Syllabic Size Restriction on Verb Reduplication in Brazilian Portuguese

Jamilläh Rodriguez

University at Albany

jrodriguez20@albany.edu

February 10, 2018

Overview

1. Introduction of reduplication pattern
2. Observations about the pattern
3. Comparison of two corpora
4. Predictions about outputs

Introduction

- Nominalization through total reduplication of verbs (Araújo, 2002)
- Bases are 3rd sg. present indicative

  ['pɛgə] 'catch' → ['pɛgə.'pɛgə] 'a game of tag'
- Analogous pattern in Cuban Spanish (Lederer, 2003)

  ['kome] 'eat' → ['kome.'kome] 'a lot of eating'
- Not as productive as other forms of nominalization

Introduction

- Spanish: strictly limited to disyllabic bases
- Portuguese: allows 1-3 syllables
- While verbs overall in Brazilian Portuguese are primarily trisyllabic, reduplicated verbs are primarily disyllabic.

Disyllabic Restriction

- The disyllabic restriction phenomenon is present in other research:
  - Expletive infixation (McCarthy, 1982)
  - Disyllabic word minimum requirement in several Australian languages (Downing, 2006).
- And restrictions that only apply to reduplication:
  - Reduplication in Tonkawa (Gouskova, 2007)

    Word-initial syllables are usually heavy, but reduplication is restricted to word-initial light syllables.
    [naa.to$^{\text{ʔ}}$s] → [**na**-na.to$^{\text{ʔ}}$s] 'I step on it (repeatedly)'

Word-Initial Onset

- Word-initial onset is also a factor in the reduplication pattern. While monosyllables and disyllables are C-initial, trisyllables are solely V-initial.

- Reduplication in Timugon Murut (McCarthy, 2008)
  - C-initial:

    [li.mo] 'five' → [**li**.li.mo] 'about five'
  - V-initial:

    [a.ba.lan] 'bathes' → [a.**ba**.ba.lan] 'often bathes'

My Question

Questions

1. Brazilian Portuguese verbs are not primarily disyllabic, so why are disyllabic verbs the most frequently reduplicated?
2. How can we explain the difference between consonant-initial and vowel-initial reduplication restrictions?

Answer

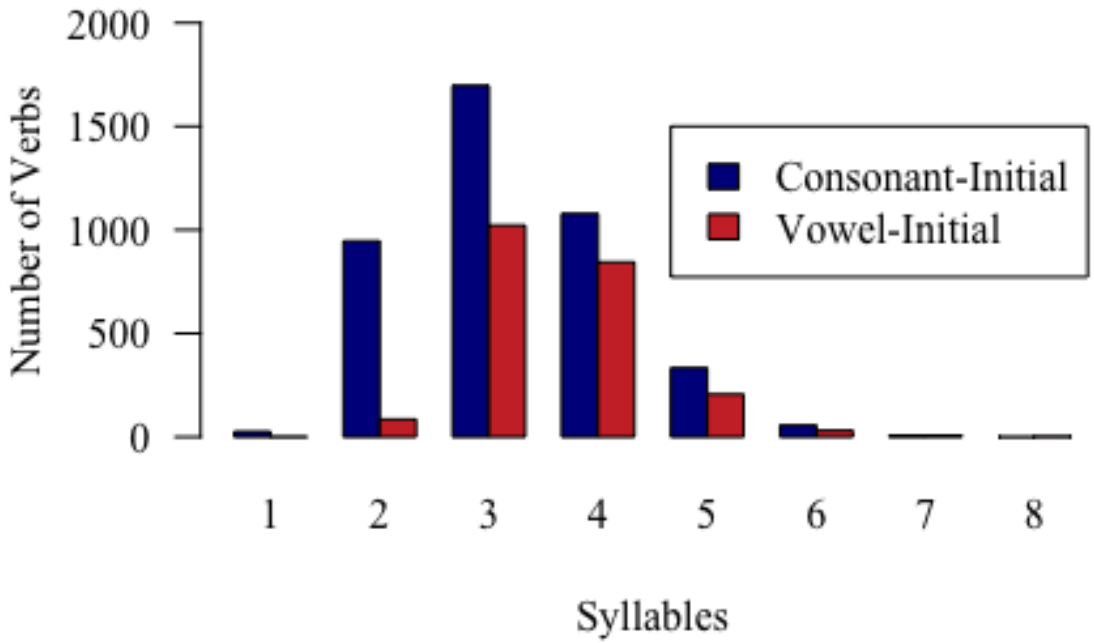
There is a size restriction on the reduplication pattern, and the difference between overall verbs and reduplicated verbs can be accounted for by mapping to the null parse (Prince and Smolensky, 2004; McCarthy and Wolf, 2010).

Corpus 1: Overall Verbs

- Corpus 1: All Brazilian Portuguese verbs
- Extracted from SUBTLEX (Tang, 2012), 61 million words from subtitles
- Separated by part of speech (R Core Team, 2013), verbs isolated and sorted by syllable count
- Manually corrected a set of verbs in which the number of syllables in the infinitive differs from the third-person singular

Corpus 1: Overall Verbs

- Total of 6,339 verbs in the corpus
- Type frequencies separated into consonant-initial and vowel-initial groups:



Corpus 2: Reduplicated Verbs

- 50 reduplicated forms collected from literature and corpus (Araújo, 2002; Gonçalves, 2004; Sempere, 2006)
  - 5 (10%) monosyllabic
  - 42 (84%) disyllabic
  - 3 (6%) trisyllabic
- Reduplicated verbs consist primarily of disyllabic bases
- Monosyllabic and disyllabic: all C-initial
- Trisyllabic: all V-initial

Null Parse Theory

- Null Parse candidate represented as O
- Output form that holds no phonological or morphological properties (McCarthy and Wolf, 2010; Prince and Smolensky, 2004)
- Accounts for unattested forms
- The only constraint that the null parse candidate violates is MParse.

MaxEnt Predictions

- MaxEnt Grammar Tool (Hayes and Wilson, 2006)
- A realistic model of the reduplication pattern should reflect the following proportions in the learner:

Table: Predicted (Overall corpus) vs. observed (Reduplicants corpus)

σ	**Initial**	**Predicted (%)**	**Observed(%)**
1σ	C	0.5	10.0
	V	0	0
2σ	C	25.0	84.0
	V	2.5	0
3σ	C	45.0	0
	V	27.0	6.0

Key Constraints

PARSE-σ

Assign one violation for every syllable that does not belong to any foot (McCarthy, 2008).

FT-BIN(σ)

Assign one violation for every monosyllabic foot (McCarthy, 2008)

MPARSE

The output must have a phonetic realization (Prince and Smolensky, 2004; McCarthy and Wolf, 2010).

Monosyllablic Verbs

- The grammar predicts that C-intial stems will have a fully faithful, totally reduplicated output.

Monosyllabic with word-initial onset, *dói* ['dɔj] 'hurt'

/'dɔj/ + **RED**	FT-Bin w = 4.23	Dep w = 8.89	MParse w = 15.75	$\mathscr{H}$	*p*
'dɔj.'dɔj	**			-8.46	**.99**
'dɔ.jə.'dɔ.jə		**		-17.78	<.01
ʘ			*	-15.75	<.01

Monosyllablic Verbs

- Unattested in the corpus
- V-initial verb stems will have a null output.

Monosyllabic without word-initial onset, *'há* [a] 'have'

/'a/ + **RED**	FT-Bin w = 4.23	Dep w = 8.89	MParse w = 15.75	Onset w = 8.99	$\mathscr{H}$	*p*
'a.'a	**			**	-26.44	<.01
'a.jə.'a.jə		****		**	-53.54	<.01
ʘ			*		-15.75	**.99**

Monosyllablic Verbs

- Potential problem because monosyllables are overrepresented in the data
- Compensated by showing violation of monosyllables in candidates of verbs that do not have only one syllable
- Without the presence of monosyllabic candidates from disyllabic/ trisyllabic inputs, $\text{FT-BIN}(\sigma)$ would receive no weight.

Disyllablic Verbs

- Like monosyllables, C-initial verb stems will have a pronounced output.

Disyllabic with word-initial onset, *pega* ['pɛ.gə] 'catch'

/'**pɛ.gə**/ + **RED**	MaxC w = 17.28	MParse w = 15.75	FT-Bin w = 4.23	$\mathscr{H}$	*p*
'pɛ.gə.'pɛ.gə				0	**.99**
'pɛ.'pɛ.gə	**		*	-38.79	<.01
ʘ		*		-15.75	<.01

Disyllablic Verbs

- Unattested in the corpus
- V-initial reduplicants will most likely be null.

Disyllabic without word-initial onset, *acha* ['a.ʃə]'find'

/'a.ʃə/ + **RED**	MAX-BR w = 5.31	MPARSE w = 15.75	FT-BIN w = 4.23	ONSET w = 8.99	$\mathscr{H}$	*p*
'a.ʃə.'a.ʃə				**	-17.98	.09
'a.ʃə.'a	**		*	*	-23.84	<.01
Θ		*			-15.75	**.91**

Trisyllablic Verbs

- Inverse of the disyllabic verbs
- Unattested in the corpus
- C-initial verbs will have a null output.

Trisyllabic with word-initial onset, *carimba* [ka.'ɾĩ.bə] 'stamp'

/ka.'ɾĩ.bə/+ RED	MaxC w = 17.28	MParse w = 15.75	Parse-σ w = 10.25	$\mathscr{H}$	p
ka.'ɾĩ.bə.ka.'ɾĩ.bə			**	-20.50	.01
ka.'ɾĩ.ka.'ɾĩ.bə	*		*	-27.53	$<$.01
ʘ		*		-15.75	**.99**

Trisyllablic Verbs

- V-initial reduplicants have a small chance of being pronounced, but will most likely have a null output.

Trisyllabic without word-initial onset, *empurra* [ĩj.'pu.hə]'shove'

/ĩj.'pu.hə/ + RED	MAX w = 0.0	MPARSE w = 15.75	PARSE-σ w = 10.25	ONSET w = 8.99	$\mathscr{H}$	p
ĩj.'pu.hə.ĩj.'pu.hə			**	**	-38.48	$<.01$
ĩj.'pu.hĩj.'pu.hə	*		*	*	-19.24	.03
Ø		*			-15.75	**.97**

- No reduplicants with more than three syllables → all have null output.

Vowel Deletion

- Vowel clusters in *agarra* → *agarra-agarra*
- What about *empurra*?

Deletion of a vowel in the base of the reduplicated form

/ĩj.ˈpu.hə/ + RED	MAX w = 0	PARSE-σ w = 10.25	MAX-BR w = 5.31	UNIFORMITY w = 10.20	$\mathscr{H}$	*p*
ĩj.ˈpu.hə.ĩj.ˈpu.hə		**			-20.50	<.01
ĩj.ˈpu.hĩj.ˈpu.hə	*	*			-10.25	**.99**
ĩj.ˈpu.həj.ˈpu.hə		*	*		-15.56	<.01
ĩj.ˈpu.hẽj.ˈpu.hə		*		*	-20.45	<.01

Conclusion

- Curated and compared two corpora
- Found differences between the corpora
  - Syllabicity
  - Word-initial segments
- Differences accounted for with null parse
- Reduplication:
  - Somewhat lenient size restriction to disyllabic verbs
  - C-initial verbs are more acceptable, except in the case of trisyllables when V-initial bases are more acceptable.

Going Forward

1. Do the intuitions of native speakers reflect the outputs predicted by the MaxEnt learner?
2. How can the over-representation of monosyllables be better accounted for?
3. Is this a real typology for disyllabic roots in reduplication?
4. How do the restrictions here compare to other processes of nominalization?

References

Araújo, G. (2002). Truncamento e reduplicação no português brasileiro. *Revista de Estudos da Linguagem*, 10(1).

Downing, L. J. (2006). *Canonical Forms in Prosodic Morphology*. Oxford University Press.

Gonçalves, C. A. (2004). Processos Morfológicos Não-Concatenativos do Português Brasileiro: Formato Morfoprosódico e Latitude Funcional.

Gouskova, M. (2007). The reduplicative template in Tonkawa. *Phonology*, 24(3):367–396.

Hayes, B. and Wilson, C. (2006). Maxent Grammar Tool. Computer software, UCLA.

Lederer, J. S. (2003). Reduplication in Romance: An example from Cuban Spanish. In *Proceedings of the 29th Berkeley Linguistics Society Annual Meeting*.

McCarthy, J. J. (1982). Prosodic Structure and Expletive Infixation. *Language*, 58(3):574–590.

McCarthy, J. J. (2008). *Doing Optimality Theory*. Blackwell Publishing.

McCarthy, J. J. and Wolf, M. (2010). *Less than Zero: Correspondence and the null output*, chapter 2, pages 17–66. Equinox Publishing.

References (cont.)

Prince, A. and Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Wiley-Blackwell.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sempere, A. G. (2006). *Conflicting quantity patterns in Ibero-Romance prosody*. PhD thesis, University of Texas at Austin.

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

Acknowledgments

I am extremely grateful to Michael Becker, Lee Bickmore, Lauren Clemens, and the faculty and graduate students of the Anthropology department at the University at Albany for their continued support, comments, and edits. Any remaining errors are my own.

Questions and Comments