This paper proposes a novel analysis of stress in Portuguese based on weight gradience. Previous analyses have argued that weight effects in Portuguese are categorical and restricted to the word-final syllable. I show that weight effects in the Portuguese lexicon are **gradient**, and can be found in all three positions in the stress domain (trisyllabic window). These effects are statistically significant whether we consider syllables or intervals as the domain of weight computation. I suggest that a statistical analysis based on intervals is more economical, more empirically motivated and more accurate. Finally, I demonstrate how this analysis can be mapped into a Maximum Entropy (MaxEnt) grammar. Crucially, the vast majority of words in the lexicon are accounted for (≈ 80%) by one single predictor, namely, weight—including a portion of irregular words, where stress is unpredictable. Unlike previous analyses, no abstract notions such as extrametricality and catalexis are used.

**Background**

**Weight:** In languages where stress is weight-sensitive, syllables with greater weight are more likely to be prominent, i.e., to attract stress. In interval theory, greater weight entails greater duration in a given interval, defined as the rhythmic unit that spans from a vowel up to (but not including) the following vowel. Segments preceding the leftmost vowel are not included in any interval—though such segments may have an impact on the interval if we assume the p-center model. Intervals have no a priori constituency, and predict different rhythmic units when compared to syllables: onset segments in a given syllable are computed as part of the preceding interval. For example, the syllable string CVC\_CCVC\_CCVC\_ seems to be better captured by intervals, as they are negatively correlated with stress.

**Portuguese:** Previous analyses have argued that weight effects on stress in Portuguese non-verbs are restricted to the word-final syllable (stress in verbs is not phonologically conditioned): stress is final if the word-final syllable is heavy. Otherwise, stress falls on the penultimate syllable (regardless of weight). Both final and penultimate stress patterns are (mostly) regular. Antepenultimate stress is irregular, and no pre-antepenultimate stress is allowed. Approximately 72% of the non-verbs in the language (N=163,626) have regular/predictable stress. Scholars have employed different factors to account for stress regularities (e.g., foot binarity, foot type, metrical alignment) and irregularities (e.g., extrametricality, catalexis, theme vowel influence) in the language.

**Methodology**

I examine a large subset of the Houaiss corpus of Portuguese (n = 137,970), and model stress likelihood with Ordinal Regressions in R based on syllables and intervals. The stress domain was treated as a 3-point scale, where the antepenultimate (3) and final (1) positions demarcate the end points of the scale. For comparison, two binomial Logistic Regressions in R were also used in a separate analysis, which confirmed the results present in the Ordinal Regressions used in this study. Phonotactic patterns were controlled for to avoid unnatural words. As well, borrowings and monosyllables were excluded. Intervals and syllables are compared side by side, each with its own statistical model. **onset**, **nucleus** and **coda** were included as (binary) predictors.
in the syllable-based model ($\times 3 = 9$ predictors). The interval-based model included interval size as the predictor ($\times 3 = 3$ predictors): $\text{int3}$ (antepenult), $\text{int2}$ (penult) and $\text{int1}$ (final).

**Results**

Both syllable- and interval-based models confirmed the gradual effect of weight on stress. Results show that weight effects are highly significant in all positions in the stress domain ($p < 0.00001$), and weaken as we move away from the right edge of the word. For the interval model, effect sizes were: $\text{int3}$ ($\hat{\beta} = 0.315$), $\text{int2}$ ($\hat{\beta} = -0.216$) and $\text{int1}$ ($\hat{\beta} = -2.184$); positive coefficient values indicate increased likelihood of antepenult stress. In other words, the more segments an interval contains, the more likely that interval is to attract stress.

Onset effects in the syllable model confirm the negative correlation observed in the data, indicating that the interval-based model is more empirically motivated—these models also have fewer predictors, by definition, since intervals have no internal structure *a priori*. The interval-based model also presents (i) a lower AIC value, which indicates a better fit, and (ii) higher accuracy (79.32% vs. 76.71%) when compared to the syllable-based model.

Finally, the interval-based analysis proposed in this paper can be mapped into a MaxEnt grammar, where constraints are weighted. I assume a modified version of the Weight-to-Stress Principle [9, p. 3]: wsp$_n$, where $n$ represents a given position in the stress domain. wsp$_n$ assigns one violation mark to each segment in $n$ **if and only if** $n$ is unstressed. In this grammar, inputs represent unique sequences of intervals in the lexicon, and outputs represent possible stress patterns ($N = 3$). Constraint weights were determined using the MaxEnt Grammar Tool [13]: wsp$_3 = 0.14$, wsp$_2 = 1.46$ and wsp$_1 = 2.05$. As expected, this analysis also captures the weight gradience found in the lexicon. Importantly, predicted probabilities and observed frequencies are highly correlated: $r(85) = 0.871, p < 0.00001$.

Therefore, this novel analysis shows that (i) weight is *not* constrained to the word-final position; (ii) weight is gradient; and (iii) weight alone is able to account for stress in the vast majority of words in the language. As a result, this proposal is more economical, more empirically motivated and more accurate than previous analyses, which had to employ different mechanisms to account for the regular and irregular patterns found in the language.

**References**