend is the PSED between it and its polysyllabic base word, and  $\Delta$  is the difference between the two PSEDs, which is greater than 0.05.

Table 5: MONOL and FAITHPOLY conflict

	MONOL wins		FAITHPOLY wins		$\Delta$	gloss
a.	bóybo	0.18	bímboy	0.03	0.144	boy & bimbo
b.	cóonfrican	0.28	áfricoon	0.07	0.206	African, derogatory
c.	cuntmerican	0.29	américunt	0.15	0.142	American, derogatory
d.	jewfugée	0.18	refujéw	0.05	0.133	Jewish refugee
e.	nerdtreprenéur	0.14	entreprenérd	0.09	0.052	nerd & entrepreneur
f.	pornreprenéur	0.19	entrepornéur	0.12	0.078	entrepreneur of porn
g.	tótstitute	0.16	próstitot	0.07	0.081	heavily made up toddler

Not all cases of variable base-word positioning in the sub-corpus of synonymous blend doublets can be attributed to a conflict between FAITHPOLY and MONOL, as in Table 5. Consider the eight doublets in Table 6, where the two PSEDs are more or less equal (the arbitrary cutoff point is 0.05 or less). In such cases, FAITHPOLY has no effect, because both blend alternatives yield around the same PSED. Evidently, MONOL is not the sole determining factor, because it loses in the blends in the “MONOL loses” column. In what follows, we discuss what other factors might be at play, causing MONOL to lose.

Table 6: Why does MONOL lose?

	MONOL wins		MONOL loses		$\Delta$	gloss
a.	átpersand	0.10	ámpersat	0.15	0.050	the '@' symbol
b.	bárstaurant	0.19	réstaurbar	0.20	0.007	restaurant & bar
c.	beergaríta	0.12	margabéerta	0.16	0.041	margarita & beer (mixed)
d.	bushbáma	0.30	obábush	0.29	0.013	era of Bush & Obama
e.	dógmonaut	0.10	cósmodog	0.10	0.000	cosmonaut dog
f.	foolósopher	0.06	philósofool	0.11	0.049	philosopher & fool
g.	momtreprenéur	0.11	entrepremóm	0.11	0.003	entrepreneur & mom
h.	plúmpricot	0.33	ápriplum	0.29	0.042	plum-apricot hybrid

At this point, it is fruitful to explicate the status of Pāṇini’s law (and MONOL by extension), specifically what it means to violate it. Cooper & Ross [1975] found its activity in binomials such as *kit and caboodle* vs. \**caboodle and kit*. But what is the status of the latter, i.e. what does the asterisk mean? Recall that our criteria for entering the sub-corpus of variable base-word positioning is relatively liberal: the blend must have at least one webpage occurrence where (i) it is not a proper noun with two common nouns as base words, and it occurs (ii) in a full sentence (iii) in a non-metalinguistic context. By that criteria, both *kit and caboodle* and *caboodle and kit* are attested, because the latter is attested in “The whole **caboodle and kit** stopped in the middle of the road for a minute or two while nothing much else happened.” [Marsh, 2002: 230]. In light of that, blends that violate MONOL are not ungrammatical. Rather, they are predicted to intuitively sound worse than those that obey MONOL, all other things being equal. Thus, *ápriplum* is predicted to sound worse than *plúmpricot*, if nothing else were at play besides MONOL.

- 29 To check what other factors might be at play besides MONOL, we subjected the 8 blend doublets in Table 6 to a blend preference experiment. Participants were given definitions reworded from Table 6 (e.g. *a bar and restaurant*), and had to choose from between two blends which one sounded better to them (e.g. *barstaurbar* or *restaurbar*). In all definitions, the monosyllabic base word occurred before the polysyllabic one, and in all blend options the MonoL blend occurred first. 30 self-reported native monolingual English speakers took part in the experiment, and the results are shown in Table 7.

Table 7: Blend preference experiment results (30 participants)

	MonoL	%	MonoElse	%	
a.	bushbáma	0.90	obábush	0.10	era of Bush and Obama (2000-2016)
b.	plúmpricot	0.80	ápriplum	0.20	plum-apricot hybrid
c.	montreprenéur	0.80	entrepremóm	0.20	a mom entrepreneur
d.	beergáríta	0.77	margabéerta	0.23	beer mixed with margarita
e.	bárstaurbar	0.67	réstaurbar	0.33	a bar and restaurant
f.	átpersand	0.57	ámpersat	0.43	the @ 'at' symbol, analogous with 'ampersand'
g.	foolósopher	0.47	philósfool	0.53	a fool philosopher
h.	dógmonaut	0.37	cósmodog	0.63	a dog that's a cosmonaut (Russian astronaut)

Three of the eight results in Table 7, specifically (b-d), show an 80% or so preference for the MonoL blend. Thus, we stipulate that MONOL is the sole determining factor in these cases from among the constraints considered here. The task is to explain the five results that deviate from 80%-20%.

- 30 When the preference for MonoL exceeds 80%, as in (a), there are two strategies to explain this deviance from 80%-20%: (i) the MonoL blend has an additional advantage (other than satisfying MONOL), or (ii) the MonoElse blend has an additional disadvantage (other than violating MONOL). For (a), we opt for strategy (i), and suggest that the increased preference for *bushbáma* is influenced by the correlation between temporal and linear order (President Bush came before President Obama).
- 31 Similarly, when the preference for MonoElse exceeds 20%, as in (e-h), there are two strategies to explain this deviance from 80%-20%: (i) the MonoL blend has a disadvantage that detracts from it obeying MONOL, or (ii) the MonoElse blend has an advantage that compensates for it violating MONOL. Thus, the increased preference (above 20%) for the MonoR *réstaurbar* in (e) might be influenced by the existing term *restobar* (strategy ii). As for the increased preference for the MonoR *ámpersat* in (f), perhaps this is due to dispreferring *átpersand*, because the [t] at the end of [æt] is unaspirated, making this base word harder to recognize (strategy i). Next, the increased preference for the MonoR *philósfool* in (g) might be because the monosyllabic base word occupies its own syllable, unlike in *foolósopher*, where the string [ful] is split across two syllables (strategy i). Finally, the increased preference for *cósmodog* in (h) is likely due to it being analyzable not just as a blend, but also as a prefixed form including *cosmo-*, which must be initial.
- 32 In conclusion, when FAITHPOLY has no effect (because both blend alternatives are more or less equally (dis)similar to the polysyllabic base word), variable base-word positioning can arise from MONOL interacting with several other factors. Of these, we argue that MONOSYLLABIC INTEGRITY explains further data in our corpus, reviewed next.

- 33 The third major group of blends in the sub-corpus of variable base-word positioning (n=26) are the eight doublets in Table 8, where the monosyllabic base word is most similar to the initial part of the polysyllabic base word. Thus, FAITHPOLY and MONOL both influence the monosyllabic base word to be left-aligned, but it is also attested as right-aligned.

Table 8: Why do MONOL and FAITHPOLY lose?

	MONOL, FAITHPOLY win		They lose		$\Delta$	gloss
a.	cábulance	0.22	ámbucab	0.32	0.100	non-emergency medical transit
b.	cóatigan	0.16	cárdicoat	0.28	0.120	coat-cardigan hybrid
c.	cucksérvative	0.06	consérvacuck	0.17	0.110	conservative & cuck
d.	cuntsérvative	0.09	consérvacunt	0.25	0.160	conservative & cunt
e.	guésstimate	0.13	éstiguess	0.20	0.070	estimate & guess
f.	rocktérnative	0.16	altérnarock	0.19	0.030	alternative & rock music
g.	soultérnative	0.23	altérnasoul	0.29	0.060	alternative & soul music
h.	spámvertise	0.26	ádverspam	0.42	0.160	to spam & advertise

Two notes on Table 8 are in order. First, the PSED difference in (f) is less than 0.05, so this doublet technically belongs in Table 6. We nevertheless include it here, because [ʤɑk] *rock* has an overlapping vowel with the initial part of [ɑltɹnətɪv] *alternative*. Second, we gloss (e-f) as coordinative, e.g. (f) *soultérnative* and *altérnasoul* denote a hybrid of soul and alternative music. However, the MonoR blends in (e-f) might turn out to be endocentric, in which case *altérnasoul* would denote alternative soul music, and (e-f) would belong in the sub-corpus of non-synonymous doublets. Lacking the musical expertise to distinguish between a hybrid of alternative and soul music and alternative soul music, we gloss the blends in (e-f) as coordinative for simplicity's sake.

- 34 The attestedness of the MonoR blends in Table 8 indicates which factors are apparently strong enough to override both FAITHPOLY and MONOL, which in these cases conspire to prefer the MonoL blend. In what follows, we suggest two such factors, the first being (5) monosyllabic integrity, which has already been mentioned with respect to (f) in Table 6.

(5) MONOINTEG: The edges of the monosyllabic base word correspond to the edges of the same syllable.

- 35 Previously, we suggested that MONOINTEG influenced participants in our experiment to exhibit a preference for *philósofool* that is greater than 20% (53%). In Table 8, MONOINTEG can explain the occurrence of (a) *ámbucab* and (b) *cárdicoat*, as the MonoL alternatives (*cábulance* and *cóatigan*) violate MONOINTEG.

- 36 The second constraint that we suggest is responsible for the existence of the MonoR blends in Table 8 is formulated in (6), where SYNT stands for *syntagmatic*.

(6) SYNT: if the base words can form a syntactic constituent, then their order in the blend mirrors their order in that constituent.

- 37 The doublets in Table 8 for which (6) is relevant are (c-d) and (f-g), with *consérvative* or *altérnative* as a base word. In these cases, the base words can form a syntactic constituent where the monosyllabic base word is right-aligned (e.g. *conservative cuck*, *alternative rock*), so only the MonoR blends obey SYNT.

- 38 The two MonoR blends in Table 8 that we have yet to account for are (e) *éstiguess* and (h) *ádiverspam*. As for the former, perhaps it is appealing enough to override both FAITHPOLY and MONOL because of its internal rhyme. Presently, we have no explanation for the existence of *ádiverspam*.
- 39 So far, we have treated 23 of the 26 doublets in the sub-corpus of synonymous blend doublets. This leaves the three doublets in Table 9, with *exercice* as a base word:

Table 9: Doublets with *exercice*

	MonoL	MonoR	gloss
a.	éxerdance	dáncercise	dancing exercise
b.	éxergame	gámercise	video game for exercising
c.	éxerjazz	jázzercise	jazz dancing exercise

We propose to account for the variable base-word positioning in Table 9 with the idea that *exercice* has yielded two SECRETED AFFIXES [Haspelmath 1995: 15, Fradin 2001: 46, Schoenfeld *forthcoming*]: the prefix *éxer-* and suffix *-ercise*. Secreted affixes (per the definition in Schoenfeld [*forthcoming*]) are affixes that exist in a language alongside a content word that has participated in their diachronic derivation, called the MODEL WORD. Secreted affixes are also known as COMBINING FORMS [Warren 1990, Bat-El 1996: 317, Tomaszewicz 2008, Miller 2014: § 13.], PRODUCTIVE SPLINTERS [Lehrer 1996, 2007], SPLINTER-ORIGINATING AFFIXES [Danks 2003] and RECURRING SPLINTERS [Bauer 2006]. Leading examples include *-holic* (model word *alcoholic*), as in *foodoholic* ‘person addicted to food’, and *franken-* (model word *Frankenstein*), as in *frankenfood* ‘genetically modified food’.

- 40 Rarely, a model word yields a prefix and affix. The only English example that we were aware of until now was *economics*, which has yielded the prefix *econo-* [1964], as in *econobabble* [1992] and *econo-nerds* [2003] (years are from the *Oxford English dictionary*), as well as the suffix *-onomics*, as in *bleakonomics* and *Trumponomics* (see Schoenfeld [*forthcoming*] for an analysis of complex Biblical Hebrew personal names as involving prefix-suffix pairs that originate from the same model word). Parallel to *economics*, we suggest that *exercice* has yielded the prefix *éxer-* and suffix *-ercise*, which explains the variable base-word positioning in Table 9.

## Conclusion

- 41 In this paper, we explored the conditions that result in variable base-word positioning in English blends, an issue that has not been addressed extensively in the literature of blends. We argued that in non-synonymous blend doublets, the main factor is that in endocentric blends, the base word that contributes the semantic head is right-aligned (HEADR). By contrast, with synonymous blend doublets (where the vast majority of blends are non-endocentric), variable base-word positioning results from at least four factors interacting: (i) maximizing segmental similarity between blend and polysyllabic base word (FAITHPOLY), (ii) MONOL, a manifestation of Pāṇini’s law (short before long), (iii) keeping the monosyllabic base word in the same syllable (MONOINTEG), and (iv) faithfulness of the order of the base words to their order in a would-be syntactic constituent (SYNT).

- 42 Variable base-word positioning in synonymous blend doublets is unique to blends, because in other word-formation processes, the order of the morphemes is rigid, determined mostly by lexical categories and selectional restrictions. That is, it is rare to find an affix that freely attaches to either the left or right of the base, depending on the phonological structure of the output (e.g. a CV affix, prefixed when the base begins with a consonant and suffixed when the base begins with a vowel; cf. Kim [2010]). In blends, however, phonology plays a major role in determining the order of the base words (albeit less so in endocentric blends), and phonologically-condition variation is thus expected.
- 

## BIBLIOGRAPHY

- ADAMS Valerie, 1973, *An introduction to Modern English word-formation*, London: Longman.
- 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* 13.3, 283-328.
- BAT-EL Outi, 2006, "Blend", in BROWN Keith (Ed.), *Encyclopedia of language and linguistics*, Oxford: Elsevier 2<sup>nd</sup> ed., 66-70.
- BAT-EL Outi & COHEN Evan-Gary, 2012, "Stress in English blends: a constraint-based analysis", in RENNER Vincent, MANIEZ François & ARNAUD Pierre (Eds.), *Cross-disciplinary perspectives on lexical blending*, Berlin: De Gruyter Mouton, 193-212.
- BAILEY Todd & HAHN Ulrike, 2005, "Phoneme similarity and confusability", *Journal of memory and language* 52(3), 339-362.
- BAUER Laurie, 1983, *English word-formation*, Cambridge: Cambridge university press.
- BAUER Laurie, 2006, "Compounds and Minor Word-formation Types", in AARTS Bas & MCMAHON April (Eds.), *The handbook of English linguistics*, Malden: Blackwell, 483-506.
- BRYANT, Margaret M., 1974, "Blends are increasing", *American Speech* 49, 163-184.
- CHOMSKY Noam & HALLE Morris, 1968, *The sound pattern of English*, New York: Harper and Row.
- COOPER William E. & ROSS John R., 1975, "World order", in GROSSMAN Robin E., SAN James L. & VANCE Timothy J. (Eds.), *Papers from the parasession on functionalism in linguistics*, Chicago: Chicago Linguistics Society, 63-111.
- FRADIN Bernard, 2002, "Combining forms, blends and related phenomena.", *Extragrammatical and marginal morphology* 12, 11-59.
- GRIES Stefan Th., 2004, "Shouldn't it be breakfunch? A quantitative analysis of blend structure in English", *Linguistics* 42, 639-667.

- GRIES Stefan Th., 2012, “Quantitative corpus data on blend formation: Psycho- and cognitive-linguistic perspectives”, in RENNER Vincent, MANIEZ François & ARNAUD Pierre (Eds.), *Cross-disciplinary perspectives on lexical blending*, Berlin: De Gruyter Mouton, 145-167.
- HASPELMATH Martin, 1995, “The growth of affixes in morphological reanalysis”, *Yearbook of Morphology 1994*, Dordrecht: Springer, 1-29.
- KIM Yuni, 2010, “Phonological and morphological conditions on affix order in Huave”, *Morphology* 20.1, 133-163.
- LEHRER Adrienne, 1996, “Identifying and interpreting blends: An experimental approach”, *Cognitive Linguistics* 7.4, 359-390.
- LEHRER Adrienne, 2007, “Blendalicious”, *Lexical creativity, texts and contexts* 19, 115-136.
- LEVENSHTEIN Vladimir I., 1966, “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet physics doklady* 10(8), 707-710.
- LÓPEZ RÚA Paula, 2004, “The categorical continuum of English blends”, *English studies* 85, 63-76.
- MILLER Gary D., 2014, *English lexicogenesis*, Oxford: Oxford University Press.
- MORETON Elliott, SMITH Jennifer L., PERTSOVA Katya, BROAD Rachel & PRICKETT Brandon, 2017, “Emergent positional privilege in novel English blends”, *Language* 93.2, 347-380.
- PIÑEROS Carlos-Eduardo, 2004, “The creation of portmanteaus in the extragrammatical morphology of Spanish”, *Probus* 16.2, 203-240.
- PRINCE Alan & SMOLENSKY Paul, 2004 [1993], *Optimality Theory: Constraint interaction in generative grammar*, Oxford: Wiley-Blackwell.
- RENNER Vincent, 2008, “On the semantics of English coordinate compounds”, *English Studies* 89:5, 606-613.
- SCHOENFELD Aviv, forthcoming, “Secreted affixes in Biblical Hebrew personal names”, in DORON Edit & NOTARIUS Tania (Eds.), *Brill’s Journal of Afroasiatic Languages and Linguistics*.
- TOMASZEWICZ Ewa, 2008, “Novel words with final combining forms in English. A case for blends in word formation”, *Poznań studies in contemporary linguistics* 44.3, 363-378.
- WARREN Beatrice, 1990, “The importance of combining forms”, in DRESSLER Wolfgang U., LUSCHÜTZKY Hans C., PFEIFFER Oskar E. & RENNISON John R. (Eds.), *Contemporary morphology*, Berlin: Mouton de Gruyter, 111-132.
- WILLIAMS, Edwin, 1981, “On the notions “lexically related” and “head of a word””, *Linguistic Inquiry* 12, 245-274.

## APPENDIXES

### APPENDIX: Proportional string edit distance (PSED)

The proportional string edit distance [Levenshtein 1966] runs between 0-1. 0 is identity, and 1 is complete dissimilarity. Segment-zero mismatches cost 1, linearity mismatches cost 1, and consonantal mismatches cost between 0.2-0.8, depending on number of differing distinctive features (place, manner, voicing and sonorant/obstruent), as detailed in Tables 10-11, adapted from Bailey and Hahn [2005]. Similarity, vocalic

mismatches cost between 0.17–0.83, depending on number of differing features (high, low, back, round and ATR), as detailed in Tables 12-13. Finally, the proportional string edit distance between two strings is the cost divided by the number of segments in the longer string.

Table 10 : Consonantal features

	pl	son	man	vc		pl	son	man	vc		pl	son	man	vc
p	lab	0	stop	0	t	alv	0	stop	0	dʒ	pal	0	aff	1
b	lab	0	stop	1	d	alv	0	stop	1	f	pal	0	fric	0
f	lab	0	fric	0	s	alv	0	fric	0	ʒ	pal	0	fric	1
v	lab	0	fric	1	z	alv	0	fric	1	j	pal	1	gld	1
m	lab	1	nas	1	n	alv	1	nas	1	k	vel	0	stop	0
w	lab	1	gld	1	l	alv	1	lat	1	g	vel	0	stop	1
θ	dent	0	fric	0	r	alv	1	rhot	1	ŋ	vel	1	nas	1
ð	dent	0	fric	1	ʃ	pal	0	aff	0	h	glot	0	fric	0

Table 11: Consonantal featural similarity

	p	b	f	v	m	w	θ	ð	t	d	s	z	n	l	r	ʃ	dʒ	f	ʒ	j	k	g	ŋ	h
p	0	.2	.2	.4	.6	.6	.4	.6	.2	.4	.4	.6	.8	.8	.8	.4	.6	.4	.6	.8	.2	.4	.8	.4
b	.2	0	.4	.2	.4	.4	.6	.4	.4	.2	.6	.4	.6	.6	.6	.6	.4	.6	.4	.6	.4	.2	.6	.6
f	.2	.4	0	.2	.6	.6	.2	.4	.4	.6	.2	.4	.8	.8	.8	.4	.6	.2	.4	.8	.4	.6	.8	.2
v	.4	.2	.2	0	.4	.4	.4	.2	.6	.4	.4	.2	.6	.6	.6	.6	.4	.4	.2	.6	.6	.4	.6	.4
m	.6	.4	.6	.4	0	.2	.8	.6	.8	.6	.8	.6	.2	.4	.4	.8	.6	.8	.6	.4	.8	.6	.2	.8
w	.6	.4	.6	.4	.2	0	.8	.6	.8	.6	.8	.6	.4	.4	.4	.8	.6	.8	.6	.2	.8	.6	.4	.8
θ	.4	.6	.2	.4	.8	.8	0	.2	.4	.6	.2	.4	.8	.8	.8	.4	.6	.2	.4	.8	.4	.6	.8	.2
ð	.6	.4	.4	.2	.6	.6	.2	0	.6	.4	.4	.2	.6	.6	.6	.6	.4	.4	.2	.6	.6	.4	.6	.4
t	.2	.4	.4	.6	.8	.8	.4	.6	0	.2	.2	.4	.6	.6	.6	.4	.6	.4	.6	.8	.2	.4	.8	.4
d	.4	.2	.6	.4	.6	.6	.6	.4	.2	0	.4	.2	.4	.4	.4	.6	.4	.6	.4	.6	.4	.2	.6	.6
s	.4	.6	.2	.4	.8	.8	.2	.4	.2	.4	0	.2	.6	.6	.6	.4	.6	.2	.4	.8	.4	.6	.8	.2
z	.6	.4	.4	.2	.6	.6	.4	.2	.4	.2	.2	0	.4	.4	.4	.6	.4	.4	.2	.6	.6	.4	.6	.4
n	.8	.6	.8	.6	.2	.4	.8	.6	.6	.4	.6	.4	0	.2	.2	.8	.6	.8	.6	.4	.8	.6	.2	.8
l	.8	.6	.8	.6	.4	.4	.8	.6	.6	.4	.6	.4	.2	0	.2	.8	.6	.8	.6	.4	.8	.6	.4	.8
r	.8	.6	.8	.6	.4	.4	.8	.6	.6	.4	.6	.4	.2	.2	0	.8	.6	.8	.6	.4	.8	.6	.4	.8
ʃ	.4	.6	.4	.6	.8	.8	.4	.6	.4	.6	.4	.6	.8	.8	.8	0	.2	.2	.4	.6	.4	.6	.8	.4
dʒ	.6	.4	.6	.4	.6	.6	.6	.4	.6	.4	.6	.4	.6	.6	.6	.2	0	.4	.2	.4	.6	.4	.6	.6
f	.4	.6	.2	.4	.8	.8	.2	.4	.4	.6	.2	.4	.8	.8	.8	.2	.4	0	.2	.6	.4	.6	.8	.2
ʒ	.6	.4	.4	.2	.6	.6	.4	.2	.6	.4	.4	.2	.6	.6	.6	.4	.2	.2	0	.4	.6	.4	.6	.4
j	.8	.6	.8	.6	.4	.2	.8	.6	.8	.6	.8	.6	.4	.4	.6	.4	.6	.4	0	.8	.6	.4	.8	.8
k	.2	.4	.4	.6	.8	.8	.4	.6	.2	.4	.4	.6	.8	.8	.8	.4	.6	.4	.6	.8	0	.2	.6	.4
g	.4	.2	.6	.4	.6	.6	.6	.4	.4	.2	.6	.4	.6	.6	.6	.6	.4	.6	.4	.6	.2	0	.4	.6
ŋ	.8	.6	.8	.6	.2	.4	.8	.6	.8	.6	.8	.6	.2	.4	.4	.8	.6	.8	.6	.4	.6	.4	0	.8
h	.4	.6	.2	.4	.8	.8	.2	.4	.4	.6	.2	.4	.8	.8	.8	.4	.6	.2	.4	.8	.4	.6	.8	0

Table 12: Vocalic features

	high	low	back	round	ATR		high	low	back	round	ATR		
eat	i	1	0	0	0	1	but	ʌ	0	0	1	0	1
it	ɪ	1	0	0	0	0	fort	ɔ	0	0	1	1	0
boot	u	1	0	1	1	1	bat	æ	0	1	0	0	1
foot	ʊ	1	0	1	1	0	not	ɑ	0	1	1	0	0
bet	ɛ	0	0	0	0	0	ate	e	0	0	0	0	1
Mc	ə	0	0	1	0	0	oat	o	0	0	1	1	1
burn	ɹ	0	0	1	0	0	bite	a	0	1	0	0	0



Table 13: Vocalic featural similarity

	i	ɪ	u	ʊ	ɛ	ə	ʌ	ɔ	æ	ɑ	e	o	a
i	0	0.17	0.33	0.50	0.33	0.50	0.33	0.67	0.33	0.67	0.17	0.50	0.50
ɪ	0.17	0	0.50	0.33	0.17	0.33	0.50	0.50	0.50	0.50	0.33	0.67	0.33
u	0.33	0.50	0	0.17	0.67	0.50	0.33	0.33	0.67	0.67	0.50	0.17	0.83
ʊ	0.50	0.33	0.17	0	0.50	0.33	0.50	0.17	0.83	0.50	0.67	0.33	0.67
ɛ	0.33	0.17	0.67	0.50	0	0.17	0.33	0.33	0.33	0.33	0.17	0.50	0.17
ə	0.50	0.33	0.50	0.33	0.17	0	0.17	0.17	0.50	0.17	0.33	0.33	0.33
ʌ	0.33	0.50	0.33	0.50	0.33	0.17	0	0.33	0.33	0.33	0.17	0.17	0.50
ɔ	0.67	0.50	0.33	0.17	0.33	0.17	0.33	0	0.67	0.33	0.50	0.17	0.50
æ	0.33	0.50	0.67	0.83	0.33	0.50	0.33	0.67	0	0.33	0.17	0.50	0.17
ɑ	0.67	0.50	0.67	0.50	0.33	0.17	0.33	0.33	0.33	0	0.50	0.50	0.17
e	0.17	0.33	0.50	0.67	0.17	0.33	0.17	0.50	0.17	0.50	0	0.33	0.33
o	0.50	0.67	0.17	0.33	0.50	0.33	0.17	0.17	0.50	0.50	0.33	0	0.67
a	0.50	0.33	0.83	0.67	0.17	0.33	0.50	0.50	0.17	0.17	0.33	0.67	0

## ABSTRACTS

In this paper, we explore the conditions that result in variable base-word positioning in English blends, where the same base words have variable order, yielding two blends, i.e. blend doublets. In non-synonymous doublets, such as (*egg* × *pregnant* >) *éggnant* ‘pregnant with egg’ and *prégeg* ‘egg that counts down pregnancy’, the main factor is that in endocentric blends, the base word that contributes the semantic head is right-aligned. In synonymous blend doublets, such as *plúmpricot* ~ *ápriplum* ‘plum-apricot hybrid’ (< *plum* × *ápricot*), variable base-word positioning results from at least four factors interacting: (i) segmental faithfulness – maximizing segmental similarity; (ii) Pāṇini’s law – positioning the short base word before the long one; (iii) monosyllabic integrity – keeping the monosyllabic base word in the same syllable; and (iv) syntagmatic faithfulness – matching the linear order of the base words to their order in a would-be syntactic constituent.

Cet article explore les conditions qui résultent du positionnement variable du terme source dans les amalgames anglais, dans lesquels les mêmes termes sources ont un ordre variable, créant ainsi deux amalgames différents, c’est-à-dire des doublets d’amalgames. Pour les doublets d’amalgames non synonymiques, tels que (*egg* × *pregnant* >) *éggnant* ‘pregnant with egg’ et *prégeg* ‘egg that counts down pregnancy’, le facteur principal est que dans les amalgames endocentriques, le terme source qui constitue la tête sémantique se place à droite. Pour les doublets d’amalgames synonymiques, tels que *plúmpricot* ~ *ápriplum* ‘plum-apricot hybrid’ (< *plum* × *ápricot*), le positionnement variable du terme source dépend d’au moins quatre facteurs qui se combinent : (i) la fidélité segmentale – qui permet l’optimalisation de la similarité segmentale ; (ii) la loi de Pāṇini – qui consiste à placer le terme source le plus court devant le terme source le plus long ; (iii) l’intégrité monosyllabique – c’est-à-dire la conservation du terme source monosyllabique dans la même syllable ; et (iv) la fidélité syntagmatique – qui ajuste l’ordre linéaire des termes sources selon l’ordre qu’ils auraient dans un constituant syntaxique potentiel.

## INDEX

**Keywords:** blends, variable base-word positioning, variation, endocentric, exocentric, coordinative

**Mots-clés:** amalgames, positionnement variable du terme source, variation, endocentrique, exocentrique, coordinatif

## AUTHORS

### **AVIV SCHOENFELD**

Tel Aviv University  
schoenfeld@mail.tau.ac.il

### **EVAN GARY COHEN**

Tel Aviv University  
evangarycohen@gmail.com

### **OUTI BAT-EL**

Tel Aviv University  
obatel@tauex.tau.ac.il