

RESEARCH

Antepenultimate stress in Spanish: In defense of syllable weight and grammatically-informed analogy

Martín Fuchs

Yale University, 370 Temple St., New Haven, CT, US
martin.fuchs@yale.edu

Spanish has a contrastive stress system with three major possibilities: antepenultimate, penultimate, and final stress. While penultimate and final stress are to some extent predictable, a major point of contention in the literature is whether antepenultimate stress assignment is rule-governed (Harris 1983; Roca 1991; *i.a.*). By examining different analogical and grammatically-informed models and their predictive power in capturing experimental data, I show that a Maximum Entropy model (Hayes & Wilson 2008) that includes syllable weight in its lexical representations is the best predictor of antepenultimate stress assignment. In doing so, I also dispute the claim that the trill in Spanish is a geminate tap (Harris 1983), and provide support for its status as a singleton consonant.

Keywords: phonotactics; Spanish stress; syllable weight; analogical models; maximum entropy models

1 Introduction

Stress in Spanish is contrastive, given that words with identical segmental content and of the same lexical category can have different stress patterns while having different meanings. Minimal pairs such as [ˈsa.βa.na] ‘sheet’ ~ [sa.ˈβa.na] ‘savannah’ show this difference between antepenultimate stress in the former case and penultimate stress in the latter.

The distribution of stress is thus said to be unpredictable, so that nonverbal word stress in Spanish must be lexically encoded.¹ However, as Harris (1983; 1991) points out, although lexically codified information is definitely necessary for cases in which stress is contrastive or unpredictable, stress assignment in Spanish is not completely free and follows important restrictions. Several generalizations about Spanish stress have been made. For instance, primary stress needs to fall on one of the last three syllables of the word (Harris 1983; Roca 1991; *i.a.*), which yields a set of three possible stress types: *antepenultimate* stress (e.g., [ˈka.ma.ra] ‘camera’), *penultimate* stress (e.g., [ka.ˈðe.na] ‘chain’), and *final* stress (e.g., [ka.ma.ˈrin] ‘dressing room’). Another well-attested pattern is the unmarked case of stress assignment (Harris 1983; 1991; Roca 1991; Lipski 1997): setting aside inflectional endings, vowel-final words are generally stressed on the penultimate syllable (e.g., [ˈma.no] ‘hand’), while consonant-final words are usually stressed on the final one (e.g., [kan.ˈsjon] ‘song’).

Antepenultimate stress, on the other hand, seems to follow a more nuanced pattern. The claim in the literature is that Spanish does not allow for stress to fall on the antepenultimate syllable when the penultimate syllable is heavy (i.e., contains a branching

¹ Spanish exhibits different stress patterns for verbal and nonverbal forms. The verbal stress system is fully predictable (e.g., Harris 1989; Roca 1991), so the focus of this paper is restricted to stress assignment in nonverbal forms.

rhyme)² by having any of the following segmental configurations (Harris 1983; Roca 1991; Baković 2009; *i.a.*):

- a. When the penultimate syllable is CVC, as in *[te.¹le.fos.no].³
- b. When the penultimate syllable contains a falling diphthong (CVG), as in *[te.¹le.boj.na].
- c. When the penultimate syllable contains a rising diphthong (CGV), as in *[te.¹le.fjo.no].⁴

Another set of restrictions regarding antepenultimate stress in Spanish is related to the nature of the onsets of the final syllable. Harris (1983) points out that antepenultimate stress in Spanish is not possible when the final syllable has a trill /r/ in its onset, as in *[te.¹le.fo.ro], which contrasts with the availability of antepenultimate stress when the final onset is a tap /r/, as in [re.¹ka.ma.ra] ‘bedroom’. He claims that this pattern is due to intervocalic trills in Spanish being underlyingly ambisyllabic geminate taps. For instance, in a word like [ka.¹ʃo.ro] ‘puppy’, the stress necessarily falls on the penultimate syllable because the underlying representation of that word is /ka.ʃor.ro/. Therefore, the closed penultimate syllable would prevent antepenultimate stress.⁵

Other authors (Roca 1988; 1991; Lipski 1990; 1997; Baković 2009; *i.a.*) point out that the geminate tap account is not well-founded to explain the restriction on antepenultimate stress dependent on final trill onsets. They present a set of seemingly related patterns that show that antepenultimate stress in Spanish is also impossible when the onset of the final syllable is a palatal nasal /ɲ/, as in *[te.¹le.fo.ɲo], a palatal lateral /ʎ/, as in *[te.¹le.fo.ʎo], or a postalveolar affricate /tʃ/, as in *[te.¹le.fo.tʃo]. Whereas the geminate analysis might be possible for the Spanish trill, there is no good reason to think of the other segments as underlying geminates. Instead, these authors propose that the conditions on the onset of the final syllable are the result of a historical gap inherited from Latin, given that all of these segments are usually derived from ambisyllabic geminates or consonant clusters in Latin, which was a quantity-sensitive language in which closed penultimate syllables prevented antepenultimate stress (Spanish /ɲ/ is usually derived from Latin /nn/, /ʎ/ was many times /ll/ in Latin, /tʃ/ is often derived from the Latin

² As a reviewer points out, antepenultimate stress is also unattested when the *final* syllable is heavy, besides some minor exceptions such as [re.xi.men] ‘regimen, diet’. These restrictions on antepenultimate stress are not taken into consideration for the purposes of this paper, which focuses on the effects of the segmental configurations of the penultimate syllable and on the onsets of the final one.

³ The rule seems to be ignored in the case of loan words, such as [‘wa.ʃiŋ.ton] ‘Washington’ or [‘man.ʃes.ter] ‘Manchester’. Stress is usually not repaired to penultimate position in these cases. In any case, loanword phonology is usually claimed to have its own set of conditions (e.g., Tsuchida 1995 for Japanese).

⁴ Although an account is needed for cases like [a.‘li.kwo.ta] ‘aliquot’, as mentioned in Roca (1991). It can be argued that these words are compounds derived from two Latin roots (Lat. *alius* ‘other’, Lat. *quot* ‘how much’), and thus fall under different stress assignment rules (see Hualde 2006/7; 2012).

⁵ Harris (1983: 62ff.) is mostly concerned with defending a geminate tap analysis of the trill and uses the facts about antepenultimate stress only as one part of his evidence. He also supports the single phoneme account for the rhotic by some other facts about Spanish. For instance, the tap and the trill only contrast in intervocalic position, but have a predictable distribution elsewhere, which would favor the claim of a purely underlying contrast between a single and a geminate segment. Another of his arguments is that in natural speech there is no difference between the utterance of [sa.‘li ‘ra.pi.ðo] ‘I left quickly’ and [sa.‘lir ‘ra.pi.ðo] ‘to leave quickly’, showing that the maximum duration of a rhotic complex is just the duration of a trill, an underlyingly geminate tap. These arguments have each their counterpart in the work of other authors. Bradley (2001) explains the availability of a contrast only at intervocalic positions by making reference to perceptual salience (i.e., the difference between the two segments can only be perceived in this position). As for the duration evidence, Hualde (2004) performs an experiment that measures the duration of the rhotic clusters across word boundaries in the production of several native speakers, and finds out that there is a difference in the duration of the trill in [sa.‘li ‘ra.pi.ðo] ‘I left quickly’ and the tap + trill sequence in [sa.‘lir ‘ra.pi.ðo] ‘to leave quickly’, supporting the idea that there are two different phonemes in the language. More arguments (and the corresponding counterarguments) can be found in Baković (2009: §3).

consonant cluster /kt/, and the Spanish trill /r/ was a geminate /rr/). Therefore, what were ambisyllabic [C.C] sequences in Latin became onsets of the second syllable in the case of Spanish, producing the lexical gaps that we observe.

These two sets of restrictions regarding antepenultimate stress in Spanish are related to more general questions of Spanish phonology that this paper intends to address. First, there is of course the question about the phonemic representation of the trill, and whether it should be considered a single phoneme or a geminate tap. But in second place, and more generally, there is the question of whether we need to include syllable weight to account for the language's stress patterns or whether these forms are just lexically stored. The restrictions on antepenultimate stress related to heavy penultimate syllables—give or take the cases where the onset of the final syllable is a trill—point to the existence of syllable weight, but the second set of restrictions based on historical facts seem to be a fact about the lexicon from which speakers could only extract some analogical generalizations. Moreover, if these restrictions on antepenultimate stress were to be applied productively to new words, how would this knowledge about restrictions be represented? By making reference to properties of the phonological grammar and its possible abstract elements or by mere analogy to the lexicon? Are these restrictions of apparent separate origins (i.e., syllable weight restrictions and historical lexical gaps) present in the synchronic grammar of Spanish speakers? And if so, are they related in any way?

Both theoretical and experimental work has explored these questions, but there seems to be a gap between these literatures. While the majority of theoretical studies assume that stress assignment is generated by phonological rules or constraints (Harris 1983; 1987; Roca 1991; Lipski 1997; Oltra-Massuet & Arregi 2005; Gibson 2011; Martínez-Paricio 2013; Baković 2016; Piñeros 2016; *i.a.*), most experimental research on Spanish stress argues for an analogical process of stress assignment that is based on forms previously stored in the lexicon (Aske 1990; Eddington 2000; Face 2000 *et seq.*). While this last set of authors argues that stress is purely listed, and that stress assignment to nonce words is made on the basis of analogical processes, they fail to make explicit claims about the nature of the lexical representations they assume, and, in some cases, about the structure of the analogical model itself or about the constraints under which analogy can operate. On the other hand, authors who claim that syllable weight interacts with stress assignment in Spanish in a direct way—i.e., by making heavy syllables attract stress—fail to characterize the fine-grained distinctions that speakers make when faced with nonce words in experimental tasks. These seemingly opposing hypotheses, however, could be reconciled if we allowed for models of analogy that are grammatically informed, and are thus able to pick up generalizations about stress patterns from structured lexical representations before extending them to new items, such as Maximum Entropy models (e.g., Hayes & Wilson 2008).

This paper investigates the restrictions on antepenultimate stress in Spanish and their relation to syllable weight in a series of incremental steps that include experimental work and posterior modeling of the obtained data. The goal is to inform both the discussion about the phonemic representation of the trill, and the nature of the stress assignment process in Spanish (including whether it needs to make reference to structured lexical representations or not). To this end, Section 2 presents previous accounts that have experimentally explored the interaction between antepenultimate stress and segmental configurations in Spanish. Given the lack of an experimental study that takes into consideration in a single task all the restrictions related to antepenultimate stress that are mentioned above, after confirming in Section 3 that these restrictions hold in the Spanish lexicon, Section 4 presents an experimental study in which native speakers had to rate nonce words that presented one of the several possible segmental configurations that disallow

antepenultimate stress. Even if there are no Spanish words that present these segmental configurations together with antepenultimate stress, we can expect native speakers to have gradient intuitions about particular nonce words, showing differences in their preferences between “accidentally” *unattested* words (such as *blick* in English) and *impossible* ones (such as *bnick* in English), ultimately reflecting patterns in their lexical statistics (e.g., Albright 2009a; Hayes & White 2013). Besides assessing the productivity of the restrictions on antepenultimate stress when the penultimate syllable is heavy, it is of interest to explore which of the other restrictions pattern together, to provide support for either a geminate tap in Spanish (in lieu of the trill) or for a historical account of the lexical gaps that are observed.

The second part of the paper deals with the mechanisms that are at play in the process of stress assignment in Spanish, and, in doing so, with the question of whether Spanish is quantity-sensitive or not. Therefore, Sections 5 and 6 deal with possible interpretations of the experimental data and different ways of modeling it. Specifically, these sections present models of phonotactic learning that presuppose that participants’ ratings are either purely based on segmental similarity with the lexicon (§5) or that hidden hierarchical structure—such as syllable weight—is part of the lexical representations that native speakers use when computing stress in nonce words (§6). A Maximum Entropy model (Hayes & Wilson 2008) that incorporates syllable weight in lexical representations most accurately captures speakers’ intuitions about patterns of antepenultimate stress in Spanish, and also strongly suggests a phonemic representation of the Spanish trill as a singleton consonant. The comparison between the different models is discussed in Section 7, together with some alternative accounts for the residual data that these models leave unexplained. Finally, Section 8 presents the general conclusions of the paper.

2 Previous experimental work on the restrictions on stress assignment in Spanish

Experimental work that investigates the relation between stress assignment and segmental conditions in Spanish is plentiful, but the results that come from those studies are not conclusive. Uncovering the mechanisms at play in the gradient intuitions that speakers show in experimental tasks has proven to be a difficult task and a controversial topic. Moreover, there is no single study that investigates every constraint that seems to play a role in antepenultimate stress assignment.

The first study that explores a subset of these problems was undertaken by Aske (1990), whose purpose was to shed light on whether the stress patterns of Spanish were based on hard generative rules or driven by a simpler kind of “analogy” with other words in the lexicon. If the stress assignment algorithm is driven by “disembodied” generative rules, the author argues, speakers would only make use of those rules when assigning stress to nonce words. On the other hand, if the stress assignment algorithm is based on an analogical process that makes reference to the lexicon, participants might be also influenced by the particular phonological shape of the word (as opposed to the specific configuration that the rule states as relevant) or by other non-phonological information that is present in the word. In a task in which subjects had to assign either penultimate or final stress to capitalized nonce words⁶ (which, by convention, are typically not orthographically marked for stress in Spanish), participants replicated the stress patterns and subpatterns in the lexicon, as opposed to following a hard rule that stated whether those nonce words

⁶ Spanish stress is always predictable from the orthography. Final stress words have an orthographic accent when they end in a vowel, an *n* or an *s*. Penultimate stress words have an orthographic accent when they end in any grapheme but a vowel, an *n* or an *s*. Finally, antepenultimate stress words always carry orthographic accent.

should have penultimate or final stress given their segmental composition. For instance, words ending with /-n/ in Spanish typically present final stress, as in [kan.'sjon] ‘song’. However, this pattern only holds when the preceding vowel is not /e/; in that case, while there are words with final stress, such as [des.'den] ‘disdain’, there is a high percentage of words with penultimate stress, such as ['mar.xen] ‘margin’. When subjects had to stress nonce words, they followed this subpattern: they overwhelmingly produced final stress for words that ended in /-n/, but when the preceding vowel was /e/, they also produced many penultimate stressed words. In summary, given that native speakers are sensitive to these kinds of subpatterns in the lexicon, Aske (1990) claims that stress is necessarily listed. His rationale is that if hard rules were followed, nonce words should be stressed according to them, and not with respect to lexical patterns and subpatterns. However, he does not consider the possibility of more complex systems in which rules can be probabilistic and replicate lexical patterns (e.g., Albright & Hayes 2003; Hayes et al. 2009; Zuraw 2010; Moore-Cantwell 2016). The question of how those rules should be represented then arises, but that does not prevent in principle this kind of explanation. Moreover, the participants in the study were from different regions of Spain and Latin America—20 students at UC Berkeley from different Hispanic origins, and 16 participants living in Spain—and 14 subjects out of the 36 were bilinguals with Basque,⁷ which overall constitutes a sample that might be too varied.

Eddington (2000) replicates Aske’s (1990) findings by modeling Spanish stress assignment under the Analogical Modeling of Language (AML) framework (Skousen 1989; 1992), a model that intends to reflect how speakers determine their linguistic behavior when faced with nonce words. With respect to stress assignment, the AML predicts that when the system encounters a new word that needs stress, it will search the whole lexicon for the most similar word(s), and then apply the same stress of that exemplar(s) to the new form. In Eddington’s (2000) study, an AML was created with a lexicon of the most frequent 4,970 forms of Spanish as its database, which in turn was able to correctly assign stress in 94% of the cases. However, the database of lexical forms was coded not only with phonemic content, but with syllable structure (i.e., whether each segment was in the onset, nucleus, or coda of a given syllable). Therefore, the “analogy” that the model is said to be doing is enriched by the positional and structural information provided by the syllabic configuration. Moreover, the study also tested the database on *real* words—the database was divided in 10 sets, where 9 sets functioned as predictors and one set functioned as the testing data—which can increase the reliability of the model in testing “novel” forms. Furthermore, even if this sort of testing (words vs. words; that is, testing 1/10th of the real words in the database against the remaining words) is standard practice for this kind of analogical model, we know that some lexical statistical regularities are productively extended to nonce words, while other regularities are not. For instance, Turkish speakers will not use vowel height or backness productively to predict vowel alternations in nonce words, even when this pattern is present in the lexicon (Becker et al. 2011). Finally, even if the AML worked at a 94% accuracy in general, it only predicted antepenultimate stress correctly in 40.1% of the cases (Eddington 2000: 100), so either pure listedness or a new rule should be proposed for these cases. In a follow-up study, Eddington (2004) analyzes the influence of different factors—phonemic information, syllable configuration, and syllable weight—in the success of analogical models of stress placement on real words, and concludes that the only crucial factor to determine stress assignment is phonemic information. However, the author also compares the performance of these models to nonce

⁷ Many varieties of Basque have fixed penultimate stress (Elordieta & Hualde 2014), which might affect the process of stress assignment in Spanish when the task is performed by Spanish-Basque bilinguals.

word stress assignment tasks (Face 2000; Waltermire 2004) and admits that when dealing with productivity “the role of CV tier and syllable weights should not be discounted” (Eddington 2004: 110).

In Face (2000), the role of syllable weight in predicting Spanish stress is assessed. The study evaluates the perception of stress in unstressed nonce words in which pitch and duration are artificially controlled. The results seem to provide evidence in favor of the cognitive reality of syllable weight, given that the rightmost heavy syllable overwhelmingly seems to attract stress. However, in later studies (Face 2003; 2004), the author recognizes that in the 2000 study only the duration of the vowels was controlled, but not the duration of the whole syllables, which can be also relevant for stress computation (Gordon 2002; Ryan 2011). In that way, longer syllables could still be giving an acoustic cue to participants about where stress would fall. When retesting the experiment by controlling the duration of the syllables, no effect of weight was found (Face 2004; Face & Alvord 2005). Face (2006), a more extensive study, retests all previous experiments and finally claims that Spanish stress placement is only affected by segmental similarity to other words, subregularities in the lexicon, and morphological category—but crucially, not by syllable weight. The same lack of effect is found by Bárkányi (2002) in another study where participants had to mark stress on orthographic nonce words that lacked orthographic accent marks, which presented different segmental configurations. The author found that speakers assigned antepenultimate stress both when the penultimate syllable was light and when it was heavy, though she also claimed that the different proportions of acceptability might reflect different subpatterns in the language. The main claim, however, is that this assignment is based on analogy, given that pure rule-based learning should not generalize antepenultimate stress to nonce words in this kind of task. The understanding of rules in this study is again within an account that does not allow for exceptions or for probabilistic rules.

The first experimental study to systematically investigate the relation between antepenultimate stress assignment and the role of the trill in the onset of final syllables in Spanish is Alvord (2003). In this study, Spanish native speakers had to provide grammaticality judgments of nonce words with antepenultimate stress that had a heavy penultimate syllable or a trill in the onset of the final one. Both conditions (i.e., antepenultimate stressed words with heavy penultimate syllables, and antepenultimate stressed words with a trill as the onset of the final syllable) were accepted at approximately 95% by native speakers. Alvord (2003) therefore concludes that Spanish is not quantity-sensitive, and that the restriction on antepenultimate stress when the final onset is a trill is not productive. However, binary acceptability judgments are usually not sensitive to the gradient intuitions reported by speakers in tasks that do allow for that variability (e.g., Daland et al. 2011; Lau et al. 2017). Moreover, the study was undertaken only by 10 subjects who provided 100 judgments of nonce words within 4 conditions each, which does not constitute a significant sample of the Spanish-speaking population.

Waltermire (2004) also argues for an analogical procedure for stress assignment of novel words, but one in which syllable weight *is* relevant. He replicates the experiment in Face (2000), though in a written production task, and finds that heavy syllables attract stress. However, this seems to work only for final and penultimate stress, given that antepenultimate stress is dispreferred across conditions. Given that the proportions of stress assignment in every other condition parallel the stress proportions in the lexicon, the author claims that subjects base their stress assignment for nonce words on the listed representations in their lexicons, but he also argues that these representations encode syllable weight. Nonetheless, the details of the analogical process are left unspecified.

Another set of more recent work by Shelton and colleagues explores the relationship between syllable phonotactics—in particular, syllable weight—and stress assignment in Spanish by means of a set of naming tasks. In these tasks, participants are presented with words on a computer screen, and they are asked to read each word as quickly and accurately as they can. The analysis is then performed both on their reaction times and on the error rates for the different kinds of (nonce) words they read. In Shelton (2007) and Shelton et al. (2009), the relation between antepenultimate stress and syllable weight is explored. The results in those studies show that speakers make more errors when naming antepenultimate stressed nonce words that have a heavy penultimate syllable than when that syllable is light. However, they also show that speakers have an intermediate error rate (between the other two conditions) when the onset of the final syllable is a palatal consonant, either lateral or nasal. It is argued that speakers are sensitive to subpatterns in the lexicon and to the accidental historical gaps in the case of the final palatal onsets. Shelton et al. (2012) further analyze the interaction between rising and falling diphthongs in penultimate syllables—which make them heavy—and the possibility of antepenultimate stress, and they find that speakers have significantly more errors in reading aloud nonce words with antepenultimate stress and a falling diphthong in the penultimate syllable, such as *['ko.taj.ɣa], than when reading nonce words with antepenultimate stress and rising diphthongs in the penultimate syllable, such as *['lo.tja.ɣo]. They also show that speakers do not produce these errors when, instead of the antepenultimate syllable, it is the penultimate syllable that is stressed, and they claim that these results show that Spanish is quantity-sensitive. However, according to the authors, the significant difference between rising and falling diphthongs points to a more fine-grained weight system than a binary classification between light and heavy syllables,⁸ or alternatively—as in their previous studies—to a concurrent effect of the segmental properties of the lexicon, given that rising diphthongs are much more common than falling diphthongs in any syllable of Spanish. Finally, Shelton (2013) also points to gradience in the syllable weight system by experimentally investigating the nature of the rhotics in Spanish in a naming task. Native speakers had to read aloud nonce words with antepenultimate stress that had either a tap or a trill in the onset of the final syllable. The author found that participants produced more errors in the trill condition than in the tap condition. However, while the tap condition patterned with the control condition (CV final syllable), the trill condition did not produce as many errors as the ceiling condition (a CVC penultimate syllable). The interpretation of the results is again that speakers are aware of very detailed patterns in their lexicons in their assignment of stress to nonce words. Moreover, the author argues that, given that the trill condition does not pattern either with the heavy penultimate condition or with the control condition, we cannot conclude what the nature of the rhotic is.

As we can see, the picture that arises from all these studies with respect to the role of syllable weight in the process of stress assignment in Spanish is not clear. While there is some consensus about the fact that some “analogical process” operates when assigning stress to nonce words, the nature of the analogical model and the nature of the lexical representations is mostly unspecified. Finally, the status of the rhotic phoneme(s) is still unresolved. The work in this paper intends to shed light on these debates, by presenting an experiment in which speakers are presented with all the conditions under discussion (i.e., the restrictions on antepenultimate stress that interact with the segmental configuration of the penultimate syllable, or with the presence of some particular final onset), in

⁸ As a reviewer points out, the need for this distinction was already suggested by Harris (1983), based on the fact that final CVG syllables attract stress more strongly than final CGV syllables.

order to understand the relevant properties of the interaction between syllable configurations and stress assignment. The results are later modeled under different analogical models that allow for representations of various levels of complexity, so as to analyze the role in Spanish stress assignment of both syllable weight and of final onsets derived from Latin consonant clusters or geminates, together with the nature of the trill as a singleton consonant or as a geminate tap in Spanish. But first I will perform a corpus search to confirm in the Spanish lexicon the claims about the absence of antepenultimate stressed words with these specific segmental configurations.

3 Corpus search: A first step testing the restrictions

A corpus search targeting the theoretical restrictions on antepenultimate stress described in §§1–2 was performed to explore whether the constraints mentioned in the literature with respect to antepenultimate stress in Spanish held in the lexicon. The corpus search was performed manually using the Constraints-to-Words function of the Latin American Spanish corpus of the EsPal lexical database (Duchon et al. 2013) developed by the Basque Centre of Cognition, Brain, and Language (BCBL), which consists of 277,771 types and 307,772,547 tokens.⁹ The search engine allows for queries conditioned by phonological structure, where stress and segmental information can be specified.

The results confirm the claims in the literature: there are no types with antepenultimate stress when the word has a heavy penultimate syllable closed by a consonantal sound, a heavy penultimate syllable with a falling diphthong, or a heavy penultimate syllable with a rising diphthong. Moreover, there are also zero types with antepenultimate stress when the word has a trill /r/ in the onset of the last syllable, confirming Harris' observation (1983). There are also zero cases of words with antepenultimate stress when the onset of the last syllable is a nasal palatal /ɲ/ or a postalveolar fricative /ʃ/.¹⁰ All these gaps are obviously not due to an absolute prohibition on antepenultimate stress in Spanish, given that a search for words with only CV syllables and antepenultimate stress returns a total of 399 types.

The corpus also shows that the determining factor in the aforementioned gaps is actually stress and not segmental material as such. Searching for words with penultimate stress, there are 1,210 types with heavy penultimate syllables closed by a consonantal sound, 67 types with a falling diphthong on the penultimate syllable, and 293 types with a rising one in that position. As for the role of the onset of the final syllable, the determining factor is also stress, since a search for words with penultimate stress returns 73 types when the onset of the final syllable is a trill, 70 types when the onset of the final syllable is a nasal palatal, and 435 types when the onset of the final syllable is a postalveolar fricative. A summary table of the corpus search results is presented in Table 1. Given that the corpus results support the theoretical claims, we can feel confident moving forward to test the productivity of these patterns experimentally.

4 Experimental evidence for a quantity-sensitive stress system: A nonce word rating task

An experimental task was designed to collect acceptability judgments from native speakers of Spanish with respect to nonce words that violated one of the constraints that seem to disallow antepenultimate stress. Seven experimental conditions governing segmental

⁹ The corpus is available online at: <http://www.bcbl.eu/databases/espal/>.

¹⁰ Given that the experimental task in §4 was undertaken by Rioplatense Spanish speakers, the palatal sound was coded as a postalveolar fricative. Rioplatense Spanish realizes what Baković (2009) calls a lateral palatal sound [ɬ], which corresponds to the orthographic letters *ll*, and *y* in all environments but word-final, as the postalveolar fricative [ʃ] (or its voiced variant [ʒ]).

Table 1: Number of tokens by condition in a corpus search. Examples are given in the attested conditions in the lexicon.

Condition/Stress	Antepenultimate	Penultimate
CV.CV. CVC .CV	0	1,210 (e.g., [lo.ya.'rit.mo] 'logarithm')
CV.CV. CVG .CV	0	67 (e.g., [de.sa.'raj.ɣo] 'uprooting')
CV.CV. CGV .CV	0	293 (e.g., [me.ri.'ðja.no] 'meridian')
CV.CV.CV. rV	0	73 (e.g., [ma.sa.'mo.ra] 'maize pudding')
CV.CV.CV. ɲV	0	70 (e.g., [mu.sa.'ra.ɲa] 'shrew')
CV.CV.CV. JV	0	435 (e.g., [pe.sa.'ði.fa] 'nightmare')
CV.CV. CV.CV	399 (e.g., [pi.'ra.mi.ðe] 'pyramid')	4,771 (e.g., [su.ri.'ka.ta] 'suricate')

Table 2: Experimental conditions and examples on nonce word rating task: syllable structure and stress placement interactions. Bold indicates segmental/syllabic configuration of interest.

Condition/Stress	Antepenultimate	Penultimate
CV.CV. CVC .CV	[da.'ti.pem.bo]	[da.ti.'pem.bo]
CV.CV. CVG .CV	[bu.'ne.ðew.ta]	[bu.ne.'ðew.ta]
CV.CV. CGV .CV	[lo.'ma.fja.ɣo]	[lo.ma.'fja.ɣo]
CV.CV.CV. rV	[li.'ko.ða.ro]	[li.ko.'ða.ro]
CV.CV.CV. ɲV	[pa.'mu.ðo.ɲo]	[pa.mu.'ðo.ɲo]
CV.CV.CV. JV	[la.'ri.mu.fa]	[la.ri.'mu.fa]
CV.CV. CV.CV	[ro.'ku.na.to]	[ro.ku.'na.to]

configuration were designed, which were crossed with two types of stress: antepenultimate and penultimate—the latter as a control condition given that all these syllabic configurations allow for penultimate stress. The segmental configuration conditions were divided into three sets: a) heavy penultimate syllables (presenting a branching rhyme whose second segment was either a consonant, a glide, or a vowel), b) final syllable onsets (a trill, a nasal palatal, or a postalveolar fricative), and c) a baseline condition with a light penultimate CV syllable. A table with sample stimuli for each condition is given in Table 2.

4.1 Methods

An experiment was carried out in which participants were asked to judge orthographically presented nonce words that targeted different conditions on stress assignment and on syllable structure by providing acceptability judgments on a Likert scale from 1 to 5.

4.1.1 Stimuli

Nonce words were created by the author—a native speaker of Rioplatense Spanish—to test the different hypotheses with respect to antepenultimate stress and syllable configurations. First, I wanted to check if the claims made in the literature (and confirmed in the corpus search) that there are no words in Spanish with antepenultimate stress and penultimate heavy syllables could be extended to novel words, assessing the productivity of

these syllable configurations. Thus, three experimental conditions that targeted heavy penultimate syllables were designed: a condition in which the penultimate syllable was closed by a consonantal segment, as in [da.'ti.pem.bo], a condition in which the penultimate syllable had a falling diphthong, as in [bu.'ne.ðew.ta], and a condition in which the penultimate syllable had a rising diphthong, as in [lo.'ma.fja.ɣo]. These three conditions were paired with conditions that had the same structure in the penultimate syllable, but that also carried the stress on that syllable (that is, that had penultimate stress instead of antepenultimate stress). The conditions with penultimate stress were expected to be more acceptable than the ones with stress on the antepenultimate syllable, given that penultimate stress is allowed with these syllabic configurations, as shown by the results in the corpus search.

Three conditions with antepenultimate stress and constraints on the last syllable were created to test the claims in Harris (1983) with respect to the nature of the trill /r/ as a geminate tap in Spanish, and the counterarguments made by Roca (1988), Lipski (1990), Hualde (2004), Bradley (2006), and Baković (2009), among others. A first condition included antepenultimate stress and a last syllable that had a trill /r/ as its onset, as in [li.'ko.ða.ro]. A second condition was designed in which nonce words had antepenultimate stress and the final syllable was formed by the palatal nasal /ɲ/ and a vowel, as in [pa.'mu.ðo.ɲo]. Finally, a third condition included nonce words with the postalveolar fricative /ʃ/ as the final syllable onset, while also presenting antepenultimate stress, as in [la.'ri.mu.ʃa]. As in the previous set of conditions on the structure of the penultimate syllable, three control conditions in which the stress was assigned to the penultimate syllable were also created in these cases. Finally, a condition in which antepenultimate stress is acceptable in Spanish (all syllables in the word have a CV structure) was created to obtain a baseline score on words with antepenultimate stress, as in [ro.'ku.na.to]. A condition for penultimate stress with all CV syllables was also included.

Each of the 14 conditions included 10 different stimuli, resulting in a total of 140 tokens. Out of the 10 stimuli per condition, and given that Spanish has five phonemic vowels (/a/, /e/, /i/, /o/, /u/), 2 items per vowel were designed in each syllable condition with respect to the stressed syllable. The vowels and onsets in the remaining non-critical syllables were distributed roughly evenly among all the consonantal and vocalic sounds of Spanish, though not in a systematic way. The sequences in falling and rising diphthongs were also controlled to have an even distribution and to present all possible combinations in Spanish. All words ended in vowels, evenly distributed between /a/ and /o/, which are the most common final vowels of Spanish nominals. All stimuli had four syllables and the syllables that were not determining the experimental conditions were all CV, so that there was at most one heavy syllable per item.

4.1.2 Participants

Participants were recruited via Facebook. A link to the questionnaire with a description in Spanish of the task to be performed was posted in the Facebook wall of the author. Access to the questionnaire was unrestricted, but the participants were asked to provide basic demographic data on the first page of the questionnaire. 37 complete answers were recorded during the last week of November 2015 (another 11 sessions were started, but abandoned halfway through), and after reviewing the data provided by the participants, it was decided to take into account only the acceptability judgments of a fairly homogeneous group: native speakers of Rioplatense Spanish, only from the city of Buenos Aires ($n = 31$), 21 female, ages between 18 and 35 (mean = 27.25, SD = 5.1). Subjects were not compensated for the task.

4.1.3 Procedure

The link on the Facebook wall directed the participant to a questionnaire built on Qualtrics (2015). All instructions and scales were written in Spanish. Once the questionnaire opened, and after the collection of personal data, subjects read the instructions that

described the task: subjects were told that they were going to read some nonce words or “novel words” of Spanish that could become the name of a new object. They were asked to imagine that a new object had been discovered or invented and that a name was needed for it—to make sure that the words were interpreted as nominals.^{11,12} Then, they were asked to judge each item on a Likert scale from 1 to 5. The scores were defined as follows, and participants had access to these definitions of the points in the scale at all times on a separate browser tab. An English translation of the scale is given below:

- i. **Definitely no:** The word sounds bad, I do not think it could be a word of Spanish.
- ii. **No:** It does not sound good, but it looks like a word that could be Spanish.
- iii. **I am not sure:** It is neither a good nor a bad sounding Spanish word; I am not sure if it could be a new word.
- iv. **Yes:** It sounds good and I think it could be a new word of Spanish, though I don’t know if I would use it.
- v. **Definitely yes:** It sounds very good, it could be a new word of Spanish without any problem, and I would use it myself.

After 4 items of practice,¹³ the task began. All 140 items described in §4.1.1 appeared in a completely randomized order.¹⁴ The task took an average time of 21 minutes, 14 seconds.

4.2 Results

Results show that speakers prefer penultimate stress over antepenultimate stress across conditions. All conditions with penultimate stress received higher ratings than their antepenultimate stress counterparts. Nonce words with a closed penultimate syllable showed higher ratings when they had penultimate stress (mean = 3.56, SE = .08) than when they presented antepenultimate stress (mean = 2.10, SE = .09). The same difference between stress conditions occurred when participants rated nonce words that contained a falling diphthong in their penultimate syllable (Antepenultimate Stress: mean = 1.92, SE = .07; Penultimate Stress: mean = 3.26, SE = .09), and nonce words that contained a rising diphthong in their penultimate syllable (Antepenultimate Stress: mean = 2.18, SE = .08; Penultimate Stress: mean = 3.23, SE = .15). In the conditions that were dependent on the final onset, participants also rated penultimate stressed nonce words higher than antepenultimate stressed ones. When the final onset was a trill, penultimate stressed nonce words were rated higher (mean = 3.59, SE = .12) than antepenultimate stressed ones (mean = 2.53, SE = .14). The same pattern between stress conditions was observed in the ratings in the nasal palatal as a final onset condition (Antepenultimate

¹¹ As stated above, some differences across grammatical categories with respect to stress assignment have been reported in the literature (e.g., Harris 1989; Roca 1991). Specifically, stress is morpho-syntactically conditioned in Spanish verbs.

¹² It could be possible—as a reviewer points out—that the participants judged the stimuli as potential names for *commercial* products. To my knowledge, there is not much research on the phonology of brand names, but the general observation is that good, successful brand names are generally easy to pronounce and conform to markedness constraints and phonotactics of a language (see e.g., Suh 2002; Kim 2017). Therefore, if participants were interpreting the stimuli as product names, this would have made them more conservative in their judgments across the board on the antepenultimate stress conditions that are unattested in the lexicon.

¹³ Practice items included both stress conditions (antepenultimate and penultimate) and two conditions on the penultimate syllable (heavy CVC and light CV).

¹⁴ A reviewer asks if, given a fully randomized procedure, certain types of items ended up being presented earlier on average while other types ended up being presented later on average. Manual inspection of the randomization orders does not show any clustering of items of a particular condition in any subject at any point during the task. Correlations between the randomization orders and the script order (in which the items are grouped by condition) are not greater than 0.13 for any subject (with an average value of 0.06), indicating that there is no effect of clustering of items due to full randomization.

Stress: mean = 2.13, SE = .06; Penultimate Stress: mean = 3.50, SE = .14), as well as in the postalveolar fricative as a final onset condition (Antepenultimate Stress: mean = 2.32, SE = .09; Penultimate Stress: mean = 3.35, SE = .19). The control condition with all CV syllables also showed this difference between ratings on penultimate stressed nonce words (mean = 3.29, SE = .10) and antepenultimate stressed nonce words (mean = 2.87, SE = .18).¹⁵ The means for each condition, after a z-score transformation to account for variance across speakers, are given in Table 3, while Figure 1 illustrates those results across stress and segmental conditions.

A linear mixed model analysis to assess the effects of stress placement and segmental conditions was performed in R (R Core Team 2015) with the *lme4* package (Bates et al. 2015). As for model selection, I followed the recommendations for linguistic analysis in Winter (2013) and performed Likelihood Ratio Tests of the full model with the effect under consideration against the model without it to obtain p-values, which indicate the probability that adding the fixed effect under consideration to the model would not be significant. All post hoc tests for multiple comparisons across conditions were run with the *multcomp* package (Hothorn et al. 2008), and p-values were adjusted by Tukey method.

A model analyzing the fixed effect of stress placement on the participants' ratings—with random intercepts for subject and item and by-subject random slopes for the effects of stress placement—showed that stress placement was a significant predictor of ratings on nonce words ($\chi^2(3) = 487.66$, $p < .0001$). Post hoc tests showed that the effect of stress placement was due to significantly higher values for items with penultimate stress over those with antepenultimate stress ($\beta = 1.104$, $p < .0001$).

After subsetting the data by stress placement to analyze the role of the different segmental and syllabic configurations on the nonce words ratings, the analysis within the penultimate stress cases showed that a model with segmental/syllabic configuration condition as a fixed effect, and random intercepts for subject and item and by-subject random slopes for condition, performs significantly better than the null hypothesis ($\chi^2(33) = 51.755$,

Table 3: Mean z-scores and standard errors for antepenultimate and penultimate stress per syllable structure/segmental condition (Bold indicates relevant syllabic/segmental configuration).

Condition/Stress	Antepenultimate	Penultimate
CV.CV. CVC .CV	-.594 (SE = .070)	.568 (SE = .065)
CV.CV. CVG .CV	-.732 (SE = .058)	.335 (SE = .075)
CV.CV. CGV .CV	-.530 (SE = .066)	.304 (SE = .116)
CV.CV.CV. rV	-.253 (SE = .110)	.589 (SE = .096)
CV.CV.CV. pV	-.566 (SE = .046)	.522 (SE = .108)
CV.CV.CV. jV	-.419 (SE = .074)	.399 (SE = .151)
CV.CV. CV .CV	.019 (SE = .142)	.356 (SE = .080)

¹⁵ A reviewer points out that voiceless stops as final onsets also disfavor antepenultimate stress in words of Latin origin, given that they were once Latin geminates (Baković 2009). Nonce words in the control condition presented both voiceless (/p/, /t/, /k/) and voiced stops (/b/, /d/, /g/) as final onsets. The results on the experimental task show a slight tendency reflecting this claim: the mean for antepenultimate stressed nonce words with voiceless final onsets was 2.58, while the mean for antepenultimate stressed nonce words with voiced final onsets was 2.86. However, given that there was only one stimulus for each of the voiceless/voiced stops as final onsets, the results are far from conclusive. Moreover, the results show no clear pattern, given that nonce words with final onset /d/ are rated higher than words with final onset /t/, but nonce words with final onset /p/ are rated higher than words with final onset /b/, for instance. I leave the exploration of this pattern for further experimental research.

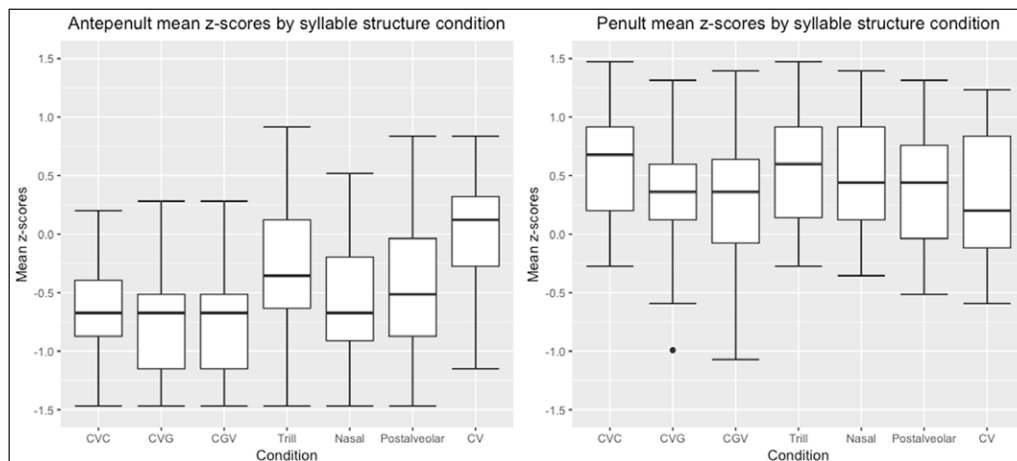


Figure 1: Box plots for mean z-scores for antepenultimate and penultimate stress per syllable structure/segmental conditions. Boxes indicate the data between the 25th to the 75th percentile in each condition, while whiskers extend to 1.5 times the size of the inter-quartile range. Black dots represent outliers beyond those parameters.

$p < .05$), but post hoc tests showed no significant differences across segmental/syllabic configuration conditions.

Within the cases with antepenultimate stress, a model with segmental/syllabic configuration condition as a fixed effect, and random intercepts for subject and item and by-subject random slopes for condition, was significantly better than the null hypothesis ($\chi^2(33) = 146.58$, $p < .0001$). A full set of the statistical comparisons across all conditions is presented in Table 4.

As we can see in Table 4, post hoc tests showed significant differences favoring the control condition over the following syllabic/segmental conditions: a) penultimate CVC syllable ($\beta = 0.771$, $p < .001$); b) falling diphthong in penultimate syllable ($\beta = 0.945$, $p < .001$); c) rising diphthong in penultimate syllable ($\beta = 0.690$, $p < .005$); d) nasal palatal in the onset of the final syllable ($\beta = 0.735$, $p < .001$); and e) postalveolar fricative in the onset of the final syllable ($\beta = 0.552$, $p < .05$). Crucially, they did not show a significant effect between the control condition and the condition with a trill in the onset of the final syllable ($\beta = 0.342$, $p = .304$). In summary, all the conditions that encompass restrictions on antepenultimate stress, but the condition on the trill, are significantly different from the control condition.

On the other hand, the condition with the trill as a final onset—while somewhat different from the other conditions that prevent antepenultimate stress—does not reach significance levels in its difference from most of those conditions (only being significantly different from the condition with a falling diphthong on the penultimate syllable; $\beta = 0.603$, $p < .05$). What does this tell us? How can the condition that has a trill as a final onset pattern at the same time both with the control condition (which would show that it does not prevent antepenultimate stress) and with all the other conditions (which would make it a segmental condition that precludes antepenultimate stress)?

In a nutshell, my proposal is that this *intermediate status*—which is also shown by the mean score that the final trill onset condition displays in its ratings—needs to be understood at face value. We should not equate absence of evidence (which is what failing to reject the null hypothesis stands for) with evidence of absence (see Altman & Bland 1995; Alderson 2004, *i.a.*). The statistical comparisons have failed to reject the hypothesis that this condition behaves equally to all the other conditions that reflect restrictions on antepenultimate stress, but they have also failed to reject the hypothesis that it behaves equally

Table 4: Estimates, standard errors, z-values and p-values of statistical comparisons across all conditions in the experimental task. P-values have been adjusted by Tukey (Bold indicates particularly relevant comparisons).

Contrast	Estimate	SE	z-value	p-value
Control – CVC	.771	.159	4.836	<.001
Control – CVG	.945	.170	5.545	<.001
Control – CGV	.690	.179	3.847	<.005
Control – Trill	.342	.157	2.180	.304
Control – Nasal	.735	.171	4.311	<.001
Control – PostAlv	.552	.163	3.390	<.05
CVC-CVG	-.081	.167	-.483	.999
CVC – CGV	.174	.153	1.138	.915
CVC – Trill	-.429	.167	-2.587	.129
CVC – Nasal	-.036	.156	-.227	.999
CVC – PostAlv	-.219	.164	-1.339	.831
CVG – CGV	-.255	.160	-1.593	.683
CVG – Trill	-.603	.178	-3.387	<.05
CVG – Nasal	-.210	.155	-1.349	.826
CVG – PostAlv	-.394	.172	-2.285	.249
CGV – Trill	-.348	.175	-1.992	.416
CGV – Nasal	.045	.158	.286	.999
CGV – PostAlv	-.138	.168	-.828	.982
Trill – Nasal	.394	.169	2.334	.226
Trill – PostAlv	.210	.153	1.373	.814
Nasal – PostAlv	-.184	.158	1.161	.907

to the control condition. This is not the same as confirming that they behave in the same way, a conclusion that can not be achieved by statistical comparisons based on p-values. On the other hand, the statistical comparisons have been able to reject the hypothesis that all the other conditions that reflect restrictions on antepenultimate stress—both the ones with heavy penultimate syllables and the conditions with nasal palatals and postalveolar fricatives as final onsets—behave in an equal fashion to the control condition.

The lack of significance in the difference between the condition on the trill and the other conditions that prevent antepenultimate stress might be a problem of the task not being sensitive enough, or a problem that arises with a reduced sample size that fails to trigger significance in statistical terms. However, this should not make us doubt about the general result that the data show: the trill condition is different from the heavy penultimate conditions and from the nasal palatal and postalveolar fricative conditions, even if not significantly. Moreover, the trill is the *only* condition that is not significantly different from the control condition (while all the other conditions are significantly different from it), so there is still some support for the status of the final trill onset as special when it comes to its role in antepenultimate stress assignment. This last bit is particularly revealing: we can reject the hypothesis that the nasal palatal, the postalveolar fricative, and the heavy penultimate syllable conditions behave in the same way as the control condition, but we cannot reject the hypothesis that the final trill onset condition and the control condition behave in the same way.

4.3 Interim discussion

The results of the experimental task showed that the observations made by Harris (1983) with respect to heavy penultimate syllables, and by Roca (1988), Lipski (1990), and Baković (2009) with respect to final syllable onsets that are nasal or lateral palatals, can be not only confirmed by a corpus search but also by an experimental nonce word judgment task. The prediction that antepenultimate stress should be prohibited in those conditions is borne out by the results, opposing previous experimental studies, such as the one in Alvord (2003).

With respect to the status of the trill, however, the results are less conclusive. They might provide support to Harris' (1983) claim that it is an underlying tap geminate, given that the final trill onset condition is not significantly different from most of the heavy penultimate syllable conditions, but they might also provide support to the historical account (Roca 1988; Lipski 1990; Baković 2009; *i.a.*), because the final trill onset condition is not significantly different from the final nasal palatal onset or from the final postalveolar fricative onset conditions. However, while all of these conditions are significantly different from the control condition, the final trill onset condition is not. All these comparisons seem to point to the trill condition as a restriction that is less penalized by the grammar than all the others restrictions on antepenultimate stress, but still not as good as the control condition.

At this point, we can think of a different way of tackling the problem of the representational status of the Spanish trill, and the question of quantity-sensitivity in Spanish. Testing different stress assignment algorithms—which are based either on purely segmental similarity to the lexicon or make reference to more abstract structures as part of the lexical representations—by assessing their reliability in predicting the experimental results can provide an insight into the nature of the phonological representations. In doing so, we can also address questions about the nature of the task that participants are performing when rating a stressed nonce word. The textbook division between existing words (*brick*), nonce words that could exist (*blick*) and nonce words that are judged to be completely ill formed (*bnick*) seems to be insufficient to capture the patterns of these ratings. When people are faced with these nonce words, which are based on combinations of syllabic configurations and stress patterns that do not exist in their lexicon, they react with different ratings to each condition. How can we account for this variance? Is it just noise? Or are participants taking into account other properties of the nonce words that make them better or worse as possible words of their language?

Most of the previous experimental work on these issues has focused only on some of these conditions and has claimed that participants are resorting to some kind of analogy with the lexicon (Eddington 2000; Bárkányi 2002; Face 2004; 2006). However, the range of analogical procedures previously reported ranges from purely similarity-based analogy (Aske 1990) to an analogy that includes hierarchical structure codified in the segments (Eddington 2000). Most of the studies have failed to define which are the subpatterns in the lexicon that participants would be recovering when they assign stress to a nonce word. Moreover, besides Shelton (2007; 2009), most studies fail to recognize the possibility of a more fine-grained distinction between the zeroes in the lexicon and their productivity.

In the following section, I evaluate different analogical algorithms to model the experimental data. The goal is to assess which properties and subpatterns of the lexicon are productive in the nonce word task that was assigned to the participants, by adding those properties and subpatterns to the different models and observing whether they produce better results in predicting the participants' ratings (see Hayes & Wilson 2008; Daland et al. 2011, on the “inductive baseline” approach to modeling experimental data). In this

way, I argue that predictors that significantly increase the performance of the model are at work when speakers assign stress to nonce words, and therefore hold some cognitive reality in lexical representations, limiting the kinds of analogy that are possible.

5 Analogical models “without structure”

5.1 A segmental similarity-based model

If we consider that participants are taking into account lexical properties to rate nonce words, the simplest analogical model that can try to capture the process that speakers are performing is one based on simple similarity between the segmental content of the words and the nonce words—a model that lacks comparisons that involve syllabic structure or access to non-local relations between distant segments, such as the final onset and the nucleus of the antepenultimate syllable.

When analyzing how similar a word is to other words in the lexicon, a common metric in the literature is to compute its Neighborhood Density (ND); that is, compute the number of words that differ in only one segment from the item that is analyzed (e.g., Luce 1986). However, when using such a metric, all of the nonce words in the experimental task would have a Neighborhood Density equal to zero: they are at least three changes away from a word in the lexicon.¹⁶ This was empirically confirmed by calculating the ND of every experimental item using the Spanish version of the CLEARPOND (Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities) Database (Marian et al. 2012). The results were that all items had a ND equal to zero; that is, there were no words that differed in only one segment from any of the experimental items. Given this specific limitation, I decided to compute similarity in a slightly different way by using an implementation of the Generalized Context Model (GCM: Nosofksy 1986; 1990), which assumes that neighbors are on a continuous scale of similarity. When applied to linguistic data, this model assumes that the rating of a nonce word is determined by calculating its similarity to a set of items, and the measure of similarity is determined by counting how many segment changes (insertions or deletions) are necessary to arrive at a word from the lexicon given a single nonce word; that is, the Levenshtein distance between the two items. As Albright & Hayes (2003) and Albright (2009b) clearly state, this analogical model constitutes the baseline of comparison to any other model, given that it only assumes that whenever assigning a novel item to a particular class (in this case, to a stress class), the model just compares that item to every existing member of that class. The similarity of that item to that class is then the sum of similarities to each class member, and the probability of assigning that item to that class is proportional to its similarity to all classes (i.e., we need to compensate for how similar an item is to all classes in general).

When dealing with stress in Spanish, this model—which necessarily assumes that classes with more members will increase the probability that a new item will join that class—makes the prediction that most items will receive penultimate stress, following the patterns of the Spanish lexicon. To implement this model, I used the *stringdist* package (van der Loo 2014) in R (R Core Team 2015), which calculates the Levenshtein distance between two strings. Given that the goal was to compare similarities between nonce words and words, I used the function *stringsim*, which calculates similarities between strings by first calculating the Levenshtein distance between them, then dividing that distance over the maximum possible distance, and finally subtracting the result from 1. In this way, the function also normalizes the Levenshtein distance measure by word length. This process provides a score between 0 and 1, where 0 corresponds to complete dissimilarity to a word and 1

¹⁶ See Yao (2011) for the general finding that long words lack lexical neighbors.

to complete identity to it. As a lexicon, I used the one provided by Davies & Perea (2005), which consists of 31,395 types. I converted the lexicon and all the experimental stimuli to phonological form before establishing their similarity.^{17,18} Given that the goal was to compare stress assignment probabilities, I decided to code stimuli as unstressed and compute an extra change (i.e., an insertion) that was needed to go from an unstressed vowel to the stressed one in the lexical form across all conditions. In this way, segmental similarity is computed independently of any particular stress pattern. I split the lexicon by stress pattern, and compared the unstressed stimuli to the antepenultimate stressed lexicon, the penultimate stressed lexicon, and the “other stress” (mostly final) stressed lexicon, as well as with the whole lexicon.¹⁹ The probability p of a nonce word to get one kind of stress was then the ratio between the sum of its similarities to each word of a particular kind of stress over the sum of its similarities to every word in the lexicon (to compensate for its similarity to all stress classes in general), as expressed in (1):

$$(1) \quad p \text{ of a nonce word to get } \alpha \text{ stress} = \frac{\Sigma \text{similarities to } \alpha \text{ stress lexicon}}{\Sigma \text{similarities to whole lexicon}}$$

For instance, if we consider the nonce word [da.'ti.pem.bo], we would first calculate the similarities of its unstressed version [da.ti.pem.bo] to each antepenultimate stressed word of the Spanish lexicon. The similarity to *antípoda* ‘antipode’ is equal to 0.11, the similarity to *somnífero* ‘sleeping pill’ is equal to 0.22 and so forth. Once we compute the similarities to the whole antepenultimate lexicon, we can sum the results, which in the case of [da.ti.pem.bo] are equal to 382.44. Repeating the same process with the penultimate lexicon gives a sum of similarities equal to 3194.56. Finally, the sum of similarities to the lexicon that includes words with “other stress” is equal to 909.5. Consequently the sum of similarities to the whole lexicon of Spanish is equal to 4486.5. To obtain the corresponding probability that the unstressed nonce word [da.ti.pem.bo] gets antepenultimate stress, we then would divide the sum of its similarities to the antepenultimate lexicon over the sum of its similarities to the whole lexicon, which yields 0.085. Its probability to get penultimate stress is of 0.712, and its probability to get another kind of stress is of 0.203. Considering the formula in (1), we would calculate the probability of [da.ti.pem.bo] to get antepenultimate stress as in (2):

$$(2) \quad \frac{\Sigma \text{similarities of [da.ti.pem.bo] to antepenult stress lexicon}}{\Sigma \text{similarities of [da.ti.pem.bo] to whole lexicon}} = \frac{382.44}{4486.5} = 0.085$$

The Spearman’s correlation of the probabilities yielded by this model²⁰ with respect to the mean z-scores of the participants’ data came out significant ($r_s = .60$, $p < .001$), so we can assume that segmental similarity is a good model to predict stress assignment. However, as we can observe in Figure 2, this model is mostly just predicting the split between antepenultimate and penultimate scores (that is, the fact that penultimate stress is pre-

¹⁷ A script written by Kevin Tang was used for the conversion from orthography to phonology. The grapheme-to-phoneme rules that the script implemented were designed together with the author.

¹⁸ The trill and the tap were coded as different phonological segments in the lexicon used for all the analogical models. A separate set of analyses with a lexicon in which the trill was coded as a geminate tap was performed. The results were very similar to the model in which both segments were codified as separate phonemes, and the correlations with the experimental data were also not significant.

¹⁹ Given that the experimental stimuli were intended to be nominal forms, comparisons to only the nominal forms of the lexicon were also made. The results and the correlations with the experimental results were not significantly different from the correlations obtained when the whole lexicon was used. For that reason, I report the original results, which correspond to comparisons to the whole lexicon.

²⁰ Given that many data points in the calculations were concentrated at the ends of the scale, we used non-parametric (Spearman) regressions for all of the models (Hayes & Wilson 2008).

ferred over antepenultimate stress across all conditions), but fails to capture the variance across segmental conditions within each stress condition.²¹ For instance, when we focus only on antepenultimate stress, the results predicted by a segmental similarity model do not correlate significantly with the mean z-scores of the participants' ratings ($r_s = .06$, $p = .598$), as we can observe in Figure 3.

5.2 A frequency-weighted Neighborhood Density model

Another property of the lexicon that seems to play a role in experimental tasks involving word recognition, such as lexical decision tasks, is token frequency (e.g., Whaley 1978; Taft 1979; Segui et al. 1982). As such, lexical frequency has been argued to be a cognitive property of the way in which our mental lexicons are organized. One can thus assume that similarity across words can be weighted by token frequency; that is, a nonce word will be perceived as more similar to a neighbor that is three segmental changes away and has a high token frequency than to a neighbor that is three segmental changes away but has a low token frequency. The main idea is that more frequent words have a higher resting activation, and hence exert a stronger effect on nonce words.

The Generalized Neighborhood Model (GNM: Bailey & Hahn 2001) is an improvement on the GCM in that it incorporates a term that encodes lexical token frequency into the calculation of Neighborhood Density. To model the experimental data, I used an adaptation of this model and decided to multiply the similarity measures for each nonce word obtained in the GCM model described in the previous subsection by the log token frequency of each word—frequency effects are usually modeled as a log function of token frequency (e.g., Luce & Pisoni 1988; Vitevitch et al. 1999). As a result, the probability p of a nonce word to get one kind of stress was then the ratio between the sum of its similarities to each word of a particular kind of stress multiplied by its log token frequency

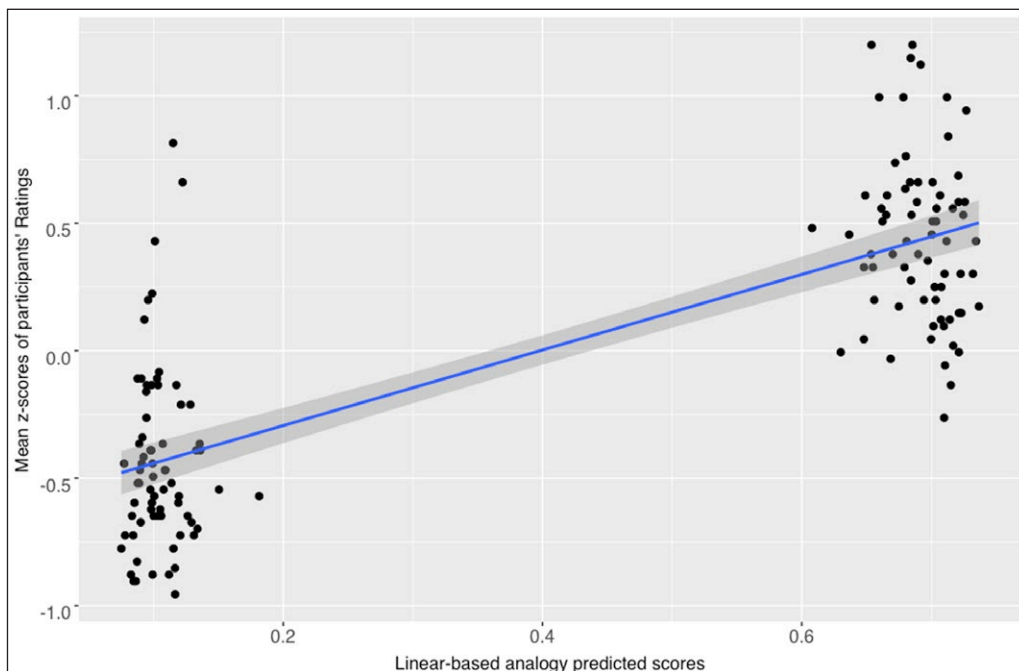


Figure 2: Linear regression of mean z-scores (+/– standard error) of participants' ratings and probabilities of stress assignment by a segmental similarity analogical model.

²¹ This is not a failure of the model itself, but a failure of any simple linear correlation at capturing a bimodal distribution.

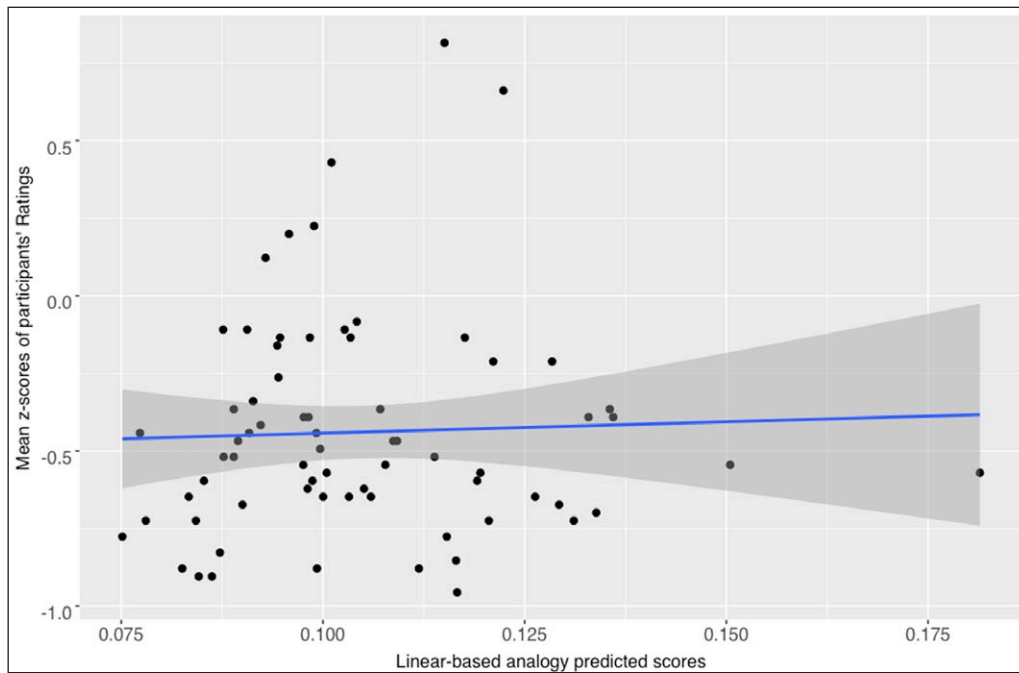


Figure 3: Linear regression of mean z-scores (+/- standard error) of participants' ratings on antepenultimate stress words and probabilities of antepenultimate stress assignment by a segmental similarity analogical model.

over the sum of its similarities to every word in the lexicon multiplied by its log token frequency, as we can see in (3). Frequencies for the Spanish lexicon were obtained from the same database in Davies & Perea (2005):

$$(3) \quad \text{p of a nonce word to get } \alpha \text{ stress} = \frac{\sum(\text{sim.} \times \log(\text{Freq})) \text{ to } \alpha \text{ stress lexicon}}{\sum(\text{sim.} \times \log(\text{Freq})) \text{ to whole lexicon}}$$

For instance, with the same example as in the previous subsection, the unstressed word [da.ti.pem.bo] would have a similarity score to *antípoda* 'antipode' equal to 0.11, and a similarity score to *somnífero* 'sleeping pill' of 0.22. Each of those values would be multiplied by the log token frequency of the corresponding word; for instance, in the case of *antípoda* 'antipode', which has a frequency of 0.18 per million, the frequency-weighted similarity would be of 0.0198, and in the case of *somnífero* 'sleeping pill', which has a frequency of 1.61 per million, it would be equal to 0.3542. Adding all the frequency-weighted similarities of [da.ti.pem.bo] to the antepenultimate lexicon equals 4,580, while the sum of all the frequency-weighted similarities to the whole lexicon yields a value of 55,183. The ratio between those two values is 0.083; that is, there is an 8.3% probability that the string [da.ti.pem.bo] receives antepenultimate stress. When we do the same calculations for penultimate stress, we obtain a probability of 0.707, and a value of 0.21 for the "other stress" condition. Taking into account the formula in (3), we would calculate the probability of [da.ti.pem.bo] to get antepenultimate stress as in (4):

$$(4) \quad \frac{\sum(\text{freq-weighted sim. of [da.ti.pem.bo] to antepenult stress lexicon})}{\sum(\text{freq-weighted sim. of [da.ti.pem.bo] to whole lexicon})} = \frac{4580}{55183} = 0.083$$

The Spearman's correlation of the probabilities yielded by this model with respect to the mean z-scores of the participants' data came out significant again and improved with respect to a model without frequency coded into it ($r_s = .67$, $p < .001$), so we can argue that not only segmental similarity is needed in the model, but also some measurement

of token frequency of the words that are in the analogical base. However, this significant correlation is again the result of a model that is able to predict the split between antepenultimate and penultimate scores, without capturing the internal variance in each of the stress conditions. Given that the goal in this paper is to capture the variance in the antepenultimate stress conditions to better understand the effects of segmental properties on antepenultimate stress assignment, I looked again only at the results for antepenultimate stress. In this case, and as shown in Figure 4, the results predicted by a frequency-weighted segmental similarity model do not correlate significantly with the participants' data ($r_s = .05$, $p = .677$).

6 Maximum entropy models

In order to refer to non-local properties of the lexicon, such as syllable weight or the effects that non-adjacent segments can have on each other, it is crucial to build analogical sets, both of the lexicon and the nonce words, that include structural information. To this end, some analogical models (e.g., AML in Eddington 2000) encode the position of each segment in the word and rely on this information to pick up non-local dependencies between segments. For instance, the restriction on antepenultimate stress when the penultimate syllable is heavy can be expressed as a restriction on the nucleus of the antepenultimate syllable (i.e., do not bear stress) when there is a branching rhyme in the penultimate syllable. On the other hand, Hayes & Wilson (2008) present a phonotactic learning model that is based on a maximum entropy (MaxEnt) grammar—a grammar that uses weighted constraints to assign probabilities to outputs. Essentially, the phonotactic learner receives a lexicon with token frequencies,²² where every segment in a word is also coded for a set

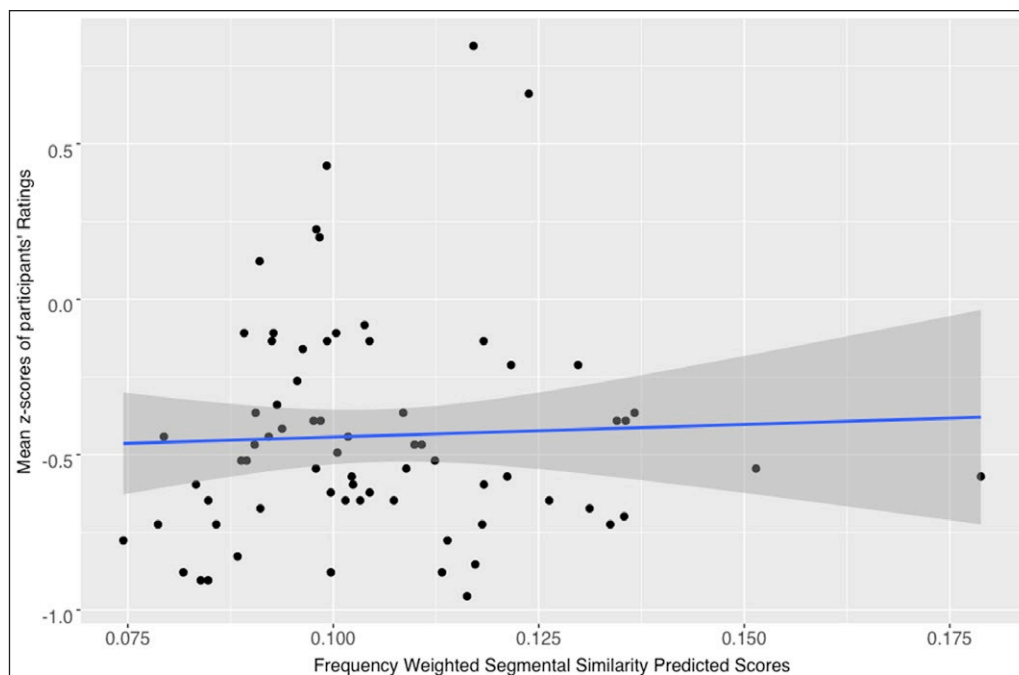


Figure 4: Linear regression of mean z-scores (+/– standard error) of participants' ratings on antepenultimate stress words and probabilities of antepenultimate stress assignment by a frequency-weighted analogical model.

²² It has been claimed (e.g., Daland et al. 2001; Albright 2009b) that type frequency is a better predictor than token frequency with respect to phonotactic judgments. On this basis, all the maximum entropy models presented in this paper have been implemented both with and without token frequency weighting. I did not find a significant difference between the two kinds of models, so the results reported in here correspond to the initial models, which include token frequency weighting.

of features, and induces a set of weighted constraints that is able to assign well-formedness scores to the words in the lexicon (for details on how the constraints are obtained and weighted, see Hayes & Wilson 2008: §§ 3–4). The weights can be thought of as a measure of the importance of each constraint. The set of constraints that are obtained can then be used to evaluate any given set of novel items. Given that constraints are weighted, this is a useful model to account for gradient intuitions like the ones native-speakers report with respect to nonce words that present different phonotactic violations (e.g., Hayes et al. 2009; Hayes & White 2013; Colavin et al. 2014).

The model allows for flexibility in terms of how the researcher defines the lexicon and which properties are encoded in it. The lexicon can thus be defined on a set of features that encode stress, syllable weight, etc. The model also allows the user to expand or restrict the constraint search space—that is, how many adjacent segments (i.e., matrices of features) a constraint can refer to. For instance, to express that in English a consonant cluster **bn* is impossible, we only need a window size of 2, which would generate a constraint of the form **[+ consonantal, + labial] [+ nasal]*. However, to express that the consonant cluster **spk* is unattested in English (as opposed to the well-formedness of *spl*, or *spr*), we need to make reference to three adjacent segments, in a constraint such as **[+ strident] [+ consonantal, + labial] [+ consonantal, + dorsal]*. To express constraints that make reference to more adjacent segments, the model allows for window sizes up to 4. However, increasing the window size above 4 makes the search space for constraints exponentially bigger so that the computations become hard to implement (Hayes & Wilson 2008: §4.1) Finally, these models permit the encoding of autosegmental tiers (Goldsmith 1979), such as a vowel tier or *projection*, which in turn allows constraints that only refer to the sequence of vowels without taking into account the consonantal sounds in between. These projections are defined on a subset of the features that are specified for all segments (for instance, on [+ syllabic] for a vowel projection).

6.1 Inductive baseline: A linear feature-based MaxEnt model

Hayes & Wilson (2008) argue that the kind of algorithmic procedure they propose can be used to evaluate and test different theories about the nature of the representations that are involved in an analogical process. However, we first need an *inductive baseline*, a very simple model against which more complex ones can be compared. The main claim is that if introducing some structural element to the lexical representation results in the learning of phonotactic patterns that could not be learned without it, the structural element is supported.

I consider the inductive baseline to be a purely linear approach in which segments are specified for a bundle of features. I also assume a vowel projection (defined as [+ syllabic]) in the inductive baseline, given that it has proven to be relevant for different languages and language families (e.g., for Finnish and Hungarian, see Goldsmith 1985; for Shona, see Beckman 1997). The lexicon I used was a constructed phonological lexicon based on Davies & Perea (2005), and I also used the token frequencies per million reported there. I constructed a feature chart based on a standard feature set for Spanish phonemes (Harris 1967), and I also coded stress in the corresponding vowel (i.e., vowels had two versions on the feature chart, unstressed and stressed, which only differed in a feature [Stress], which had a binary value defined only for vowels, left unspecified for consonants). The feature chart is presented in Table 5.

I set the maximum constraint size at 3 (i.e., only three adjacent feature matrices/segments could participate in a constraint), limited the number of constraints to 100,²³ and ran the phonotactic learner 10 separate times. Constraint selection is stochastic (for practical

²³ The results show that across all models about 60% of the constraints have a weight of 0 (that is, they do not penalize novel forms). This indicates that limiting the number of constraints to 100 is sufficient to pick up both high and low weighted phonotactic constraints in Spanish.

Table 5: Feature chart for Spanish consonants and vowels.

	p	b	m	f	t	d	s	l	r	ʀ	n	ɲ	ʎ	ɟ	k	g	x	a	e	i	o	u	w	j
SYLL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-
CONS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-
SON	-	-	+	-	-	-	-	+	+	+	+	+	-	-	-	-	-	+	+	+	+	+	+	+
CONT	-	-	-	+	-	-	+	+	+	+	-	-	+	-	-	-	+	+	+	+	+	+	+	+
DELR	-	-	0	+	-	-	+	0	0	0	0	0	+	+	-	-	+	+	+	+	+	+	+	+
APPR	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	+	+	+	+	+	+	+
TAP	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TRILL	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
NASAL	-	-	+	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
VOICE	-	+	+	-	-	+	-	+	+	+	+	+	-	-	-	+	-	+	+	+	+	+	+	+
SPR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CONSTR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LAB	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-
ROUND	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-
LABDEN	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
COR	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
ANT	0	0	0	0	+	+	+	+	+	+	+	-	-	-	0	0	0	0	0	0	0	0	0	0
DIST	0	0	0	0	-	-	-	-	-	-	-	+	+	+	0	0	0	0	0	0	0	0	0	0
STRID	0	0	0	0	-	-	+	-	-	-	-	-	+	+	0	0	0	0	0	0	0	0	0	0
LAT	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DORS	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	+	+	-	-	-	-	-	+	+
HIGH	0	0	0	0	0	0	0	0	0	0	0	+	0	0	+	+	+	-	-	+	-	+	+	+
LOW	0	0	0	0	0	0	0	0	0	0	0	-	0	0	-	-	-	+	-	-	-	-	-	-
FRONT	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	-	+	+	-	-	-	+
BACK	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	-	-	-	+	+	+	-
TENSE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+	+	+	+	+

reasons; see Hayes & Wilson 2008: §§4.1 and 4.2), so the model learned slightly different constraints and weighted them slightly differently in each run. However, the generated grammars were very similar—as we can observe in the correlation matrix in Table 6—so I decided to average over the penalty scores assigned to each of the forms in the testing data (i.e., the nonce words).²⁴ To obtain predictions from these penalty scores, Hayes & Wilson (2008) propose that the maxent value of every penalty score is obtained by negating the score, and raising the base of natural logarithm *e* to the result. The probability of the nonce word (i.e., the predicted rating by a speaker) is then obtained by dividing that maxent value for a value *Z*, which is a free parameter whose value is determined on a best-fit to the participant’s data basis.²⁵

A Spearman’s correlation of the predicted ratings obtained by this model with the mean z-scores of the participants’ data reached significance ($r_s = .54, p < .001$), so we can assume that a model that encodes featural information on the segments is relevant for

²⁴ Correlation matrices were run for all MaxEnt models reported in this paper. All runs within each model were highly correlated, so in all cases the penalty scores were averaged over 10 runs.

²⁵ Given the use of Spearman’s rank correlations to fit the predicted scores by a MaxEnt model and the participants’ ratings in this paper, the best-fit procedure using a *Z* value becomes irrelevant. Once the predicted scores are ranked, there is no possible effect on the ranking by dividing the predicted scores by a *Z* value.

Table 6: Correlation matrix of 10 separate runs of the linear feature-based MaxEnt model.

	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Run1	1.00	0.96	0.99	0.98	0.98	0.98	0.97	0.98	0.99	0.98
Run2	0.96	1.00	0.95	0.98	0.98	0.99	0.97	0.96	0.97	0.99
Run3	0.99	0.95	1.00	0.98	0.98	0.98	0.97	0.99	0.99	0.98
Run4	0.98	0.98	0.98	1.00	0.99	0.99	0.99	0.98	0.99	0.99
Run5	0.98	0.98	0.98	0.99	1.00	0.99	0.97	0.98	0.98	0.98
Run6	0.98	0.99	0.98	0.99	0.99	1.00	0.98	0.98	0.99	0.99
Run7	0.97	0.97	0.97	0.99	0.97	0.98	1.00	0.97	0.98	0.98
Run8	0.98	0.96	0.99	0.98	0.98	0.98	0.97	1.00	0.99	0.97
Run9	0.99	0.97	0.99	0.99	0.98	0.99	0.98	0.99	1.00	0.98
Run10	0.98	0.99	0.98	0.99	0.98	0.99	0.98	0.97	0.98	1.00

predicting stress assignment in nonce words, even if just at the level of predicting the split between antepenultimate and penultimate stress, as in the purely analogical models. When we consider only the predicted ratings for antepenultimate stress, the model improves with respect to the naive analogical models described in §5, even if it still does not reach significance in its correlation with the participants' data ($r_s = .22$, $p = .07$). Figure 5 illustrates the results of this correlation.

6.2 A MaxEnt model with syllable boundaries

A possible structural element that could help to predict stress assignment would be to encode syllable boundaries in the lexicon. The argument would be that if we allow for the model to pick up constraints that refer to syllable boundaries, these could refer to different coda and onset sequences. For instance, a constraint could penalize sequences of two heterosyllabic vowels where the first one is stressed in cases where the first syllable lacks a coda and the second one contains one (i.e., *[V.VC]). A constraint of that form would explain the regular lack of hiatus in Spanish nominal forms when the second syllable is closed.²⁶ With that purpose, I created a new version of the lexicon that included syllable boundaries, codified as an extra-segment (#) defined in the feature table with a feature [+ SyllBound]. All other segments were left unspecified for this feature, and this segment was left unspecified for all the other features in the chart.

In this case, I set the maximum constraint size at 4, so that the learner could pick up constraints that refer, for instance, to sequences such as V#VC.²⁷ I again limited the number of constraints that the learner could generate to 100, and I ran the model 10 times and averaged over the scores obtained for the nonce words. I then performed the transformation described in §6.1, which allows for obtaining predicted ratings from this model.

²⁶ Both reviewers point out some exceptions to this rule such as ['ka.os] 'chaos'. Under this account, however, this can be explained by claiming that the constraint would have a high penalization score, but it could still be violated by some items. There are of course also some verbal forms that display this pattern, such as the subjunctive forms ['ri.an] 'laugh.SBJV.3.PL' or ['le.an] 'read.SBJV.3.PL'.

²⁷ It is possible that this is not a completely fair comparison to the other MaxEnt models given the increase in window size. However, most constraints will only refer to a 3-segment window, given that constraints that refer to dependencies between the nucleus of a heavy syllable and the following onset are unlikely. Moreover, they would not affect the correlations with the participants' data, because there are no conditions with that segmental configuration. In any case, running this specific model with the maximum constraint size set a 3 predicts ratings that have the same correlations both with respect to the whole set of conditions in the experimental data and with respect to the antepenultimate stress conditions. This shows that the set of constraints obtained under that window size produces the same results that when the constraint search space is increased.

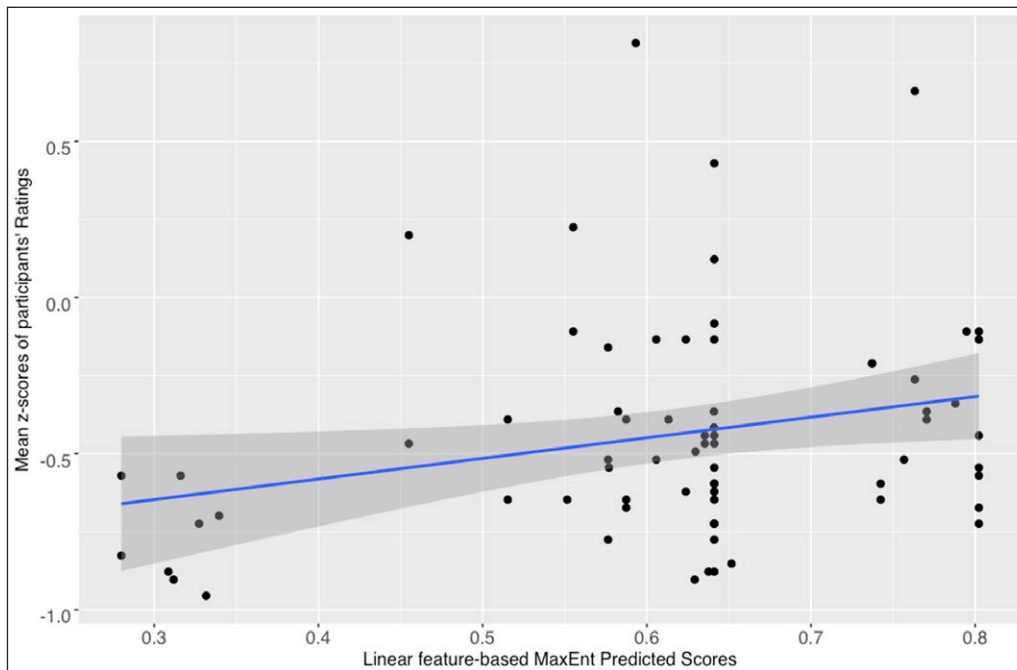


Figure 5: Linear regression of mean z-scores (+/– standard error) of participants’ ratings on antepenultimate stress words and predicted scores on antepenultimate stress words by a linear feature-based maximum entropy model.

The results of a Spearman’s correlation of the predicted ratings to the mean z-scores of the participant’s ratings showed an improvement with respect to the inductive baseline ($r_s = .67$, $p < .001$), which shows a better grasp at explaining the split between antepenultimate and penultimate stress participants’ ratings. However, as presented in Figure 6, the predicted ratings on the antepenultimate conditions failed again to reach significance ($r_s = .02$, $p = .844$), and were significantly worse than in the model that did not encode syllable boundaries.

6.3 A MaxEnt model with syllable weight

Relations defined locally seem to improve the performance of the model in general, but antepenultimate stress assignment may be constrained by some “hidden structure”—structure that is not detectable in the phonetic string, but that is phonologically present and that provides order and systematicity in the data pattern (Hayes & Wilson 2008: 425). Phonological weight is claimed to be a structural component that affects stress assignment. Latin, for instance, had a stress rule that made reference to syllable weight: stress falls on the penultimate syllable if it is heavy (i.e., it ends on a consonant or a long vowel), or on the antepenultimate syllable otherwise. In past accounts of the antepenultimate stress patterns of Spanish (Harris 1983; Roca 1991; Lipski 1997; Baković 2016; *i.a.*), it is debated whether Spanish is weight-sensitive or not, so I decided to code syllable weight (instead of syllable boundaries) in the lexicon so that the phonotactic learner could pick up constraints that made reference to it. To that end, I added a new feature [+/- Heavy] to the feature chart, and coded every vowel in the lexicon and in the stimuli as heavy or light in direct relation to the syllable it is a part of. The feature was coded only in the vowels so it projected to the vowel tier. The learner could thus pick up constraints that made reference to sequences of light/heavy vowels (i.e., syllables).

I ran the model 10 times with the maximum constraint size set at 3 and the maximum number of constraints that the learner could generate limited to 100. I averaged over the 10 runs to obtain the penalty scores generated by the grammar for each of the nonce words. I again transformed the results under a best-fit analysis to obtain the predicted ratings by the model.

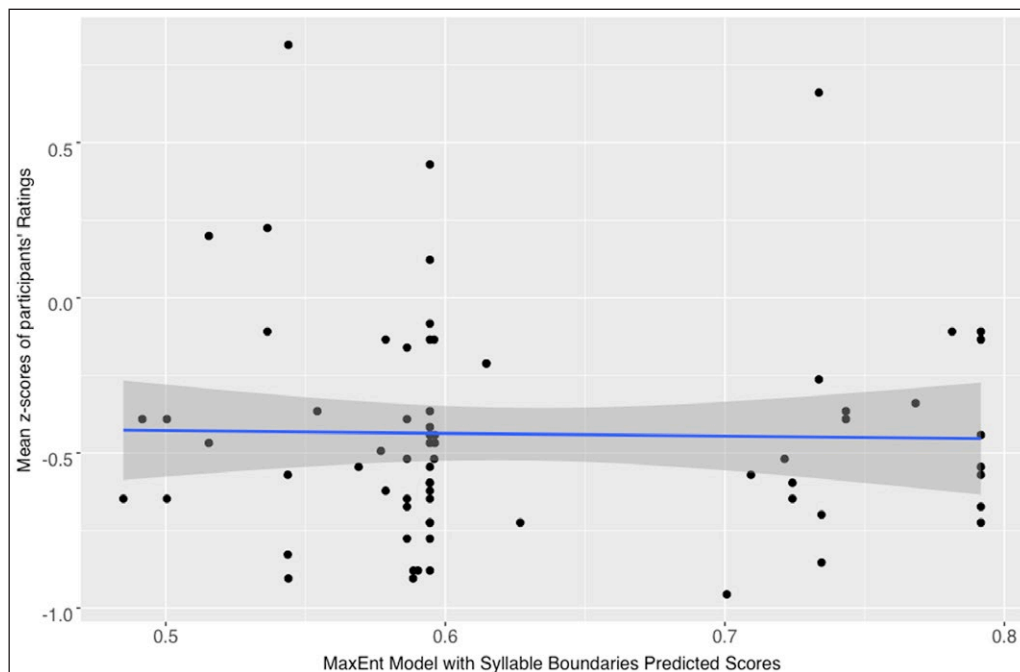


Figure 6: Linear regression of mean z-scores (+/– standard error) of participants’ ratings on antepenultimate stress words and predicted scores on antepenultimate stress words by a maximum entropy model with syllable boundaries.

A Spearman’s correlation shows significance between the predicted ratings and the mean z-scores of the participants’ data ($r_s = .57$, $p < .001$), capturing the split between antepenultimate and penultimate ratings. Crucially, with this model, which encodes syllable weight, the predicted ratings on the antepenultimate stress conditions reach a significant correlation with the mean z-scores of the participants’ data in the same conditions ($r_s = .49$, $p < .001$), as shown in Figure 7.²⁸ These results point to the relevance of syllable weight as a structural element that helps predicting the stress assignment patterns of Spanish.

6.4 On the nature of the trill in Spanish

One of the debates in the literature that antepenultimate stress patterns of Spanish can help elucidate is whether the trill is a distinct phoneme of the language or whether it is an ambisyllabic geminate tap. As discussed earlier in the paper, the argument for the latter is introduced by Harris (1983), who shows that there are no words with antepenultimate stress in the language that have a trill in the onset of the final syllable. His claim is that the trill is an ambisyllabic geminate tap, which in this case would close the penultimate syllable. On the other hand, Roca (1998), Lipski (1990) and Baković (2009) point out that there are no antepenultimate stressed words that have a palatal lateral or a palatal nasal on the onset of the final syllable either, and that these facts are all due to historical reasons (i.e., all these consonants usually developed from geminates or consonant clusters in Latin), and not because these sounds make the previous syllable heavy.

I addressed this issue in the modeling by having an extra set of runs of the phonotactic learner in which the heterosyllabic vowels before a trill were coded as [+ Heavy]—treating /r/ as an ambisyllabic geminate /r.r/closing the previous syllable—to see whether the

²⁸ It is noticeable that this maximum entropy model predicts several different conditions to have the same rating (see the column of dots around 0.7 in the figure), given that they are penalized only by one constraint which penalizes antepenultimate stress when the penultimate syllable is heavy (i.e., *[+ Stress][+ Heavy] [–Heavy] defined in the vowel projection). Participants’ ratings show much more variability in those cases, so the remaining variance within that data is yet to be explained.

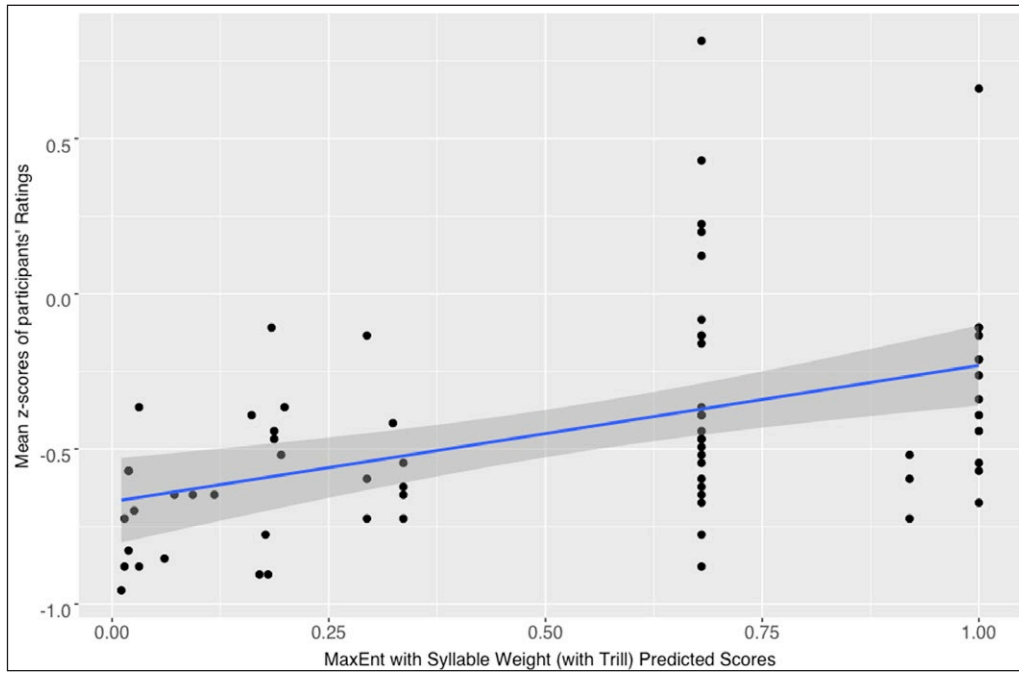


Figure 7: Linear regression of mean z-scores (+/– standard error) of participants’ ratings on antepenultimate stress words and predicted scores on antepenultimate stress words by a maximum entropy model with syllable weight.

model performed better than the one in the previous subsection, in which those syllables were coded as [– Heavy]. I again set the maximum constraint size at 3, and the maximum number of constraints that the learner could pick up was set to 100. I averaged over the 10 runs to obtain the penalty scores generated by the grammar for each of the nonce words, and performed the transformation on the results to obtain predicted ratings.

With this set up, while the predicted ratings still correlate significantly with the mean z-scores of the participants’ data ($r_s = .63$, $p < .001$), explaining the split between antepenultimate and penultimate ratings, when it comes to the subset of antepenultimate stress conditions, its explanatory power decreases ($r_s = .45$, $p < .001$). A significant correlation is still found, but it could be due to the other conditions correlating in the same way than in the previous model. Therefore, considering that the Spanish trill is a single phoneme of Spanish—and that it does not contribute to syllable weight as an onset—has better correlations with experimental data than taking the trill to be a geminate tap. Moreover, an Akaike Information Criterion (AIC) model comparison (see Shih 2017 for an account of why AIC-based model comparisons allow for comparing different competing grammars) favors the model that considers the trill a singleton consonant (AIC = 42.92) over the model that considers it a geminate tap (AIC = 47.62). The author argues that generally any difference greater than 10 in AIC between two grammars is considered large. This can be translated into an evidence ratio E by the formula in (5), where a 10-point difference between two grammar is equivalent to about a 150 to 1 odds that the highest AIC model has no evidential support of being as good as the lowest AIC model:

$$(5) \quad E_{i,j} = \frac{1}{e^{-(1/2)\Delta_j}} \text{ for models } i \text{ and } j \text{ where } \Delta_j \text{ is } AIC_j - AIC_i.$$

Even if in this case the difference is smaller than 10 points between the AIC scores, the 4.7-point difference in AIC scores indicates about a 10.5 to 1 odds ($E = 10.486$) that there is more support for a grammar that includes the trill as a singleton consonant in predicting the participant’s ratings. This strongly suggests that the Spanish phonological system includes the trill as a singleton consonant.

7 Discussion of modeling procedures

I have explored in detail a specific part of the phonotactics of Spanish; namely, the mechanisms that take part in the process of stress assignment in the case of antepenultimate stress. The experimental data does not support a system in which stress assignment of novel forms is performed under exceptionless rules, but makes salient the gradient intuitions that native speakers have when faced with this task. Moreover, this gradient seems to be dependent on particular segmental and syllabic configurations, so that an analogical procedure that bases its predictions on sublexical patterns seems to be at work. These results are in line with most of the experimental work that has been done in this area (e.g., Aske 1990; Eddington 2000; Bárkányi 2002; Face 2003 *et seq.*; Shelton 2007 *et seq.*).

However, while most of the experimental literature does not define the nature of the analogical mechanism that they claim to be operating in Spanish stress assignment, I tried to model the experimental results by making explicit the kind of task that speakers perform when they are faced with a novel form and rely on “analogy” to rate it. To this end, I defined a series of analogical models that ranged in the amount of structure they were able to make reference to. Given that I intended to capture the variance in the antepenultimate stress conditions, a summary comparison of the correlations that the predicted ratings of each model obtained with respect to the participants’ ratings in the antepenultimate conditions, together with R^2 values, is given in Table 7.

We can see that the model that by far best fits the participants’ data is one in which syllable weight is encoded. We can argue thus for the cognitive reality of this structural element and for its relevance in the process of stress assignment in Spanish. With respect to the nature of the trill, the modeling has provided evidence that supports its existence as a singleton consonant, as opposed to a single rhotic phoneme that considers the trill a geminate tap. The model that encodes the trill as a singleton consonant (i.e., that does not make the preceding syllable heavy) has a higher correlation with the participants’ data than the model that considers the trill to be a geminate tap making the preceding syllable heavy. An AIC-based model comparison also favors the model that considers the trill a singleton.

Even if this study provides support for the role of syllable weight in stress assignment in Spanish, the results still leave somewhat unexplained the variance observed in the different conditions manipulating the onset of the final syllable. A trill as a final onset does not seem to preclude antepenultimate stress in the same way that a penultimate heavy syllable does. The nasal palatal and the postalveolar fricative segments as onsets of the final syllable do not make the previous syllable heavy, but still are overwhelmingly dispreferred by native speakers. The different phonotactic models that we have explored fail to capture the internal variance that native speakers are showing with respect to those conditions.

Table 7: Comparison of performance by each model—Spearman’s rank correlation and R squared—with respect to the experimental results on the antepenultimate stress conditions. Bold indicates the best performing model.

Model	r_s	R^2
Segmental similarity (GMC)	.06 (p = .598)	.0015
Frequency-weighted similarity (GNM)	.05 (p = .677)	.0018
MaxEnt – inductive baseline	.22 (p = .07)	.0675
MaxEnt – with syllable boundaries	.02 (p = .844)	.0001
MaxEnt – with syllable weight (trill)	.49 (p < .0001)	.1897
MaxEnt – with syllable weight (geminate tap)	.45 (p < .001)	.1335

An attempt to explain this variance might rely on a non-binary system of syllable weight. Gordon (2002; 2006) shows that most weight-sensitive stress systems are binary, but there are systems that include a three-way weight hierarchy. For instance, in Klamath (Barker 1964) or Telugu (Brown 1981), CVV syllables are heavier than CVC syllables, which in turn are heavier than CV syllables. These systems, according to Gordon (2002: 68), match phonological weight with duration.²⁹ In light of this, Spanish could also have a more fine-grained weight distinction based on duration, which could explain the exceptionality of the trill condition.

One way of incorporating duration and the role of final onsets into the computation of phonological weight would be to rely on interval theory, which considers that the domain of weight computation is an interval—that is, a rhythmic unit that spans from a vowel up to (but not including) the next vowel (i.e., a V-to-(V) interval)—and not the syllable (Steriade 2012; Hirsch 2013). Moreover, given that duration is a continuous property, syllable weight also becomes gradient under such an account (Ryan 2011; García 2017, for an implementation). However, based on different durations reported in the Spanish literature for the segments under consideration (Borzzone de Manrique & Signorini 1983; Del Barrio & Torner Castells 1999; Lavoie 2001), a durational account based on intervals would not capture the experimental results. Another way of incorporating duration into weight computation would be by acknowledging that the nature of a consonantal sound can affect the duration of the preceding vowel. For instance, voiced consonants in English make preceding vowels longer than their voiceless counterparts (e.g., Locke & Heffner 1940; Peterson & Lehiste 1960). We can hypothesize that different final onsets could make the preceding vowel in the penultimate syllable longer. If we consider that weight computation is made on syllables, but that it takes the duration of the nucleus into account, then a possible weight system based on vowel duration could be at work for Spanish. However, preliminary results from a production task in Rioplatense Spanish ($n = 12$), which measured the duration of the onsets under discussion in both stress conditions, do not provide an explanation for the variability in the experimental data.

Finally, another possible explanation for the remaining variance would be that speakers are sensitive to other segmental dependencies. Wilson (2016) shows that recombination errors—that is, changes in a segmental string that is supposed to be remembered—support the existence of both consonantal and edge dependencies. In particular, there seems to be a dependency between the segment at the beginning and the segment at the end of any given word. However, coding a consonantal tier into a maximum entropy model by means of a consonantal projection, so that the model could pick up consonantal dependencies, does not provide better correlations to the experimental data presented in this paper. Further investigation should address the effects of encoding edge dependencies into maximum entropy models to check whether those dependencies can account for some of the unexplained variance.

8 Conclusions

This paper has explored the properties of antepenultimate stress in Spanish and whether there are rules that govern its distribution. We have looked at some of the different restrictions on antepenultimate stress in Spanish; specifically, we have considered what is the relation between these restrictions and structural properties such as syllable weight, and we also have considered how these restrictions relate to historical facts about the

²⁹ CVV heavy and CVX heavy systems match duration with phonological weight. Other systems that take into consideration vowel quality for weight distinctions correlate with more complex measures, such as total perceptual energy (see Gordon 2002: §6 for details).

language. In doing so, we have also analyzed how evidence coming from the properties of these restrictions on antepenultimate stress can shed light into the nature of the phonological representation of the trill in Spanish. We have pursued these goals by using both an experimental task and a data-modeling procedure that intended to capture the stress assignment process.

The experimental data provided support for a quantity-sensitive system for Spanish stress assignment, given that heavy penultimate syllables precluded antepenultimate stress. As for the nature of the trill, it generally provided support for its phonemic status (as opposed to a geminate tap account): when the trill appeared in the final onset position, it did not fully prevent antepenultimate stress. Moreover, the experimental data showed that speakers have gradient intuitions when it comes to assessing the productivity of specific restrictions in their grammars, providing support for weighted constraints in phonological grammars.

The second part of the paper dealt with modeling the mechanisms that could be at play when speakers assign stress to (novel) words. The results of the comparisons between models also provided support for a quantity-sensitive system for Spanish and for a phonemic representation of the trill, given that the model that best correlated with the experimental data was a maximum entropy model that encoded syllable weight and the trill as a singleton consonant in its lexical representations.

In summary, this paper has provided several pieces of evidence that converge in supporting that Spanish is quantity-sensitive and that the trill is a singleton consonant of the language. In doing so, it has also shown the utility of experimental methods and of modeling procedures in testing the phonotactics of a given language, and it has provided support for a grammatically-informed model of analogy to reproduce the stress assignment algorithm.

Abbreviations

AIC = Akaike information criterion, AML = Analogical Modeling of Language, ANT = anterior, APPR = approximant, CONS = consonantal, CONSTR = constricted glottis, CONT = continuant, COR = coronal, DELR = delayed release, DIST = distributed, DORS = dorsal, GCM = Generalized Context Model, GNM = Generalized Neighborhood Model, LAB = labial, LABDEN = labiodental, LAT = lateral, ND = Neighborhood Density, PL = plural, SBJV = subjunctive, SON = sonorant, SPR = spread glottis, STRID = strident, SYLL = syllabic

Acknowledgements

I would like to thank Eric Baković and an anonymous reviewer for their insightful and constructive suggestions, which have definitely improved this paper. I would also like to thank Ryan Bennett for his continuous support and for the thorough discussion along all stages of this project. Many thanks to Kevin Tang for comments on earlier drafts of this work and for his assistance with the grapheme-to-phoneme conversion for the lexicon in the analogical models. Thanks also to Claire Moore-Cantwell and Jason Shaw for valuable remarks and insights. Finally, I want to express my gratitude to the audiences at PhoNE 2017 at UMass Amherst, AMP 2017 at NYU, and to my colleagues at the Yale Phonology Reading Group for their observations. All errors and omissions remain of course my own.

Competing Interests

The author has no competing interests to declare.

References

- Albright, Adam. 2009a. Feature-based generalisation as a source of gradient acceptability. *Phonology* 26(1). 9–41. DOI: <https://doi.org/10.1017/S0952675709001705>
- Albright, Adam. 2009b. Modeling analogy as probabilistic grammar. In James Blevins & Juliette Blevins (eds.), *Analogy in grammar: Form and acquisition*, 185–213. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199547548.003.0009>
- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90(2). 119–161. DOI: [https://doi.org/10.1016/S0010-0277\(03\)00146-X](https://doi.org/10.1016/S0010-0277(03)00146-X)
- Alderson, Phil. 2004. Absence of evidence is not evidence of absence: We need to report uncertain results and do it clearly. *British Medical Journal* 328(7438). 476–477. DOI: <https://doi.org/10.1136/bmj.328.7438.476>
- Altman, Douglas & John Martin Bland. 1995. Absence of evidence is not evidence of absence. *British Medical Journal* 321(7003). 485. DOI: <https://doi.org/10.1136/bmj.311.7003.485>
- Alvord, Scott. 2003. The psychological unreality of quantity sensitivity in Spanish: Experimental evidence. *Southwestern Journal of Linguistics* 22(2). 1–12.
- Aske, Jon. 1990. Disembodied rules versus patterns in the lexicon: Testing the psychological reality of Spanish stress rules. *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society*, 30–45. DOI: <https://doi.org/10.3765/bls.v16i0.1685>
- Bailey, Todd & Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44(4). 568–591. DOI: <https://doi.org/10.1006/jmla.2000.2756>
- Baković, Eric. 2009. Abstractness and motivation in phonological theory. *Studies in Hispanic and Lusophone Linguistics* 2(1). 183–198. DOI: <https://doi.org/10.1515/shll-2009-1041>
- Baković, Eric. 2016. Exceptionality in Spanish stress. *Catalan Journal of Linguistics* 15. 9–25. DOI: <https://doi.org/10.5565/rev/catjl.182>
- Bárkányi, Zsuzsanna. 2002. A fresh look at quantity sensitivity in Spanish. *Linguistics* 40(2). 375–394. DOI: <https://doi.org/10.1515/ling.2002.016>
- Barker, Muhammad Abd-al-Rahman. 1964. *Klamath grammar* (University of California Publications in Linguistics 32). Berkeley, CA: University of California Press.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Becker, Michael, Nihan Ketrez & Andrew Nevins. 2011. The Surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language* 87(1). 84–125. DOI: <https://doi.org/10.1353/lan.2011.0016>
- Beckman, Jill. 1997. Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology* 14(1). 1–46. DOI: <https://doi.org/10.1017/S0952675797003308>
- Borzzone de Manrique, Ana María & Ángela Signorini. 1983. Segmental duration and rhythm in Spanish. *Journal of Phonetics* 11. 117–128.
- Bradley, Travis. 2001. A typology of rhotic duration contrast and neutralization. In Min Joo Kim & Uri Strauss (eds.), *Proceedings of the 31st North East Linguistics Society*, 79–97. Amherst, MA: GSLA Publications. DOI: <https://doi.org/10.1515/PROBUS.2006.001>
- Bradley, Travis. 2006. Spanish rhotics and Dominican hypercorrect /s/. *Probus* 18(1). 1–33. DOI: <https://doi.org/10.1515/PROBUS.2006.001>
- Brown, Charles. 1981. *A grammar of the Telugu language*. New Delhi: Asian Educational Services.

- Colavin, Rebecca, Roger Levy & Sharon Rose. 2014. Modeling OCP-Place in Amharic with the maximum entropy phonotactic learner. *Proceedings from the annual meeting of the Chicago Linguistic Society* 46. 27–41.
- Daland, Robert, Bruce Hayes, James White, Marc Garellec, Andrea Davis & Ingrid Normann. 2011. Explaining sonority projection effects. *Phonology* 28(2). 197–234. DOI: <https://doi.org/10.1017/S0952675711000145>
- Davis, Colin & Manuel Perea. 2005. BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods* 37(4). 665–671. DOI: <https://doi.org/10.3758/BF03192738>
- Del Barrio, Laura & Sergio Torner Castells. 1999. La duración consonántica en castellano. *ELUA: Estudios de Lingüística Universidad de Alicante* 13. 9–35. DOI: <https://doi.org/10.14198/ELUA1999.13.01>
- Duchon, Andrew, Manuel Perea, Nuria Sebastián-Gallés, Antonia Martí & Manuel Carreiras. 2013. EsPal: One stop shopping for Spanish word properties. *Behavioral Research Methods* 45. 1246–1258. DOI: <https://doi.org/10.3758/s13428-013-0326-1>
- Eddington, David. 2000. Spanish stress assignment within the Analogical Model of Language. *Language* 76(1). 92–109. DOI: <https://doi.org/10.1353/lan.2000.0022>
- Eddington, David. 2004. A computational approach to resolving certain issues in Spanish stress placement. In Timothy Face (ed.), *Laboratory approaches to Spanish phonology*, 95–116. Berlin: Mouton de Gruyter.
- Elordieta, Gorka & José Ignacio Hualde. 2014. Intonation in Basque. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 405–463. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199567300.003.0014>
- Face, Timothy. 2000. The role of syllable weight in the perception of Spanish stress. In Héctor Campos, Elena Herburger, Alfonso Morales-Front & Thomas Walsh (eds.), *Hispanic linguistics at the turn of the millennium: papers from the 3rd Hispanic Linguistics Symposium*, 1–13. Sommerville, MA: Cascadilla Press.
- Face, Timothy. 2003. Where is stress? Synchronic and diachronic Spanish evidence. In Paula Kempchinsky & Carlos Piñeros (eds.), *Theory, practice, and acquisition*, 21–39. Sommerville, MA: Cascadilla Press.
- Face, Timothy. 2004. Perceiving what isn't there: Non-acoustic cues for perceiving Spanish stress. In Timothy Face (ed.), *Laboratory approaches to Spanish phonology*, 117–141. Berlin: Mouton de Gruyter.
- Face, Timothy. 2006. Cognitive factors in the perception of Spanish stress placement: Implications for a model of speech perception. *Linguistics* 44(6). 1237–1267. DOI: <https://doi.org/10.1515/LING.2006.040>
- Face, Timothy & Scott Alvord. 2005. Descriptive adequacy vs. psychological reality: The case of two restrictions on Spanish stress placement. *Studies in the Linguistic Sciences* 32(2). 1–16.
- García, Guilherme. 2017. Weight gradience and stress in Portuguese. *Phonology* 34(1), 41–79. DOI: <https://doi.org/10.1017/S0952675717000033>
- Gibson, Mark. 2011. A typology of Stress in Spanish non-verbs. *Ianua: Revista Philologica Romanica* 11. 1–30.
- Goldsmith, John. 1979. *Autosegmental phonology*. New York, NY: Garland.
- Goldsmith, John. 1985. Vowel harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology* 2(1). 253–275. DOI: <https://doi.org/10.1017/S0952675700000452>
- Gordon, Matthew. 2002. A phonetically driven account of syllable weight. *Language* 78(1). 51–80. DOI: <https://doi.org/10.1353/lan.2002.0020>
- Gordon, Matthew. 2006. *Syllable weight: Phonetics, phonology, typology*. New York, NY: Routledge Press.

- Harris, James. 1967. *Spanish phonology*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Harris, James. 1983. *Syllable structure and stress in Spanish: A nonlinear analysis*. Cambridge, MA/London: The MIT Press.
- Harris, James. 1987. The accentual patterns of verb paradigms in Spanish. *Natural Language and Linguistic Theory* 5(1). 61–90. DOI: <https://doi.org/10.1007/BF00161868>
- Harris, James. 1989. How different is verb stress in Spanish? *Probus* 1(3). 241–258. DOI: <https://doi.org/10.1515/prbs.1989.1.3.241>
- Harris, James. 1991. With respect to accentual constituents in Spanish. In Héctor Campos & Fernando Martínez Gil (eds.), *Current studies in Spanish linguistics*, 447–474. Washington DC: Georgetown University Press.
- Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39(3). 379–440. DOI: <https://doi.org/10.1162/ling.2008.39.3.379>
- Hayes, Bruce & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44(1). 45–75. DOI: https://doi.org/10.1162/LING_a_00119
- Hayes, Bruce, Kie Zuraw, Péter Siptár & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85(4). 822–863. DOI: <https://doi.org/10.1353/lan.0.0169>
- Hirsch, Aron. 2013. What is the domain for weight computation: The syllable or the interval? *Proceedings of the Annual Meeting on Phonology* 1(1). DOI: <https://doi.org/10.3765/amp.v1i1.21>
- Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346–363. DOI: <https://doi.org/10.1002/bimj.200810425>
- Hualde, José Ignacio. 2004. Quasi-phonemic contrasts in Spanish. In Benjamin Schmeiser, Vineeta Chand, Ann Kelleher & Angelo Rodríguez (eds.), *WCCLF 23 proceedings*, 374–398. Somerville, MA: Cascadilla Press.
- Hualde, José Ignacio. 2006/7. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5/6. 59–89.
- Hualde, José Ignacio. 2012. Stress and rhythm. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 153–171. Malden, MA: Blackwell. DOI: <https://doi.org/10.1002/9781118228098.ch8>
- Kim, Jong-mi. 2017. Phonology of brand naming. *Studies in Phonetics, Phonology and Morphology* 23(1). 3–26.
- Lau, Jey Han, Alexander Clark & Shalom Lappin. 2017. Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive Science* 41(5). 1202–1241. DOI: <https://doi.org/10.1111/cogs.12414>
- Lavoie, Lisa. 2001. *Consonant strength: Phonological patterns and phonetic manifestations* (Outstanding Dissertations in Linguistics). New York, NY/London: Garland.
- Lipski, John. 1990. Spanish taps and trills: Phonological structure of an isolated opposition. *Folia Linguistica* 24(3–4). 153–174.
- Lipski, John. 1997. Spanish word stress: The interaction of moras and minimality. In Fernando Martínez Gil & Alfonso Morales-Font (eds.), *Issues in the phonology and morphology of the major Iberian languages*, 559–593. Washington DC: Georgetown University Press. DOI: <https://doi.org/10.1515/flin.1990.24.3-4.153>
- Locke, William & Roe-Merril Secrist Heffner. 1940. Notes on the length of vowels (II). *American Speech* 15(1). 74–79. DOI: <https://doi.org/10.2307/452732>
- Luce, Paul. 1986. *Neighborhoods of words in the mental lexicon*. Bloomington, IN: Indiana University Bloomington dissertation.

- Luce, Paul & David Pisoni. 1988. Recognizing spoken words: The Neighborhood Activation model. *Ear and Hearing* 19. 1–36. DOI: <https://doi.org/10.1097/00003446-199802000-00001>
- Marian, Viorica, James Bartolotti, Sarah Chabal & Anthony Shook. 2012. CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE* 7(8). e43230. DOI: <https://doi.org/10.1371/journal.pone.0043230>
- Martínez-Paricio, Violeta. 2013. The intricate connection between diphthongs and stress in Spanish. *Nordlyd* 40(1). 166–195. DOI: <https://doi.org/10.7557/12.2505>
- Moore-Cantwell, Claire. 2016. *The representation of probabilistic phonological patterns: Neurological, behavioral, and computational evidence from the English stress system*. Amherst, MA: University of Massachusetts dissertation.
- Nosofksy, Robert. 1986. Attention, similarity, and the identification – categorization relationship. *Journal of Experimental Psychology: General* 115(1). 39–57. DOI: <https://doi.org/10.1037/0096-3445.115.1.39>
- Nosofksy, Robert. 1990. Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology* 34(4). 347–370. DOI: [https://doi.org/10.1016/0022-2496\(90\)90020-A](https://doi.org/10.1016/0022-2496(90)90020-A)
- Oltra-Massuet, Isabel & Karlos Arregi. 2005. Stress-by-structure in Spanish. *Linguistic Inquiry* 36(1). 43–84. DOI: <https://doi.org/10.1162/0024389052993637>
- Peterson, Gordon & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6). 693–703. DOI: <https://doi.org/10.1121/1.1908183>
- Piñeros, Carlos Eduardo. 2016. The phonological weight of Spanish syllables. In Rafael Nuñez-Cedeño (ed.), *The syllable and stress: Studies in honor of James W. Harris*, 271–314. Boston, MA: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9781614515975-011>
- Qualtrics. 2015. Qualtrics system, version: November 2015. Yale University license. <http://yalesurvey.qualtrics.com>.
- R Core Team. 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>.
- Roca, Iggy. 1988. Theoretical implications of Spanish word stress. *Linguistic Inquiry* 19(3). 393–423.
- Roca, Iggy. 1991. Stress and syllables in Spanish. In Héctor Campos & Fernando Martínez Gil (eds.), *Current studies in Spanish linguistics*, 599–635. Washington DC: Georgetown University Press.
- Ryan, Kevin. 2011. *Gradient weight in phonology*. Los Angeles, CA: University of California dissertation.
- Segui, Juan, Jacques Mehler, Uli Frauenfelder & John Morton. 1982. The word frequency effect and lexical access. *Neuropsychologia* 20(6). 615–627. DOI: [https://doi.org/10.1016/0028-3932\(82\)90061-6](https://doi.org/10.1016/0028-3932(82)90061-6)
- Shelton, Michael. 2007. *An experimental approach to syllable weight and stress in Spanish*. State College, PA: Pennsylvania State University dissertation.
- Shelton, Michael. 2013. Spanish rhotics: More evidence of gradience in the System. *Hispania* 96(1). 135–152. DOI: <https://doi.org/10.1353/hpn.2013.0003>
- Shelton, Michael, Chip Gerfen & Nicolás Gutiérrez-Palma. 2009. Proscriptions... gaps... and something in between: An experimental examination of Spanish phonotactics. In Pascual José Masullo, Erin O’ Rourke & Chia-Hui Huang (eds.), *Romance linguistics 2007: Selected papers from the 37th Linguistic Symposium on Romance Languages*, 261–275. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.304.17she>
- Shelton, Michael, Chip Gerfen & Nicolás Gutiérrez-Palma. 2012. The interaction of subsyllabic encoding and stress assignment: A new examination of an old problem in Spanish. *Language and Cognitive Processes* 27(10). 1459–1478. DOI: <https://doi.org/10.1080/01690965.2011.610595>

- Shih, Stephanie. 2017. Constraint conjunction in weighted probabilistic grammar. *Phonology* 34(2). 243–268. DOI: <https://doi.org/10.1017/S0952675717000136>
- Skousen, Royal. 1989. *Analogical modeling of language*. Dordrecht: Kluwer. DOI: <https://doi.org/10.1007/978-94-009-1906-8>
- Skousen, Royal. 1992. *Analogy and structure*. Dordrecht: Kluwer. DOI: <https://doi.org/10.1007/978-94-015-8098-4>
- Steriade, Donca. 2012. Intervals vs. syllables as units of linguistic rhythm. *Handouts, EALING*, Paris.
- Suh, Kyung-hee. 2002. The Grammar of the naming of brands in Korean: The case of cosmetics. *사회언어학 [Sociolinguistics]* 10(2). 283–310.
- Taft, Marcus. 1979. Recognition of affixed words and the word frequency effect. *Memory and Cognition* 7(4). 263–272. DOI: <https://doi.org/10.3758/BF03197599>
- Tsuchida, Ayako. 1995. English loans in Japanese: Constraints in loanword phonology. *Working Papers of the Cornell Phonetics Laboratory* 10. 145–164.
- van der Loo, Mark. 2014. The stringdist package for approximate string matching. *The R Journal* 6. 111–122.
- Vitevitch, Michael, Paul Luce, David Pisoni & Edward Auer. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language* 68(1–2). 306–311. DOI: <https://doi.org/10.1006/brln.1999.2116>
- Waltermire, Mark. 2004. The effect of syllable weight on stress in Spanish. In Timothy Face (ed.), *Laboratory Approaches to Spanish Phonology*, 171–191. Berlin: Mouton de Gruyter.
- Whaley, C. P. 1978. Word-nonword classification time. *Journal of Verbal Learning and Verbal Behaviour* 17(2). 143–154. DOI: [https://doi.org/10.1016/S0022-5371\(78\)90110-X](https://doi.org/10.1016/S0022-5371(78)90110-X)
- Wilson, Colin. 2016. Memory for phonological structure. *Talk at 4th Annual Meeting on Phonology*. University of Southern California, Los Angeles, CA. October 23rd, 2016.
- Winter, Bodo. 2013. Linear models and linear mixed effects models in R with linguistic applications. Available at: <http://arxiv.org/pdf/1308.5499.pdf/>.
- Yao, Yao. 2011. *The effects of phonological neighborhoods on pronunciation variation in conversational speech*. Berkeley, CA: University of California dissertation.
- Zuraw, Kie. 2010. A model of lexical variation and the grammar with application to Tagalog nasal substitution. *Natural Language and Linguistic Theory* 28(2). 417–472. DOI: <https://doi.org/10.1007/s11049-010-9095-z>

How to cite this article: Fuchs, Martín. 2018. Antepenultimate stress in Spanish: In defense of syllable weight and grammatically-informed analogy. *Glossa: a journal of general linguistics* 3(1): 80. 1–34, DOI: <https://doi.org/10.5334/gjgl.433>

Submitted: 10 May 2017 **Accepted:** 09 May 2018 **Published:** 19 July 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS