Stress and gradient weight in Portuguese

Guilherme D. Garcia
guilhermegarcia.github.io

McGill University

WCCFL 33 | March 2015
Portuguese stress: previous analyses

- Weight effects on stress: only **word-finally**
- Falling diphthongs or coda segments → heavy
- Weight-sensitivity is categorical
  - Word-final syllables: either H or L
- **Syllable-based** generalizations (rules and constraints)

See Araújo (2007) for a comprehensive review
Portuguese stress: previous analyses

- **Weight effects on stress**: only *word-finally*
  - Falling diphthongs or coda segments $\rightarrow$ heavy
  
- **Weight-sensitivity is categorical**
  - Word-final syllables: either H or L

- **Syllable-based** generalizations (rules and constraints)
  
  See Araújo (2007) for a comprehensive review
Today: A novel analysis

- Weight effects → across the whole stress domain ✓
- Weight-sensitivity is gradual ✓
- Probabilistic interval\(^1\)-based account ✓
  - Economy, accuracy, empirical motivation

\(^1\)Steriade 2012
I

Portuguese stress
Intervals and syllables

II

Questions
Methods
Analysis

III

Summary
Final remarks
Portuguese main stress

- **Non-verbs** ⇒ phonological factors + exceptional cases
  - Trisyllabic window: $\sigma\sigma\sigma]_{PWd}$
  - *ma.rí.ti.mo, ca.vá.lo, fa.ról* ‘maritime, horse, lighthouse’
  - final and penult stress: (mostly) regular
  - antepenult stress: irregular/unpredictable
  - Weight effects only word-finally (See Araújo (2007) for a review)

- Verbs ⇒ morphological factors (Wetzels 2007)
Non-verbs

Regular patterns: 72% of the Portuguese Stress Corpus (Garcia 2014; Houaiss et al. 2001, n=154,083)

- Final stress (≈15%): papél ‘paper’
- Penult stress (≈57%): caválo ‘horse’

- Standard view: Heavy σs only attract stress word-finally
Non-verbs
Irregular cases: 28% of the lexicon

\[ \sigma \sigma \sigma \approx 13\% \quad \text{pérola} \quad \text{‘pearl’} \]
\[ \sigma \sigma \sigma \approx 11\% \quad \text{café} \quad \text{‘coffee’} \]
\[ \sigma \sigma \sigma \approx 4\% \quad \text{nível} \quad \text{‘level’} \]

▶ Previous work: Marked stress, extrametricality, catalexis and morphological factors (Roca 1999; Araújo 2007)
Portuguese stress

- Previous approaches:
  i. weight domain $= \sigma_{\text{rhyme}}\#$
  ii. weight is categorical, word-final and binary: H or L
  iii. onsets do not contribute to weight

- Possible issues:
  - Antepenult stress restricted to $\text{XLL}_{PWd}$ words. Why?
    i.e., *XHL, *XLH, *XHH
  - Recent studies on onset effects
    (Gordon 2005, Topintzi 2010, Ryan 2014)

Do onsets affect stress in Portuguese? If so, how?
Intervals and syllables
Onsets and weight

- Prediction: no effect or a *positive* correlation $CC\checkmark \geq C\checkmark$

- *Unexpected*: $C\checkmark > CC\checkmark$, i.e., *negative* correlation

- Preview: *This* is found in the Portuguese lexicon (Houaiss et al. 2001)
Intervals and syllables

(LLL words)

Penult onset

Stress: • Antepenult ▲ Penult ■ Final

Penult onset size

Garcia 2015 Stress and gradient weight in Portuguese
Intervals and syllables

(LLL words)

Penult onset

Final onset

Stress: ■ Antepenult ▲ Penult ■ Final

Stress: ■ Antepenult ▲ Penult ■ Final

Penult onset size

Final onset size
Intervals and syllables

Defining intervals (Steriade 2012)

Facts are consistent with *interval* theory

- Interval: rhythmic unit that spans from one V up to the next V: no constituency assumed

  \[ \text{CVC}_{\sigma_2} \text{CCVC}_{\sigma_1} \Rightarrow \langle C \rangle \text{VCCC}_{\iota_2} \text{VC}_{\iota_1} \]

- More segments in one \( \iota \Rightarrow \) longer duration \( \Rightarrow \) more weight

- Intervals argued to be the weight domain in unrelated languages: Finnish, Norwegian, Greek, Latin, Bhojpuri and Estonian

*How about Portuguese?*
Questions

i How does weight-sensitivity affect stress in Portuguese? 
   categorical or gradual? 
   word-final or broader in the stress domain?

ii How do syllables and intervals differ in answering question i?
Methods

The Portuguese lexicon (Houaiss et al. 2001)

▶ Houaiiss Dictionary \((n = 442,000)\)
▶ Non-verbs: \(n = 154,083\)
▶ Representativeness: spoken vs. written Portuguese two frequency corpora: similar proportions of stress patterns (penult > final > antepenult)

▶ What factors best predict stress position?
Methods

Factors: stress domain ($\sigma\sigma\sigma | \iota\iota$)

Syllable

- Onset: 0-2
  - Nucleus: 1-2
  - Coda: 0-2

Interval

- VCCC 4
- VCC 3

Garcia 2015

Stress and gradient weight in Portuguese
Methods

Data: **final** syllable by stress position

### Onset

- **Stress:** Final, Penult, Antepenult
- **Size:** 0%, 25%, 50%, 75%, 100%
- **Graph:**

### Nucleus

- **Stress:** Final, Penult, Antepenult
- **Size:** 1, 2
- **Graph:**

### Coda

- **Stress:** Final, Penult, Antepenult
- **Size:** 0, 1
- **Graph:**

Garcia 2015

Stress and gradient weight in Portuguese
Methods

Data: penult syllable by stress position

Onset

Nucleus

Coda

Garcia 2015

Stress and gradient weight in Portuguese
Methods

Data: **antepenult** syllable by stress position

Onset

Nucleus

Coda

Garcia 2015

Stress and gradient weight in Portuguese
Methods

Data: stress pattern by interval

3 (ant)

2 (pen)

1 (fin)

Garcia 2015

Stress and gradient weight in Portuguese 20 / 40
Methods

Variables predicting stress: Ordinal Regression* (clm() in R)

Syllable model
9 predictors \( (3/\sigma) \)

- onset (y/n)
- nucleus (monoph-/diphthong)
- coda (y/n) (binary)

Interval model
3 predictors \( (1/\iota) \)

- int.0
- int.1
- int.2

Response:
\[ 3 > 2 > 1 \] _P_{wd}
Methods: Model evaluation

- **Accuracy**: What % of the lexicon is accurately predicted?
- **Consistency**: Which model is more theoretically consistent?
- **Fit**: Which model has a better fit (AIC, fit vs. df)?
Analysis: Syllable model

In what positions does weight affect stress?

What has been argued: CVC.CVC.CVC

Actual patterns: CV.C. CVC. CVC
Analysis: Syllable model

- Onset behaviour: Inconsistent with syllable theory

CVC—C_2 VC—C_1 VC#

1. Final onsets $\rightarrow$ penult stress $\hat{\beta} = 1.52 \quad p < 0.00001$
2. Penult onsets $\rightarrow$ antepenult stress $\hat{\beta} = 0.50 \quad p < 0.00001$
Analysis: Interval model

- Longer intervals $\rightarrow$ higher stress likelihood
- Weight gradience: **between** and **within** interval

<table>
<thead>
<tr>
<th>Interval</th>
<th>Stress Type</th>
<th>$\hat{\beta}$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>int1</td>
<td>final stress</td>
<td>$-2.01$</td>
<td>$&lt; 0.00001$</td>
</tr>
<tr>
<td>int2</td>
<td>penult stress</td>
<td>$-0.25$</td>
<td>$&lt; 0.00001$</td>
</tr>
<tr>
<td>int3</td>
<td>antepenult stress</td>
<td>$0.32$</td>
<td>$&lt; 0.00001$</td>
</tr>
</tbody>
</table>
Question i: Weight gradience (regression weights by model)

Syllables

Intervals

Constituent: Coda Nucleus Onset

Predicted effect

Antepenult Penult Final

β

Garcia 2015

Stress and gradient weight in Portuguese
Question ii: Syllables vs. intervals

<table>
<thead>
<tr>
<th>Domain</th>
<th>Predictors</th>
<th>AIC</th>
<th>$\kappa$</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable</td>
<td>9</td>
<td>186433.21</td>
<td>40.79</td>
<td>74.75%</td>
</tr>
<tr>
<td>Interval</td>
<td>3</td>
<td>180035.58</td>
<td>15.38</td>
<td>78.29%</td>
</tr>
</tbody>
</table>

Intervals:
- More accurate
- More empirically motivated
- More parsimonious
Summary of proposal

- Weight effects on stress are gradual, **not** categorical
- Intervals offer a better analysis of stress in Portuguese
- Onset effects are accurately captured

- Crucially: only **one** predictor, namely, **weight**
- **Probabilistic:** compatible with constraint-based grammars
Final remarks

> Weight effects in Portuguese are *gradient*
> Onset effects suggest that weight is computed in intervals
> Advantages of a probabilistic interval-based approach:
  > Economy, empirical motivation and accuracy
  > Nearly 80% accuracy vs. 72% regular cases
  > Only one positionally-defined predictor/constraint
> Probabilistic constraint-based implementation (E.g., MaxEnt)
Main references


Main references


Many thanks to Heather Goad, Morgan Sonderegger, Kie Zuraw, Michael Wagner and Colin Wilson.

Thank you | Obrigado

guilherme.ca
### Appendix I: Syllable model

| predictor   | $\hat{\beta}$ | $e^{|\hat{\beta}|}$ | se($\hat{\beta}$) | z value | p value   |
|-------------|----------------|----------------------|-------------------|---------|-----------|
| onset.fin   | 1.52           | 4.59                 | 0.02              | 53.97   | < 0.00001 |
| nucleus.fin | -2.86          | 17.55                | 0.02              | -137.99 | < 0.00001 |
| coda.fin    | -4.68          | 108.10               | 0.02              | -194.62 | < 0.00001 |
| onset.pen   | 0.50           | 1.65                 | 0.02              | 23.75   | < 0.00001 |
| nucleus.pen | -1.32          | 3.75                 | 0.02              | -46.18  | < 0.00001 |
| coda.pen    | -1.09          | 3.00                 | 0.01              | -63.85  | < 0.00001 |
| onset.ant   | 0.24           | 1.27                 | 0.02              | 11.64   | < 0.00001 |
| nucleus.ant | -1.06          | 2.90                 | 0.02              | -40.50  | < 0.00001 |
| coda.ant    | -0.02          | 1.02                 | 0.01              | -1.646  | 0.0999    |

$\theta = \{-1.64, 3.34\}$

AIC: 186433.21
Accuracy: 74.75% $\kappa = 40.79$
Appendix II: Interval model

| predictor | $\hat{\beta}$ | $e|\hat{\beta}|$ | se($\hat{\beta}$) | $z$ value | $p$ value |
|-----------|----------------|--------------------|-------------------|-----------|-----------|
| int1      | $-2.01$       | $7.46$             | $0.01$            | $-196.60$ | $< 0.00001$ |
| int2      | $-0.25$       | $1.28$             | $0.01$            | $-35.33$  | $< 0.00001$ |
| int3      | $0.32$        | $1.38$             | $0.01$            | $43.82$   | $< 0.00001$ |

$\theta = \{-2.21, 2.63\}$  \hspace{1cm} \text{AIC: 180035.58} \hspace{1cm} \text{Accuracy: 78.29\%} \hspace{1cm} \kappa = 15.38$
Appendix III: Edge effects

\[ \langle C \rangle VC \text{ vs. } [CVC] \rightarrow \text{No onset extrametricality in the lexicon} \]

- Onsets at the left edge have a *positive* effect on stress. Why?
- Hypothesis: Parsing is *exhaustive*
- Test: words with 2, 3 and 4 syllables
If onsets have to be parsed...

Overall, penult onsets have a **negative** effect on penult stress.

How about disyllabic words? Do we find a null effect?

\[
2\sigma ([CV\ldots]) \quad \hat{\beta} = 0.18 \quad p < 0.002 \\
> 2\sigma \quad \hat{\beta} = -1.39 \quad p < 0.00001
\]

Onsets have a **positive** effect at the left edge of the domain.
Appendix IV: Constraint-based approach

An adaptation of the Weight-to-Stress Principle (Prince 1990)

Let $\iota_u$ be an unstressed interval:

$WSP_{\iota_u}$ Assign one $*$ to every segment in $\iota_u$

Example: MaxEnt Grammar (Hayes & Wilson 2008)
MaxEnt: Lexical modelling

▶ Lexicon = \{\text{different combinations of interval sequences}\}
▶ So, amostragem ‘sampling’ ⊆ \{4-2-2\} (VCCC-VC-VC)
▶ Assume inputs = unique sequences
▶ Each input includes a non-empty set of words
▶ Sets of words (inputs) have different stress patterns
▶ Problem: sets → very skewed lexical distribution
MaxEnt
Learned weights for all words in Portuguese

<table>
<thead>
<tr>
<th>Constraint</th>
<th>WSP&lt;sub&gt;3&lt;/sub&gt;</th>
<th>WSP&lt;sub&gt;2&lt;/sub&gt;</th>
<th>WSP&lt;sub&gt;1&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>0.00</td>
<td>0.27</td>
<td>0.83</td>
</tr>
</tbody>
</table>

- Weight gradience across the stress domain
- Antepenult interval: No weight effect learned
  - Lexical skewness: 37 out of 86 inputs $\rightarrow$ 99% of the lexicon
  - Model treats each input *equally*
MaxEnt

Example

<table>
<thead>
<tr>
<th>/VCCCVCVC/</th>
<th>WSP₁</th>
<th>WSP₂</th>
<th>WSP₃</th>
<th>H(x)</th>
<th>MaxEnt</th>
<th>P(x)</th>
<th>Freq(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[VCCC•VC•VC]</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0.54</td>
<td>0.58</td>
<td>0.66</td>
<td>0.63</td>
</tr>
<tr>
<td>[VCCC•VC•VC]</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>1.66</td>
<td>0.19</td>
<td>0.22</td>
<td>0.36</td>
</tr>
<tr>
<td>[VC•CC•VC]</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2.2</td>
<td>0.11</td>
<td>0.13</td>
<td>0.01</td>
</tr>
</tbody>
</table>