Brazilian Portuguese reduplication optimizes foot structure

Word count: 5036

Abstract This paper uses a corpus alongside a large-scale acceptability judgement experiment of nonce reduplicants and real reduplicants to understand phonological and non-phonological restrictions in a nominalizing reduplication process of Brazilian Portuguese. Based on responses to the reduplicants of real verbs and nonce verbs, we analyze acceptability according to foot structure of the reduplicant and vowel hiatus. Reduplicants are dispreferred when there are lapses between feet and when vowel hiatus is not resolved. Furthermore, verbs with higher token frequency have higher rates of acceptability in the reduplication process. We attribute the difference between the distribution of reduplicants and verbs overall in the language to a null parse output, and show the probability of a pronounced output based on foot structure and vowel hiatus in a maximum entropy model.

Keywords: Brazilian Portuguese, reduplication, nominalization, nonce task, acceptability judgement task, null parse, foot structure, maximum entropy

1 Verb reduplication in Brazilian Portuguese

Verbs in Brazilian Portuguese undergo a process of nominalization through total reduplication of the third person singular indicative form (Araújo 2002; Sempere 2006; Rodriguez 2018). Examples of this pattern are given in (1). Note that (1a)-(1b) show examples of consonant-initial disyllabic reduplicants, while (1c)-(1d) show examples of vowel-initial trisyllabic reduplicants.

(1) Nominalizing reduplication in Brazilian Portuguese

a.	/kɛbɾə/	'break'	→	[kɛbɾə-kɛbɾə]	'riot'
b.	/pulə/	'jump'	→	[pulə-pulə]	'trampoline'
c.	/iskõʤi/	'hide'	→	[iskõʤ-iskõʤi]	'hide-and-seek'
d.	/ĩpuhə/	'push'	→	[ĩpuh-ĩpuhə]	'pushing'

These reduplicated verbs differ from the properties of verbs overall in the language. Figure 1 below shows the distribution of overall verbs in the SUBTLEX corpus (Tang 2012). Verbs in Brazilian Portuguese are primarily trisyllabic but nominalized reduplicants, shown in Figure 2, mostly consist of disyllabic bases. In addition, disyllabic reduplicants are exclusively consonant-initial while trisyllabic reduplicants are exclusively vowel-initial. This is not true of verbs overall; both consonant-initial and vowel-initial forms are attested across all syllable counts. Consonant-initial verbs are more common regardless of phonological size in the language, but trisyllabic forms are only vowel-initial in the reduplication process (Rodriguez 2018).

Verbs in Brazilian Portuguese are primarily trisyllabic but nominalized reduplicants, shown in Figure 2, mostly consist of disyllabic bases. In addition, disyllabic reduplicants are exclusively consonant-initial while trisyllabic reduplicants are exclusively vowel-initial. This is not true of verbs overall; both consonant-initial and vowel-initial forms are attested across all syllable counts. Consonant-initial verbs are more common regardless of phonological size in the language, but trisyllabic forms are only vowel-initial in the reduplication process (Rodriguez 2018).

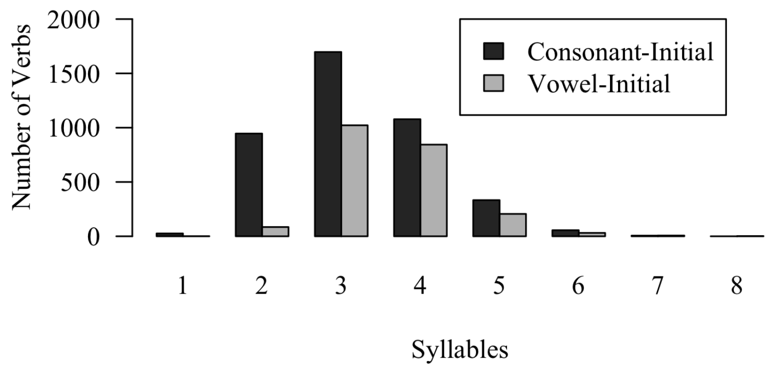

Figure 1: Distribution of verbs in the SUBTLEX corpus.

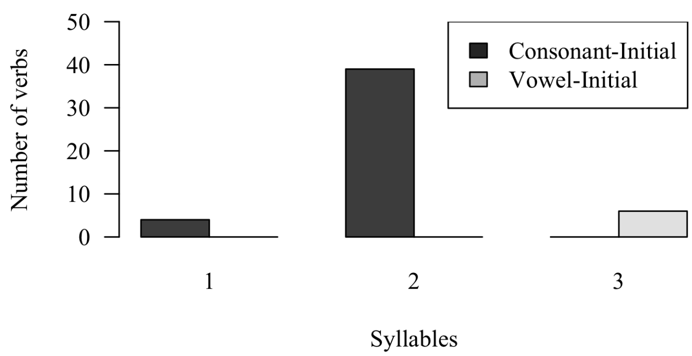

Figure 2: Distribution of verbs in the reduplicant corpus.

Based on existing reduplicants, we predict that when speakers are presented with nonce forms (*wugs*; Berko Gleason 1958) they will prefer consonant-initial disyllables and vowel-initial trisyllables. We use the constraints described in this section to demonstrate this preference in an Optimality Theory framework (Prince & Smolensky 1993; Kager 1999). Throughout this paper, we analyze Brazilian Portuguese in terms of feet. Duarte Garcia & Goad (2020) proposes that Portuguese does not have feet. The points here are compatible with both views, but we utilize constraints that are sensitive to foot structure.

We posit that consonant-initial disyllables are most preferred because they do not violate *Lapse (Green & Kenstowicz 1995), NoHiatus or Have-κ (Topintzi 2016), as shown in Table 1. The null parse candidate, represented with ⊙, only violates MParse but fails to be the output as the pronounced candidate has no violations. MParse, as detailed in (2), prohibits outputs that are not pronounced. Null parse is harmonically bounded in this tableau.

(2) MParse: Outputs must have a phonological and morphological realization (McCarthy & Wolf 2009).

/(ˈkɛ.brə)-RED/	NoHiatus	*Lapse	Have-κ	MParse
☞ a. (ˈkɛ.brə)(ˈkɛ.brə)				
b. ⊙				∗!

Table 1: Tableau of reduplication for consonant-initial disyllables.

For vowel-intial disyllables, we assume the foot structure for the base and reduplicant given in (3).

(3) Foot structure of the base and reduplicant for vowel-initial disyllables
  a. Base: (ˈV.CV)
  b. Reduplicant: (ˈV.CV) + (ˈV.CV) → V(ˈCV.CV)

In Brazilian Portuguese, vowel hiatus is resolved through the deletion of the first vowel in the sequence. This results in an unparsed vowel after reduplication. Vowel-initial disyllables violate NoHiatus when vowel deletion does not occur, as shown in candidate (a) of Table 2.

(4) NoHiatus: Outputs must not have two adjacent vowels.

Vowel-initial disyllabic reduplicants are unattested in the lexicon. Candidate (b) violates Have-κ which requires that the output minimally contain one colon, a prosodic unit composed of two adjacent feet (Stowell 1979; Topintzi 2016; Lionnet 2019). Because one vowel was deleted to resolve the hiatus, one foot is absent from the needed two to form a colon.

(5) Have-κ: Each word must minimally contain one colon (Topintzi 2016).

/('V.CV)-RED/	NoHiatus	*Lapse	Have-κ	MParse
a. ('V.CV)('V.CV)	*!			
b. V('CV.CV)			*!	
☞ c. ⊙				*

Table 2: Tableau of reduplication for vowel-initial disyllables (unattested).

Consonant-initial trisyllables are unattested in reduplication, and we predict that they are least preferred based on their foot structure, given in (6). There are two unparsed syllables in the reduplicant; one creates a lapse between two feet which is shown with an underline.

(6) Foot structure of the base and reduplicant for C-initial trisyllables
  a. Base: CV('CV.CV)
  b. Reduplicant: CV('CV.CV) + CV.('CV.CV) → CV.('CV.CV)CV('CV.CV)

Consonant-initial trisyllables are least preferred because they violate a highly ranked *Lapse constraint, which prohibits lapses between feet. These candidates additionally violate Have-κ due to the lapsed syllable creating a break between adjacent feet that prevents the formation of a colon. Although the null parse candidate violates MParse, its ranking is lower than *Lapse and Have-κ.

(7) *Lapse: Assign a violation for every unparsed syllable between feet (Green & Kenstowicz 1995).

/CV('CV.CV)-RED/	NoHiatus	*Lapse	Have-κ	MParse
a. CV('CV.CV)CV('CV.CV)		*!	*	
☞ b. ⊙				*

Table 3: Tableau of reduplication for consonant-initial trisyllables (unattested).

Finally, vowel-initial trisyllables are more acceptable in reduplication than consonant-initial trisyllables. The foot structures of the base and reduplicant for these forms, which are attested in the reduplicant corpus, are given in (8).

(8) Foot structure of the base and reduplicant for vowel-initial trisyllables
  a. Base: V('CV.CV)
  b. Reduplicant: V('CV.CV) + V.('CV.CV) → V.('CV.CV)('CV.CV)

Because vowel hiatus resolution occurs in some candidates of vowel-initial reduplicants, the unparsed syllable that creates the lapse between feet that we saw in candidate (a) of Table 3 does not appear in candidate (b) of Table 4. The pronounced reduplicant with vowel deletion does not violate any constraints. Candidate (a), which is fully faithful to the input, violates NoHiatus, *Lapse, and Have-κ. Finally, the null parse candidate violates MParse and as a result can not be the output.

/ĩ('pu.hə)-RED/	NoHiatus	*Lapse	Have-κ	MParse
a. ĩ('pu.hə)ĩ('pu.hə)	*!	*	*	
☞ b. ĩ('pu.hĩ)('pu.hə)				
c. ⊙				*!

Table 4: Tableau of reduplication for vowel-initial trisyllables.

To summarize, consonant-initial disyllables and vowel-initial trisyllables are the reduplicants that are attested and we attributed this pattern to *Lapse, Have-κ, and NoHiatus. All three constraints must be ranked higher than MParse, but do not have a definite ranking amongst themselves. When speakers are presented with an input of vowel-initial disyllables or consonant-initial trisyllables, the unpronounced null parse will be the output.

2 Acceptability judgement experiment

To test the hypothesis that reduplicant acceptability is based on the initial segment and phonological size of the base, an online acceptability judgement experiment was conducted. The experiment consisted of two blocks: one block with nonce verbs conforming to the phonotactics of Brazilian Portuguese and one block with reduplicants created from existing verbs of Brazilian Portuguese. Judgements of nonce reduplicants shed light on the phonological determinants of reduplicant acceptability, such as phonological size of the base and initial segment, while judgements of real reduplicants show non-phonological factors of acceptability, such as frequency of a verb. This experiment focuses on the acceptability judgements of reduplicants with disyllabic and trisyllabic bases, as monosyllabic stems are uncommon in the language and reduplicants larger than three syllables are unattested in the reduplicant corpus.

2.1 Participants

A total of 154 speakers of Brazilian Portuguese accessed the experiment. The results from participants who did not complete both blocks of the experiment were excluded, resulting in 100 final participants. Each participant was assigned a random Participant ID, providing anonymity to all participants.

Of the 100 participants, 89 participants self-reported all of the following demographic information. Eleven did not supply demographic information. All participants lived in Brazil at the time of taking the survey. The age of participants ranged from 18 to 67 years, with a median age of 31 years and a mean age of 33 years. Of the 89 participants who reported demographic information, 63 identified as female and 26 identified as male. The education level of participants ranged from a highschool degree (15 participants), a bachelors degree (50 participants), a masters degree (15 participants), and a doctoral degree (9 participants). When asked about any other languages that they speak, 61 participants reported varying levels of English, 30 reported some ability to speak Spanish, and 13 reported some ability to speak French.

2.2 Materials

For the nonce reduplicant block, a total of 20 nonce items following the phonotactics of Brazilian Portuguese were created and confirmed not to be words of the language by a native speaker[1]. The 20 nonce items consisted of 10 reduplicants with disyllabic bases: 5 consonant-initial and 5 vowel-initial; and 10 reduplicants with trisyllabic bases: 5 consonant-initial and 5 vowel-initial. These items were presented to all 100 participants, resulting in a total of 2000 responses. All nonce items were in the third person singular indicative form and conjugated with the [-ə] ending that indicates the as *-ar* conjugation class. The nonce items are listed in Appendix A with their total acceptability rate across all participants. Items were presented in frame sentences which were all created by the second author, who is a native speaker of Brazilian Portuguese.

A total of 20 existing verbs in the language that were not attested in the reduplicant corpus were chosen at random from the SUBTLEX corpus (Tang 2012) and presented as the items for acceptability judgement in the real

[1] In the comment portion of the demographic survey, some younger participants mentioned that the nonce item *hitar* could be used in Brazilian Portuguese as a loanword from English 'hit', meaning 'to be successful' on social media platforms and having the opposite meaning of *flopar*, also from English 'to flop'. The second author followed up with these participants and found that some pronounce *hita* with one consonant at the beginning ([hitar]) and some of them with a vowel ([itar]). Older participants were unaware of this form and meaning, and so we analyze it as [ita] for the purposes of this experiment.

reduplicant block of the experiment. These forms are listed in Appendix B. Mirroring the nonce reduplicants, 10 were disyllabic with 5 consonant-initial forms and 5 vowel-initial forms, and 10 were trisyllabic with 5 consonant-initial forms and 5 vowel-initial forms. All verbs in the present indicative tense in Brazilian Portuguese have trochaic penultimate stress (Pizzini 1983) and so stress can be eliminated as a possible influence on perceived acceptability.

2.3 Methodology

The judgement acceptability task was conducted through `Experigen` (Becker & Levine 2013) over an internet connection. Participants were recruited by word of mouth and through social media, and completed the experiment either with a laptop or their mobile phone. Participants volunteered their time and effort, and the entire experiment was presented in Brazilian Portuguese.

Each participant was presented with two frame sentences for each of the 40 items (20 nonce, 20 real). An example of a frame sentence given in the survey is shown in (9). The sentence in (9a) gives the nonce verb in an unreduplicated infinitive form to provide context. Each sentence with an unreduplicated infinitive nonce verb was then followed by a sentence with the reduplicated nominalized form, as in (9b).

(9) Frame sentences for item *vapezar*

a.

Consegui	*vapezar*	as	plantas	hoje.
get	*wug*-INF	FEM.PL	plants	today

‘I managed to *wug* the plants today’.

b.

O	*vapeza-vapeza*	durou	quase	a	manhã	toda.
MASC.SG	*wug*-NOM	last	almost	FEM.SG	morning	all

‘The *‘wugging’* lasted almost all morning’.

Participants were then asked to decide whether the reduplicated form was acceptable (*Aceitável*) or unacceptable (*Inaceitável*) as a potential word of Brazilian Portuguese. An example of the screen shown to participants for each question is given in Figure 3.

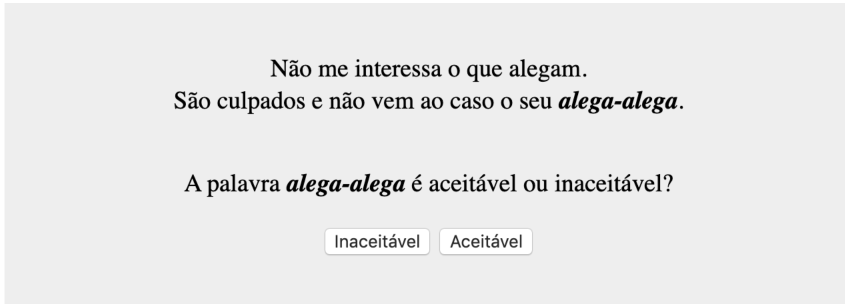

Figure 3: Example of a screen presented to participants during the real word block of the acceptance judgement task.

Immediately upon completing the nonce reduplicant block, participants were moved on to the experiment block with existing Brazilian Portuguese verbs. The procedure for the acceptability judgement task of real reduplicants was identical to that of nonce reduplicants.

At the end of the task, participants were asked to complete a demographic survey. The demographic survey included questions about participants' year of birth, sex, language profile, place of residence, and education level, as mentioned in §2.1.

2.4 Results

Overall, the results from both blocks show that consonant-initial disyllables are most favorable, with an average of 65% acceptability; followed by vowel-initial disyllables (53%), vowel-initial trisyllables (52%), and consonant-initial trisyllables (31%). Real reduplicants have higher rates acceptability than nonce reduplicants, except for when the base is a vowel-initial trisyllable.

Length	Initial	Nonce	Real	Average
2σ	C	62	69	65
2σ	V	50	57	53
3σ	C	28	35	31
3σ	V	58	46	52

Table 5: Acceptability rates for nonce and real items by length of the base and initial segment.

The acceptability rate for each nonce reduplicant is shown in Figure 4 below. With the exception of *acarra-acarra*, forms with an acceptability rate higher than 50% had disyllabic bases. Similarity of a nonce item to existing words has been found to influence the judgement of speakers, such as in Spanish stress (Face 2004). A similar reduplicant *agarra-agarra*, exists in the SUBTLEX corpus (Tang 2012) and could account for the higher acceptability rate of this form when compared to other trisyllabic bases.

It could also be the case that vowel hiatus is more easily avoided in *acarra-acarra* due to the fact that the vowels are identical and vowel coalescence occurs. In the other cases the vowels in contact are different and to avoid hiatus one of them should be deleted. However, as shown in (1), /ẽpuh-ẽpuhə/ 'a lot of pushing and shoving' is an attested reduplicant in the language and does not undergo vowel coalescence.

Forms with lower acceptability rates were almost always trisyllabic, with the exception of *empa-empa*. Overall, speakers preferred reduplicants with disyllabic bases over reduplicants with trisyllabic bases.

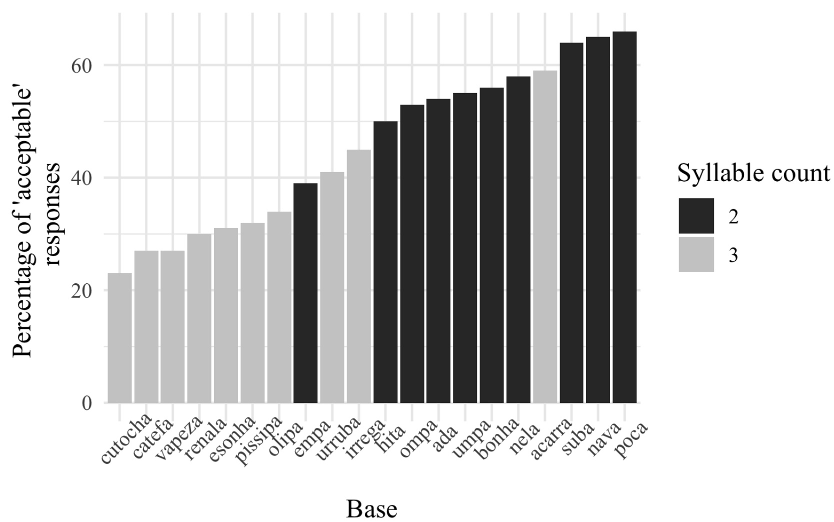

Figure 4: Acceptability rate of nonce items by phonological size of the base.

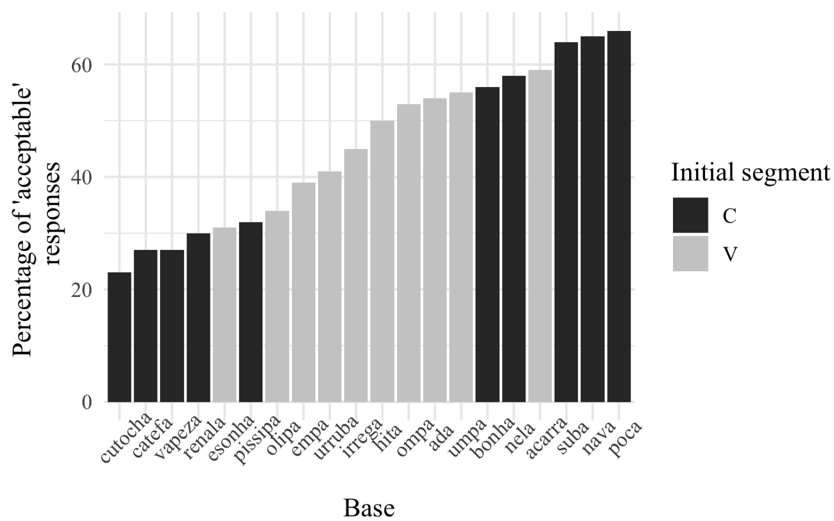

Figure 5: Acceptability rate of nonce items by initial segment of the base.

Similarly, Figure 6 shows the acceptability rate of real items. Although less distinct, real items follow the trend of being more acceptable when disyllabic and less acceptable when trisyllabic.

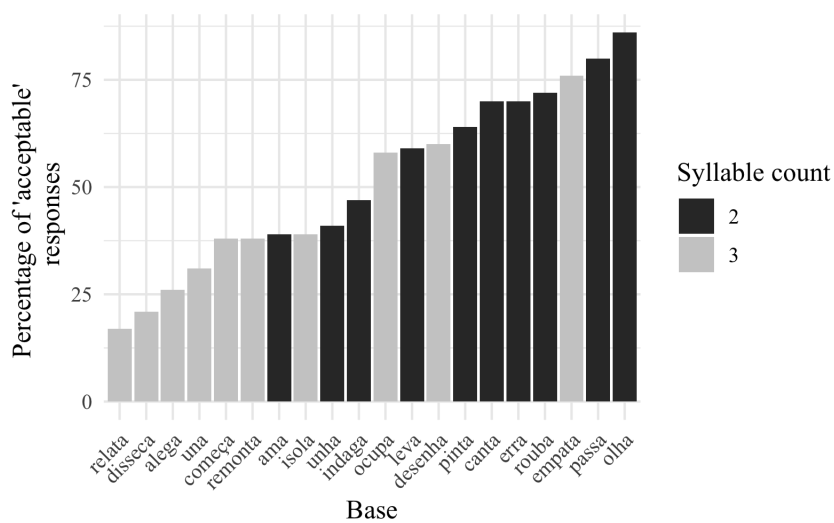

Figure 6: Acceptability rate of real item by phonological size of the base.

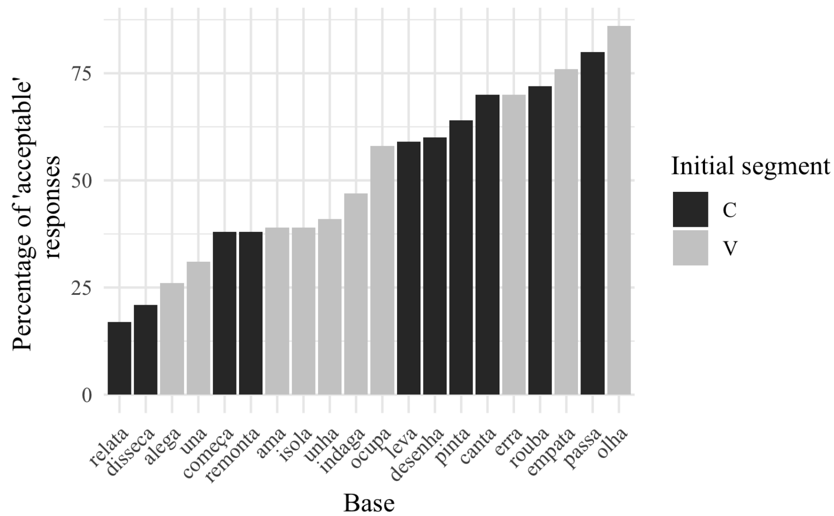

Figure 7: Acceptability rate of real items by initial segment of the base.

To fully understand the impact of base length and initial segment on acceptability, a mixed-effects logistic regression was fitted in R (R Development Core Team 2008) using the *glmer()* function in the *lme4* package (Bates et al. 2015). Random intercepts were included for participants and item in the model. Acceptability of an item was the dependent variable; phonological size in syllables, initial segment, and the interaction between phonological size and initial segment were the independent variables. Because token frequency of a verb was included as an independent variable in only real verbs, two regressions were fitted separately for nonce and real items.

The results for the nonce acceptability task are given in Table 6. Syllable count, initial segment, and the interaction between syllable count and initial segment all contribute significantly ($p < .001$) to the acceptability of a reduplicant. Reduplicants that are vowel-initial are dispreferred as well as trisyllabic reduplicants. The positive coefficient for the *syllable:initialV* interaction indicates that speakers prefer consonant-initial bases when the stem is disyllabic, and prefer vowel-initial bases when the stem is trisyllabic.

	β	SE(β)	z	p-value
(Intercept)	5.35	.52	10.37	<.001
trisyllable	–2.35	.18	–12.89	<**.001**
initialV	–4.43	.61	–7.27	<**.001**
trisyllable:initialV	1.81	.24	7.54	<**.001**

Table 6: A summary of the results from a mixed effects logistic regression using the following function: `glmer(acceptable ∼ syll + initial + syll*initial + (1|participant) + (1|item))`.

An identical model was fitted for the real items, with token frequency of the items as an additional independent variable. The toke frequency of each item was collected from the SUBTLEX corpus as an additional independent variable. The results are given in Table 7. The token frequency of a verb in the SUBTLEX corpus is statistically significant in the model (p<.001). Verbs that have higher token frequency are more acceptable in the reduplication process than verbs that have lower token frequency in the corpus. The results from the regression fitted for real items are almost identical to the results from the regression fitted for nonce items. Phonological size of the base, initial segment, and the interaction between the two are all statistically significant.

	β	SE(β)	z	p-value
(Intercept)	2.97	.57	5.22	<.001
trisyllable	–1.57	.16	–9.77	<**.001**
initialV	–2.75	.57	–4.80	<**.001**
trisyllable:initialV	1.18	.22	5.29	<**.001**
log(frequency)	–0.28	.07	–3.94	<**.001**

Table 7: A summary of the results from a mixed effects logistic regression using the following function in R: `glmer(acceptable ∼ syll + initial + syll*initial + freq + (1|participant) + (1|item))`.

3 Predicting outputs in reduplication

According to the results from the acceptability judgement tasks, consonant-initial disyllables are the most acceptable bases for reduplication, with an average of 65% acceptability. The second-most acceptable bases are vowel-initial disyllables (53%), followed by vowel-initial trisyllables (52%). Consonant-initial trisyllables are the least acceptable (31%). The results from the acceptability judgements of real and nonce words follow the same gradient of acceptability according to initial segment and phonological size of the base, but these results differ from those in the lexicon study of reduplicants from Rodriguez (2018). A comparison of acceptability is given in Table 8 below.

Base	**Lexicon**	**Nonce**	**Real**
2σ C-initial	100	62	69
2σ V-initial	0	50	56
3σ C-initial	0	28	34
3σ V-initial	100	58	46

Table 8: Acceptability of reduplicants in the lexicon vs. acceptability judgement tasks.

An examination of reduplicants in the SUBTLEX corpus reveals that only consonant-initial disyllabic bases and vowel-initial trisyllabic bases are attested in reduplication. The results from the acceptability judgement task follow this for the most part. Consonant-initial disyllabes and vowel-initial trisyllabes have higher rates of acceptability as expected, and consonant-initial trisyllables have low rates of acceptability as expected. However, as indicated in red, vowel-initial disyllables have high rates of acceptability in the judgement task but are unattested in the lexicon. We hypothesize that this is due to the methodology of the acceptability task, where participants did not hear the reduplicated forms, but rather read them. In Brazilian Portuguese, vowel hiatus is avoided in reduplicants through a deletion of the first vowel in a sequence. However, if participants are introduced to novel forms and are reading them, it is likely that vowel hiatus resolution does not occur. Furthermore, vowel-initial trisyllables are less commonly attested in the lexicon than consonant-initial disyllables, as presented in Appendix C, which can account for lower rates of acceptability of these forms despite being attested in existing reduplicants.

The range of acceptability from the judgement task according to foot structure are presented in Figure 8. We analyze the acceptability in terms of four constraints: *Lapse, Have-κ, NoHiatus, and MParse.

Acceptability ⇓

C-initial disyllable	('σσ)	('pu.lə)	→	('σσ)('σσ)	('pu.lə)('pu.lə)
V-initial trisyllable	σ('σσ)	ĩ.('pu.hə)	→	σ('σσ)('σσ)	ĩ.('pu.hẽ)('pu.hə)
V-initial disyllable	('σσ)	('ũ.pə)	→	σ('σσ)	ũ('pũ.pə)
C-initial trisyllable	σ('σσ)	pi.('si.pə)	→	σ('σσ)**σ**('σσ)	pi.('si.pə.)**pi**.('si.pə)

Figure 8: Acceptability gradient based on foot structure of reduplicants with underlined colons, and lapse syllables in bold.

As in §1, we utilize four constraints in our analysis to account for this range of acceptability: NoHiatus, *Lapse, Have-κ and MParse. NoHiatus prevents vowel-initial outputs that do not undergo vowel deletion. Have-κ makes vowel-initial disyllabic reduplicants unpreferred as they undergo vowel deletion and as a result will not have two feet to make a colon. The Lapse constraint penalizes gaps between feet, making consonant-initial trisyllabic reduplicant unfavorable. Finally, we use the MParse constraint to mandate that candidates must be phonetically pronounced (McCarthy & Wolf 2009)[2].

The analysis presented here follows Rodriguez (2018) in accounting for less acceptable reduplicants through the Null Parse theory (McCarthy & Wolf 2009) in a Maximum Entropy model (Goldwater & Johnson 2003; Hayes & Wilson 2008; White 2014). A null parse candidate, often represented as ⊙, does not have any morphological characteristics and is not phonetically pronounced. As a result, it only ever violates the MParse constraint described above. Using the MaxEnt Grammar Tool (Hayes & Wilson 2008), weights were assigned to constraints based on the acceptability of real items. The weights of each constraint are given in (9). The combined total of the violations is multiplied by the constraint weight for each candidate to get the harmony ($\mathscr{H}$). The probability (p) is calculated by dividing the exponentiated harmony by the sum of the exponentiated harmony of all the candidates.

Constraint	Weight
*Lapse	0.42
Have-κ	0.99
NoHiatus	0.98
MParse	0.66

Table 9: Weights of constraints generated from the MaxEnt Grammar Tool.

Consonant-initial disyllables are most acceptable as reduplicant bases because they do not violate *Lapse, Have-κ, NoHiatus or MParse. In the tableau in Table 10 the pronounced form does not incur any violations, but the null parse candidate violates MParse. As a result, consonant-initial disyllables have a large probability of being accepted as bases for the reduplication process, as is reflected in the results from both blocks of the acceptability judgement task.

[2] We assume that the Trochee constraint also exists, given that all verbs in this form have trochaic stress in Brazilian Portuguese. We exclude this constraint from the OT tableaux and restrict candidates to trochees for the purposes of our analysis.

/ˈpɔkə-RED/	*LAPSE $w = .42$	HAVE-κ $w = .99$	NOHIATUS $w = .98$	MPARSE $w = .66$	$\mathscr{H}$	p
a. (ˈpɔkə)(ˈpɔkə)					0	**.66**
b. ⊙				*	–.66	.34

Table 10: Consonant-initial disyllable.

Vowel-initial disyllabic bases are less favorable. In Table 11, candidate (a) violates NOHIATUS while candidate (b) violates HAVE-κ. The null parse candidate violates MPARSE, which has a lower weight than either NOHIATUS or HAVE-κ and as a result will have a higher probability of being the output than either of the pronounced candidates.

For vowel-initial candidates, speakers had two possible options: one that undergoes vowel hiatus resolution and one that does not undergo vowel hiatus resolution. Because speakers read the reduplicants, both options were possible and we separate the observed proportions applied to both candidates in the MaxEnt Grammar Tool using a hidden structure learner (Staubs 2011).

/ˈũpə-RED/	*LAPSE $w = .42$	HAVE-κ $w = .99$	NOHIATUS $w = .98$	MPARSE $w = .66$	$\mathscr{H}$	p
a. (ˈũpə)(ˈũpə)			*		–.98	.30
b. ũ(ˈpũpə)		*			–.99	.29
c. ⊙				*	–.66	**.41**

Table 11: Vowel-initial disyllable.

When a vowel-intial base is trisyllabic, the fully faithful candidate violates both *LAPSE, HAVE-κ, and NOHIATUS, and has a very low chance of being the output. In contrast, candidate (b), where vowel hiatus resolution occurs, satisfies all of these conditions and will be the most likely output. The null parse candidate only violates MPARSE.

/əˈgahə-RED/	*LAPSE $w = .42$	HAVE-κ $w = .99$	NOHIATUS $w = .98$	MPARSE $w = .66$	$\mathscr{H}$	p
a. ə(ˈgahə)ə(ˈgahə)	*	*	*		–2.38	.06
b. ə(ˈgahə)(ˈgahə)					0	**.62**
c. ⊙				*	–.66	.32

Table 12: Vowel-initial trisyllable.

Finally, consonant-initial trisyllables have the lowest rate of acceptability. Because consonant-initial forms do not have adjacent vowels, vowel resolution does not occur and an unparsed syllable remains between two feet. Consequently, the pronounced candidate violates *LAPSE, prohibiting unparsed syllables between feet, and HAVE-κ, requiring minimally one colon.

/pi'sipə-RED/	*LAPSE $w = .42$	HAVE-κ $w = .99$	NOHIATUS $w = .98$	MPARSE $w = .66$	$\mathscr{H}$	p
a. pi('sipə)pi('sipə)	*	*			–1.41	.32
b. ⊙				*	–.66	**.68**

Table 13: Consonant-initial trisyllable.

In conclusion, we have used a maximum entropy model containing the constraints from §1 to predict whether an output will be pronounced or null parse. Weights from these constraints show the gradient of acceptability displayed by speakers in the acceptability judgement task where consonant-initial disyllables were most acceptable, followed by vowel-initial trisyllables, vowel-initial disyllables, and consonant-initial trisyllables.

4 Conclusion

Through the attested reduplicants in the lexicon, we can summarize that there is a preference for consonant-initial disyllables and vowel-initial trisyllables as bases for the reduplication process. We account for this through constraints that are sensitive to the foot structure (*LAPSE and HAVE-κ) and vowel hiatus resolution (NOHIATUS). An acceptability judgement task on existing verbs and nonce verbs revealed that the acceptability gradient for reduplicants follows the pattern in the lexicon. Disyllables are preferred and consonant-initial forms are preferred overall. Trisyllabic reduplicants are preferred if the base is vowel-initial. A maximum entropy model can predict whether an output will be pronounced or null using the NOHIATUS, *LAPSE, and HAVE-κ constraints.

Reduplication is not the sole method of nominalization in Brazilian Portuguese. There are two suffixes, /-sɐ̃ʊ̃/ and /-mẽĩtu/, which nominalize verbs when attached to third person singular indicative bases, and are more productive than the reduplication pattern. A study by Silveira & Schwindt (2016) found that although the domains for the two nominalizing suffixes are not mutually exclusive, there is a general preference for /-sɐ̃ʊ̃/. The /-mẽĩtu/ suffix is preferred with verbs that end in *-er* in the infinitive form, and with verb stems containing a final sibilant.

With our approach, the three nominalization strategies compete both among themselves and with the null parse output and are expected to be sensitive to phonological size. Our analysis utilizing MaxEnt as a tool for output prediction can extend towards these ouputs as well, using constraints such as *Lapse, OCP(sib), and *mẽĩtʊ. When given all three nominalization strategies as candidates, the reduplicated form is unlikely to win due to a highly weighted violation against *Lapse.

A similar nominalizing reduplication process exists in Cuban Spanish with a categorical restriction to consonant-initial disyllabic stems, as shown in (10) (Lederer 2003).

(10) Nominalizing reduplication in Cuban Spanish

a. /ˈtira/	‘throw’	→	[ˈtira-ˈtira]	‘an instance of lots of throwing’
b. /ˈkore/	‘run’	→	[ˈkore-ˈkore]	‘an instance of running around’
c. /ˈkome/	‘eat’	→	[ˈkome-ˈkome]	‘an instance of lots of eating’
d. /ˈtoka/	‘touch’	→	[ˈtoka-ˈtoka]	‘an instance of lots of touching’

Both Cuban Spanish and Brazilian Portuguese are historically influenced by speakers of Bantu languages like Kikongo (Lederer 2003, Ferriera 2002), which has similar reduplication patterns and can explain their similar restrictions to this pattern in comparison to other Portuguese and Spanish varieties which are not known to have this reduplication pattern.

The analysis we present here shows that Brazilian Portuguese reduplication is sensitive to vowel hiatus and phonological size of the base. Both of these restrictions can be represented by *Lapse, Have-κ, and NoHiatus constraints. Furthermore, the presence of a null parse candidate in a maximum entropy model allows for gradient variation in acceptability based on these constraints, rather than strict adherence to them.

Abbreviations

c = consonant, fem = feminine, inf = infinitive, nom = nominative, pl = plural, red = reduplicant, sg = singular, v = vowel,

Acknowledgements

Redacted for submission.

References

Aluísio, Sandra, Jorge Pelizzoni, Ana Raquel Marchi, Lucélia de Oliveira, Regiana Manenti & Vanessa Marquiafável. 2003. An Account of the Challenge of Tagging a Reference Corpus for Brazilian Portuguese. In *International Workshop on Computational Processing of the Portuguese Language*, 110–117. https://doi.org/10.1007/3-540-45011-4_17.

Araújo, Gabriel. 2002. Truncamento e reduplicação no português brasileiro. *Revista de Estudos de Linguagem* 10(1). https://doi.org/10.17851/2237-2083.10.1.61-90.

Barbosa, Plínio A. & Eleonora C. Albano. 2004. Brazilian Portuguese. *Journal of the International Phonetic Association* 34(2). 227–232.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Becker, Michael & Jonathon Levine. 2013. Experigen – an online experiment platform. http://becker.phonologist.org/experigen.

Berko Gleason, Jean. 1958. The Child's Learning of English Morphology. *Yearbook of Morphology 1993* 14(2–3). 150–177.

Downing, Laura. 2006. *Canonical Forms in Prosodic Morphology*. Oxford University Press.

Duarte Garcia, Guilherme & Heather Goad. 2020. Weight effects and the parametrization of the foot: English vs. Portuguese. *Ms* https://ling.auf.net/lingbuzz/005221.

Face, Timothy L. 2004. *Perceiving what isn't there: Non-acoustic cues for perceiving spanish stress* 117– 141. Mouton de Gruyter.

Goldwater, S. & M. Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In J. Spenader, A. Eriksson & O. Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. 111–120.

Gouskova, Maria. 2007. The reduplicative template in Tonkawa. *Phonology* 24(3). 367–396. https://doi.org/10.1017/S0952675707001261.

Green, Thomas & Michael Kenstowicz. 1995. The lapse constraint. In Leslie Gabriele, Debra Hardison & Robert Westmoreland (eds.), *FLSM VI: Proceedings of the sixth annual meeting of the Formal Linguistics Society of Mid-America*. Bloomington, Indiana: Indiana University Linguistics Club.

Hayes, Bruce & Colin Wilson. 2006. Maxent Grammar Tool. http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39. 379–440. https://doi.org/

10.1162/ling.2008.39.3.379.
Ito, Junko, Yoshihisa Kitagawa & Armin Mester. 1996. Prosodic faithfulness and correspondence. *Journal of East Asian Linguistics* 5. 194–217. https://doi.org/10.1007/BF00132604.
Kager, Rene. 1999. *Optimality Theory*. Cambridge University Press.
Lederer, Jenny Simone. 2003. Reduplication in Romance: An Example from Cuban Spanish. In *Proceedings of the Twenty-Ninth Annual Meeting of the Berkeley Linguistics Society*, 95–106. Berkeley: Berkeley Linguistics Society. http://linguistics.berkeley.edu/bls/previous_proceedings/bls29S.pdf.
Lionnet, Florian. 2019. The Colon as a Separate Prosodic Category: Tonal Evidence from Paicî (Oceanic, New Caledonia). In Richard Stockwell, Maura O'Leary, Zhongshi Xu & Z.L. Zhou (eds.), *Proceedings of the 36th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project.
McCarthy, John J. 1982. Prosodic Structure and Expletive Infixation. *Language* 58(3). 574–590. https://doi.org/10.2307/413849.
McCarthy, John J. 2008. *Doing Otpimality Theory: Applying Theory to Data*. Wiley-Blackwell.
McCarthy, John J. & Alan Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory* 8(2). 209–283. https://doi.org/10.1007/BF00208524.
McCarthy, John J. & Matthew Wolf. 2009. Less than Zero: Correspondence and the null output. In *Modeling Ungrammaticality in Optimality Theory*, 17–66. Equinox Publishing.
Pizzini, Quentin. 1983. Verb stress in Portuguese. *Hispania* 66(3). 395–403. https://doi.org/10.2307/342315.
Prince, Alan. 1980. A Metrical Theory for Estonian Quantity. *Linguistic Inquiry* 11(3). 511–562.
Prince, Alan & Paul Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Wiley-Blackwell.
R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. http://www.R-project.org.
Rodriguez, Jamilläh S. 2018. Syllabic size restrictions on verb reduplication in Brazilian Portuguese. In Karee Garvin, Noah Hermalin, Myriam Lapierre, Yevgeniy Melguy, Tessa Scott & Eric Wilbanks (eds.), *Proceedings of the Fourty-Fourth Annual Meeting of the Berkeley Linguistics Society*, 257–270. Berkeley: Berkeley Linguistics Society.
Schwegler, Armin. 2000. On the (sensational) survival of Kikongo in 20th-century Cuba. *Journal of Pidgin and Creole Languages* 15. 159–164. https://doi.org/10.1075/jpcl.15.1.10sch.

Sempere, Antonio Garu. 2006. *Conflicting quantity patterns in Ibero-Romance prosody*: University of Texas at Austin dissertation.

Silveira, Luciana Morales da & Luiz Carlos Schwindt. 2016. Alternância do uso de -ção e -mento em nominalizações no português do sul do Brasil. *ReVEL, edição especial* (13). 395–403. www.revel.inf.br.

Staubs, Robert. 2011. Harmonic Grammar in R (hgR). Software package. http://blogs.umass.edu/hgr/.

Stowell, Timothy A. 1979. Stress systems of the world, unite! In *MIT Working Papers in Linguistics 1*. Somerville, MA.

Tang, Kevin. 2012. A 61 Million Word Corpus of Brazilian Portuguese Film Subtitles as a Resource for Linguistic Research. *UCL Work Pap Linguist* 24. 208–214.

Topintzi, Nina. 2016. *Iquito: The prosodic colon and challenges to OT stress accounts* 123–167. Cambridge University Press. https://doi.org/10.26262/istal.v22i0.6010.

Valdés Acosta, Gema. 2000. La herencia Bantú en el centro de Cuba: Los hechos lingüísticos. *Islas* 42(124). 22–31.

White, James. 2014. Evidence for a learning bias against saltatory phonological alternations. *Cognition* 130(1). 96–115. https://doi.org/10.1016/j.cognition.2013.09.008.

A Experiment 1: Nonce items

Item	IPA	Initial	σ	Acceptability
bonha	[ˈbõɲə]	C	2	.56
nava	[ˈnavə]	C	2	.65
nela	[ˈnɛlə]	C	2	.58
poca	[ˈpɔkə]	C	2	.66
suba	[ˈsubə]	C	2	.64
catefa	[kaˈtɛfə]	C	3	.27
cutocha	[kuˈtɔʃə]	C	3	.23
pissipa	[piˈsipə]	C	3	.32
renala	[heˈnalə]	C	3	.30
vapeza	[vaˈpɛzə]	C	3	.27
ada	[ˈadə]	V	2	.54
empa	[ˈẽpə]	V	2	.39
hita	[ˈitə]	V	2	.50
ompa	[ˈõpa]	V	2	.53
umpa	[ˈũpə]	V	2	.55
acarra	[əˈkahə]	V	3	.59
esonha	[eˈzõɲə]	V	3	.31
irrega	[iˈhɛgə]	V	3	.45
olipa	[oˈlipə]	V	3	.34
urruba	[uˈhubə]	V	3	.41

Table 14: Acceptance rates of nonce words from acceptability judgement task (n = 100).

B Experiment 2: Real items

Item	IPA	Gloss	Initial	σ	Accept	Freq
canta	[ˈkɐ̃tə]	‘sing’	C	2	.70	69
leva	[ˈlɛvə]	‘take’	C	2	.59	96
passa	[ˈpasə]	‘iron’	C	2	.80	79
pinta	[ˈpĩtə]	‘paint’	C	2	.64	51
rouba	[ˈhɔbə]	‘steal’	C	2	.72	73
ama	[ˈɐ̃mə]	‘love’	V	2	.39	77
erra	[ˈɛhə]	‘miss’	V	2	.70	42
olha	[ˈɔlʎə]	‘look’	V	2	.86	81
una	[ˈunə]	‘unite’	V	2	.31	1
unha	[ˈuɲə]	‘scratch’	V	2	.41	3
começa	[koˈmɛsə]	‘start’	C	3	.38	88
desenha	[deˈzɛ̃ɲə]	‘draw’	C	3	.60	49
disseca	[diˈsɛkə]	‘dissect’	C	3	.21	21
relata	[heˈlatə]	‘give an account’	C	3	.17	36
remonta	[heˈmõtə]	‘raise’	C	3	.38	14
alega	[aˈlɛgə]	‘plead’	V	3	.26	36
empata	[ẽˈpatə]	‘tie’	V	3	.76	17
indaga	[ĩˈdagə]	‘inquire’	V	3	.47	23
isola	[iˈzɔlə]	‘isolate’	V	3	.39	33
ocupa	[oˈkupə]	‘occupy’	V	3	.58	64

Table 15: Acceptance rates of real words from acceptability judgement task (n = 100) and token frequency from SUBTLEX corpus.

C Reduplicants attested in the lexicon

Verb	IPA	Initial	σ	Base gloss	Reduplicant gloss
cai	[ˈkaj]	C	1	‘fall’	‘the act of multiple players in a sport faking an injury and falling in order to delay the game’
dói	[ˈdɔj]	C	1	‘hurt’	‘boo-boo (child’s speech)’
põe	[ˈpõj]	C	1	‘put’	‘act of putting something somewhere repetitively’
rói	[ˈhɔj]	C	1	‘gnaw’	‘act of continuosuly gnawing’

bate	[ˈbatʃi]	C	2	‘hit’	‘bumper cars’
beija	[ˈbɛjʒə]	C	2	‘kiss’	‘a lot of kissing’
bole	[ˈbɔle]	C	2	‘move’	‘shaking’
canta	[ˈkɐ̃ntə]	C	2	‘sing’	‘a lot of singing’
cheira	[ˈʃɛjɾə]	C	2	‘smell’	‘a flatterer’
chupa	[ˈʃupə]	C	2	‘suck’	‘lollipop’
coça	[ˈkɔsə]	C	2	‘scratch’	‘back scratcher’
come	[ˈkɔ̃mi]	C	2	‘eat’	‘Cookie Monster’
corre	[ˈkɔhi]	C	2	‘run’	‘a rush’
corroi	[koˈhɔj]	C	2	‘corrode’	‘chew toy for dogs’
fecha	[ˈfɛʃə]	C	2	‘close’	‘closure of businesses when a disturbance occurs’
foge	[ˈfɔʒi]	C	2	‘run away’	‘an escape due to panic’
gira	[ˈʒiɾə]	C	2	‘turn’	‘amusement park ride’
lambe	[ˈlɐ̃mbi]	C	2	‘lick’	‘a lot of licking’
leva	[ˈlɛvə]	C	2	‘take’	‘great agitation or fuss’
lufa	[ˈlufə]	C	2	‘blow’	‘anxiety’
luze	[ˈluzi]	C	2	‘shine’	‘firefly’
marche	[ˈmaɾʃi]	C	2	‘march’	‘marching hastily’
mela	[ˈmɛlə]	C	2	‘make dirty’	‘street carnival game where people throw water, talc and gum on each other’
mexe	[ˈmɛʃi]	C	2	‘mix’	‘Scrabble’
mija	[ˈmiʒə]	C	2	‘pee’	‘a lot of peeing’
pega	[ˈpɛgə]	C	2	‘catch’	‘a game of tag’
pinga	[ˈpĩgə]	C	2	‘drip’	‘something that dribbles constantly’
pisca	[ˈpiskə]	C	2	‘flash’	‘turn signal’
pula	[ˈpulə]	C	2	‘jump’	‘trampoline’
puxa	[ˈpuʃə]	C	2	‘pull’	‘toffee’
quebra	[ˈkɛbrə]	C	2	‘break’	‘a riot; jigsaw puzzle’
range	[ˈhɐ̃ŋʒi]	C	2	‘squeak’	‘continuously grinding’
raspa	[ˈhaspə]	C	2	‘scratch’	‘shaved ice with fruit syrup’
rela	[ˈhɛlə]	C	2	‘touch’	‘the lining of somehting’
rema	[ˈhɛ̃mə]	C	2	‘row’	‘a bike that requires peddling to move’
roça	[ˈhɔsə]	C	2	‘rub against’	‘the lining of something’
rola	[ˈhɔlə]	C	2	‘exchange’	‘doing acrobatics’
serra	[ˈʃɛhə]	C	2	‘saw’	‘a lot of sawing’

tange	[ˈtẽ̃ŋʒi]	C	2	‘play; pluck’	‘continuous noise caused by instruments’
treme	[ˈtɾɛ̃mi]	C	2	‘shake’	‘a lot of shaking’
troca	[ˈtɾɔkə]	C	2	‘exchange’	‘bartering’
vira	[ˈviɾə]	C	2	‘turn’	‘twirling’
agarra	[əˈgahə]	V	3	‘grab’	‘inappropriate grabbing’
empina	[ẽˈpĩnə]	V	3	‘raise’	‘a lot of raising’
empurra	[ẽˈpuhə]	V	3	‘shove’	‘a lot of pushing and shoving’
encosta	[ẽˈkostə]	V	3	‘put down’	‘slope’
esconde	[esˈkõndʒi]	V	3	‘hide’	‘hide-and-seek’
espreme	[esˈpɾɛ̃mi]	V	3	‘squeeze’	‘a lot of squeezing’

Table 16: Attested reduplicants from the SUBTLEX corpus.