WEIGHT EFFECTS ON STRESS:

LEXICON AND GRAMMAR

Guilherme Duarte Garcia

Department of Linguistics

McGill University

Montréal, Québec, Canada

June 2017

A thesis submitted to McGill University
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy

'... if narratives without statistics are blind,

statistics without narratives are empty.'

(Pinker 2012, p. 193)

Contents

List of Figures in Chapters

List of Tables in Chapters

Weight effects on stress: lexicon and grammar

by

GUILHERME DUARTE GARCIA

Submitted to the Department of Linguistics
on June 9th 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This thesis examines weight effects on stress and proposes a probabilistic approach based on the notion that weight is gradient, not categorical. Arguments for this proposal are divided into three main chapters, which examine and statistically model weight in the lexicon (Chapter 1), weight in the grammar (Chapter 2), and the interaction of weight and footing (Chapter 3). The statistical analyses in Chapters 2 and 3 also discuss how our linguistic expectations regarding weight effects can be incorporated in statistical models through the use of mildly informative priors, and to what extent the fit of such models compare with that of models based on non-informative priors.

In Chapter 1, I examine weight effects in the Portuguese lexicon, and show that they are considerably more intricate than what is assumed in the literature. Previous weight-based studies consider that weight only affects stress in word-final syllables, and that weight is categorical (e.g., Bisol 1992, Lee 2007). In other words, syllables in Portuguese are traditionally classified as heavy or light. I show that weight should not be seen as categorical. By exploring a comprehensive lexicon (Houaiss et al. 2001), I demonstrate that heavy syllables have a gradient effect on stress. This effect is modulated by the position of a given heavy syllable in the stress domain as well as its segmental count. This entails that weight effects are not restricted to word-final syllables. Rather, *all* syllables in the stress domain present some weight effect on stress. One such effect is in fact puzzling: antepenultimate light syllables are more stress-attracting than antepenultimate heavy syllables. This contradicts the typology of weight and stress, since heavy syllables are not expected to repel stress in weight-sensitive languages (Gordon 2006). Given the non-categorical patterns observed in the lexicon, I propose a probabilistic approach to stress in the language. To demonstrate the empirical advantage of such an approach, I show that the accuracy of probabilistic predictions is substantially higher than that of categorical predictions.

In Chapter 2, I examine to what extent these lexical patterns in Portuguese are captured by speakers' grammars. First, I show that speakers do generalise the weight gradience in the lexicon to novel words. The effects monotonically weaken as we move away from the right edge of the word, which mirrors what is found in the lexicon (Chapter 1). Second, I show that speakers do *not* generalise the typologically contradictory pattern found in antepenultimate syllables in the lexicon. Instead, speakers assign positive weight effects to all syllables in the stress domain; i.e., they repair the negative weight effect in question. Previous findings in the literature on phonological (under)learning have shown that unnatural (or contradictory) patterns are harder to learn, and are often ignored by speakers (e.g., Hayes and Londe 2006, Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012). Chapter 2 shows that speakers can go beyond ignoring such patterns: they can in fact repair them.

The probabilistic approach presented in Chapters 1 and 2 raises the question of how footing impacts stress in a language such as Portuguese, where weight effects are gradient. Indeed, a non-categorical weight-based approach poses important challenges to footing. In Chapter 3, I argue that Portuguese does not offer compelling evidence for the foot. First, the gradient weight effects found in the lexicon and in speakers' behaviour cannot be captured with any foot type, given that even antepenultimate stress is directly affected by weight. Second, no phonological process (e.g., truncation, reduplication, hypocorisation) makes reference to the foot. Third, different foot types have been proposed across the literature because of contradictory patterns of stress location—patterns which are mirrored in truncation, reduplication, and hypocorisation in the language. Fourth, sub-minimal words are not only common in the Portuguese lexicon, but are also productive in the language.

As discussed in Chapter 3, the evidence *against* footing in Portuguese is compelling, and I therefore conclude that the language does not have feet. To further strengthen this argument, I turn to English, where the evidence for footing is robust. Even though English and Portuguese present similar stress patterns on the surface, I show that these two languages are fundamentally different. Unlike in Portuguese, stress patterns in the English lexicon and in speakers' grammars exhibit weight effects that are predicted if one assumes moraic trochees and extrametricality in

the language (e.g., Hayes 1982). The probabilistic weight-based approach to stress adopted in this thesis thus concludes that feet are parametric (following, e.g., Özçelik (2013, 2014)), and are therefore present in English, but absent in Portuguese.

Abrégé

Cette thèse examine les effets du poids syllabique (aussi appelé la quantité syllabique) sur l'emplacement de l'accent tonique et propose une approche probabiliste dans laquelle est incorporée la notion que le poids est gradient plutôt que catégoriel. Les arguments en faveur de cette proposition sont répartis entre trois chapitres principaux, qui examinent — et qui modèlent de façon statistique — les tendances par rapport au poids dans le lexique (chapitre 1), au poids dans la grammaire (chapitre 2) et à l'interaction du poids et du pied phonologique (chapitre 3). Les analyses statistiques dans les chapitres 2 et 3 expliquent également comment nos attentes linguistiques concernant les effets de poids peuvent être incorporées dans des modèles statistiques en utilisant des distributions *a priori* légèrement informatives et comment les modèles qui incorporent ces distributions *a priori* se comparent à ceux basés sur des distributions *a priori* non informatives.

Dans le chapitre 1, j'examine les effets du poids dans le lexique portugais et je montre qu'ils sont considérablement plus complexes que ce que l'on suppose dans la littérature. Les études antérieures qui abordent le poids proposent que le poids n'affecte que l'accent tonique dans les syllabes finales et que le poids est catégoriel (par exemple, Bisol 1992, Lee 2007). Autrement dit, on classe traditionnellement les syllabes en portugais comme lourdes ou légères. Je montre plutôt que le poids ne doit pas être conçu comme étant catégoriel. En explorant un lexique complet (Houaiss et al. 2001), je démontre que les syllabes lourdes ont un effet gradient sur l'accent tonique. Cet effet est modulé par la position d'une syllabe lourde quelconque dans le domaine de l'accent tonique ainsi que par le nombre de segments dans le mot. Cela suggère que les effets du poids ne se limitent pas aux syllabes finales. Ce sont plutôt *toutes* les syllabes dans le domaine de l'accent tonique qui présentent un effet de poids sur l'accent tonique. En effet, un tel résultat est déroutant : une antépénultième légère est donc plus attirante — en matière de l'accentuation —

que les antépénultièmes lourdes. Cela contredit la typologie du poids et de l'accent tonique car on ne s'attend généralement pas à ce que les syllabes lourdes repoussent l'accent tonique dans les langues qui sont sensibles au poids (Gordon 2006). Compte tenu des tendances non-catégoriques observées dans le lexique, je propose une approche probabiliste de l'accentuation dans la langue. Pour démontrer les bénéfices empiriques d'une telle approche, je démontre que l'exactitude des prédictions probabilistes est sensiblement supérieure à celle des prédictions catégoriques.

Dans le chapitre 2, je vérifie si ces tendances lexicales surprenantes en portugais font bien partie des grammaires des locuteurs. Premièrement, je démontre que les locuteurs généralisent les effets de poids non-catégoriel dans le lexique à des mots nouveaux. Les effets s'affaiblissent de façon monotonique lorsque nous nous éloignons de la frontière à la droite du mot, ce qui reflète les tendances dans le lexique (chapitre 1). Deuxièmement, je démontre que les locuteurs ne généralisent *pas* le modèle typologiquement contradictoire trouvé dans les antépénultièmes dans le lexique. Plutôt que de faire cela, les locuteurs attribuent des effets de poids positifs à toutes les syllabes dans le domaine de l'accent tonique — c'est-à-dire qu'ils renversent cet effet négatif de poids. La littérature sur l'apprentissage phonologique montre que les tendances non naturels (ou contradictoires) sont plus difficiles à apprendre et ne sont souvent pas apprises par les locuteurs (par exemple, Hayes and Londe 2006, Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012). Le chapitre 2 montre que les locuteurs ne font pas qu'ignorer les tendances problématiques : ils rectifient ces tendances en généralisant.

L'approche probabiliste présentée dans les chapitres 1 et 2 soulève la question de l'influence du pied sur l'accent tonique dans une langue où les effets de poids sont gradients, comme en portugais. En effet, une approche non-catégorielle basée sur le poids crée des défis importants pour la formation du pied. Dans le chapitre 3, je propose qu'il manque de preuves que le pied existe en portugais. Premièrement, les effets de poids dégradés que l'on retrouve dans le lexique et dans le comportement des locuteurs ne peuvent être représentés avec aucun type consistent de pied car même chez les antépénultièmes voit-on des effets de poids. Deuxièmement, aucun processus phonologique (par exemple la troncation, la réduction et l'hypocorisation) fait référence au pied. Troisièmement, différents types de pieds ont été proposés dans la littérature en raison de tendances

contradictoires de l'emplacement de l'accent tonique —- un défi qui est reflété dans la troncation, la réduction et l'hypocorisation dans la langue. Quatrièmement, les mots sous-minimaux sont à la fois fréquents et productifs dans le lexique portugais.

Le chapitre 3 propose que les preuves *contre* le pied en portugais sont convaincantes, menant à la conclusion que la langue n'a pas de pieds. Pour renforcer cet argument, je discute de l'anglais, où la preuve pour le pied est robuste. Même si en anglais et en portugais on trouve des tendances d'accentuation superficiellement similaires, je montre que ces deux langues sont fondamentalement différentes. Contrairement au portugais, les tendances de l'accent tonique en anglais — dans le lexique et dans les grammaires des locuteurs — suggèrent des effets de poids qui sont prévus selon une analyse proposant des trochées moraïques et des éléments extramétriques dans la langue (par exemple, Hayes 1982). L'approche probabiliste de l'accentuation basée sur le poids que j'adopte dans cette thèse conclut donc que la présence des pieds est un paramètre langagier (suivant, par exemple, Özçelik 2013, 2014) et que le pied est présent en anglais, mais absent en portugais.

Acknowledgments

This thesis concludes the most enriching and challenging period of my life. I won't deny that much of what happened in the past five years was the result of planning and hard work, but I'd like to highlight another factor, namely, *chance*. While I do not like to use the word 'luck', the truth is that most people forget how much of their lives is determined by random factors over which they have little or no control. I did plan to come to McGill, and I did plan to work as hard as possible to succeed if accepted. But it was chance that made me work with **Heather Goad** during my entire doctoral programme.

It is extremely hard for me to quantify Heather's importance during the five years I spent at McGill. Her support, knowledge, and enthusiasm provided the most solid, enriching, and pleasant experience a graduate student could ask for—and I am sure that her previous graduate students will say exactly the same thing. As a supervisor, Heather was always available to meet and talk: our numerous meetings were always invigorating, and this played a crucial role on those days when I questioned my own sanity. As a researcher and a teacher, Heather has always been a role model to me. I was always impressed with her ability to see problems that other people miss—and I am particularly impressed with her ability to ask 5-minute long questions in colloquia and conferences. As a second language speaker, I aspire to *write* with that level of clarity and completeness.

I think that most graduate students will agree that feedback from a supervisor is one of the most important factors that define your (successful) progress in any graduate programme. Heather's frequent feedback, attention to detail, and speed made me confident in my own progress. I have no idea how many times she has read my drafts, and how many comments she has made in them—I sometimes dream of PDFs with highlighted sentences in them. I am extremely grateful for that—and so is my writing. I will certainly miss our meetings and chats, her emails at 4 AM, and her

energy. I will never forget all those things.

Another person I was lucky to work with is **Morgan Sonderegger**. If I now identify myself as someone who does quantitative data analysis, it's because of Morgan. His research methods classes introduced me to R, which now plays a central role in most of what I do in Phonology. Morgan was also the co-supervisor in my first evaluation paper, but he *always* played an important role in my five years at McGill. His numerous comments significantly improved my manuscripts, and always made me learn something new involving statistics, phonology, phonetics, R or Python.

Morgan came to McGill in 2012, the same year I joined the PhD programme. Even in his first years at McGill, he always had this calm and analytical way of doing things, which the vast majority of young scholars lack. Morgan has been a great example to me, and I have always admired his teaching, his research, and his sense of humour. I am extremely grateful to him for having taught me so much.

I am also very grateful to **Lydia White** for serving on my thesis committee. Lydia was a committee member for my second evaluation paper as well, and her thoughtful input on second language acquisition substantially improved my paper. She may not remember this, but the first time I contacted her was back in September 2010—two years before I came to McGill. I sent her an email where I discussed the possibility of applying for the PhD programme, and briefly explained what I did at the time. Her response to my email was encouraging, and I decided to apply the following year. It is an *immense* pleasure to work on projects with her today, and I am very grateful for everything she taught me during my time at McGill.

After emailing Lydia back in 2010, I also emailed a couple of graduate students who were in the programme at the time. One of them, **Öner Özçelik**, deserves a special thanks. Öner was patient enough to answer several of my questions about the programme and the department. I'm sure he was busy at the time working on his thesis, but he was incredibly helpful. In 2013, we met at GASLA, in Florida, and I had the chance to thank him in person for all the help.

I would also like to thank **Meghan Clayards** and **Michael Wagner** for their thoughtful feedback on my evaluation papers and thesis. Meghan's phonetics course taught me a lot, and having taken it early in the programme made a big difference in my research. Michael's suggestions

were always very helpful—no matter the topic. Finally, I wish to thank **Jessica Coon**, who read my post-doctoral applications and provided valuable tips on how to improve them.

Three other professors played important roles at different moments in my PhD. First, **Kie Zuraw**, who was kind enough to meet and talk about my research on multiple occasions—I benefitted greatly from her insights. Second, **John Kruschke**, whose workshop on Bayesian methods I attended in 2016. His help during that workshop (and by email afterwards) played an important role in the statistical analysis presented in this thesis—as did his great book on the subject. Third, **Harry van der Hulst**, who took the time to reply to my emails on Accent-First Theory.

I also wish to thank the anonymous reviewers of WCCFL 33 and NELS 47, whose suggestions greatly improved different stages of my research. Likewise, I would like to thank the anonymous reviewers at *Phonology* for their thoughtful feedback, which significantly improved (and shortened) the first chapter of this thesis.

When I moved to Montreal back in 2012, Heather told me she'd had another Brazilian student under her supervision: **Walcir Cardoso**. I was glad to know Walcir actually lived and worked in Montreal. We met during my first term at McGill, and he's now a good friend—who also happens to be a phonologist. Even though Walcir argues that Windows is a good operating system, he gave me important advice throughout my PhD. It's always nice to sit with him and talk about life for hours. I am grateful for his support and friendship.

At the beginning of my programme at McGill, people told me that the first year is usually the worst. And they were right. All five years were intense, but the first year is intense in a 'special' way, given all the courses we have to take in different sub-fields. But the workload was balanced by the great professors in the department. I am especially grateful to **Luis Alonso-Ovalle** for being available to talk about all the questions a phonologist may have about semantics. His friendliness and support were incredibly important in my first year. Likewise, I thank **Lisa Travis** for her clarity while teaching Syntax 3—I learned *a lot* in that course.

The pressures of my first year were also reduced thanks to my cohort back in 2012. In particular, **Hye-Young Bang** and **Daniel Goodhue**, with whom I shared numerous hours of hard work. I

can't remember how many times Hye-Young and I discussed the uncertainties of graduate school. As a phonetician, Hye-Young was also very helpful when I needed a more acoustic take on phonemes. Above all, Hye-Young is *funny*—in a unique way. For example, she has always told me that my wife is too good-looking for me. But 'it's OK', she says, because 'this is actually a compliment'.

I also had the pleasure to make good friends outside of my cohort. One example is **Liz Smeets**. Liz and I talked a lot about life as graduate students during these last years. She helped me prepare for job talks, and her feedback definitely helped me improve my slides. Another example is **Donghyun Kim**. If you like food, Don is the person to talk to. I always enjoyed our talks about academia, publication strategies, and the differences in display quality between LG and Samsung.

I also wish to thank **Jiajia Su** for recruiting an amazing number of Mandarin speakers, and the Montreal Language Modelling Lab (MLML), in particular **Jeffrey Lamontagne**. Jeff was my go-to French speaker, and he also helped with the recording of stimuli in different experiments.

A special thanks goes to **Andria de Luca** and **Giuliana Panetta**. Many of my visits to the main office during these five years ended up taking much longer than necessary, simply because I enjoyed talking to them both about life in general. I have to say that my five years at McGill would have been much more complicated without Andria's effective help. I'm not sure I would ever understand how my funding works without her.

I received a lot of support while at McGill. The funding package offered by the Department of Linguistics was generous, and I am extremely grateful for that. From 2012 to 2014, I was fully funded by the department; from 2015 to 2017, my funding also came from a doctoral fellowship awarded by the *Fonds de recherche du Québec - Société et culture* (**FRQSC**).

I also benefitted from financial support destined for conference travel. Going to conferences is an extremely important part of an academic life, and I have been fortunate to go to several conferences throughout my PhD. In particular, I benefitted from the *Social Sciences and Humanities Research Council* (**SSHRC**) and FRQSC, through grants awarded to Heather Goad and Lydia White. I also received travel awards from the *Centre for Research on Brain, Language and Music* (**CRBLM**) and the Faculty of Arts at McGill.

My five years at McGill involved several experiments, and I am grateful to all the participants who came to our labs: speakers of Brazilian Portuguese, English, Mandarin, and Québec French. I had the opportunity to meet several people during these experiments. Most of them were more than happy to share a lot of interesting stories about their immigration to Canada—and many were curious about the kind of questions I wanted to investigate. Some experiments were also run in Brazil, and I thank **Maria Gabriel** for her help recruiting and running participants. Likewise, I wish to thank **Zwinglio Guimarães** and **Mário Viaro** for their help with MatLab scripts and word lists.

∘ ∘

I obviously wish to thank my parents, **José** and **Traudimar**, who not only provided all I needed during my childhood, but also never questioned my decision to be in academia. Even though they never went to college, they always valued education above all. I also wish to thank my sister, **Cecília**, for taking care of things back in Brazil. The decision to immigrate is never easy, and it's inevitable that important moments will be missed. I will always be grateful to her for just being there when I couldn't.

Finally, I want to thank my lovely wife **Natália Brambatti Guzzo**, who was always by my side during my PhD. Natália and I used to be classmates in a Phonology course back when we lived in Brazil—I certainly had no idea we'd be married and living in Canada in 2017. I am incredibly lucky to have found such a wonderful person, and it still amazes me how much compatibility there is between us. Being a phonologist, Natália carefully read numerous drafts of my papers, questioned my arguments, and suggested important changes. More importantly, we both truly enjoy talking about what we study, and this had a huge impact on this thesis. It's really a privilege to share my life with her.

Preface

The studies presented in this thesis were prepared as manuscripts for publication elsewhere, and constitute original scholarship. Chapter 1 was published in *Phonology* (Garcia 2017b)—a complementary analysis to Chapter 1 comparing syllables and intervals was published in the *Proceedings of the 33rd West Coast Conference on Formal Linguistics* (Garcia 2016). Chapter 2 was published in *Language* (Garcia 2019)—a shorter and (considerably) simpler version of the article has appeared in the *Proceedings of the 47th Annual Meeting of the North East Linguistic Society* (Garcia 2017a). Chapter 3 was co-authored with Heather Goad, and will be submitted to a peer-reviewed journal (Garcia and Goad, *in prep*). Goad and I collaborated in parts of the experimental design and in the theoretical implications of the chapter. The lexicon analysis, data collection, as well as the statistical analyses presented in the chapter were carried out by myself.

Note to the reader: The PDF version of this thesis contains active hyperlinks for bibliographic references and URLs (in **maroon**), as well as for numbers of chapters, sections etc. (in **sepia**).

Introduction

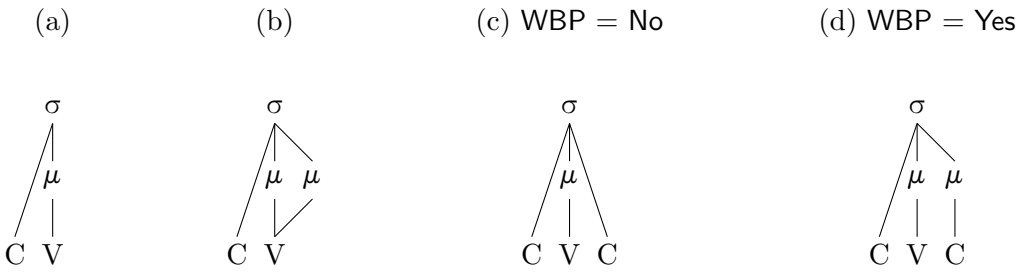
In many languages, the location of stress is determined at least in part by weight. For example, of the 310 languages surveyed in Gordon (2006), 136 (or 43.9%) show some effect of weight on stress. Most such languages (118) have a binary weight distinction: heavy syllables attract stress relative to light syllables. Since the late 1970s, these effects have typically been formalised in two main frameworks, namely, (i) onset-rhyme theory coupled with × theory, and (ii) moraic theory, both of which extended the autosegmental view (Goldsmith 1976) that phonological representations are highly articulated. In onset-rhyme theory (e.g., Fudge 1969, McCarthy 1979, Steriade 1982 , Levin 1985, Lowenstamm and Kaye 1986, Fudge 1987), long and short vowels are differentiated through the presence or absence of branching in the nucleus: a long vowel contains two timing units (××), whereas a short vowel contains only one (×). Likewise, VV and VC rhymes are distinguished from V rhymes by the presence of branching somewhere inside the rhyme in the former, and its absence in the latter.

Unlike onset-rhyme theory, moraic theory (Hyman 1985, Hayes 1989, McCarthy and Prince 1995) captures length and weight distinctions by assuming that the only sub-syllabic constituent is the mora (μ); i.e., no skeletal tier is present in the representation adopted. In this theory, short vowels (light) project one mora, whereas long vowels (heavy) project two moras. Thus, moraic theory represents a light CV syllable as monomoraic (CV_{μ}), and a long CV**:** syllable as bimoraic ($CV_{\mu\mu}$).

Under moraic theory, onset consonants are not moraic (i.e., do not project moras), and coda consonants *may* be moraic, depending on the setting of a parameter, namely, Weight by Position (WBP)—see figure below. As a result, moraic theory achieves a more nuanced formalisation of weight than × theory, insofar as it directly expresses in the representation the cross-linguistic dif-

ferences in the behaviour of VC rhymes. It also straightforwardly captures cross-linguistic processes that are sensitive to weight, such as compensatory lengthening (Hayes 1982). First, moraic theory correctly predicts that compensatory lengthening should only occur in languages with a weight distinction, i.e., where syllables can be bimoraic. Second, because onsets project no moras, moraic theory also predicts that compensatory lengthening can only result from coda deletion, not onset deletion.

Moraic representation of CV, CVː and CVC syllables



In spite of the explanatory advances made by moraic theory in weight-related phonology, it holds the view that weight distinctions are categorical. Recent research has uncovered important empirical challenges for this view. Furthermore, onsets have been shown to trigger compensatory lengthening (Yun 2010), and to contribute to weight (Ryan 2011). Neither effect can be captured by moraic theory, given that onsets do not project moras, as mentioned above. And critically, because the weight effect of onsets represents only a fraction of the weight effect found in codas, this problem cannot be rectified by simply adding a mora to onsets.

The challenges faced by a categorical weight distinction also include syllable rhymes. In a language where both long vowels and coda consonants contribute to weight, CVC and CVV syllables are represented with two moras each, and are therefore formally indistinguishable. However, long vowels are well-known to be heavier than coda consonants (e.g., Gordon 2016). For example, in Klamath, CVV syllables are heavier than CVC syllables, which are heavier than CV syllables (Barker 1964)—this three-way distinction is also observed in Cairene Arabic (Mitchell 1960), Chickasaw (Munro and Willmond 1994), and Kashmiri (Kenstowicz 1994; see also Morén (2000) on variable

weight in CVC syllables). Therefore, the weight distinction between CVV and CVC syllables should somehow be encoded in their (moraic) representation (i.e., CVV ≻ CVC ≻ CV).

One option to deal with ternary weight distinctions in moraic theory would be to represent CVV syllables as bearing three moras. However, this would incorrectly predict that CVV syllables are always heavier than CVC syllables across languages (Gordon 2006, ch. 4). Furthermore, if each vowel projects one mora, it is unclear why three moras would be projected from a long vowel. This would also incorrectly entail that the glide in a falling diphthong projects two moras, and is therefore heavier than a vowel. Indeed, even if ternary weight distinctions could be appropriately encoded in moraic theory, we would still need a theory that accommodates more nuanced weight systems than mora count can capture in a principled way. One such system is Portuguese.

In this thesis, I focus primarily on weight effects on Portuguese stress, and then turn to a comparison of weight effects between English and Portuguese. I show that weight distinctions are gradient *within* and *across* syllables in Portuguese, which poses further challenges to the categorical notion of weight discussed above. Thus, the effect that a heavy syllable has on stress can also depend on which position it occupies in the stress domain. These gradient effects can be observed in the lexicon and, crucially, in speakers' grammars when stress patterns are generalised to novel words.

To account for gradient weight effects on stress found in Portuguese, I argue for a probabilistic approach, whereby a given stress pattern is *more* or *less likely*, rather than categorically defined as *regular* or *irregular*. As will be shown, statistical models based on this framework generate predictions which are empirically better motivated given the lexical patterns in the language. Importantly, even though this probabilistic approach is based on a single parameter, namely, weight, it is more accurate than previous approaches to Portuguese stress, which employ not only weight, but also footing.

By only including weight as a predictor of stress, this thesis also argues that Portuguese does not build feet. To strengthen this argument, I contrast English and Portuguese. As will be shown, in spite of apparent similarities in stress patterns, these two languages are fundamentally different: Whereas English stress offers robust evidence for the foot, Portuguese stress offers compelling evidence against the foot.

Background

Stress in Portuguese is constrained to the three last syllables in any given word. Final stress is assigned if the final syllable is heavy, i.e., contains a VV or VC rhyme. If the final syllable is light, then stress falls on the penultimate syllable. Irregular cases include antepenultimate stress, penultimate stress when the final syllable is heavy, and final stress when the final syllable is light.

Given the generalisations above, most previous analyses of stress in Portuguese agree that weight plays a role in the language. These studies, however, constrain the effect of weight on stress to the word-final syllable. In other words, they assume that syllables always behave as light in penultimate and antepenultimate positions (Bisol 1992, Lee 2007 among others). These assumptions are stated in (1).

(1) **Weight-sensitivity traditionally assumed in Portuguese**

a. Syllables are either heavy or light

b. Weight only affects stress in word-final syllables

In Chapter 1, I show that the weight effects found in the Portuguese lexicon are significantly more intricate than what is traditionally assumed. First, these effects are found across all three syllables in the stress domain. Second, the strength of weight effects on stress monotonically weakens as we move away from the right edge of the word. Third, antepenultimate syllables actually show a *negative* weight effect on antepenultimate stress, an observation which contradicts the typology of weight, given that heavy syllables should not repel stress. In other words, a word where all syllables are light (LLL) is more frequently assigned antepenultimate stress than a word where the antepenultimate syllable is heavy (HLL). These observations are summarised in (2).

To detect the gradient weight effects discussed in Chapter 1, I model stress in a comprehensive lexicon of Portuguese (Houaiss et al. 2001, Garcia 2014). The question of interest in the statistical models examined is how accurately we can predict the location of stress given a set of quantitative parameters, namely, onset, nucleus and coda size of each syllable in the stress domain. As we will see, the level of accuracy of such a probabilistic approach is considerably higher relative to the

predictions of a categorical approach.

(2) **Weight effects in the Portuguese lexicon (Chapter 1)**

a. Weight affects stress in all syllables of the stress domain

b. Weight effects monotonically weaken as we move away from the right edge of the word

c. Weight effects are negative in antepenultimate syllables

A question that emerges from Chapter 1 is whether speakers' grammars mirror not only the gradient weight effects observed in the Portuguese lexicon, but also the negative weight effect in antepenultimate syllables. For example, the effects in (2) could be restricted to the lexicon. The negative weight effects found in the lexicon are of particular interest, given that previous studies have shown that unnatural patterns are harder to learn (e.g., Hayes and Londe 2006, Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012, Jarosz 2017).

In Chapter 2, I show that native speakers generalise the gradient weight effects present in the lexicon to novel words. Crucially, however, the negative weight effect found in antepenultimate syllables is not generalised: instead, speakers 'repair' this effect, and favour antepenultimate stress in words where the antepenultimate syllable is heavy.

(3) **Weight effects in the Portuguese grammar (Chapter 2)**

a. Weight affects stress in all syllables of the stress domain

b. Weight effects monotonically weaken as we move away from the right edge of the word

c. Weight effects are *positive* in all positions, including antepenultimate syllables

We can see from (3) that speakers' behaviour is consistent with what we would expect from a weight-sensitive language where weight is the most important factor affecting stress in all three syllables in the stress domain. This, however, poses a major challenge for foot-based approaches in Portuguese, as no foot type can capture weight effects in antepenultimate syllables. The fact that weight effects are gradient further complicates a foot-based analysis.

Antepenultimate weight effects and gradient weight are not the only challenge faced by foot-

based approaches in Portuguese. Indeed, the evidence for footing in Portuguese is questionable at best. First, the language has several lexical words that violate word-minimality, insofar as they are monomoraic (McCarthy and Prince 1995). Indeed, monomoraic words are not only present in the lexicon, but they are also productive in hypocorisation, for example. Second, the empirical data in Portuguese is ambiguous with regard to footing: even though most words in the language motivate trochees (*caválo* 'horse', *sapáto* 'shoe'), some words are better captured with iambs (*jacaré* 'alligator', *café* 'coffee'). Other words are better captured with dactyls (*patético* 'pathetic', *prático* 'practical'), as proposed in Wetzels (1992). Given the different metrical patterns observed in Portuguese, it is not surprising that different analyses have assumed different foot types as well as additional mechanisms such as exceptional extrametricality (see Collischonn (2010) for a review).

In Chapter 3, I argue that Portuguese does not have feet. The central argument is based on the observation that weight effects are gradient in the language (as alluded to above), and can be found in all syllables in the stress domain. To better assess the proposal that the foot cannot be motivated for Portuguese, I turn to English, a language which provides robust evidence for footing. I show that English and Portuguese are fundamentally different with regard to the formal system that regulates stress in these two languages—in spite of their similarities in stress patterns on the surface. Crucially, the weight effects found in English, including in antepenultimate position, are predicted by a foot-based approach, unlike the effects in Portuguese. A summary of Chapter 3 is provided in (4).

Given the evidence forwarded against the foot in Portuguese, I assume that the status of the foot is not universal (*contra* Selkirk (1984) and Nespor and Vogel (1986)). Rather, the foot is parametric (McCarthy and Prince 1995, Özçelik 2013). One important question then is how the stress domain is constrained in a language like Portuguese without feet, as well as what determines the location of stress within this domain. In Chapter 3 I explore an alternative to footing, namely, Accent-First Theory (van der Hulst 2012). This theory makes use of parameters to establish the typology of stress and weight observed cross-linguistically. Crucially, Accent-First Theory can be employed to delimit the stress window in Portuguese without making reference to the foot. The actual location of stress is then determined not only by the parameters within Accent-First Theory,

but crucially by weight. I end the chapter by discussing how gradient weight can be formalised, and how it can interact with Accent-First Theory in a constraint-based probabilistic approach (e.g., Hayes and Wilson 2008).

(4) **Footing and weight in English and Portuguese (Chapter 3)**

a. EVIDENCE AGAINST FOOTING IN PORTUGUESE

(i) Gradient weight effects are also found in antepenultimate syllables

(ii) Existing words require various foot types to capture stress location

(iii) Violations of word-minimality are found in lexical words and hypocorisation

(iv) Truncated and reduplicated forms yield outputs consistent with various foot types

b. EVIDENCE FOR FOOTING IN ENGLISH

(i) Antepenultimate heavy syllables pattern with light syllables

(ii) Moraic trochees and extrametricality capture the attested stress patterns

(iii) No lexical word violates word-minimality; i.e., all words are minimally bimoraic

(iv) Truncation and hypocorisation respect word-minimality

Throughout Chapters 2 and 3, I employ Bayesian models to statistically analyse data from two different experiments. Unlike traditional statistical models, Bayesian models provide the most credible parameter values (i.e., weight effects in Portuguese and English) given the data. I also discuss how our theoretical assumptions can be incorporated in statistical models through (mildly) informative priors, which constrain the space of possible weight effects on the basis of our understanding of what native speakers may know about weight and stress in their respective languages. Finally, I compare such informed models with naïve models, which assume that all weight effects in both languages are null *a priori*. Because Bayesian models are not as common as frequentist models in linguistic research, I provide an overview of Bayesian methods in Chapter 2, as well as a short glossary of relevant terms in Appendix A.

Chapter 1

Weight Gradience and Stress in Portuguese

ABSTRACT

This paper examines the role of weight in stress assignment in the Portuguese lexicon, and proposes a probabilistic approach to stress. I show that weight effects are gradient and monotonically weaken as we move away from the right edge of the word. Such effects depend on the position of a syllable in the word as well as the number of segments the syllable contains. The probabilistic model proposed in this paper is based on a single predictor, namely, weight, and yields more accurate results than a categorical analysis, where weight is treated as binary. Finally, I discuss implications for the grammar of Portuguese.

Keywords: stress, weight, onsets, probabilistic grammar, Portuguese

1.1 Introduction

This paper examines Brazilian Portuguese (BP) primary stress in non-verbs,[1] and proposes a probabilistic analysis based on weight gradience in the language. Portuguese stress is constrained to the final three syllables of the word ('trisyllabic window'), although only final and penultimate stress

[1]BP and European Portuguese (EP) are nearly identical vis-à-vis primary stress. Crucially, stress in verbs is morphologically conditioned, whereas stress in non-verbs is affected by phonological factors (§1.2). Therefore, most of what follows can in principle be applied to both varieties. The main differences between the two lie in phonetics (see Frota and Vigário (2001) for a comprehensive comparison). Phonologically, both BP and EP have an almost identical phonemic inventory (see Mateus and d'Andrade (2000)), even though they respect different syllabification constraints. All transcriptions are in BP, but I use 'BP' and 'Portuguese' interchangeably in this paper, as the lexicon examined here is not limited to Brazilian Portuguese.

are typically analysed as regular and productive (Hermans and Wetzels 2012). Previous research has proposed that weight-sensitivity in the language is constrained to the word-final syllable, i.e., that stress is influenced by the weight of the final syllable, but not the weight of syllables located earlier in the word (Bisol 1992). Additionally, weight-sensitivity is seen to be categorical and binary (a syllable is either heavy or light according to the shape of its rhyme).

Primary stress placement in Portuguese non-verbs can be largely explained by weight, in terms of the following generalisations (Bisol 1992): stress is final (U) if the word-final syllable is heavy—where *heavy* is defined as containing a diphthong, a nasal vowel or a coda consonant (1a)—Portuguese has no long vowels. If the word-final syllable is light, stress falls on the penultimate (PU) syllable (1b). Taken together, these are the regular stress patterns in the language, which are found in 72% of the lexicon (Houaiss et al. 2001). Phonetically, stress in Portuguese is highly correlated with duration (Major 1985).

(1) **Regular stress in Portuguese non-verbs**

a.	*cacau* [kaˈkaw] ‘cocoa’	*anã* [aˈnã] ‘dwarf’ (f)	*pomar* [poˈmaɾ] ‘orchard’
b.	*boca* [ˈbokɐ] ‘mouth’	*tonto* [ˈtõntʊ] ‘dizzy’	*pátio* [ˈpatʃjʊ] ‘patio’

There are, however, three types of irregular stress patterns in the language: final stress when the word-final syllable is light (2a); penultimate stress when the word-final syllable is heavy (2b); and antepenultimate (APU) stress (2c).

(2) **Irregular stress in Portuguese non-verbs**

a.	*café* [kaˈfɛ] ‘coffee’	*sofá* [soˈfa] ‘sofa’
b.	*nível* [ˈnivew] ‘level’	*míssil* [ˈmisiw] ‘missile’
c.	*fósforo* [ˈfɔsfoɾʊ] ‘match’ *n*	*pérola* [ˈpɛɾolɐ] ‘pearl’

Researchers have employed different mechanisms in order to accommodate the cases in (2) (Bisol 1992, Lee 2007). For example, cases (2b) and (2c) have been accounted for by segmental and syllabic extrametricality, respectively (discussed in §1.2). The pattern in (2a) has been explained via

consonantal catalexis: *café* [ka$_\mu$ˈfɛ$_\mu$C$_\mu$]. Even though the catalectic consonant is only phonetically realised in derived forms (e.g., *cafe**t**-eira* [kafeˈtejɾa] 'coffee maker'), it bears its own mora (Hyman 1985), and stressed light word-final syllables are thus underlyingly heavy according to such analyses.

Cases such as (2a) have motivated some researchers to propose that morphological factors govern the location of stress—as an alternative to catalexis. In particular, the presence or absence of theme vowels has been argued to play an important role in determining where stress falls: most non-verbs in Portuguese are composed of a stem and a theme vowel (TV) (3b), but words such as (3a) are exceptions to that pattern, in that no theme vowel is present. By positing that regular stress in Portuguese falls on the stem-final vowel, such forms are no longer irregular.

(3) a. *jacaré* [ʒakaˈɾɛ]$_{stem}$ 'alligator'

b. *boca* [ˈbok]$_{stem}$[-ɐ]$_{\text{TV}}$ 'mouth'

Thus, existing accounts explain the location of stress in most of the lexicon (regular stress) largely by a single phonological factor, namely, syllable weight, with exceptions generally accounted for by mechanisms not directly involving weight. When we examine the lexicon of the language more closely, however, the relationship between weight and stress becomes less clear than what is traditionally assumed. As will be shown in §1.3, weight seems to affect stress in all syllables in the stress domain, including the irregular cases in (2), though to different degrees. For instance, antepenultimate stress is almost always found in words that contain light penultimate and light final syllables. If penultimate syllables are not sensitive to weight, this is an unexpected correlation. Furthermore, onsets seem to affect stress location in the lexicon (§1.3), which indicates that weight computation in Portuguese may not be restricted to the rhyme.

A more accurate measure of how weight is computed in Portuguese is naturally important if one wishes to have a more comprehensive understanding of how stress and weight interact in said language. In this paper, I present a probabilistic analysis that accounts for the vast majority of cases that fall into the patterns in (1) and (2). I propose that weight in Portuguese has a gradient effect on stress, which is positionally[2] and quantitatively determined. In other words, the weight

[2]For positional weight, see Gordon (2004) and Ryan (2014).

effects of a given syllable depend on the position of said syllable within the word as well as the number of segments present in the syllable. As we will see, weight effects in Portuguese go beyond 'heavy' and 'light' syllables.

The analysis in this paper is developed by addressing three questions, provided in (4). Question (4a) examines whether weight in fact only plays a role word-finally in Portuguese. In the lexicon investigated here, weight seems to have some influence on all three syllables in the stress domain. Statistical models (§1.4) indicate that weight-sensitivity gradiently weakens as we move away from the right edge of the word. The observation that final, penultimate and antepenultimate stress are sensitive to weight (4a) shows that antepenultimate stress is not as idiosyncratic as one might think, *contra* standard views on Portuguese.

(4) a. Is weight-sensitivity only found word-finally in Portuguese?

b. Is weight-sensitivity categorical or gradient?

c. Do onsets contribute to weight, affecting stress likelihood in Portuguese?

Question (4b) refers to whether weight is categorical, as assumed in standard views. I show that weight is in fact *gradient*: how much each syllable is affected varies considerably, but the effects are statistically significant. Weight-sensitivity depends on the position of a given syllable within the word, and, crucially, its effect on stress monotonically weakens as we move away from the right edge of the word.

Previous research in BP is based on the assumption that onsets do not influence stress—following the traditional view that weight is a property of the rhyme (Chomsky and Halle 1968, Liberman and Prince 1977, Halle and Vergnaud 1987a, Halle and Kenstowicz 1991, Hayes 1995, among many others). Question (4c) investigates whether this assumption is appropriate for Portuguese, and, if not, how onsets might affect stress in the lexicon. Onsets do show statistically significant effects in Portuguese (§1.4), but not in the way we would expect from more recent studies, which have shown that onsets have a positive effect on stress in other languages (Gordon 2005, Topintzi 2010, Ryan 2014).

This paper is organised as follows: in section 1.2, I discuss Portuguese stress in detail and revisit

analyses proposed to account for both the regular and irregular patterns found in the language. In section 1.3, I analyse weight and stress in a comprehensive lexicon in order to answer the questions in (4). In section 1.4, I model the patterns in the lexicon using Binomial Logistic Regressions. Crucially, given their probabilistic nature, the predictions of the models presented here are more consistent with the actual lexical patterns than are previous analyses, which assumed categoricity. Finally, section 1.6 summarises the findings of the paper, and discusses directions for future work.

1.2 Stress in Portuguese

In this section, I discuss stress in Portuguese non-verbs, and examine both morphological (§1.2.1) and phonological approaches (§1.2.2) previously proposed to account for irregular patterns in the language. I argue that there is no compelling argument for morphological influence on non-verb stress, and therefore the analysis presented in this paper is solely based on phonological factors.

Stress in many Indo-European languages is constrained to the final three syllables of the word.[3] This is the case in Romance languages such as Italian, Portuguese, Catalan and Spanish—a trait inherited from Latin. Unlike Latin, however, stressed word-final syllables are relatively common in modern Romance languages, including Portuguese (Roca 1999). Stress in German, English and Dutch monomorphemic words also falls within a trisyllabic window (Domahs et al. 2014).

Several studies on stress in Portuguese (Câmara Jr. 1970, Major 1985, Bisol 1992, Lee 1994, Collischonn 1994, Araújo 2007, Wetzels 2007, among others) agree that primary stress in the language is relatively predictable in non-verbs with final or penultimate stress. On the other hand, antepenultimate stress is regarded as idiosyncratic (i.e., unpredictable), and represents less than 15% of all non-verbs in the Houaiss Dictionary (Houaiss et al. 2001), the most comprehensive dictionary of the Portuguese language. Words with antepenultimate stress have always existed in Portuguese, and although their stress profile is not regular in the language, there is no evidence suggesting that such forms are completely avoided (Araújo et al. 2007, p. 58)—though in some

[3]In this paper, 'word' is to be equated with Prosodic Word (ω), defined as 'a single root plus any additional morphemes within the 'grammatical word' such that the resulting constituent exhibits the properties determined to be the crucial ω domain properties for the language in question [...]' (Vogel 2008, p. 212). Theme vowels, for example, fall within the ω.

northeastern varieties 'this pattern has completely vanished in non-verbs' (Wetzels 2007, p. 29). Finally, antepenultimate stress is sometimes 'repaired' via syncope and resyllabification—as long as the resulting form obeys the phonotactic patterns in the language (see Amaral 2000): *fósforo* → [ˈfɔsfɾʊ] 'match' *n.* This type of repair is found in most dialects of Brazilian Portuguese.

Antepenultimate stress is therefore phonologically more peripheral in the language when compared to final and penultimate stress, which are more common and much more productive (≈18% and ≈68% in the Houaiss Dictionary, respectively). As a result, it is normally assumed that a new word in the language is not likely to have antepenultimate stress (Hermans and Wetzels 2012). Rather, new words tend to have either final or penultimate stress, aside from some borrowings. The words *penalty* [ˈpenaltʃi] and *performance* [pɛɾˈfɔɾmãnsi], for example, are present in Portuguese dictionaries with the original stressed syllable, even though this results in stress on the antepenultimate syllable in both cases (once the final cluster in 'performance' is repaired). This preservation of the source language's stressed syllable is respected in the spoken language as well, despite following a disfavoured pattern in Portuguese.

Across the entire Portuguese lexicon (Houaiss et al. 2001), primary stress has both morphological and phonological components: whereas stress in verbs is lexically defined by mood, tense, person and number morphemes (see Wetzels (2007) for a review), stress in non-verbs is heavily influenced by weight (cf. Mateus and d'Andrade 2000). The morphological aspect of stress in verbs is undisputed, but some researchers have suggested that morphological factors also play a role in stress in non-verbs (Pereira (2007) and Lee (2007), among others). These researchers assume stress in non-verbs is sensitive to both morphological and phonological factors.

Table 1.1 summarises the stress patterns in non-verbs. As mentioned earlier, heavy syllables ('H') may have a nasal vowel, a coda consonant, and/or a complex nucleus: *pagã* [paˈgã] 'pagan'; *valor* [vaˈloɾ] 'value'; *funil* [fuˈniw] 'funnel'. Light syllables ('L') are open and contain only one segment in the nucleus: *abacaxi* [abakaˈʃi] 'pineapple' ('X' stands for either 'H' or 'L').

Note that very few words have antepenultimate stress and a heavy penultimate or final syllable (also noted in Wetzels (2007) for a subset of cases, discussed in §1.2.2)—this situation is similar to what we find in Dutch (van Oostendorp 2012). Almost all these cases consist of borrowings,

such as *performance* [peɾˈfɔɾmãnsi] and *propolis* [ˈpɾɔpʊlis] 'propolis'. Some of these words undergo syncope in spoken BP: *óculos* [ˈɔkʊlʊs] → [ˈɔklʊs] 'glasses'.

Table 1.1: Portuguese stress patterns (> 1σ non-verbs) in the Houaiss Dictionary (N= 163,625)

Stress pattern	**Regular**	N	%	**Irregular**	N	%
Final (U)	XH́]$_\omega$	24,060	14.7%	XĹ]$_\omega$	5,662	3.46%
Penultimate (PU)	X́L]$_\omega$	93,715	57.27%	X́H]$_\omega$	18,546	11.33%
Antepenultimate (APU)				X́LL]$_\omega$	21,367	13.05%
				X́LH]$_\omega$	233	0.14%
				X́HL]$_\omega$	35	0.02%
				X́HH]$_\omega$	7	< 0.01%
		117,775	≈ 72%		45,850	≈ 28%

1.2.1 Morphological approaches to Portuguese stress in non-verbs

In this section, I review the arguments for morphological influence in non-verb stress, and argue that there is no unambiguous evidence for such an influence. Previous research has proposed that morphology plays an important role in Portuguese stress, in that theme vowels are never stressed. I show that, whether or not theme vowels have an active role in the synchronic grammar of Portuguese non-verbs, there is no convincing evidence suggesting that such vowels actually influence stress: effects often attributed to theme vowels can be accounted for by phonological factors alone.

Morphological influence on Portuguese non-verb stress has been proposed by Mateus (1983), Lee (1995, 2007) and Pereira (2007). These analyses assume that the stress domain in Portuguese non-verbs is the stem—that is, number, gender and theme vowels are not visible to stress, and therefore these morphemes are never stressed in Portuguese.

(5) a. jacaré -s [ʒakaˈrɛs] (singular: *jacaré*)
STEM PL
alligators

b. boc -a -s ['bokas] (singular: *boca*)
STEM FEM.TV PL
mouths

As a result, irregular final stress in Table 1.1 is accounted for in the following way: in a word like *jacaré*[4] (5a), for example, stress falls on the stem-final vowel (/ɛ/)—this approach entails that all words with irregular final stress are monomorphemic. A word like *boca* (5b), on the other hand, has a theme vowel (/a/), and therefore stress falls on /o/, the only vowel in the stem.

The main argument for this proposal lies in derived forms. If we add a suffix to both words above, the theme vowel is typically deleted, whereas the stem-final vowel cannot be. In (6), the diminutive suffix *-inho* [-iɲʊ] is attached to *pato* and *sofá*. In (6a), the theme vowel is deleted, yielding *patinho*; in (6b), since the word-final vowel is part of the stem, an epenthetic consonant (/z/) is inserted to avoid hiatus (Bachrach and Wagner 2007).

(6) a. pat -o -inh -o *patinho* (cf. **patoinho*) [paˈtʃiɲʊ]
STEM MASC.TV DIM MASC
'Small duck'

b. sofá -inh -o *sofazinho* (cf. **sofinho*) [ˌsofaˈziɲʊ]
STEM DIM MASC
'Small sofa'

However, example (7) shows that the situation is not as straight-forward as implied by (6). Whereas /livr-o/ should pattern exactly like /pat-o/, two forms are instead accepted, indicating the optionality of TV deletion. Such cases are less common but not rare. In addition, they seem to be more acceptable with certain lexical items than others (de Freitas and Barbosa 2013).

[4]The use of a diacritic (´ or ˆ) in BP orthography denotes stress irregularity—hence all words with an irregular stress pattern in Table 1.1 will be orthographically marked, except for words with final stress ending in /u/ or /i/, as these vowels cannot be thematic. Note that the diacritics employed in Table 1.1 do not mirror the orthography, as they are used to indicate primary stress in all stress patterns in the language, regular or irregular.

(7) a. livr -o -inh -o *livrinho* or *livrozinho* [liˈvɾiɲʊ] ~ [livɾʊˈziɲʊ]
STEM MASC.TV DIM MASC
'Small book'

A stem-based analysis of stress seems to be more comprehensive than a purely phonological analysis, in that it accounts for more patterns: XXˈL]$_{\omega}$ words are no longer irregular, as they are in phonological approaches—rather, they simply lack a theme vowel. However, the assumptions of such an analysis are problematic. The argument in question is circular: a given vowel is stressed because it is not thematic, and it is not thematic because it is stressed. Note that there is nothing in the pair presented in example (6) that motivates the presence/absence of TV in present-day Portuguese—except for the location of stress. In addition, the three nominal TVs in Portuguese {a, e, o} also appear stem-finally in words like *sofá*, *dendê* and *metrô*, which have word-final stress ('sofa', 'palm oil', 'metro'). Thus, stress placement is the only way to determine whether a given vowel is (or is not) thematic.

A purely phonological alternative to theme vowels follows from the observation that, cross-linguistically, prominent segments are more likely to be preserved (Harris 2011). In Portuguese, stressed vowels are never deleted in monomorphemic or derived forms. Consequently, a word like 'sofá' could not possibly lose its stressed vowel in any derived form (see (6)). Theme vowels, on the other hand, are not stressed, which explains why they may be deleted.

There are other phonological processes in BP often said to be associated with theme vowels, such as vowel raising and external sandhi.[5] Theme vowels may raise in the language, whereas stem-final vowels cannot: *mergulh-o* [meɾˈguʎo] → [meɾˈguʎʊ] 'dive' (n), but *robô* [xoˈbo] ↛ *[xoˈbu] 'robot'. Likewise, external sandhi is only allowed in words with a theme vowel: *camisa usada* [kaˈmizɐ uˈzada] → [kaˈmizuˈzadɐ] 'used shirt', but *jacaré amarelo* [ʒakaˈɾɛ amaˈɾɛlʊ] ↛ *[ʒakaˈɾamaˈɾɛlʊ] 'yellow alligator'. Like vowel deletion in derived forms ((6a) and (7)), both vowel raising and sandhi can be accounted for without additional mechanisms: stressed vowels are protected, and therefore they cannot raise, be deleted in derivations, nor undergo external sandhi.

The question, thus, is whether stressed vowels are maintained because they are more prominent or because they are part of the stem. Given the facts, it is not possible to tell these two alternatives

[5] Vowel deletion across word boundaries.

apart. The same question can be posed for other Romance languages, where the problem also arises. In fact, Roca (1999, p. 673) proposes an extrametricality rule for all Romance languages to capture the observation that theme vowels are 'invisible' to stress.

(8) **Romance Extrametricality Rule**:

Assign extrametricality to the (metrical projection of the) desinence

Roca prefaces the rule as follows: 'In the absence of evidence to the contrary, however, it is reasonable to assume that final stressless vowels are desinential'. What motivates the rule in (8) is exactly the fact that theme vowels seem to be frequently deleted in Romance languages (unlike stressed stem-final vowels).

Whether or not theme vowels exist in present-day Portuguese is beyond the scope of this paper, but their alleged relevance to stress clearly bears on the questions examined here. In this section, however, I argued that there is no solid evidence that such vowels have a role in Portuguese stress. Therefore, this paper is based only on phonological factors, discussed in the next section.

1.2.2 Phonological approaches to Portuguese stress in non-verbs

Even if we assumed that morphological factors did impact stress in Portuguese, we would still need to consider phonological factors, which heavily influence stress in the language. In this section, I examine such factors in more detail, focusing on weight and how it affects the stress patterns found in the language. I briefly review previous analyses of stress in Portuguese, which employ different mechanisms to account for stress irregularities. Finally, I provide independent evidence for weight effects in Portuguese.

Previous analyses of Portuguese stress all make reference to syllabic constituency. In view of this, I first describe syllable shape in the language. In Brazilian Portuguese, up to two segments can occupy the onset position (see Fig. 1.1). Onset clusters consist of stop+liquid or labial fricative+liquid sequences. A word such as *macabro* 'macabre', for example, can only be syllabified as [ma.ˈka.bɾo] (cf. *[ma.ˈkab.ɾo]). In other words, stop+liquid clusters in Portuguese are not ambiguous vis-à-vis their syllabification (Cristófaro-Silva 2005). Finally, rhymes in Portuguese normally

contain up to four segments.

Very few words violate these syllabic restrictions (borrowings, proper names etc.), some of which are listed in the Houaiss Dictionary (Houaiss et al. 2001). These cases, however, are phonotactically adapted in spoken BP, mostly via epenthesis (e.g., the borrowing *skate* is produced as [is.ˈkej.t͡ʃi]). Recent borrowings are not the only words that are repaired: well-established words also undergo epenthesis and resyllabification if they violate the syllabic template in Portuguese: *advogado* 'lawyer' and *obstetra* 'obstetrician', for example, are normally produced as [ad͡ʒi.vo.ˈga.dʊ]/[ade.vo.ˈga.dʊ] and [o.bis.ˈtɛ.tɾɐ] in BP, respectively.

The syllabification algorithm in Portuguese is straightforward and unambiguous, given the restricted number (and quality) of segments in complex onsets and codas (see Thomas (1974) for a comprehensive description and Neto et al. (2015) for a computational implementation). A nonce word such as *pantridocra*, for example, is unambiguously syllabified as /pan.tɾi.do.kɾa/—regardless of stress position. How such a word is actually produced will vary considerably between BP and EP (Mateus and d'Andrade 1998). Take the word *devedor* 'debtor', which is syllabified as /de.ve.ˈdoɾ/. In colloquial EP, where vowels are frequently deleted, such a word is often produced as [ˈdvdor]. This type of reduction never happens in BP (as we have seen, certain coda-onset sequences often undergo epenthesis).

In Fig. 1.1, on-glides and off-glides are treated identically: both contain two × slots. Under a categorical view, this entails that both rising and falling diphthongs are heavy. This contrasts with standard approaches to stress in Portuguese, which treat rising diphthongs as light (Bisol 2013). The present paper, however, does not distinguish rising and falling diphthongs. One reason is empirical: As Harris (1983, p. 11) shows for Spanish, rising diphthongs also seem to contribute to weight in some positions in the word. Antepenultimate stress in Spanish (e.g., *teléfono* 'telephone') is blocked when another syllable in the stress domain is heavy. As a result, a word such as **teléfosno* or **teléfoino* is not found in Spanish. Interestingly, a word such as **teléfiono* ([te.ˈle.fjo.no]) is not found in the language either. All these generalisations, and critically the last one, also hold for Portuguese. This indicates that, like coda consonants, both types of diphthongs affect weight to some degree, a fact which is consistent with the representation in Fig. 1.1. Even though the overall

findings of the present paper do not hinge on the particular choice of on-glide treatment, I return to this discussion in §1.4.1.3, where I provide another reason for not differentiating on-glides and off-glides in the probabilistic approach proposed in this paper.

Figure 1.1: Syllabic structure in Portuguese

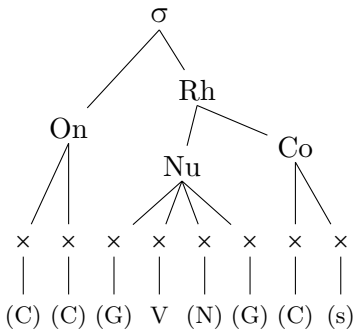


Fig. 1.1 implies that at most four segments can occupy the nucleus of the syllable in Portuguese: *gato*, *gaita*, *guaipeca*, *bastião* ('cat', 'harmonica', 'mongrel', 'bastion'). This, however, depends on how one treats nasality (*fã* 'fan'), in particular nasal diphthongs (*patrão* 'boss') and triphthongs (*bastião*). This paper assumes the standard approach to nasality in Portuguese, according to which a word such as *fã* is underlyingly bisegmental: /faN/ (see Battisti (1997) for a review). Since nasality is realised on the vowel, if the same assumption is extended to diphthongs and triphthongs, then certain syllable nuclei in Portuguese may contain up to four segments (e.g., *bastião*: /bas.ˈtjaNw/). Assuming this standard representation, minimal pairs such as *mão* ('hand') and *mau* ('bad') can be quantitatively differentiated: the former contains three nuclear segments, while the latter contains two nuclear segments.[6]

Traditionally, the concept of weight has been seen as relevant only for the presence of rhyme segments—thus excluding onsets from the computation of weight (Halle and Vergnaud 1980, Hyman 1985, Hayes 1989, among others). Portuguese is an example of a language that is analysed as such: as mentioned earlier, a heavy syllable contains a diphthong, a nasal vowel or a coda consonant;

[6]The analysis proposed in §1.4 takes this quantitative difference into account. However, this is not a crucial aspect of the probabilistic approach presented in this paper for two reasons. First, very few words contain nasal triphthongs (n=15). Second, nasal triphthongs are only found in word-final syllables, where weight effects are already known to be robust.

onset structure is seen to be irrelevant. However, some studies show that onsets also have an impact on stress in several languages, suggesting at least some contribution to the calculation of weight (Davis 1988, Gordon 2005, Topintzi 2010, Ryan 2011, 2014).

To my knowledge, thus far no researcher has proposed a role for onsets in Portuguese stress. However, in southeastern varieties of BP, onset clusters are often simplified in unstressed syllables (Harris 1997): *prato* [ˈpɾatʊ] *-inho* [iɲʊ] → [paˈt͡ʃiɲʊ] 'plate', 'small plate'. In other words, complex onsets are preferred in more prominent positions. This simplification is relatively common in some spoken BP varieties: words such as *próprio* [ˈpɾɔpɾjʊ] are sometimes produced as [ˈpɾɔpjʊ] 'proper'. In addition, onset metathesis is observed in words such as *obstetra* [ob(i)sˈtɛtɾɐ] → [ob(i)sˈtɾɛtɐ] 'obstetrician'. Despite the apparent correlation between onset clusters and stressed syllables in such processes, Cristófaro-Silva (2005) argues that cluster reduction is not in fact phonologically conditioned. She shows that cluster simplification may occur in both stressed and unstressed syllables, which suggests that word-level prominence is not the underlying cause for the process in question.

The proposal that onset cluster simplification is not related to stress does not necessarily mean that onsets do not affect stress. Since no study has directly examined the impact of onsets on Portuguese stress, all weight-based analyses thus far only focus on rhymes (Bisol 1992, Lee 2007, Wetzels 2007, Bisol 2013, among others), given the assumptions of standard Moraic Theory (Hyman 1985, Hayes 1989). Under such a view, a CV.CV.CV word (*macaco* 'monkey') and a CV.CCV.CV word (*catraca* 'turnstile') are both predicted to bear penultimate stress, as they have exactly the same moraic representation: $\sigma_{\mu}.\sigma_{\mu}]_{\omega}$. As onsets are outside the rhyme, these constituents are not moraic, and therefore are not predicted to affect stress likelihood. However, in §1.3 I show that the onset patterns found in the Portuguese lexicon deviate from these predictions.

As mentioned in §1.2, previous studies of Portuguese stress only consider weight in word-final rhymes (Bisol 1992, Araújo 2007, Collischonn 2010, and others). The standard claim that only word-final syllables are weight-sensitive is mostly based on the observation that antepenult, penultimate and final syllables behave very differently from one another regarding syllable shape (open

vs. closed) and stress, as can be seen in Table 1.2. If Portuguese is in fact only weight-sensitive word-finally, its weight profile could be classified as *combined*. Combined systems have distinct weight computations for different positions or circumstances. There are 42 languages (out of 500) in the WALS database with a combined weight system (Goedemans and van der Hulst 2013). Among these languages, we find Spanish and Romansch, both closely related to Portuguese.

Wetzels (2007), however, argues that weight may also play a role word-internally, given the behaviour of palatal consonants in Brazilian Portuguese.[7] Although consonantal quality does not have an evident effect on stress in the language, {[ɲ], [ʎ]} are an exception. These consonants are never found in final onsets in words with antepenultimate stress (≈ 3.8% of the Houaiss Dictionary contain such onsets). Wetzels (2007, p. 25) analyses such consonants as geminated, which therefore occupy both onset and (preceding) coda slots: *baralho* → [ba.ˈɾaʎ.ʎo] (*[ˈba.ɾaʎ.ʎo]) 'deck of cards' (see Fig. 1.1). This analysis is consistent with the fact that very few words with antepenultimate stress have a heavy penultimate syllable: in both cases, weight in the penultimate syllable would block antepenultimate stress.

Standard views on stress in Portuguese non-verbs tend to rely on more frequent/robust patterns in the lexicon, such as the distribution of open *vs.* closed syllables across stress locations. Table 1.2, for instance, provides a clear positive correlation between final closed syllables and final stress: 80.98% of all words with final stress have a closed word-final syllable. On the other hand, antepenultimate closed syllables and antepenultimate stress show a negative correlation, as only 20.33% of words in that category have a closed antepenultimate syllable. A similar pattern is found for penultimate stress, given that only 35.4% of stressed penultimate syllables are heavy. These facts have been the motivation for most phonological analyses of Portuguese stress. Such analyses typically conclude that weight-sensitivity is only present word-finally.

What is missing from Table 1.2 is whether or not the unstressed syllables in a given word are closed or open. In other words, what do the penultimate syllables look like in words with final stress? This is an important gap in traditional analyses of weight in BP. If penultimate syllables

[7] See Wetzels (1997) for a comprehensive discussion on the distribution of final and penultimate syllabic shapes. A similar discussion for Spanish is found in Harris (1983).

Table 1.2: Stressed syllable profiles by stress pattern in the Houaiss Dictionary (N=164,291)

	Open σ		Closed σ	
Pattern	*N*	%	*N*	%
Final stress	5780	19.02%	24608	80.98%
Penultimate stress	72531	64.60%	39730	35.40%
Antepenultimate stress	17242	79.67%	4400	20.33%

are not weight-sensitive, then having heavy or light syllables in that position should not alter the probability of antepenultimate stress for a given word. However, we have just seen that the weight profile of penultimate syllables does affect how likely antepenultimate stress is. §1.3 explores this and other patterns in detail.

Thus far, we have seen that weight clearly has an impact on the distribution of stress patterns in Portuguese. However, stress is not the only context where weight plays a role in the language: weight also influences mid vowel contrasts when stress is held constant on the penultimate syllable. This is known as spondaic lowering (SL), and was first formalised by Wetzels (1992). SL neutralises the mid vowel contrast in the stressed syllables of non-verbs with penultimate stress. Crucially, SL is conditioned by weight—more specifically, by the weight of the word-final syllable (see Table 1.3), a fact which is consistent with the claim that weight effects in the language are restricted to this position. Therefore, the relevance of weight to Portuguese goes beyond stress.

Table 1.3: Spondaic lowering (Wetzels 1992)

(X)ĹL]$_{\omega}$	Gloss	**(X)ĹH]$_{\omega}$**	Gloss
[ˈɛli] *vs.* [ˈeli]	‘letter L’, ‘he’	[ˈfɛzis] *vs.* ∅	‘feces’
[ˈsɛd͡ʒi] *vs.* [ˈsed͡ʒi]	‘head office’, ‘thirst’	[eˈlɛtɾoŋ] *vs.* ∅	‘electron’
[ˈbɔxa] *vs.* [ˈboxa]	‘bird species’, ‘sediment’	[ˈdɔɾis] *vs.* ∅	‘Doris’
[ˈmɔʎʊ] *vs.* [ˈmoʎʊ]	‘bundle’, ‘sauce’	[ˈmɔvew] *vs.* ∅	‘furniture’

Given the positional bias of weight effects discussed above, Bisol (1992) proposes that BP builds moraic and syllabic trochees (the former applying only word-finally). Thus, *papel* [paˈpɛw] 'paper' is parsed as [pa('pɛ$_\mu$w$_\mu$)] and *sapato* [saˈpatʊ] 'shoe' is parsed as [sa('pa$_\sigma$tʊ$_\sigma$)]. Let us now briefly look into how the moraic approach[8] deals with irregularities in stress, and what issues arise from such an approach.

Previous approaches to stress in Portuguese are categorical; that is, a set of rules or constraints generates predictable patterns only. As a result, 'exceptions' are explained with mechanisms such as catalexis (§1.1) and extrametricality: Bisol (1992), d'Andrade (1994) and Massini-Cagliari (1999) employ exceptional syllable extrametricality to account for antepenultimate stress, in which case final syllables are skipped and a syllabic trochee is built from the right edge of the word: ('σ σ) ⟨σ⟩. Likewise, words with penultimate stress and a heavy final syllable (...'XH]$_\omega$) are explained with segment extrametricality, which makes the (heavy) final syllable light: 'CV.CV⟨C⟩.

In sum, we have seen that stress as well as spondaic lowering provide strong evidence for the role of weight in Portuguese. To investigate in detail *how* weight affects stress in the language, I now turn to §1.3, which explores the patterns found in the Portuguese lexicon. We will see that the weight effects on stress are considerably more intricate than previously thought.

1.3 Data

This section probes the Portuguese lexicon in an attempt to answer the three questions posed in §1.1, repeated in (9) for convenience.

(9) a. Is weight-sensitivity only found word-finally in Portuguese?

b. Is weight-sensitivity *categorical* or *gradient*?

c. Do onsets contribute to weight, affecting stress likelihood in Portuguese?

The questions in (9) are clearly connected, since (9a) examines where weight-sensitivity is found and (9b) examines how it affects stress. Likewise, question (9c) also affects the answer to question (9b).

[8] See Lee (2007) and Hermans and Wetzels (2012) for recent moraic approaches to stress in Portuguese.

The data examined in this paper is based on the most comprehensive lexicon available in the Portuguese language: the Houaiss Dictionary (Houaiss et al. 2001). The Houaiss Dictionary contains 442,000 entries/lemmas, of which 164,291 are non-verbs, including monosyllables. Even though such a lexicon contains nearly all words in the language, it lacks the necessary components needed for a thorough phonological analysis, e.g., syllabification, stress location, segmental information etc. As a result, the list of words present in Houaiss et al. (2001) was used as a starting point for the elaboration of a stress lexicon in Portuguese (see below). The final lexicon (*Portuguese Stress Lexicon*) is freely available (Garcia 2014),[9] and contains over fifty analysable variables, which range from syllabification, stress location, weight and segmental profiles to neighbourhood density and bigram probabilities.

Given its large size, the Houaiss Dictionary also includes many words that are rarely used in spoken Portuguese. Some words are borrowings whose phonotactic patterns do not match those found in the language—e.g., German words such as *schnitzel* and *Bretschneidera* (the sequences [ʃn] and [tʃn] are not allowed in Portuguese, and undergo [i]-epenthesis). Words with more than two onset segments or two coda segments, as well as words that violate the phonotactic patterns in the language were excluded from the Portuguese Stress Lexicon ($\approx 5.6\%$), as were monosyllables ($\approx 0.4\%$).

No constraints were imposed on word length (aside from a lower bound of two syllables). The median number of syllables in the whole lexicon is four, but spoken Portuguese contains very few words with more than five syllables. If we examine the FrePOP database of spontaneous speech (Frota et al. 2010), for example, more than 90% of the words listed ($N = 188,269$) contain fewer than four syllables. Thus, a separate analysis was implemented where only words with fewer than six syllables were considered. The results of this separate analysis did not differ significantly from the results presented in this paper. Therefore, the more comprehensive analysis was preferred, where no length constraints were imposed.

One further adaptation was necessary: approximately 0.12% of the words in the lexicon have

[9] http//guilhermegarcia.github.io/psl.

antepenultimate stress *and* word-final hiatus, which is always resolved through diphthongisation in Portuguese: ˈCV.CV.V → ˈCV.CGV. For example, *terráqueo* /te.ˈxa.ke.o/ is realised as [te.ˈxa.kjʊ] 'earthling'. Diphthongisation is not categorical when the second V in a VV sequence is stressed: *piada* [pi.ˈa.da] ~ [ˈpja.da] 'joke'. This directly affects stress, since the diphthongisation yields penultimate stress in a word such as *terráqueo*. As a result, these data could potentially bias the analysis.[10] Thus, words such as *terráqueo* were removed from the data. Finally, words with more than one coda segment in any syllable in the stress domain (n = 1216) were removed, given that the vast majority of such words are either borrowings or contain a prefix such as *trans-*. The final version of the Portuguese Stress Lexicon (Garcia 2014) contains 154,610 entries (Table 1.4).

Grapheme-phoneme conversion was done by different scripts and regular expression substitutions. Some cases, however, are ambiguous. For example, the grapheme *x* can be realised as [s], [z], [k.s] and[ʃ]: *máximo* 'maximum', *exato* 'exact', *oxigênio* 'oxygen', *coxa* 'thigh', respectively—note that in all four examples *x* is in intervocalic position. Besides a qualitative difference, this grapheme is particularly important because one of its phonemic realisations involves a different syllabic configuration ([k.s]), i.e., a quantitative difference. All words containing this type of mismatch (n=2399), as well as other grapheme-phoneme idiosyncrasies, were manually checked and corrected.

Among the rare words in the lexicon, many are technical terms, which often have antepenultimate stress. This could mean that the lexicon used here is not representative of everyday Portuguese vis-à-vis stress patterns. Although the analysis in this paper is concerned with the lexicon *per se*, it would be ideal if the distribution of stress patterns in the Portuguese Stress Lexicon did not deviate much from what speakers would normally experience in their language use. To verify this, two additional word lists were examined, both of which contain only the most frequent words in the language: the Invoke Limited (IL; Dave 2012) and the LaPS (*Laboratório de Processamento de Sinais*; Klautau 2013), from the Federal University of Pará, in Brazil—unlike the Houaiss Dictionary, the IL and LaPS lists are based solely on Brazilian Portuguese. In all three word lists, the proportions of each pattern are relatively similar across all non-verbs considered. More importantly,

[10] In fact, statistical models were run with and without such words, and the predicted negative correlation was confirmed between antepenultimate stress and word-final hiatus. No other effects were influenced by these forms.

the order *penultimate* > *final* > *antepenultimate* is observed in all three cases.

Table 1.4: Portuguese word lists. IL and LaPS are based on BP.

Stress pattern	**Houaiss**	**IL**	**LaPS**
Final	18%	21%	27%
Penultimate	69%	71%	62%
Antepenultimate	13%	8%	11%
	N=154,610	N=39,705	N=8,468

1.3.1 Weight-sensitivity: the Portuguese lexicon

In this subsection, I examine how weight-sensitivity affects stress placement in the Portuguese Stress Lexicon (Garcia 2014). Firstly, I show that segmental quality does not have a clear correlation with stress in Portuguese. Secondly, I explore how the size of each syllabic constituent (§1.3.1.1) may affect stress: both subtle and robust effects are found in all three syllabic positions, namely, onset, nucleus and coda. In section 1.4, I present statistical models that capture such trends in the lexicon.

1.3.1.1 Segmental quality and stress

The lexicon described above was analysed in terms of stress patterns based on number of segments as well as consonantal quality for all three possible positions, namely, final, penultimate and antepenultimate syllables. Consonantal quality in codas or onsets does not seem to affect stress likelihood in a consistent way. Even though correlations do exist, their effects are not systematic. For example, [ɲ], which is only possible in onset position, is significantly correlated with penultimate stress when in penultimate position ($p < 0.0001$), but negatively correlated with final stress when in final position ($p < 0.0001$). On the other hand, (onset) [k] is negatively correlated with final stress in final position ($p < 0.0001$), and also negatively correlated with penultimate stress in penultimate position ($p < 0.0001$). Different trends are found for other consonants, and no systematic pattern is observed—the same can be said for vowel quality.

When we observe the distribution of the most frequent consonants in onset and coda position, we also see no consistent pattern (Table 1.5). In fact, the distribution of such consonants is as unsystematic as their correlation with stress mentioned above. For example, it could be the case that more sonorous onset segments appear more frequently in stressed syllables (shaded cells in Table 1.5). In other words, stressed positions could be more frequently occupied by more sonorous segments. That is simply not the case when we look at consonantal distributions (in Table 1.5) or consonantal correlations with stress.

Table 1.5: Most frequent onset and coda segments by stress pattern

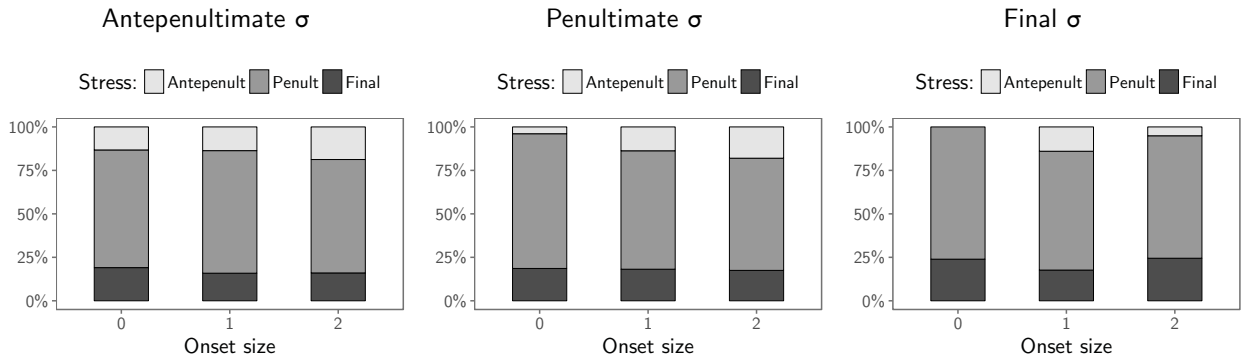
	Final σ		**Penultimate σ**		**Antepenultimate σ**	
Stress pattern	Onset	Coda	Onset	Coda	Onset	Coda
Final	/d,s,r/	/ɾ,l,s/	/k,t,ɾ/	/n,ɾ,m/	/k,t,s/	/n,ɾ,s/
Penultimate	/t,d,s/	/l,m,s/	/t,d,n/	/n,s,ɾ/	/l,k,m/	/n,ɾ,s/
Antepenultimate	/k,l,ɾ/	/s,n,ɾ/	/t,f,n/	/n,ɾ,l/	/t,l,n/	/s,n,ɾ/

1.3.1.2 Onset size effects

Let us now explore the data by examining the impact of onset size on stress. Onsets may be absent (0), as in *árvore* 'tree', singleton (1), as in *cólica* 'spasm', or complex (2), as in *prático* 'practical'—all three words have antepenultimate stress in this particular case, and are therefore represented by the darker bars in Fig. 1.2 (Antepenultimate σ). The primary focus of the exploratory data analysis that follows is to visualise how properties of a given syllable affect stress on that syllable, as opposed to stress on the other two syllables in the stress domain. The plots in Fig. 1.2 show the percentage of words with a given stress pattern according to the onset size in each syllable. All three stress patterns are shown in the top legend.

Fig. 1.2 suggests that onsets are positively correlated with stress in the antepenultimate and final syllables. The number of words with antepenultimate stress does not seem to be affected in

Figure 1.2: Onset size effects by syllable and stress pattern



different ways when the antepenultimate onset size is either 0 or 1. Rather, the difference in the Antepenultimate σ plot lies between {0,1} and 2 segments. For both antepenultimate and final syllables, onset effects on stress are not clear from the figures.

We can see in Fig. 1.2 that onset size is negatively correlated with stress in penultimate syllables. In other words, as we increase the number of onset segments in the penultimate syllable, we observe a decrease in the number of words with penultimate stress. Interestingly, it is the number of words with *antepenultimate* stress that increases as a function of penultimate onset size. As we will see below, these effects become clearer once we control for coda size. The importance of these effects will be examined in §1.4.

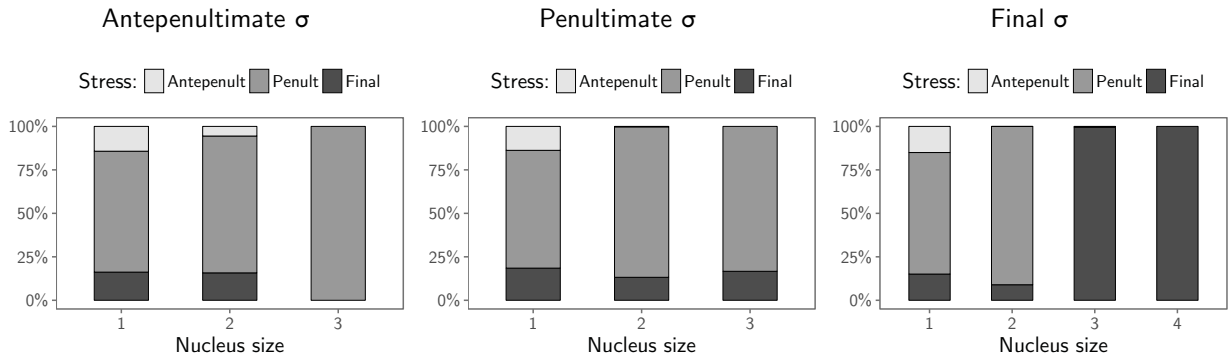
1.3.1.3 Nucleus size effects

Nuclei and codas are expected to have stronger effects on stress than onsets. In Fig. 1.3, we can see that words with penultimate and final stress seem to be affected by penultimate and final nucleus size, respectively. Longer nuclei seem to have a strong effect on stress, which is consistent with typological weight distinctions, where complex nuclei are heavier than V nuclei. For example, words such as *bastião* [bas'tjãw̃] 'bastion' always bear final stress. In these cases, the final nucleus is coded as [`ja~w`], and contains four segments, which means nasality is counted as a segment in and of itself (as mentioned in §1.2.2).

Note that the effect of nuclei on stress is visible not only word-finally, but also for the penultimate syllable, contrary to what we would expect if weight-sensitivity were constrained to the right edge

of the word in Portuguese (according to the traditional view discussed in §1.2). Interestingly, the size of antepenultimate nuclei seems to have a *negative* effect on antepenultimate stress, which is clearly unexpected.

Figure 1.3: Nucleus size effects by syllable and stress pattern

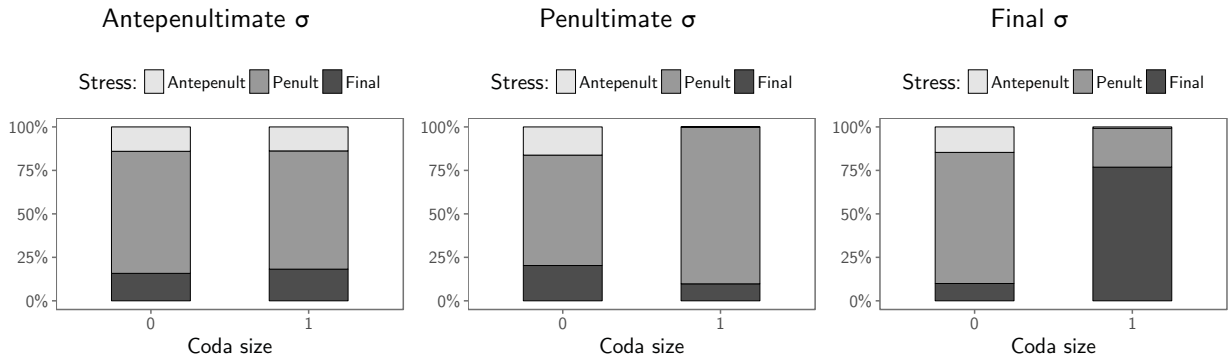


1.3.1.4 Coda size effects

Let us now examine the effect of coda size on stress placement. Fig. 1.4 shows a very strong effect of the presence of a final coda on stress placement, consistent with the standard approaches to stress in Portuguese discussed in §1.2: final stress is far more frequent when the final syllable has a coda. Like final syllables, penultimate syllables also show an effect of coda size on stress. The presence of a coda in the antepenultimate syllable, on the other hand, does not seem to strongly affect stress placement. Recall, however, that in almost all words with antepenultimate stress, only the antepenultimate syllable can be heavy (see Table 1.1). In other words, though the antepenultimate rhyme may not affect the likelihood of antepenultimate stress, the presence of penultimate and final codas is expected to have a very strong (negative) effect on antepenultimate stress. In sum, we can see that the effect of coda size on stress weakens as we move away from the right edge of the stress domain.

The trends observed above suggest that the effect of syllable weight is gradient, not categorical: coda effects are overall stronger than nucleus effects, which is unexpected, but both seem to have a substantial impact on stress. One of the possible reasons for the weaker effect of nuclei may

Figure 1.4: Coda size effects by syllable and stress pattern

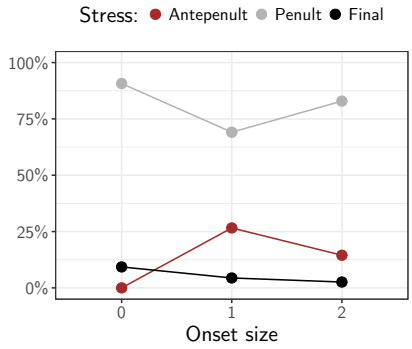


be the fact that rising diphthongs are traditionally considered to be light in Portuguese, but such cases count as complex nuclei in Fig. 1.3 (see §1.2.2; I return to this discussion in §1.4.1.3). How much weight influences stress also depends on which syllable one examines: final stress is more strongly affected by nuclei and codas than penultimate stress. In other words, weight effects seem to vary considerably across (and within) syllables, and are not only found word-finally. Onsets also show some effect on stress, though the trends observed here indicate that these segments may be *negatively* correlated with stress in a given syllable. These trends are statistically analysed in §1.4 below.

Given the trisyllabic window in which stress falls in Portuguese, we can verify the onset-stress relation in the two final syllables. Considering the coda effects in Fig. 1.4, XLL]$_{\omega}$ words will most likely have pre-final stress regardless of onset size, as the absence of a final coda will definitely impact stress on that syllable. Still, how much stress is affected could vary as the number of onset segments increases. Thus, let us examine whether final onset size affects penult/final stress.

Fig. 1.5 suggests that larger final onset sizes (specifically from 1 to 2) are more highly correlated with penultimate stress than final stress. It should be noted that singleton onsets are much more frequent in the Portuguese Stress Lexicon than complex onsets: 84.3% *vs.* 4.7% in words with final stress, 89.1% *vs.* 2.1% in words with penultimate stress, and 81% *vs.* 10.2% in words with antepenultimate stress. These data refer to stressed syllables in each pattern, but unstressed syllables also have more singleton onsets than complex onsets—Portuguese, like other Romance languages, has a relatively low frequency of onset clusters.

Figure 1.5: Stress patterns by final onset size in LLL words



The trend in Fig. 1.5 is problematic, given that onsets of a given syllable are not expected to negatively impact stress on that syllable (see, for example, Ryan (2014)). If this particular trend is statistically credible, however, the data would be consistent with a different theory of weight computation, namely, Interval Theory (Steriade 2012).[11] Unlike syllables, intervals are rhythmic units that span from a given vowel up to (but not including) the following vowel (i.e., a V-to-(V) interval). Since all intervals begin with a vowel, it follows that the string CCVCCVC is parsed into intervals as ⟨CC⟩VCC.VC (word-initial consonants are treated as extrametrical in this theory). The longer an interval, the heavier it is—and, as a result, the more likely it is to attract stress.

The crucial parsing difference between syllables and intervals lies with onset segments: an onset which would be parsed into a syllable i would be parsed into an interval $i-1$. A coda and its preceding nucleus, which in Syllable Theory belong to the same syllable, also belong to the same interval in Interval Theory. It follows that, if we transition from syllables to intervals, more onset segments in the final syllable result in a longer (and therefore heavier) *penultimate* interval—all else being equal. Consequently, the negative onset effects observed in Fig. 1.5 would be predicted by Interval Theory,[12] even though the negative effect of antepenultimate nucleus size in Fig. 1.3 would still be unexpected.

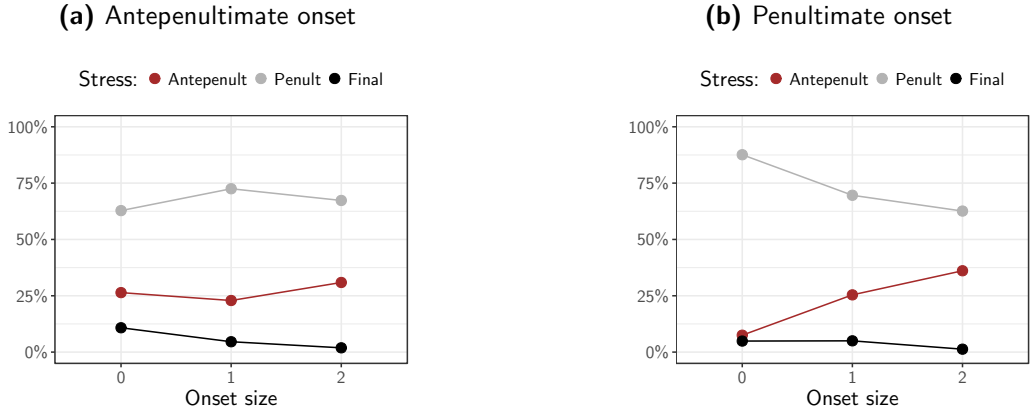
Penultimate and antepenultimate syllables are locations where coda effects are less apparent (standard analyses assume there is no such effect in these positions, as discussed in §1.2). Fig. 1.6

[11] A statistical comparison between syllables and intervals for Portuguese can be found in Garcia (2016).

[12] This assumes that complex onsets are indeed longer than singleton onsets, given that the relevant dimension of interval weight is duration (Steriade 2012).

presents the proportion of such words for different onset sizes in the penultimate syllable. Under Syllable Theory, increases in onset size in the penultimate syllable should increase the amount of material in that constituent, positively impacting its duration, which should in turn affect the likelihood of penultimate stress (assuming onsets play a role in stress assignment). Interval Theory, on the other hand, predicts an increase in the likelihood of antepenultimate stress.

Figure 1.6: Stress patterns by antepenultimate and penultimate onset size in LLL words



We see in Fig. 1.6b that the likelihood of antepenultimate stress increases when the penultimate syllable contains onset segments. Figs. 1.5 and 1.6 show a clear pattern, which is consistent with intervals. Antepenultimate onset size (Fig. 1.6a), on the other hand, presents a less clear pattern (recall that antepenultimate onsets are assumed to be extrametrical, thus no particular pattern is expected): the presence of onset clusters in this syllable seems to favour antepenultimate stress when compared to singleton onsets, but not when compared to onsetless syllables. The high degree of unpredictability of antepenultimate syllables (relative to penultimate and final syllables) might be one of the reasons behind the unexpected patterns that we find. The onset effects observed in Fig. 1.6 are also unexpected given more recent work by Kelly (2004) and Ryan (2011), for example, who show a positive effect of onset size on stress—note, however, that these studies focus on word-initial onsets.[13] As we will see in the next section, antepenultimate syllables show a pattern that is not accounted for under syllables nor under intervals.

[13] A recent study by Olejarczuk and Kapatsinski (2013) shows that stress preference in English is also affected by the phonotactic profile of word-medial clusters.

1.4 Statistical analysis

In the previous section, we observed that the patterns in the Portuguese Stress Lexicon show gradient weight effects concerning stress location in the language. In this section, I test whether the correlations in the data are supported (i.e., are significant) using statistical models that predict the location of stress based on the different syllabic constituents in the stress domain. In §1.4.1, I describe each statistical model proposed, analyse the results, and examine how they relate to the main questions in this paper, stated in (9). In §1.5, these models are compared to previous approaches, which serve as the baseline for the present analysis. Even though some of the patterns observed support intervals, the statistical models proposed in this paper are based on syllables, as the representational assumptions encoded in the predictors of syllable-based models are, by definition, a superset of those encoded by interval-based predictors, given the structural differences that hold between the two theories. In other words, one can evaluate intervals by considering syllable-based results, but not vice-versa (the reader can easily make direct comparisons between the two theories by examining the effect sizes of onsets in the models provided).

The factors examined in §1.3 are listed in Table 1.6. The number of predictors in the statistical models proposed in this paper is proportional to the size of the stress domain (3 syllables × 3 constituents per syllable = 9 predictors). Antepenultimate constituents are coded as `NA` in disyllabic words.

The analysis presented in this section employs two Binomial Logistic Regressions to model the Portuguese lexicon. Given that the stress patterns found in the language involve more than two levels, a Multinomial Logistic Regression could be employed. However, goodness of fit and diagnostics become more intricate in such a model; i.e., it is less straight-forward to assess the model's accuracy and interpret the meaning of coefficients, for instance, since outcomes are interpreted in relation to a reference level. Furthermore, the literature on multinomial models applied to linguistic data is scarce when compared to binomial models.

A more parsimonious alternative would be to model the data using *Ordinal Regression* (see Agresti 2010), also known as *Cumulative Link Model*. In this case, the stress domain in the data

Table 1.6: Predictor and predicted variables

Syllables	`onset.fin`	Number of onset segments in the **final** σ (`0-2`)
	`nucleus.fin`	Number of segments in the nucleus of the **final** σ (`1-4`)
	`coda.fin`	Number of coda segments in the **final** σ (`0,1`)
	`onset.pen`	Number of onset segments in the **penultimate** σ (`0-2`)
	`nucleus.pen`	Number of segments in the nucleus of the **penultimate** σ (`1-3`)
	`coda.pen`	Number of coda segments in the **penultimate** σ (`0,1`)
	`onset.ant`	Number of onset segments in the **antepenultimate** σ (`0-2`)
	`nucleus.ant`	Number of segments in the nucleus of the **antepenultimate** σ (`1-3`)
	`coda.ant`	Number of coda segments in the **antepenultimate** σ (`0-1`)
Stress		`antepenultimate`, `penultimate`, `final`

would need to be treated as a three-point scale, where final (1) and antepenultimate (3) positions demarcate the end-points of the domain: $3 > 2 > 1]_{\omega}$. This scale mirrors the stress domain, in terms of ordering as well as end-points (i.e., stress cannot be later than final nor earlier than antepenult). A single Ordinal Regression for the stress domain in Portuguese can be understood as equivalent to two (Binomial) Logistic Regressions. Another advantage of ordinal regressions is that predictors in such models tend to have lower standard errors when compared to equivalent binomial regressions (Christensen 2013, p. 6).

Despite the advantages of Ordinal Regressions, their interpretation is also less trivial (much like Multinomial Regressions). Because a single coefficient is generated, its interpretation depends on multiple thresholds, which act as intercepts along the scale assumed. More importantly, it is not clear that the stress domain should be treated as a scale. In other words, it is not intuitive why penultimate stress should be a higher (or lower) point in the scale when compared to final stress.

A third option is to analyse the data using Logistic Regressions (`glm()` in R (R Core Team 2020)). As mentioned above, this is the option employed in this paper. Because standard logistic models involve binary predicted variables (i.e., binomial), two such models are necessary to accommodate the stress domain in Portuguese. As a result, interpreting the effect of individual predictors

is more straight-forward, and no scale needs to be assumed (cf. Ordinal Regressions). In fact, all three options just described were compared, and the results did not differ substantially with regard to the central focus of the present study, i.e., weight gradience and its effect on stress.

In the analysis proposed in this paper, two statistical models will be used: one model (`antPenFin`) will predict `antepenultimate` *vs.* `penultimate` / `final` stress, and another model (`penFin`) will predict `penultimate` *vs.* `final` stress ('Stress' in Table 1.6). This division is aligned with traditional analyses, which classify antepenultimate stress as irregular, and penult/final stress as (mostly) regular (§1.2).

Logistic Regressions predict the log-odds of $y = 1/0$ based on a set of predictors. In this case, $y = antepenult$ *vs.* $\{penult, final\}$ in one model and $y = penult$ *vs.* $final$ in another model. The fitted model is given in (10), where $Pr(y_i = 1)$ denotes the probability that stress $y = 1$; β^0 represents the intercept, which can only be interpreted when all other variables are set to zero (this is not meaningful for the purposes of the present analysis); $(\beta^{1\dots n})$ represents the regression coefficients for each predictor; and X_i stands for the values of the i^{th} data point (i.e., the segmental count at each syllabic constituent). For example, assume we have a CVCCVCV word such as *martelo* 'hammer', which is syllabified as CVC.CV.CV (mar.te.lo). In the `antPenFin` model, we would predict the probability of antepenultimate stress (*vs.* penult/final stress) based on the segmental count in each syllable in the stress domain: $Pr(y_i = APU \text{ } vs. \text{ } \{PU, U\}) = logit^{-1}(\beta^0 + [1_m \cdot \beta^1 + 1_a \cdot \beta^2 + 1_r \cdot \beta^3]_\sigma + [1_t \cdot \beta^4 + 1_e \cdot \beta^5 + 0_\varnothing \cdot \beta^6]_\sigma + [1_l \cdot \beta^7 + 1_o \cdot \beta^8 + 0_\varnothing \cdot \beta^9]_\sigma)$. In this case, we are interested in how much each predictor in the set $\{\beta^{1\dots 9}\}$ affects such a probability.

(10) **Logistic Regression**

$$Pr(y_i = 1) = logit^{-1}(\beta^0 + X_i^1 \cdot \beta^1 + X_i^2 \cdot \beta^2 + ... + X_i^n \cdot \beta^n)$$

As we will see, both models (`antPenFin` and `penFin`) capture the weight gradience in Portuguese. Besides, given the probabilistic nature of the approach proposed in this paper, the models are more accurate than previous categorical analyses in predicting the weight-stress patterns present in the lexicon. In the subsection that follows, I present the models and examine their results and predictions. In §1.5, I contrast these predictions with the actual patterns in the lexicon, and discuss

how the present analysis differs from previous approaches.

1.4.1 Models of stress

1.4.1.1 Model A: `antPenFin`

In this model, stress (antepenultimate or penultimate/final) is predicted based on syllabic constituents in all positions in the stress domain. The `antPenFin` model is presented in Table 1.7, where we can see that all nine predictors have a highly significant effect on stress ($p < 0.00001$), which confirms that weight effects are not limited to word-final syllables. In addition, we can see that effect sizes weaken as we move away from the right edge of the word. All coefficient values in Table 1.7 have been centred and scaled, and are therefore directly comparable to one another (each $\hat{\beta}$ unit corresponds to one standard deviation of a given predictor).

Table 1.7: Scaled (and unscaled) coefficient values for `antPenFin` model ($\hat{\beta} > 0 \rightarrow$ higher likelihood of antepenultimate stress), with associated odds ratio (**OR** $= e^{|\hat{\beta}|}$), standard errors, Wald z values and significances

Predictor	`scale(`$\hat{\beta}$`)`	$\hat{\beta}$	`scale(`**OR**`)`	**OR**	**SE**	z **value**	p **value**
`onset.ant`	0.109	0.27	1.115	1.31	0.009	12.540	< 0.00001
`nucleus.ant`	-0.220	-1.22	1.246	3.38	0.012	-18.380	< 0.00001
`coda.ant`	-0.051	-0.14	1.052	1.15	0.008	-5.974	< 0.00001
`onset.pen`	0.334	0.89	1.396	2.43	0.009	36.755	< 0.00001
`nucleus.pen`	-1.107	-4.71	3.025	111.05	0.047	-23.460	< 0.00001
`coda.pen`	-2.724	-6.89	15.241	982.40	0.119	-22.840	< 0.00001
`onset.fin`	0.624	2.31	1.870	10.07	0.014	45.799	< 0.00001
`nucleus.fin`	-2.773	-5.82	16.007	336.97	0.132	-20.968	< 0.00001
`coda.fin`	-1.169	-3.68	3.219	39.65	0.026	-44.450	< 0.00001
							κ = 28.19

The results in Table 1.7 indicate key trends. First, we find divergent weight effects between rhymes and onsets across all three predictor positions in question. For example, whereas both the nucleus size and the coda size in the antepenultimate syllable negatively affect the likelihood of

antepenultimate stress, the size of antepenultimate onsets *positively* affects antepenultimate stress.

The onset effects we observe in Table 1.7 are consistent with the data trends discussed in §1.3.1.1, i.e., increasing the penultimate onset size increases the likelihood of antepenultimate stress. It is also possible to see that onset effects (in absolute terms) weaken as we move away from the right edge of the word—the same is true for nucleus effects. If we combine the penultimate onset effect with the antepenultimate rhyme effect discussed above, we can conclude that a word such as CV.CCV.CV could likely be the optimal candidate for antepenultimate stress (multiple onset clusters in the same word are uncommon in Portuguese).

Unsurprisingly, both penultimate and final rhymes negatively affect antepenultimate stress. In other words, LLL is the ideal weight profile for this particular stress pattern. Interestingly, the effect size of nuclei and codas is different when penultimate and final syllables are compared: in final position, nuclei have a stronger effect than codas, while in penultimate position codas have a stronger effect. In fact, the presence of a word-final coda reduces the odds of antepenultimate stress by a factor of nearly 3.2, whereas the presence of a penultimate coda reduces the odds of antepenultimate stress by a factor of 15.2 (see §1.4.1.2 for a discussion of nucleus-coda effects). These observed differences in effect size also capture a particular lexical pattern in the language, namely, that ẋHL is less common than ẋLH (Table 1.1).

One important characteristic of an optimal data set is that the predictors involved are orthogonal, i.e., uncorrelated—although this is rare in practice, predictors should ideally be as uncorrelated as possible. The more non-orthogonal predictors are, the more difficult it becomes to explain exactly which predictors are responsible for a given effect—this is a phenomenon known as *collinearity*[14] (Belsley et al. 1980). The predictors included in the model in Table 1.7 have medium-high collinearity ($\kappa = 28.19$).

The syllabic shapes found in Portuguese explain why collinearity is not low between onsets, nuclei and codas: although both heavy nuclei and codas are allowed, GVC/VGC syllables are rare in the language—i.e., syllabic predictors are not completely orthogonal. Furthermore, words with coda segments in multiple syllables are uncommon in the Portuguese lexicon. A Spearman ρ^2 test

[14]Represented here by κ. A model with $\kappa \leq 6$ has no collinearity; $\kappa \approx 15$ indicates moderate collinearity; and $\kappa \geq 30$ points to high collinearity (Baayen 2008, p. 182).

reveals that the most collinear pair of predictors included in the `antPenFin` model is `onset.pen` and `coda.ant` ($\rho^2 = 0.15, p < 0.00001$). Higher collinearity does not affect the model's coefficients; rather, it increases standard errors, which in turn lower the significance of a given effect (Baayen 2008). However, all the effects in question are highly significant ($p < 0.00001$), and therefore even relatively high collinearity should not pose problems for the analysis.

1.4.1.2 Model B: `penFin`

The `penFin` model in Table 1.8 shows that only penultimate onsets have no significant effect on penultimate (*vs.* final) stress—all other predictors are highly significant ($p < 0.00001$). Let us begin by examining the three predictors in the final syllable (positive $\hat{\beta}$ values indicate a higher likelihood of penultimate stress). First and foremost, we can see that most of the trends discussed in §1.3.1.1 are also confirmed in this model. For example, final onsets do have a positive effect on *penultimate* stress. In fact, adding an onset segment to the final syllable increases the odds of penultimate stress by a factor of 1.14 (note that this effect is inconsistent with the typical representational assumptions of Syllable Theory, as mentioned in §1.3.1). We also see that both `nucleus.fin` ($\hat{\beta} = -1.103, p < 0.00001$) and `coda.fin` ($\hat{\beta} = -1.49, p < 0.00001$) have negative effects on penultimate stress, which is expected, given that this is known to be a very robust aspect of Portuguese stress (§1.2).

Surprisingly, `nucleus.fin` has a weaker effect than `coda.fin`—a pattern also found for penultimate syllables in the `antPenFin` model discussed above. This contradicts a strong typological tendency, whereby VV is heavier than VC (Gordon 2011). Recall that Portuguese has no long vowels, and, importantly, not all complex nuclei in the language are assumed to affect stress, as rising diphthongs are traditionally treated as light. The model presented in Table 1.8 makes no distinction between rising and falling diphthongs (as discussed in §1.2.2), since `nucleus.fin` and `nucleus.pen` simply count the number of segments (× slots in Fig. 1.1) in the domain (this is further motivated in §1.4.1.3).

This could explain why the effect of final nuclei is smaller than that of final codas in this model. To check whether this was the case, alternative models (`*`) were run where only falling diphthongs

were considered to be heavy. In the `penFin*` model, `nucleus.fin` ($\hat{\beta}$ = 1.00) still has a smaller effect size than coda.fin ($\hat{\beta}$ = 1.39), and `nucleus.pen` ($\hat{\beta}$ = 0.08) still has a smaller effect size than `coda.pen` ($\hat{\beta}$ = 0.17). The same pattern is found in the `antPenFin*` model.

Table 1.8: Scaled (and unscaled) coefficient values for `penFin` model ($\hat{\beta} > 0 \rightarrow$ higher likelihood of penultimate stress), with associated odds ratio (**OR** $= e^{|\hat{\beta}|}$), standard errors, Wald z values and significances

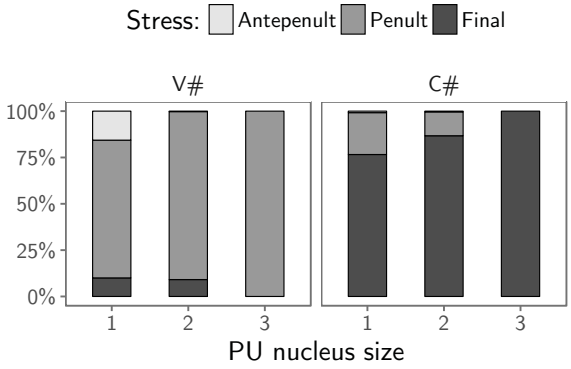
σ **predictor**	`scale(`$\hat{\beta}$`)`	$\hat{\beta}$	`scale(`**OR**`)`	**OR**	**se**$(\hat{\beta})$	z **value**	p **value**
`onset.pen`	0.010	0.03	1.01	1.03	0.01	1.09	0.259
`nucleus.pen`	−0.084	-0.33	1.09	1.39	0.01	−8.04	< 0.00001
`coda.pen`	0.141	0.33	1.15	1.39	0.01	12.28	< 0.00001
`onset.fin`	0.134	0.45	1.14	1.57	0.01	14.61	< 0.00001
`nucleus.fin`	−1.103	-2.14	3.01	8.50	0.01	−135.75	< 0.00001
`coda.fin`	−1.490	-4.29	4.43	72.97	0.01	−181.14	< 0.00001
							κ = 18.23

Let us now examine the results of `nucleus.pen` and `coda.pen`. First, `nucleus.pen` shows a negative effect on penultimate stress, which is unexpected. This, again, could be connected to the distinction between rising and falling diphthongs discussed above: if most diphthongs in `nucleus.pen` happen to be *rising* diphthongs, this pattern could be explained. However, that is not the case. In fact, if we only examine words with a complex penultimate nucleus, 52% of such words contain the falling diphthong [ej], almost all of which have penultimate stress.

One potential reason behind the negative effect of `nucleus.pen` is another variable in the model: `coda.fin`. These two variables are negatively correlated, and removing `coda.fin` makes the effect of `nucleus.pen` turn positive—which is what we would expect given the trends in Fig. 1.3. The interaction between these two variables, however, is not captured in Fig. 1.3, since nuclei are plotted independently. Once we visually inspect these two variables (Fig. 1.7), we can clearly see that penultimate diphthongs have different effects depending on whether the word-final syllable contains a coda consonant (C#) or not (V#). Particularly, once the word-final syllable contains a coda consonant, the more segments a word has in its penultimate nucleus, the less likely penultimate

stress becomes (dark bars in Fig. 1.7). For example, words such as *fácil* 'easy' are more frequent in the Portuguese lexicon than words such as *lêucon* ['lew.koŋ] 'leucon'[15] (23.2% *vs.* 12.4%). In other words, if we only look at disyllables that contain no penultimate coda but which do contain a word-final coda (n=1,871), those with a monophthong in penultimate position are two times more likely to bear penultimate stress when compared to those with a diphthong in penultimate position.

Figure 1.7: Penultimate nucleus size by word-final profile (V# *vs.* C#)



Because this paper assumes that theoretical premises should guide the statistical analysis, the model presented in Table 1.8 does not include the interaction in question. A syllabic representation does not predict that nuclei and codas in different syllables should interact. In other words, there is no principled reason to believe these two variables should affect each other (see §1.4.1.3 for a discussion)—in fact, other interactions could also exist in the language. The objective of the present analysis is not to build the best statistical model, which could include a number of unprincipled interactions. Rather, the objective is to build a theoretically principled model that is able to best capture weight gradience in Portuguese.

Let us now turn to `coda.pen`, which had a significant effect in the `penFin` model. The positive coefficient value of this predictor ($\hat{\beta}$ = 0.141) indicates that adding a coda segment to the penultimate syllable increases the odds of penultimate stress by a factor of 1.15. This is naturally a much smaller effect than, for example, `coda.fin`, but it is highly significant. The effect sizes listed

[15]Interestingly, almost all CV́G.CVC words are borrowings, and are rarely used in spoken Portuguese.

in Table 1.8 clearly show a gradient effect, whereby predictors in the final syllable have a greater absolute effect than predictors in the penultimate syllable.

In Table 1.8, `onset.pen` had no significant effect on stress. A relevant question is whether this null effect is also found once we model only disyllabic words. Indeed, if we restrict the `penFin` model to disyllables only (N=11,475), we do find that `onset.pen` has a positive effect on penultimate stress ($\hat{\beta} = 0.16, p < 0.00001$).

1.4.1.3 Model assessment

The models above have both expected and unexpected results. In the `penFin` model, for example, the effects of `nucleus.pen` and `onset.fin` go against what a syllabic representation would predict. On the other hand, the expected strong effect of final nuclei and codas possibly explains why previous analyses of Portuguese stress have constrained weight effects to the right edge of the word: such analyses have concentrated on word-final syllables only most likely because of the considerably different coefficient values between final and penultimate syllables ($\frac{\hat{\beta}_{\texttt{coda.fin}}}{\hat{\beta}_{\texttt{coda.pen}}} \approx 10$ in the `penFin` model). Therefore, though the structure of earlier syllables does affect stress placement, these effects are small compared to the structure of the final syllable, and may not be noticed unless a large enough subset of the Portuguese lexicon is examined.

In §1.2.2, we saw that, contrary to most analyses of Portuguese stress, rising diphthongs may not always pattern as light (following Harris (1983)). In particular, given that both CV́.CVG.CV and CV́.CGV.CV words are unattested in the language, it is not clear that a categorical weight difference can be determined for on-glides *vs.* off-glides. This is one of the reasons why the models above do not differentiate rising and falling diphthongs. A second reason, discussed below, is conceptual.

Should a model that only includes quantitative predictors be sensitive to the difference between rising and falling diphthongs? Should this distinction be 'visible' to the model? How rich a model is has to do with the types of theoretical and representational assumptions said model should encode. We are interested in a model that predicts stress based on quantitative information. One of the main objectives of the model is to determine how weight affects stress. Such a model should

be as unbiased as possible. By differentiating rising and falling diphthongs, we would be adding information to the model that goes beyond a neutral segmental count—in fact, this would inform the model of a specific weight effect in the language (an effect which should be unknown *a priori*). In other words, we would be telling the model that a specific sequence of segments is light, even though the purpose of the model is to inform us about weight effects.

The two models presented and discussed above show that the weight patterns in the Portuguese lexicon are much more intricate than one would expect—and far from categorical. Firstly, such effects go in two directions. Whereas in the `penFin` model penultimate stress becomes less likely when final syllables are heavy, in the `antPenFin` model antepenultimate stress becomes less likely when *antepenultimate* syllables are heavy. In fact, we also see positive and negative weight effects in the penultimate rhyme (`penFin` model), where `nucleus.pen` and `coda.pen` have opposite effects on penultimate stress.

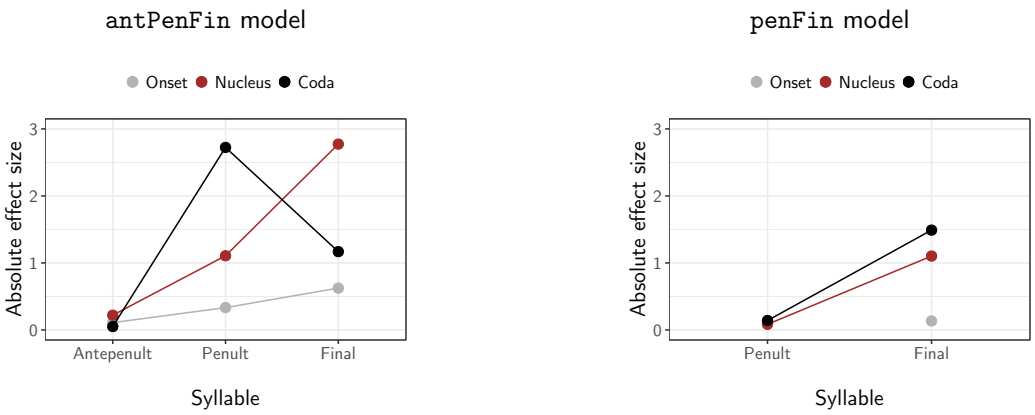
One could argue that some of these facts may be related to the footing patterns in Portuguese. The language is traditionally classified as trochaic (see Bisol (2000) for a review), and therefore (ĹL) and (Ĥ) feet should be preferred (Hayes 1995). In addition, recall that previous analyses have argued that the final syllable is extrametrical in words with antepenultimate stress (Bisol (1992) and many others). If we now combine these two facts, we can partially explain why both `nucleus.ant` and `coda.ant` are negatively correlated with antepenultimate stress: given that ĹL trochees are preferred to ĤL trochees, (ĹL)⟨X⟩ is better than (ĤL)⟨X⟩, and therefore the former should be more likely than the latter. A third footing option, namely, (Ĥ)L⟨X⟩, is preferred to (ĤL)⟨X⟩. However, this leaves a syllable unparsed in the middle of the stress domain, which not only contradicts traditional foot-based analyses of Portuguese, but is also highly marked. This approach thus assumes that light antepenultimate syllables are more stress-attracting due to footing, and not to weight *per se*.

Not all facts are accounted for by extrametricality and footing patterns, however. For example, whereas the negative effect of `nucleus.pen` would be explained, the positive effect of `coda.pen` would not. Furthermore, the onset effects found in both models would require an additional explanation, as one would not expect such effects in a standard foot-based analysis. Indeed, there does

not seem to be a theoretically unified way of accounting for all the effects found in the syllable models under discussion.

Let us now turn to the main focus of the present analysis, namely, weight gradience. The absolute coefficient values in the `antPenFin` and `penFin` models argue for a clear *gradient* notion of weight-sensitivity in Portuguese. Contrary to what previous analyses assume, the models discussed above show that weight is not a categorical phenomenon in the language. In Fig. 1.8, the absolute effect size of each predictor (i.e., syllable constituent) is plotted for each of the two models (`antPenFin` and `penFin`). These figures provide a more evident gradient trend (dotted lines): predictors in the final syllable have a stronger effect on stress when compared to predictors in the penultimate syllable (`penFin` model), which in turn have stronger effects on stress than predictors in the antepenultimate syllable (`antPenFin` model).

Figure 1.8: Absolute effect sizes of onset, nucleus and coda



As can be seen in Fig. 1.8, the effects of predictors in the penultimate syllable are relative to the statistical model. In other words, the absolute difference between penultimate and final predictors is smaller than that of penultimate and antepenultimate predictors. This trend indicates that the antepenultimate syllable is the least weight-sensitive position in the stress domain in Portuguese. In addition to the weight gradience across syllables, we also observe gradual effects within syllables: Coda > Nucleus > Onset for final syllables in the `penFin` model and penultimate syllables in the `antPenFin` model, but Nucleus > Coda > Onset for final syllables in the `antPenFin` model. For antepenultimate syllables, the absolute effect sizes indicate a different trend, namely,

Nucleus > Onset > Coda. Although this trend is highly significant, it is surprising and difficult to interpret; that is, it is not clear how such a pattern could be accommodated by any representational assumptions regarding rhythmic units.

1.5 Discussion

In this section, I summarise and discuss the main results presented in this paper. Section 1.5.1 evaluates the accuracy of the probabilistic analysis I propose, and section 1.5.2 briefly explores the implications of the approach adopted here for the grammar of Portuguese.

The models discussed in §1.4 clearly answer the questions in (9). First, weight-sensitivity is found in all positions in the stress domain, not only word-finally. Second, weight effects are gradient, not categorical. These two facts are evident in both statistical models discussed above. Third, onsets do contribute to weight in the Portuguese lexicon. However, the latter effect manifests itself in an unexpected way, given that in penultimate and final syllables onset size is *negatively* correlated with stress.

Both models examined in this paper clearly show that the relationship between stress and weight in Portuguese is far more intricate than previously assumed. Inconsistencies and surprising effects are not only limited to onsets: (i) penultimate codas have a stronger effect than penultimate nuclei; (ii) final codas have a stronger effect than final nuclei in predicting penultimate stress (`penFin` model); (iii) heavy antepenultimate rhymes disfavour antepenultimate stress.

The most important characteristic of the present approach is its probabilistic nature. A categorical approach cannot predict that a certain irregular pattern exists (e.g., ĹLL), given that it deviates from traditional generalisations about the language (XXH́ else XX́L). The present proposal, however, predicts that all licit stress patterns are possible (including so-called irregular cases), that some are more likely than others. Crucially, and perhaps most importantly, it is no longer the case that all irregular forms are *equally* unlikely, an implication of standard analyses. As we will see below, the probabilistic nature of the present approach results in a more accurate characterisation of stress in the Portuguese lexicon.

1.5.1 Accuracy

In this section, I briefly compare the predictions of the present approach with those of traditional categorical analyses. First, let us examine the predictions of the `antPenFin` model in Fig. 1.9, which plots the proportion (or probability) of words with antepenultimate stress (*vs.* penultimate or final stress) across sets of words that mirror the different weight profiles (i.e., sequences of light (L) and heavy (H) syllables) in the language.[16] The dotted line represents the predicted probability of antepenultimate stress based on traditional (categorical) approaches (i.e., 0%, since antepenultimate stress is considered to be irregular). Actual lexical proportions are represented by ● (where the size of the circle corresponds to lexical representativeness). These proportions are based on the set of words being modelled in each model: `antPenFin`: N=143,135; `penFin`: N=134,599. Finally, ◯ represents the mean predicted probability of antepenultimate stress based on the present approach.

As we can see in Fig. 1.9, in some cases (e.g., HHH, HHL, LHL), categorical predictions accurately match the actual lexical proportions. Note that a clear mismatch is observed for HLL and LLL words. These cases, however, are accurately approximated by the predicted proportions (◯) in Fig. 1.9. In other words, given a new LLL word, the present analysis predicts that there is a ≈25% probability that such a word will be assigned antepenultimate stress, and a 75% probability that stress will be either penultimate or final. Traditional approaches, on the other hand, would not predict antepenultimate stress in this (or any other) case.

Assuming that a word is not assigned antepenultimate stress, we now need to consider penultimate *vs.* final stress, which account for the vast majority of words in the lexicon (Table 1.1). Fig. 1.10 plots the percentage (or probability) of words with penultimate stress (*vs.* final stress) across the different weight profiles in the language. Recall that traditional approaches predict final stress for all words with a heavy final syllable (XXH́) and penultimate stress elsewhere (XX́L). Clearly, these predictions deviate considerably from the actual lexical proportions of penultimate stress (●).

Like Fig. 1.9, Fig. 1.10 shows that probabilistic predictions are substantially more accurate

[16] Predicted probabilities are averaged across all words with a given weight profile.

than a categorical approach. Even though we observe a clear distinction between XXH and XXL words, a gradient effect *within* each group is also visible. For example, ĹL words are more frequent than ĹL words—and this difference is mirrored in the models' mean predicted probabilities.

Figure 1.9: `antPenFin` model's accuracy: Mean predicted probabilities (◯) of antepenultimate (*vs.* penult/final) stress by weight profile as well as actual lexical frequencies (●) are plotted. Dotted lines indicate predicted stress based on a standard categorical analysis.

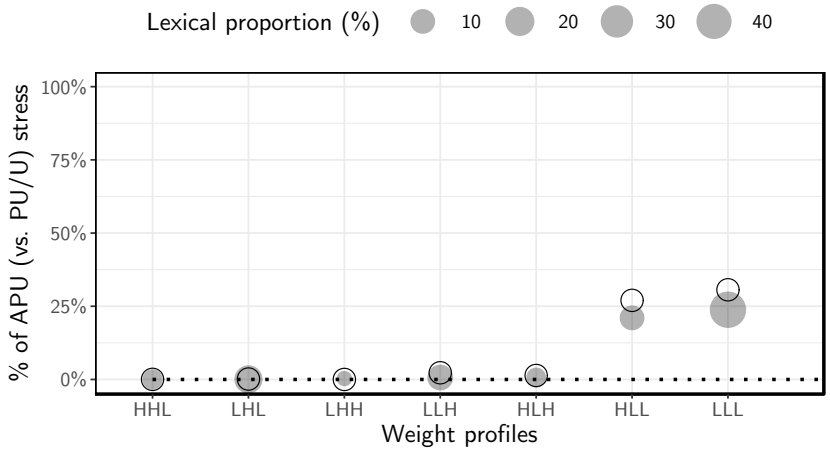
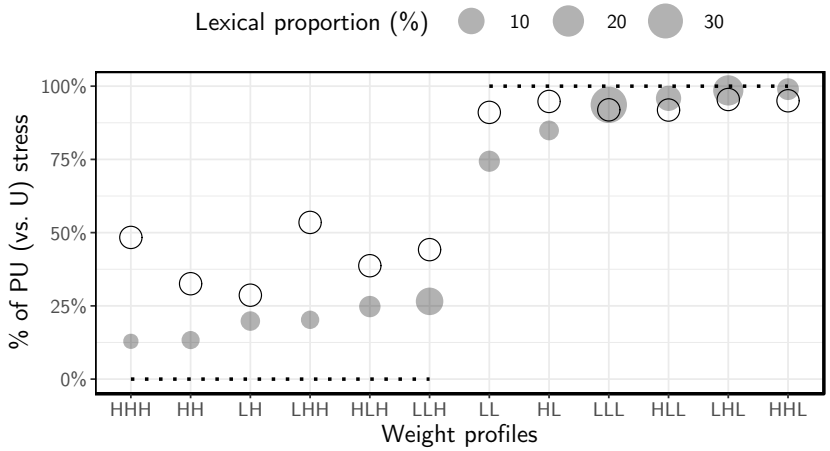


Figure 1.10: `penFin` model's accuracy: Mean predicted probabilities (◯) of penultimate (*vs.* final) stress by weight profile as well as actual lexical frequencies (●) are plotted. Dotted lines indicate predicted stress based on a categorical analysis



Whereas Figs. 1.9 and 1.10 both provide a means to visually compare the present proposal

to traditional analyses of Portuguese stress, Table 1.9 presents a numerical comparison, namely, the weighted mean deviation of predicted probabilities from actual lexical percentages. The mean deviation takes into account the representativeness of each weight profile in the lexicon (in the plots). Not only is the weighted mean deviation lower in the probabilistic approach presented here, the weighted standard deviations are also lower when compared to a categorical approach, a fact which mirrors the trends in Figs. 1.9 and 1.10.

Table 1.9: Weighted mean deviation of mean predicted probabilities from actual lexical proportions: probabilistic *vs.* categorical approaches

	Probabilistic approach		**Categorical approach**	
	Mean	SD	Mean	SD
`antPenFin`	**4.2%**	3.6%	**14.0%**	13.6%
`penFin`	**6.4%**	7.9%	**9.7%**	10.4%

1.5.2 A probabilistic grammar

Thus far, we have investigated the stress patterns in the Portuguese lexicon by employing different statistical models. Little has been said, however, about what these patterns mean for the *grammar* of Portuguese speakers. If the lexical patterns explored in this paper are psychologically real, an important question is (i) how such patterns could be implemented in a phonological grammar and (ii) how the lexicon and grammar interact. We will not construct such a grammar here, given that at present we do not know how closely speakers' grammars mirror the lexical patterns modelled in this paper, but will sketch what such a grammar (henceforth $\mathcal{G}$) could look like.

A first step towards modelling $\mathcal{G}$ would be to determine whether the patterns presented in this paper reflect what is in the minds of speakers. If that is the case, $\mathcal{G}$ could be modelled within probabilistic versions of Optimality Theory (Prince and Smolensky 1993) where constraints are weighted (Pater 2009), such as MaxEnt Grammar (Hayes and Wilson 2008) or Noisy Harmonic Grammar (Boersma and Pater 2016). A MaxEnt Grammar would make particular sense, given that constraints correspond to different predictors (Goldwater and Johnson 2003).

To map the present analysis into a MaxEnt Grammar, the predictors discussed thus far would be equivalent to Markedness constraints that enforce weight-stress mappings based on the lexical patterns observed in the language. For example, the positional constraint WSP$_n$ (WEIGHT-TO-STRESS PRINCIPLE, Prince (1990)) would penalise an unstressed syllable in position n according to the number of segments present in σ_n—where n represents the possible positions in the stress domain.

In $\mathcal{G}$ (Fig. 1.11), the lexical distributions of stress determine how $\mathcal{G}$ will assign stress probabilistically to a novel word. Once an output is selected (probabilistically), stress will remain lexically marked on the word, which correctly ensures that stress in existing words does not vary.[17] Finally, this novel word will now be part of the lexicon (③ in Fig. 1.11).[18] Therefore, only words without stress information (i.e., novel words) will be assigned stress probabilistically. This type of distinction between existing and novel words draws on Zuraw (2000, p. 48), who employs 'listedness' as a means to differentiate the two types of words vis-à-vis the application of nasal substitution in Tagalog.

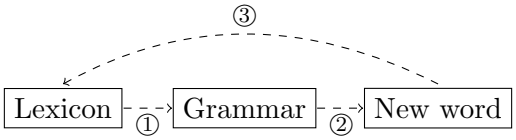
As a result of the probabilistic approach presented here, patterns are no longer treated as regular or irregular, but rather as *more* or *less* likely. For example, in a new word such as *setamira*, penultimate stress is most likely, but final (and antepenult) stress is also possible for a word of this shape. If the (less likely) candidate with final stress is chosen by the grammar, it will enter the lexicon as *setamirá*. Other constraints in $\mathcal{G}$ will ensure that (i) illicit stress patterns are not generated, e.g., pre-antepenultimate stress, and that (ii) stress does not shift once it has been assigned (i.e., stress is required to be faithfully realised in the output).

Because stress is lexically marked in the present approach, speakers need to learn a word with its particular stress position. Under categorical analyses, only irregular cases were lexically marked, since regular cases were derived based on the generalisations already discussed. The latter approach entails that speakers would memorise only the additional mechanisms responsible for irregular stress

[17] An exception is derivationally related words, where stress shifts to obey the trisyllabic window. Although an examination in stress location in such cases is beyond the scope of this paper, it appears to not be probabilistically assigned.

[18] For an alternative which assumes lexically specific constraints, see Moore-Cantwell and Pater (2016).

Figure 1.11: Relationship between lexicon and grammar ($\mathcal{G}$) assumed in the present analysis. Lexical patterns generate constraint weights ①. Stress in new words is assigned based on probabilities ②. Once stressed, a new word enters the lexicon ③.



(e.g., extrametricality). Crucially, the present approach provides an explanation as to how lexical stress is assigned to *all* words (probabilistically, based on the stress patterns already present in the lexicon). In other words, particular groups of words (e.g., words with antepenultimate stress) do not require a different explanation.

The probabilistic approach presented here is solely based on the quantitative aspect of weight, which means segmental quality was not part of the model employed in the analysis. Likewise, metrical representation is not included in the model, as the objective was to evaluate how accurate a model solely based on weight could be. Naturally, the absence of such a representation does not imply that a metrical component plays no role in the grammar of Portuguese. For example, even if feet do not play a direct role in primary stress assignment *per se*, they could still play a role in restricting the stress domain to the final three syllables in a word and in assigning secondary stress.

In addition to WSP discussed above, other constraints in $\mathcal{G}$ will play an important role if speakers' grammars in fact encode all the lexical patterns found in this paper. For example, suppose that the specific effect where antepenultimate rhymes negatively impact antepenultimate stress is indeed generalised to novel words by speakers. This effect would not be captured by a constraint such as WSP, given that constraints in a MaxEnt framework cannot have negative weights. Instead, positionally defined constraints against marked structures (e.g., *COMPLEX) could indirectly capture the observation that antepenultimate syllables are more likely to bear stress if their rhymes are minimally complex.

In sum, at present we do not know how speakers' grammars and the lexical patterns modelled in this paper compare. Previous work has shown that statistically significant trends in the lexicon are not necessarily generalised by speakers (Albright and Hayes 2006, Hayes et al. 2009, Becker et al.

2011). For example, the negative effects observed in antepenultimate syllables mentioned above may not be reflected in the minds of speakers and therefore encoded in the grammar. Likewise, the negative onset effects found for penultimate and final syllables may be restricted to the lexicon, and may not be generalised to novel forms by native speakers. Crucially, the model presented here formalises the lexicon as a hypothetical baseline, which is a necessary step if one wishes to examine whether the lexicon mirrors speakers' grammars. Future work is needed to investigate how the grammar and lexicon compare vis-à-vis the probabilistic assumptions made in this paper.

1.6 Conclusion

This paper examined the role of weight in stress assignment in Portuguese. I proposed a probabilistic model that focuses on weight as the only predictor of stress location. The objective of such a model was to show that weight effects are gradient, not categorical as assumed in previous literature (Bisol 1992, Mateus and d'Andrade 2000, Lee 2007, Wetzels 2007). Likewise, these effects are shown to be more intricate than what traditional approaches presume, given that some effects are negatively correlated with stress (e.g., antepenultimate nuclei; penultimate and final onsets).

Assuming that the lexicon does indeed reflect the grammar, the probabilistic grammar implied in this paper considers that stress is assigned based on a probability distribution derived from the patterns present in the lexicon. Stress remains lexically marked once assigned. This approach is substantially different from traditional analyses. First, a formal distinction between regular and irregular patterns no longer exists. Rather, a given stress location is more or less likely. Likewise, weight is not categorically defined (e.g., heavy or light). Instead, a weight continuum is assumed, whereby the notion of weight-sensitivity is understood as inherently gradient (cf. Albright and Hayes 2006, Ryan 2011).

Unlike previous approaches, the probabilistic analysis proposed in this paper predicts that speakers could in principle assign antepenultimate stress to a new LLL word, for instance. In contrast, categorical studies predict that so-called irregular cases are not generalisable. Future research is needed to test which of these predictions is confirmed, and whether the weight effects in the Portuguese lexicon are reflected in speakers' grammars. This will also provide a means to compare to

what extent the subtleties found in the Portuguese lexicon are in fact captured (and generalised) by speakers. As the relationship between the Portuguese lexicon and speakers' grammars becomes clearer, the probabilistic approach to weight assumed here can be further developed, and its impact on other aspects of the grammar can be evaluated.

From the lexicon to the grammar

The previous chapter examined weight effects in the Portuguese lexicon, and showed that a categorical view cannot accurately capture the patterns found in the language. Given that such patterns motivate a gradient notion of weight, the chapter also proposed a probabilistic approach to weight and stress, whereby stress is more or less likely to fall on a given syllable depending on (i) its position in the stress domain, and (ii) its segmental count. In other words, weight gradience can be observed within and across all stressable syllables in Portuguese, *contra* studies which either deny the importance of or assign a minimal role to weight in the language (e.g., Lee 1994, Santos 2001, Cantoni 2013). Furthermore, unlike traditional approaches to stress in Portuguese (e.g., Bisol 1992, Lee 2007), the probabilistic grammar assumed in Chapter 1 predicts that stress in novel words can fall on *any* of the three syllables in the stress domain—a prediction which has been recently confirmed experimentally (Benevides 2017).

In examining weight in a comprehensive corpus, the statistical analysis in Chapter 1 detected robust positive and negative rhymal weight effects in the entire stress domain, as well as subtle positive and negative effects of segmental count in onsets. Positive onset effects were found in antepenultimate syllables, which indicates that having an onset cluster in this position increases the odds of antepenultimate stress. Negative onset effects were also found: in penultimate and final syllables, an onset cluster actually decreases the odds of penultimate and final stress, respectively. The novel finding that onsets also contribute to weight in Portuguese, and thus affect stress, is consistent with studies which show similar effects in other languages (e.g., Davis 1988, Ryan 2011, 2014). Crucially, these effects lend further support to the proposal that weight effects are gradient, and thus cannot be accurately captured by a categorical approach.

A question for future research is how native speakers' grammars generalise the onset effects

found in the Portuguese lexicon, given that such effects are positive and negative, depending on which syllable in the stress domain is examined. It is possible that the role of onsets is more systematic once we probe speakers' behaviour in an experimental setting—even though detecting such effects is not a trivial task, given the small effect sizes captured in the comprehensive lexicon modelled in Chapter 1.

The next chapter focuses on a more robust pattern that emerges from Chapter 1, namely, the positive and negative gradient weight effects captured in the Portuguese lexicon. The question of interest is whether speakers' grammars mirror the weight gradience found in the lexicon, or whether speakers generalise stress patterns in a uniform fashion. In other words, weight gradience could be present in the lexicon, but not in the grammar of Portuguese.

Whether or not speakers acquire certain subtle patterns present in the lexicon has been the focus of several recent studies (Zuraw 2000, Hayes and Londe 2006, Hayes et al. 2009, Carpenter 2010, Ryan 2011, Becker et al. 2011, Becker et al. 2012, Domahs et al. 2014, Ryan 2014). One possible reason why the topic of phonological learning has gained attention may be the increasing use of more robust statistical methods to analyse sub-regular phonological patterns. Because these methods allow us to capture subtleties with great precision, researchers are now better equipped to examine whether such subtleties are acquired and generalised to novel words, which in turn sheds light on the limits of language acquisition. Furthermore, combined with the appropriate statistical tools, large data sets such as the lexicon examined in Chapter 1 allow us to uncover subtle linguistic patterns—which would otherwise be unknown.

In spite of all its advantages, a comprehensive lexicon which contains virtually all words in a given language suffers from two potential problems. First, it does not represent the input to which learners are exposed. Second, it does not represent the lexicon of any adult speaker, which naturally does not contain all the words in the language. Because our current statistical tools can detect a considerable number of subtle patterns, an important question is whether these patterns are realistically part of learners' input and, as a result, part of speakers' lexica. In other words, we first need to determine whether the effects in question can truly be understood as the 'baseline' against which speakers' grammars can be compared. For this reason, before exploring speakers' behaviour,

Chapter 2 establishes and compares different possible baselines to evaluate the representativeness of a comprehensive lexicon such as the one modelled in Chapter 1.

Among the weight effects discussed in Chapter 1, of particular interest is the negative effect found in antepenultimate syllables. Thus, besides weight gradience *per se*, the experimental approach in Chapter 2 investigates whether native speakers generalise such a negative effect to novel words.

The observation that antepenultimate stress is also affected by weight in Portuguese contradicts a long-established assumption that this stress pattern is irregular and, thus, unpredictable. Indeed, compared to penultimate and final stress, it is undeniable that antepenultimate stress is *less* predictable. The important question, however, is whether we can detect the systematic lexical effects within this class in speakers' behaviour. As will become clear throughout this thesis, weight effects in antepenultimate syllables play a crucial role in the understanding of how weight is computed in Portuguese (as well as in English, as will be shown in Chapter 3).

In summary, the following chapter examines to what extent speakers' grammars generalise (i) the weight gradience found in the Portuguese lexicon, and (ii) the negative weight effect found in antepenultimate syllables. By examining experimental data, Chapter 2 will allow us to compare the lexicon and the grammar of Portuguese, which in turn will reveal which weight patterns and sub-patterns are generalised to novel words. As a result, the discussion that follows advances the probabilistic approach in Chapter 1, and examines whether such an approach also appropriately characterises the grammar.

Chapter 2

Learn and repair:

the case of gradient weight in Portuguese

ABSTRACT

In weight-sensitive languages, stress is influenced by syllable weight. As a result, heavy syllables should attract, not repel, stress. The Portuguese lexicon, however, presents a case where weight negatively impacts stress: antepenultimate stress is more frequent in light syllables than in heavy syllables. This unnatural pattern contradicts the typology of weight and stress. This language also contains gradient, not categorical, weight effects, which weaken as we move away from the right edge of the word. In this paper, I examine how speakers' grammars capture these subtle weight effects in Portuguese, and whether the negative antepenultimate weight effect is learned or repaired. I model experimental data using Bayesian regressions and show that speakers learn the gradient weight effects in the language, but do not learn the unnatural negative effect. Instead, speakers repair this pattern, and generalise a positive weight effect for all syllables in the stress domain. This study thus provides empirical evidence that speakers may not only ignore unnatural patterns, but also learn the opposite pattern.

Keywords: stress, weight, lexicon, Bayesian analysis, probabilistic grammar, Portuguese

2.1 Introduction

Phonological learning is a central topic in phonological theory. Given the lexicon of a particular language, we are interested in determining whether (and to what extent) speakers learn robust and subtle patterns in such a lexicon from the input to which they are exposed. In this context,

the relationship between lexical statistics and speakers' grammars can help researchers understand how different patterns are learned, and, crucially, how particular phonological biases interact with linguistic information present in the input.

In the past decade, the relationship between the lexicon and the grammar has been the object of investigation of several studies (e.g., Hayes and Londe 2006, Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012, Jarosz 2017). What these studies have shown is that unnatural linguistic patterns are often underlearned by speakers. In other words, not all patterns in one's lexicon are learned, and the productivity of the patterns that are learned seems to rely on their phonological naturalness (as determined by analytic biases across languages). For example, Becker et al. (2012) examine an interesting case of a lexical pattern in English involving laryngeal alternation (e.g., *leaf* vs. *leaves*). In the English lexicon, this type of alternation is more frequent in monosyllables than in polysyllables, thus violating initial-syllable faithfulness, a cross-linguistically supported tendency to protect word-initial syllables (Steriade 1994, Beckman 1997). In a wug test, however, speakers treat both monosyllables and polysyllables equally, and therefore do not generalise the unnatural conditioning context present in the lexicon.

In the present study, I provide new evidence that speakers not only underlearn certain unnatural patterns present in their lexicon, but also *repair* such patterns. The evidence comes from weight effects on stress in Portuguese, where antepenultimate stress is *negatively* affected by weight in the lexicon (Garcia 2017b; Chapter 1).

In weight-sensitive languages, heavy syllables are not expected to repel stress, i.e., to negatively affect the likelihood of stress. In English, for example, nouns tend to have stress on a heavy penultimate syllable (*agénda, Arizóna*). If the penultimate syllable is light, stress falls on the antepenultimate syllable (*Cánada, quálity*)—this is the same stress rule found in Latin, which is therefore also classified as a weight-sensitive language. In the vast majority of such languages, the weight distinction is reported to be binary, i.e., syllables are either heavy or light (Gordon 2006).

Even though not all studies agree that weight significantly affects stress in Portuguese (e.g., Lee 1994, Cantoni 2013), most previous approaches assume that the language, like English and Latin, is sensitive to weight: heavy syllables in word-final position in nouns and adjectives attract

stress (Bisol 1992, 2013; Araújo 2007, Wetzels 2007, Garcia 2017b).[1] While weight-sensitivity in Portuguese is traditionally assumed to be constrained to the word-final syllable, the patterns found in the lexicon (Houaiss et al. 2001, Garcia 2014) contradict this assumption. As shown in Garcia (2017b; Chapter 1), once we investigate a sufficiently large word list, we find that weight effects are present in all positions in the trisyllabic stress domain in the language. More specifically, the effect of heavy syllables on stress depends on their position in the domain. Indeed, the impact of heavy syllables on stress weakens as we move away from the right edge of the word.

A second (and more surprising) characteristic of stress in the Portuguese lexicon is that heavy antepenultimate syllables negatively affect antepenultimate stress (Garcia 2017b; Chapter 1). Unlike heavy final and penultimate syllables, which positively affect final and penultimate stress, respectively, the opposite is true of the leftmost position in the stress domain. In other words, in the Portuguese lexicon, ĹLL words are more common than ĤLL words. Even though this finding contradicts the very definition of weight-sensitivity, it could be due to footing, given that ĹLL words can be analysed as having an extrametrical syllable and a moraic trochee: (ĹL)⟨L⟩. In contrast, ĤLL words can only result in more marked metrical configurations, which contain either a medial unfooted syllable, i.e., (Ĥ)L⟨L⟩, or an uneven trochee, i.e., (ĤL)⟨L⟩. The negative effect of antepenultimate syllables in the Portuguese lexicon therefore suggests that footing regulates weight effects in the language to avoid a more marked metrical structure, and that footing trumps weight.

These two facts about weight and stress in the Portuguese lexicon are the focus of the present study. The main objectives of this paper are (i) to examine whether native speakers generalise the weight gradience present in their language, and (ii) to investigate whether speakers learn the negative weight effect in the lexicon, thus favouring footing over weight, or if they favour weight over footing, and assign a positive weight effect to antepenultimate syllables. I will present empirical data from two separate experiments that show that weight gradience is indeed generalised to nonce words: not only do speakers generalise robust patterns, they are also aware of the subtle weight patterns present in the lexicon. More importantly, even though speakers' grammars capture subtle effects, they do not generalise the negative effects in antepenultimate position. Instead, (heavy)

[1] Stress in verbs, on the other hand, is determined in part by morphological factors (Wetzels 2007).

antepenultimate syllables are also shown to positively impact stress, which is what we would expect from the grammar of a language where weight is the crucial predictor of stress. These novel findings regarding weight effects on stress suggest that the effect of analytic biases can indeed go beyond the non-generalisation of a given unnatural pattern.

The analysis provided in the present paper involves a probabilistic framework, where stress is not predicted to be categorical. I employ Bayesian hierarchical models to estimate the effect of a heavy syllable in different positions in the word. This, in turn, provides the probability distribution of weight effects given the empirical data from two separate experiments. In this probabilistic approach, stress patterns are assigned probabilities on the basis of the weight profile of a given word. The flexibility of such a framework allows for a more accurate and comprehensive characterization of the experimental data analysed, which in turn helps us better understand the extent to which the empirical data reflect the lexical patterns in Portuguese.

This paper is organised as follows: in §2.2, I review the weight effects in the Portuguese lexicon, as well as previous (categorical) approaches to stress in this language. I also discuss the weight asymmetry found in the stress domain in Portuguese. In §2.3, I outline the methods used in the paper to establish a lexical baseline as well as the experimental design employed. I also present the fundamental concepts of Bayesian methods, which form the basis for the statistical analysis discussed in §2.4.

2.2 Stress and weight in the Portuguese lexicon

Weight-based analyses of Portuguese stress have traditionally assumed that weight is a categorical phenomenon in the language (see Araújo (2007) for a comprehensive review). In other words, syllables are either heavy (H) or light (L). Heavy syllables contain a coda consonant (nasal, liquid or /s/), a diphthong, or a nasal vowel: *pomár*[2] 'orchard', *gáita* 'harmonica', *aná̋* 'dwarf (fem)'. Previous analyses of stress in this language have also assumed that weight effects are restricted to the word-final syllable. As a result, the diphthong in *gáita* is not expected to increase the odds of penultimate stress relative to a word such as *gáta* 'female cat'.

[2]Throughout the paper, I use an acute accent to represent the location of primary stress.

Because previous weight-based analyses of stress in Portuguese assume that trochaic feet determine the location of stress (e.g., Wetzels 1992, Bisol 1992), the positional constraint mentioned above implies one of two alternatives: (a) either Portuguese builds moraic trochees in all positions in the stress domain, or (b) Portuguese builds moraic *and* syllabic trochees.

These widely held assumptions lie at the core of the stress rule in Portuguese, given in (1)—'X' represents either 'H' or 'L'. ASSUMPTION A entails that two heavy syllables are identical vis-à-vis their weight, i.e., a heavy (final) syllable in word W_1 cannot be heavier than a heavy (final) syllable in word W_2. ASSUMPTION B entails that no weight effects should be found in penultimate or antepenultimate syllables.

(1) **Regular stress in Portuguese** (e.g., Bisol 1992, Collischonn 2010)

ASSUMPTION A: weight is binary. Syllables are either heavy (H) or light (L)

ASSUMPTION B: weight effects are restricted to the word-final syllable

GENERALISATION: **XXH́ else XX́L**

If the word-final syllable is heavy, assign final stress. *papél* 'paper'

Else, assign penultimate stress. *cavalo* 'horse'

Words that do not follow the generalisation in (1) are considered to be irregular. For example, antepenultimate stress is traditionally deemed to be unpredictable, regardless of the weight of the antepenultimate syllable, given ASSUMPTION B in (1). In (2), I provide all combinations of weight and stress that do not follow (1). The three weight profiles listed in (2) are ordered by their lexical proportion in Houaiss et al. (2001).

(2) **Exceptional stress in Portuguese**

X́XX (13%), XX́H (11%), XXĹ (3%)

Even though the rule in (1) accounts for most words in the lexicon (72%, Houaiss et al. 2001), it does not capture important facts about the so-called exceptional cases in (2). For example, within the class of X́XX words (13%), X́LL words are much more common than X́HL, X́LH, and X́HH

words combined (99.2% *vs.* 0.8%). Indeed, once we examine the entire lexicon (Garcia 2014), we find that weight effects are considerably more intricate than previously assumed.

As I show in Garcia (2017b; Chapter 1), weight effects in the lexicon are neither categorical (*contra* ASSUMPTION A) nor restricted to the word-final syllable (*contra* ASSUMPTION B). Instead, weight-sensitivity is gradient and can only be understood in relative terms. For example, weight-sensitivity is weaker in penultimate syllables relative to final syllables. Crucially, heavy syllables affect stress differently depending on the position they occupy in the stress domain. As a result, in this paper I refer to heavy syllables in isolation according to their position in the stress domain, defined as $[\sigma_3\sigma_2\sigma_1]$, where σ_1 demarcates the syllable at the right edge of the word. Therefore, final heavy syllables will be represented as H_1. Penultimate and antepenultimate heavy syllables will be represented as H_2 and H_3, respectively. If a heavy syllable in penultimate position is heavier (i.e., has a stronger effect on penultimate stress) than a heavy syllable in antepenultimate position, we can represent this relation as $H_2 > H_3$.

2.2.1 Weight asymmetry: the case of antepenultimate stress

In the vast majority of weight-sensitive languages, syllable weight is reportedly binary (see Gordon (2006) for a typological review of weight). As we saw in (1), this generalisation also applies to Portuguese insofar as the language has been traditionally analysed as having a two-way weight distinction. Furthermore, in any given weight-sensitive language, heavy syllables are by definition expected to attract stress. In the Portuguese lexicon, however, heavy antepenultimate syllables (H_3) actually significantly *lower* the odds of antepenultimate stress (Garcia 2017b; Chapter 1).

The typologically inconsistent weight effect observed for antepenultimate syllables in the Portuguese lexicon means that LLL words are more likely to bear antepenultimate stress than HLL words—indeed, 83% of all words with antepenultimate stress are LLL. In contrast, nearly 80% of all words with final stress have a heavy final syllable (H_1)—hence the generalisation in (1). As we can see, the edges of the stress domain in the Portuguese lexicon present a remarkable asymmetry regarding weight effects. In (3), I summarise the three central observations regarding weight-sensitivity in the language.

(3) **Weight asymmetry in the Portuguese lexicon (Garcia 2017b; Chapter 1)**

OBSERVATION 1: All three syllables in the stress domain contribute to weight

OBSERVATION 2: Weight effects weaken as we move away from the right edge of the word

OBSERVATION 3: H_3 has a negative effect on stress

WEIGHT GRADIENCE: $H_3 < H_2 < H_1$

It is important to note that the effect of a given heavy syllable is relative not only to its position in the stress domain, but also to which stress patterns are being compared. For example, the effect of H_1 is stronger when final stress is compared to antepenultimate stress than when final stress is compared to penultimate stress. This is expected, given that ẊXH words are much less common than XẊH words; see (2). Likewise, H_2 has a much stronger effect in antepenultimate *vs.* penultimate stress comparisons than in penultimate *vs.* final stress comparisons. These contrasts are crucial for interpreting estimates in statistical models, where a reference stress level is used.

One possible motivation for the negative weight effect (OBSERVATION 3 in (3)) in the Portuguese lexicon stems from footing optimization, as alluded to in §2.1. Let us assume that the language builds moraic trochees across-the-board (*contra* Bisol (1992)). In that case, a ĤLL word is not optimal, given that the foot will either (a) bear three moras, (b) result in a lapse, or (c) leave a syllable unparsed even if final extrametricality is assumed: (ĤL)L, (Ĥ)LL, or (Ĥ)L⟨L⟩, respectively. As a result, ĹLL words, parsed as (ĹL)⟨L⟩, would be preferred to ĤLL words for footing reasons (assuming extrametricality). This would imply that footing regulates weight-sensitivity in the lexicon, which, in turn, would explain why we observe an apparent typological inconsistency vis-à-vis weight. A similar preference for a stressed light syllable is found in Fijian (Hayes 1995), where some stressed syllables shorten in order for the word to achieve an optimal parsing into feet. Unlike Fijian, however, Portuguese offers no empirical evidence for trochaic shortening.

In summary, the weight effects found in the Portuguese lexicon (Garcia 2017b; Chapter 1) contradict traditional assumptions insofar as they are (positionally) gradient, not categorical. Furthermore, the lexicon presents a typologically unexpected effect, namely, H_3, which negatively impacts antepenultimate stress. The objective of the present paper is thus to investigate whether the lexical facts described above are reflected in speakers' grammars. Crucially, under the assumption that

weight effects are generalised as gradient, this paper also examines whether native speakers learn the *negative* effect of H_3, thus favouring footing over weight, or if they learn a *positive* weight effect instead, thus favouring weight over footing in the language. The questions investigated in the present study are given in (4).

(4) **Questions**

a. To what extent do speakers learn the weight gradience present in the lexicon?

b. How do speakers generalise the effect of H_3?

(i) Do they favour footing over weight, and thus mirror the negative lexical effect?

(ii) Do they favour weight over footing, and thus learn the *opposite* pattern?

2.3 Methods

To investigate the questions in (4), this paper first revisits the Portuguese lexicon to establish a realistic baseline to which experimental data can be compared. Secondly, I provide data from two forced-choice experiments. I will refer to these experiments as 'Version A' and 'Version B'—Version B is a replication of Version A. Below I explain in detail how a lexical baseline is defined, the experimental design, and the data analysis employed in the paper.

2.3.1 Lexical baseline

Our starting point to define a lexical baseline is the Portuguese Stress Lexicon (Garcia 2014), which contains virtually all non-verbs in Portuguese (n = 154,610). The first important question we need to ask is whether results based on such a comprehensive word list are a realistic reflection of speakers' lexica. Because speakers' lexica are, by definition, a subset of all the words in the language, we could hypothesize that a subset with a realistic number of words might not present the same weight effects found for the entire lexicon. For example, learned words and anachronisms may follow slightly different patterns, and may therefore not be present in the lexica of individual speakers. Even though the gradient weight patterns are overall robust (4a), the subtle effect of H_3 could be an artifact of an unrealistically large lexicon (4b).

One way to evaluate the weight effects in the lexicon modelled in Garcia (2017b; Chapter 1) is to simulate native speakers' lexica. For example, we can generate smaller (i.e., more realistic) lexica and model stress in each resulting subset. After n simulations, we can then observe the distribution of H_3 effects across these subsets. If the vast majority of such smaller lexica still present weight effects that are consistent with those observed in the comprehensive lexicon in Garcia (2017b; Chapter 1), then we have a more reliable lexical baseline that approximates a possible lexicon of an adult native speaker.

I will assume a realistic lexicon size of 10,000 words (non-verbs), which represents approximately 6% of the entire Houaiss dictionary. This is in fact a considerably more conservative number compared to the estimate in Nagy and Anderson (1984) of 45,000 words for an average English-speaking high school graduate. The authors arrive at this number by sampling from 88,533 words that included both verbs and non-verbs. By assuming a considerably smaller lexicon, I intentionally lower the chances of finding the same weight effects that we observe for the entire lexicon.[3] At the same time, if even 10,000-word lexica show a negative effect of H_3, then we can be confident that speakers' lexica are highly likely to have such an effect as well.

Another simulation presented in §2.4.1 consists of a subset of frequent words in the Portuguese Stress Lexicon. This simulation helps us approximate the learners' input with regard to weight effects in Portuguese, which in turn allows us to determine how likely it is that learners are exposed to the negative effect of H_3 when building their own lexicon.

In §2.4.1, I show the results for 10,000 simulated lexica, as well as the lexicon filtered by frequency alluded to above, both of which confirm the negative effect in question. The issue, thus, is whether speakers' grammars generalise such an unexpected weight effect (4b). In the following section, I describe the experimental design employed in the present study.

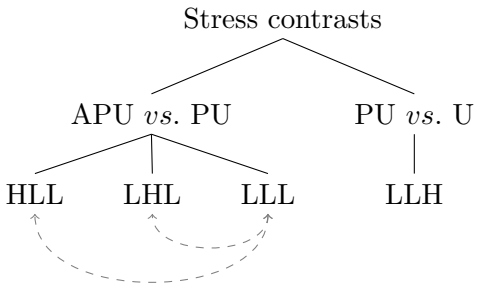
2.3.2 Experimental design

To examine speakers' behaviour, an auditory forced-choice task was designed using Praat (Boersma and Weenink 2020) in which native Portuguese speakers (n = 27 (Version A), n = 32 (Version B))

[3]Note that, unlike Nagy and Anderson (1984), the simulated sublexica in question do not contain verbs.

were presented with pairs of trisyllabic nonce words that differed only in the position of stress. No orthographic forms were provided in order to keep participants from considering alternative pronunciations on the basis of vowel quality, which could bias their stress preference.[4] Target pairs contrasted antepenultimate (APU) and penultimate stress (PU). Final (or ultimate) stress (U) was also included (PU *vs.* U) to verify whether speakers' judgments mirror the well-known robust effects of weight in word-final position. Participants were asked which word in each pair sounded more natural. They were also asked to judge how confident they were in their responses on a 6-point scale (1 = Not confident; 6 = Confident).

Figure 2.1: Stress patterns and weight profiles of stimuli



The weight profiles used in both Version A and Version B of the experiment are HLL, LHL and LLL (weight baseline) for APU *vs.* PU, and LLH for PU *vs.* U (see Fig. 2.1). All nonce words (n = 240) contained at most one heavy syllable, and were generated by an R script (R Core Team 2020, Garcia 2015). For each weight profile, approximately 200 nonce words were initially generated. These words were then ordered by their phonotactic naturalness on the basis of their bigram probabilities. The words with the highest phonotactic probabilities in each weight profile group were selected for the experiment. In addition, segmental quality was randomised to include a large set of phonotactic combinations. The syllabic shapes were constrained to C(C)V(C). As a result, all heavy syllables in the stimuli are either CVC or CCVC.

Fig. 2.1 graphically presents the different stress patterns and weight profiles of the stimuli used. Note that the questions of interest are (i) whether penultimate stress is preferred in LHL words *relative to* LLL words, and (ii) whether antepenultimate stress is *dispreferred* in HLL relative

[4]The vowel inventory of (Brazilian) Portuguese is provided in Appendix B.

to LLL words (hence the dashed arrows in Fig. 2.1). By examining (i) and (ii), we address the question in (4a), the extent to which speakers learn the weight gradience present in the lexicon; by examining (ii), we address the question in (4b), namely, whether speakers' grammars generalise or repair the typologically inconsistent weight effect of H_3. Finally, LLH words (PU *vs.* U) acted as 'controls', and thus allow us to confirm the well-known word-final weight effects in Portuguese. More importantly, we can also examine to what extent speakers' judgements for these words will be modulated by the fact that XX́H words are relatively common in the language (2).

2.3.2.1 Participants

All participants in the present study are native speakers of Brazilian Portuguese. Participants in Version A (n = 27) were tested in Montreal, Canada (n = 14), and in southern Brazil (n = 13). Those tested in Brazil had zero or very little exposure to a foreign language. Those tested in Canada had higher levels of proficiency in English and/or French. This difference in linguistic background, however, had no effect on the results of Version A. Participants in this version of the experiment were selected on the basis of their performance on a short lexical decision task run before the actual experiment.[5] A threshold of 80% accuracy was used, which reduced the original sample size from 51 to 27. This selection criterion considerably increases the reliability of the data (e.g., as a proxy for participant attention during the experiment).

Participants in Version B (i.e., the replication of Version A) were all tested in southern Brazil (n = 32). Like the participants in Version A who were tested in Brazil, none of these participants declared having fluency in any other language besides Portuguese at the time of the experiment. Importantly, Version B was not preceded by a lexical decision task: responses from all participants were analysed, which results in lower reliability relative to Version A.[6] Thus, we can be certain that if Version A results are also replicated in Version B, the effects of interest are indeed reliable. Information on the profile of the participants is provided in Table 2.1.

[5]The lexical decision task was part of a different experiment, and contained trisyllabic nonce words with different stress patterns.

[6]In Brazil, subjects cannot be compensated for their participation in experiments. As a result, their motivation is expected to be affected, which poses a problem for longer experiments.

Table 2.1: Participants in Version A and Version B

	Version A (n = 27)	Version B (n = 32)
Age	$\bar{x}$ = 30, s = 8.3	$\bar{x}$ = 26, s = 5.8
Gender	female = 15	female = 18

2.3.2.2 Stimuli

All the nonce words used in the experiment were preceded by a definite article: *o, a*, 'the' (MASC, FEM). This ensured that the stimuli would be unambiguously interpreted as nouns. To avoid typical utterance-final effects, all [article + nonce word] sequences were recorded in a carrier sentence by a female native speaker of Brazilian Portuguese with training in linguistics. This eliminates word-final lengthening and pitch falls, which could lead speakers to perceive light final syllables as heavy. The use of a carrier sentence also eliminates a list effect, which could result in similar problems. A template is provided in (5).

(5) **Stimulus template**

O/A [*palavra*] *também.* 'The [word] too'.

Each nonce word was recorded multiple times with different stress patterns (*as per* Fig. 2.1). The stimuli were then extracted from the carrier sentences (rectangle in (5)) and manually checked to ensure that no low-mid vowels were present, given that these vowels are only found in stressed position in standard Portuguese (Appendix B). For example, a nonce word such as *sostrole* was recorded as [ˈsos.tɾo.lɪ] and [sos.ˈtɾo.lɪ]. If /ɔ/ had been present in either version of the word in question, both stress *and* vowel quality would vary in these particular stimuli: [ˈsɔs.tɾo.lɪ] *vs.* [sos.ˈtɾɔ.lɪ]. By not having any low-mid vowels in the stimuli, stress was the only difference between both versions of each nonce word in the experiment—a complete list of the stimuli is provided in Appendix C.

2.3.3 Statistical analysis

In this section, I describe the statistical methods employed in the paper. In §2.4.1, where a lexical baseline is provided, the 10,000-word lexica are modelled using traditional logistic regressions. As mentioned above, this results in a distribution of estimates for H_3, i.e., one $\hat{\beta}$ for each simulation (n = 10,000). I also provide Bayesian estimates (see below) of credible parameter values (H_3) for the entire lexicon and for the word list containing only the most frequent words in the language (Tang 2012).[7] As we will see, all distributions of $\hat{\beta}_{H_3}$ are very similar.

The empirical data collected are analysed using Bayesian hierarchical logistic regressions. All models reported in §2.4.2 include by-speaker intercepts as well as random (weight) effects, and by-item random intercepts. Below, I provide a brief overview of Bayesian data analysis, given that Bayesian methods are not widely used in linguistic research, and differ considerably from traditional statistical analysis.

Bayesian data analysis

Before providing the motivation for Bayesian analysis, it is useful to briefly discuss two central concepts in traditional (i.e., Frequentist) statistics, namely, p values and confidence intervals. In Null Hypothesis Significance Testing (NHST), we are provided the probability (p value) of observing data that are at least as extreme as the data we observe given a parameter value (assuming that the null hypothesis is true). In other words, NHST provides the probability of the data given a specific statistic (e.g., a z value) for a particular parameter value θ. This is traditionally represented as $p(data|\theta)$. If $p(data|\theta)$ is above a certain threshold (e.g., α = 0.05), we fail to reject the null hypothesis (e.g., that $\theta = 0$).

NHST also provides confidence intervals, which are based on hypothetical future sampling: if a given experiment were repeated on several samples, the confidence interval would encompass the true population parameter x% of the time (where x is normally set to 90% or 95%). Confidence intervals are frequently misinterpreted as 'the probability that the *true* parameter value lies within

[7] I do not provide a Bayesian analysis for the simulated 10,000 lexica due to the highly demanding computation involved (each simulation takes approximately 14 minutes to run using parallel processing).

two values'. Importantly, confidence intervals are not a probability distributions (unlike credible intervals in Bayesian analysis reviewed below), and would be different for every sample. Let us now turn to Bayesian data analysis.

Bayesian reasoning is the re-allocation of credibility across possibilities (Kruschke 2015). The possibilities in question are parameter values in a given model of data. Re-allocation of credibility implies a previous state of knowledge which is updated as new evidence is observed. This previous state is referred to as *prior*. The new evidence, i.e., the data, is modelled through a distribution, which is referred to as *likelihood*. Finally, the actual re-allocation of credibility is the *posterior*. This relationship is mathematically expressed in Bayes' rule (6), where the posterior is the product of the prior and the likelihood divided (normalised) by the evidence for the data observed.

The prior, represented in (6) as $p(\theta)$, is a crucial component in Bayesian data analysis.[8] The intuition is as follows: if previous research has consistently shown an effect of a particular condition, we can incorporate this body of knowledge into our model by using informative priors. When priors are strongly informative, more data are needed for the posterior to be affected, i.e., shifted from what is expected *a priori*. This is intuitive to the extent that if a single study wishes to challenge an entire body of consistent previous work, it will require an immense amount of data to do so. On the other hand, priors can be made non-informative, in which case their effects on the posterior are negligible.

(6) **Bayes' rule**

$$p(\theta|data) = \frac{p(data|\theta)p(\theta)}{p(data)}$$

As we can see in (6), Bayesian data analysis provides the probability of a parameter value given the data, or $p(\theta|data)$, i.e., the posterior. Normally, this is in fact the question we are most interested in examining. In other words, assuming the data collected, what are the most credible parameter values? Instead of single estimates, a complete distribution of credible values is provided. The researcher can then specify a given credible interval, whose interpretation is straight-forward:

[8] See Gelman (2008) for responses to common criticisms of the subjectivity of priors in Bayesian analysis.

the values within that interval are more probable than the values outside that interval. Crucially, credible intervals in Bayesian estimation are probability distributions, unlike CIs, which means that parameter values that are located at the edges of the interval are less credible than parameter values in the centre of the interval given the data (assuming that the posterior distribution is unimodal).

In realistic applications, where n parameters need to be estimated, the posterior distribution cannot be analytically calculated, given that the parameter space is n-dimensional. For example, if we have eight parameters, each of which has 1,000 values, then the joint distribution has $1{,}000^8$ combinations of parameter values, which is too large a number to be computed. Instead, we approximate the distribution by randomly sampling several parameter values from it. To do that, we use sampling methods such as Markov Chain Monte Carlo (MCMC) or Hamiltonian Monte Carlo (HMC), which explore through different chains the possible values of a parameter (or combinations of parameters) in an n-dimensional space.[9] A typical distribution that results from a Monte Carlo simulation is demonstrated in Fig. 2.2.

Figure 2.2: Example of a hypothetical posterior distribution of θ

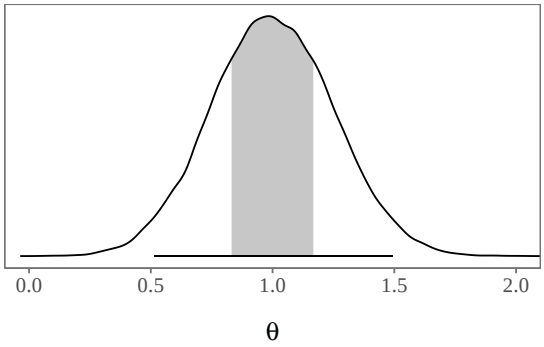


Fig. 2.2 illustrates the posterior distribution of a hypothetical parameter (θ). The mean of the distribution is $\theta = 1$. Because the distribution in question is practically normal, the mean is almost identical to the mode, and thus defines the most credible value in the distribution. Naturally, neighbouring values such as 0.99 are also highly credible. For that reason, examining a distribution

[9]For further information on sampling methods and Bayesian data analysis more generally, see Gelman et al. (2014a), Kruschke (2015) and McElreath (2016). For a general introduction to Bayesian data analysis as well as a comparison between Bayesian estimation and NHST, see Kruschke (2010, 2013).

of credible parameter values is more informative (and realistic) than considering a single estimate: clearly 0.99 is just as credible as the mean in Fig. 2.2. In other words, examining a distribution provides a more comprehensive understanding of the credible parameter values—and it also reminds the researcher that a categorical answer tends to oversimplify the analysis.

Underneath the distribution in Fig. 2.2, we find the 95% credible interval (CI), represented here with a horizontal line. The area in grey represents the 50% CI. By definition, parameter values within a given CI are more credible than parameter values outside of it. As a decision tool, we can establish that values that are located outside of the CI are rejected (Kruschke et al. 2012).[10] In this particular case, all parameter values within the 95% CI exclude zero, i.e., we conclude that $\theta > 0$—indeed, all values in the entire distribution exclude zero, given that the simulated range in question is [0.05, 2.02]. The posterior distribution for θ can therefore be reported as $\theta = 1$, 95% CI = [0.51, 1.49]. Throughout the paper, I will represent the posterior distribution of parameter values ($\hat{\beta}$s) through figures such as Fig. 2.2.

All Bayesian models reported in the present paper were diagnosed for chain convergence and Effective Sample Size (ESS; Kass et al. 1998; see Appendix A). As well, the Gelman-Rubin statistic (Brooks and Gelman 1998) was checked to ensure that between- and within-chain variance were the same. The Monte Carlo models were run using Stan (Carpenter et al. 2017) in R.

In summary, Bayesian methods allow us to examine the credible parameter values given the data, which is often a more meaningful and informative output than Frequentist estimates, p values and confidence intervals. The interpretation of posterior distributions is also more intuitive, in that the CIs provide the parameter values that are most consistent with the data modelled. In addition, CIs consist of actual probability distributions, unlike confidence intervals in NHST. As a result, Bayesian density intervals better estimate our uncertainty regarding parameter values given the data at hand: the wider the posterior distribution, the more uncertain we are about the parameter being modelled. Importantly, as will be shown below, the possibility of incorporating informative priors in a statistical model allows for more flexible analyses and simulations. Finally, a Bayesian framework

[10] Note that any cut-off value used as the CI is *arbitrary*. In other words, there is no special reason to choose 95% over 87%, just like there is no special reason to choose $\alpha = 0.05$ over $\alpha = 0.06$ in Frequentist approaches. For that reason, categorical decisions should be interpreted with care.

is also more intuitively translated into a probabilistic grammar, where constraints weights are learned or adjusted given the input (e.g., Boersma 1998, Goldwater and Johnson 2003, Hayes and Wilson 2008; see also Griffiths and Tenenbaum (2006) and Lee and Wagenmakers (2014) for Bayesian applications in cognitive science).

2.4 Data and analysis

The previous section briefly introduced Bayesian methods and their advantages over Frequentist statistics. In this section, I model the experimental data in this study using Bayesian logistic regressions and two different sets of priors, which correspond to two different assumptions regarding native speakers' linguistic knowledge. As will be shown, both Version A and Version B confirm that speakers generalise the gradient weight effects in Portuguese. Crucially, the results show that the negative effect of H_3 has not been learned. Rather, speakers' responses indicate a positive effect of H_3 in both experiments. Before examining these data, however, let us first inspect the lexicon simulations discussed in §2.3.1, which establish a baseline to which native speakers' responses can be compared—these simulations also allow us to assess the reliability of the weight effects in the Portuguese Stress Lexicon.

2.4.1 Lexicon simulation

Two simulations will be discussed. First, to approximate speakers' lexica, stress will be modelled in 10,000 sublexica containing 10,000 words each. As discussed in §2.3.1, this is a conservative estimate of a speakers' lexicon size. Second, to approximate learners' input, stress will be modelled in frequent words.

Approximating speakers' lexica

Each simulated lexicon was modelled using a binomial logistic regression, where the response was either APU or PU stress. In each model, the probability of APU stress was predicted in terms of the weight profile of each word (7). The effect of LHL (*vs.* LLL) is expected to disfavour antepenultimate stress (Garcia 2017b; Chapter 1). Indeed, LHL significantly disfavours antepenultimate stress

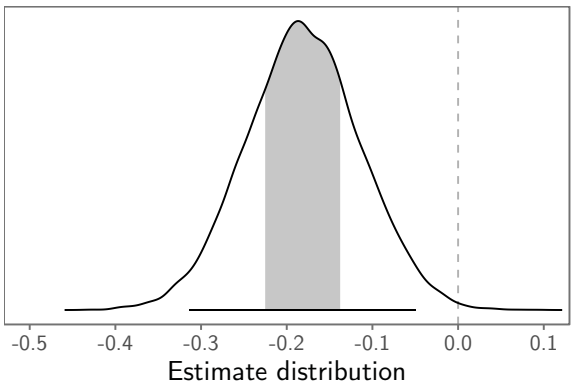
in all simulated lexica (Mean $\hat{\beta}$ = -12.36).

(7) **Simple Logistic Regression (β_0 = intercept = LLL)**

$$Pr(y_i = 1) = logit^{-1}(\beta_0 + HLL_i \cdot \beta_1 + LHL_i \cdot \beta_2)$$

The crucial weight comparison in the simulated lexica is HLL *vs.* LLL. Recall that, in the Portuguese lexicon, HLL words are *less* likely to have antepenultimate stress relative to LLL. In other words, the estimate of HLL is negative, which is consistent with observation 3 in (3).

Figure 2.3: Simple logistic regression $\hat{\beta}_{H_3}$ for 10,000 lexicon simulations (n = 10,000). Relative to LLL, HLL has a negative estimate in all simulations ($p < 0.0001$). The shaded area and horizontal line represent the 50% and 95% most frequent estimates, respectively



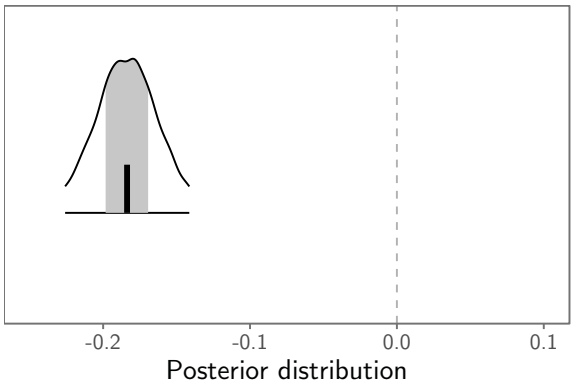
In Fig. 2.3, we can see a density plot of $\hat{\beta}_{H_3}$ effect sizes for all 10,000 simulated lexica. On the x-axis, we see a range of $\hat{\beta}$ values. The horizontal line at the bottom of the plot shows the interval that contains the 95% most frequent estimates; the shaded area represents the 50% most frequent estimates.[11] Note that the estimate of HLL is negative for nearly all simulated lexica. Considering the mean of the distribution in Fig. 2.3 (–0.18), $\hat{\beta}_{H_3}$ lowers the odds of antepenultimate stress by a factor of 1.2 ($= e^{|-0.18|}$).

If we now compare the distribution in Fig. 2.3 to the posterior distribution of $\hat{\beta}_{H_3}$ effects for the entire lexicon in Fig. 2.4, what we see is practically the same pattern (Mean $\hat{\beta}_{H_3}$ = -0.18), but

[11] Recall that these are not the CIs of a posterior distribution, given that the lexica were simulated using Frequentist regressions (§2.3.3).

a narrower distribution. The different widths in the distributions result mainly from the different samples of data used (several smaller lexica *vs.* a single large lexicon). As expected, when modelling the entire lexicon, we are more certain about the credible estimates of $\hat{\beta}_{H_3}$.

Figure 2.4: Posterior distribution of $\hat{\beta}_{H_3}$ for the entire Portuguese lexicon and associated 50% and 95% CIs



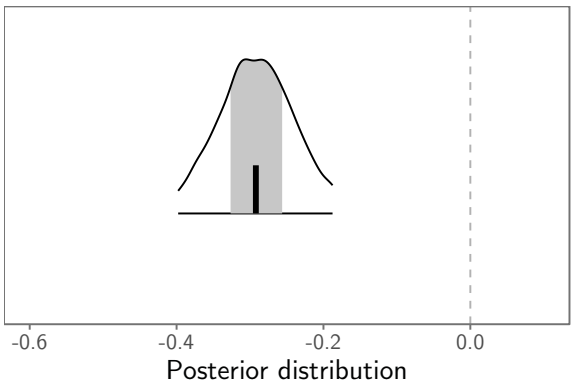
Approximating learners' input

We can also model only the most frequent words in the lexicon, in an attempt to approximate the input to which learners are exposed. To extract these words from the Portuguese lexicon, a frequency list was used (Tang 2012) as a filter, and the resulting frequency lexicon consisted of 22,634 words.[12] We can see in Fig. 2.5 that the negative effect of $\hat{\beta}_{H_3}$ is not only present, but is actually stronger (Mean $\hat{\beta}_{H_3}$ = −0.29) relative to the effect found in the entire lexicon or in the simulated sublexica discussed above.

In summary, whether we model (a) the entire lexicon or thousands of smaller and presumably more realistic lexica to approximate adults' lexica, or (b) only the most frequent words to approximate the input to learners, we find the same negative effect of $\hat{\beta}_{H_3}$. Therefore, the typologically inconsistent weight effect in antepenultimate syllables is very likely to be reliably detectable in Portuguese. In the remainder of the paper, I examine how native speakers deal with such an effect (question (4b)), which in turn will help determine whether (and to what extent) speakers acquire

[12] These were the words in the Portuguese Stress Lexicon that were also present in the frequency list in question.

Figure 2.5: Posterior distribution of $\hat{\beta}_{H_3}$ for the most frequent non-verbs in the Portuguese lexicon (based on Tang 2012), and associated 50% and 95% CIs



the weight gradience referred to in §2.2, the subject of question (4a).

2.4.2 Experimental data

In this section, I explore and model the empirical results from Version A and Version B. As previously mentioned, these data are modelled using Bayesian hierarchical logistic regressions with by-speaker random slopes for weight effects, as well as random intercepts; and by-item random intercepts.

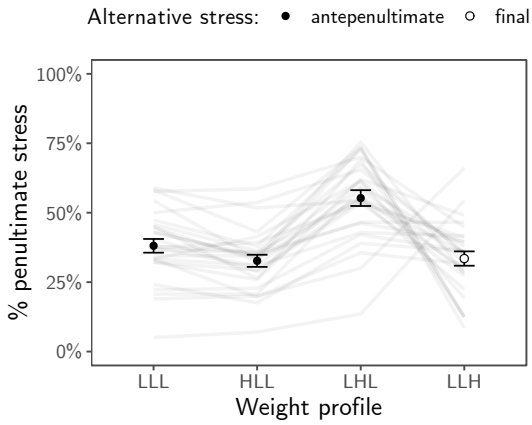
2.4.2.1 Version A

In Fig. 2.6, we can see the mean percentage of participants' preference for penultimate stress (and corresponding standard error bars) across the different weight profiles under consideration—grey lines represent the mean preference of each participant. As expected, speakers clearly favour final stress (over penultimate stress) in LLH words, confirming the well-known robust effect of H_1.

If we now turn to LHL *vs.* LLL, we can see that penultimate stress is favoured by the presence of H_2. This can be contrasted with the preference for antepenultimate stress in LLL words. Such a preference may be surprising given the literature on Portuguese stress, which often reports avoidance of antepenultimate stress (Hermans and Wetzels 2012; though see Araújo et al. (2011)).

It is possible that the preference for words with antepenultimate stress is associated with the more learned status of such words in the language. Given that these words are more commonly

Figure 2.6: Experimental results (Version A).
Mean response percentages by stress pattern and weight profile



found in the speech of more educated speakers, we could ask ourselves whether participants' preferences were biased by extralinguistic factors.

Even if novel words are associated with being more learned and, hence, with a preference for antepenultimate stress, the crucial question is whether antepenultimate stress is more frequently favoured in HLL words relative to LLL words. In both cases antepenultimate stress is favoured, but the data show a bias towards HLL words. If this is the case, then no extralinguistic explanation can account for such a difference: because the presence of H_3 is the only difference between HLL and LLL words, weight must be driving the stronger preference for antepenultimate stress in HLL words.

To model the data in question, mildly informative priors were used, as defined in (8). Because no previous experimental data exist regarding speakers' judgments of weight effects on stress in Portuguese, I assume that the regression coefficients for H_{1-3} are all normally distributed around ± 1, with a standard deviation of 1, and let the data obtained in the experiment determine what effects (i.e., parameter values) are more credible. These priors provide a less vague (expected) parameter space, and also constrain parameters to more realistic values, given the empirical data discussed thus far.[13] Finally, note that H_1 is emulated by β_0 in the intercept-only model in (8).

[13]Models were also run with priors normally distributed around zero, and the same results were found. In fact, slightly lower WAIC values were achieved by the models reported here (Widely Applicable Bayesian Information Criterion, Watanabe (2010)).

In such a model, the intercept (β_0) indirectly reveals the effect of a final heavy syllable on stress location, and can therefore emulate the effects of β_{H_1}

(8) **Mildly informative priors for βs** (* represents which stress is predicted in the model)

$$\text{MODEL: final* } vs. \text{ penultimate} = \beta^0(\text{LLH}) \sim \mathcal{N}(0, 10^6)$$

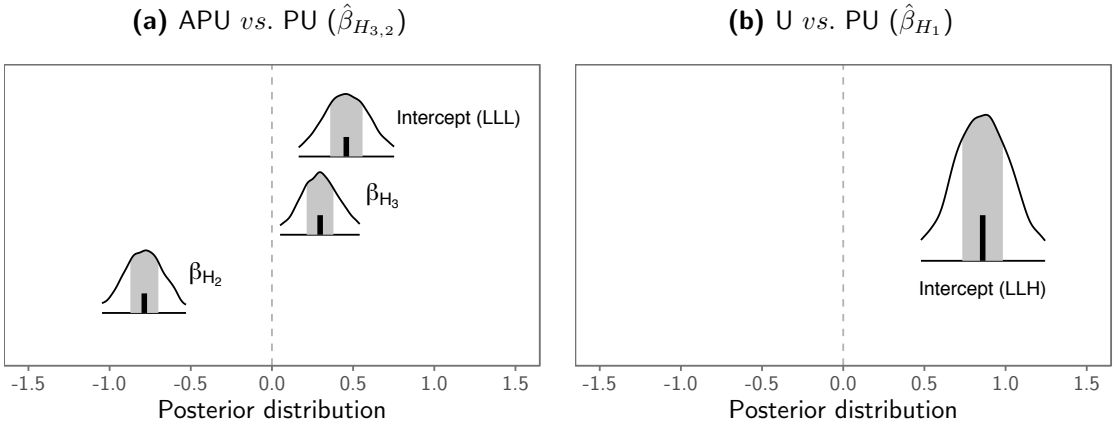
$$\text{MODEL: antepenultimate* } vs. \text{ penultimate} = \begin{cases} \beta_0 \sim \mathcal{N}(-1, 1) \\ \beta_{H_2} \sim \mathcal{N}(-1, 1) \\ \beta_{H_3} \sim \mathcal{N}(1, 1) \end{cases}$$

As shown by the posterior distributions of parameters $\hat{\beta}_{H_{1-3}}$ in Fig. 2.7, all weight effects are positive and statistically credible, i.e., all CIs exclude zero. The results show a clear gradient weight effect (i.e., $\hat{\beta}_{H_3} < \hat{\beta}_{H_2} < \hat{\beta}_{H_1}$).[14] Note that the distribution of $\hat{\beta}_{H_1}$ (Fig. 2.7b) is considerably wider when compared to $\hat{\beta}_{H_2}$ and $\hat{\beta}_{H_3}$. This is exactly what we would predict if speakers' judgments mirrored the lexical patterns involving penultimate and final stress in (L)LH words, where penultimate stress is relatively common in spite of its exceptional status (e.g., *jóvem* 'young', *nível* 'level', *fácil* 'easy')—as seen in (2). These results show that speakers capture the fact that the most robust weight effect in the domain (H_1) is also the most variable, as previously implied by (2), where XX́H words account for 11% of the lexicon.

In summary, the results from Version A clearly show that speakers are aware of the gradient weight patterns in the language. In other words, the weight of a heavy syllable depends on its position in the stress domain (i.e., $\hat{\beta}_{H_3} < \hat{\beta}_{H_2} < \hat{\beta}_{H_1}$). Importantly, $\hat{\beta}_{H_3}$ has a *positive* effect on stress, which is consistent with the fact that Portuguese is sensitive to weight. Indeed, these results show that regularities can (and do) emerge from (arguably) exceptional patterns such as antepenultimate stress. I conclude that even though speakers have a negative weight pattern in

[14] The weight effect in LLH words is not directly comparable to words with other weight profiles, given the stress options available to participants (Fig. 2.1). However, I treat H_1 as the strongest effect given its robust weight status in the literature, which has been used to motivate the generalisation in (1).

Figure 2.7: Posterior distributions with associated CIs for $\hat{\beta}_{H_{1-3}}$ in Version A



their lexica (approximated in §2.4.1 above), their grammars seem to override such a pattern in favour of a typologically consistent weight effect, to which I will briefly return below.

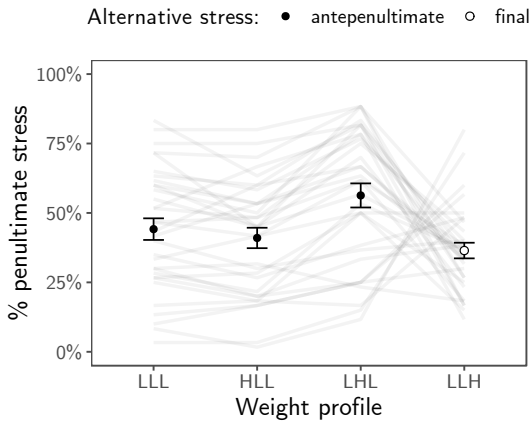
2.4.2.2 Version B

As mentioned in §2.3.3, the statistical methods employed in this paper provide a complete posterior distribution of credible parameter values given the data. Importantly, we obtain an intuitive interpretation regarding the level of uncertainty involved in the estimation of these parameter values (i.e., CIs). To test the reliability of the results discussed thus far, I now turn to a replication of the experiment presented above. The replication (Version B) includes the same experimental design and statistical analysis as Version A, but consists of a new sample of native speakers of Portuguese (n = 32), all of whom had had little or no exposure to foreign languages at the time of the experiment (cf. Version A). In addition, because exposure to a foreign language and level of instruction are correlated, this group of participants also had lower formal education.

Version B results are shown in Fig. 2.8. We can see that speakers' responses are very similar to the responses we observed in Version A (Fig. 2.6). As expected from the literature, LLH words favor final stress. Crucially, as in Version A, HLL words seem to favor antepenultimate stress, and LHL words clearly favor penultimate stress. Note that the standard errors from the mean in Fig. 2.8 are higher relative to Version A, which is likely due to the fact that the group of speakers in question was not pre-tested and then filtered on the basis of their accuracy on a lexical decision

task (as discussed in §2.3.2).

Figure 2.8: Experimental results (Version B).
Mean response percentages by stress pattern and weight profile



To model the data in Version B, the same mildly informative priors discussed above were used. In fact, we could use the posterior distributions in Version A as the priors for Version B, given that now we have at least some data on which to base our expectations. This, however, would not substantially affect the model presented here, given that the standard deviations in (8) are sufficiently wide to allow the posterior to be easily informed by the actual experimental data.

Figure 2.9: Posterior distributions with associated CIs for $\hat{\beta}_{H_{1-3}}$ in Version B

(a) APU *vs.* PU ($\hat{\beta}_{H_{3,2}}$) **(b)** U *vs.* PU ($\hat{\beta}_{H_1}$)

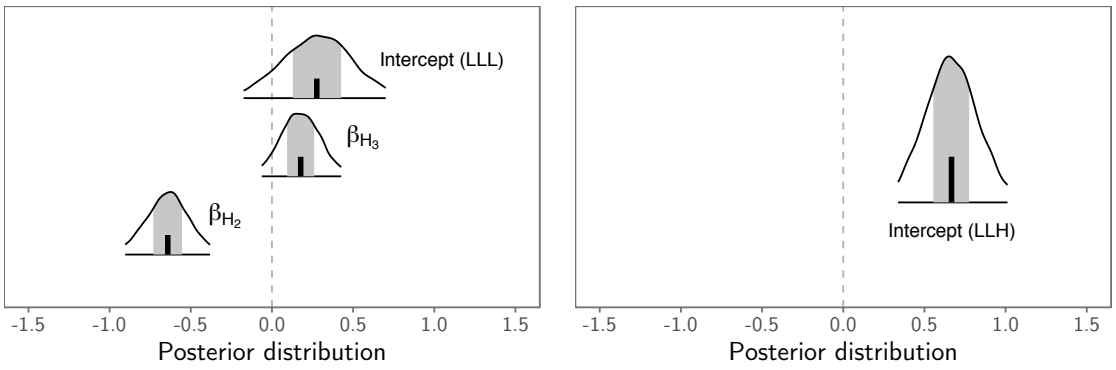


Fig. 2.9 provides the posterior distributions of all three weight estimates, namely, $\hat{\beta}_{H_{1-3}}$. As in Version A, the 50% CIs in all distributions in Version B exclude zero. The 95% CI for $\hat{\beta}_{H_3}$ is

almost entirely positive, which is consistent with the results found in Version A—recall that these estimates take into account by-speaker and by-item variation. In addition, the CI for $\hat{\beta}_{H_1}$ is again wider relative to $\hat{\beta}_{H_2}$ and $\hat{\beta}_{H_3}$, which mirrors the fact that LĹH words are indeed not uncommon in the language. As we can see, Version B replicates the same statistically credible weight effects observed in Version A.

In summary, the empirical results from both Version A and Version B address the questions in (4), namely, (a) the extent to which speakers learn the gradient weight-sensitivity present in the lexicon, and (b) whether speakers' grammars generalise or repair the typologically inconsistent weight effect of H_3. First, speakers clearly generalise the weight gradience in the language to novel words. Second, the typologically inconsistent weight effect of H_3 shows a *positive* effect on antepenultimate stress, unlike what we see in both the entire Portuguese lexicon and in the simulated smaller lexica discussed in §2.3.1.

Let us assume for a moment that speakers in Version B showed a null effect of H_3, whereby ĹLL and ĤLL words were statistically equally likely (i.e., the 50% CI would unquestionably include zero in Fig. 2.9a). In that case, the results for Version B would mirror what Becker et al. (2012) found for English laryngeal alternations, where polysyllables and monosyllables are treated equally by speakers in a wug test, even though alternations are more frequent among monosyllables in the English lexicon. We have seen, however, that speakers go beyond a null effect and learn the opposite pattern. Speakers' grammars therefore show a predictable pattern of generalisation, whereby stress is always positively and probabilistically affected by weight in the language.

Finally, recall that footing in Portuguese was discussed in §2.2.1. We entertained the possibility that the inconsistent effect of H_3 was only apparent, given that footing could be driving the effect in the lexicon (4b-ii). However, the results presented and discussed above show that, even if footing plays a role in the lexicon, that role is overridden by weight effects in speakers' grammars. Indeed, as will be discussed in Chapter 3, the status of the foot in Portuguese is uncertain, as no compelling direct evidence for it exists in the language.

2.5 Conclusion

In this paper, I have shown that the gradient weight effects present in the Portuguese lexicon are indeed acquired and generalised by native speakers. Such effects were previously unknown, and contribute to a more accurate understanding of weight and its effects on stress in the language. In that regard, the approach proposed in this paper implies a probabilistic representation of weight, in line with what is assumed in Garcia (2017b; Chapter 1).

This paper has also shown that even though speakers are able to capture subtle effects in their language, they do not generalise typologically inconsistent weight patterns (H_3). Instead, weight in antepenultimate syllables is generalised as one would predict if Portuguese is sensitive to weight in all positions in the stress domain. In other words, not only did speakers not generalise a contradictory pattern, they in fact learned the opposite pattern.

Weight and metrical structure

Thus far, this thesis has established two central points, namely, that weight effects on stress are gradient in the Portuguese lexicon, and that speakers' grammars capture this gradience when generalising stress patterns to novel words. Thus, the probabilistic approach proposed in Chapter 1 is consistent with native speakers' behaviour in Chapter 2, not only because weight effects are found in all syllables in the stress domain, but also because such effects monotonically weaken as we move away from the right edge of the word. Indeed, Chapters 1 and 2 argue for a probabilistic grammar that encodes weight distinctions which are considerably more intricate than the traditional binary distinction assumed in the literature.

Besides the overall gradient weight effects observed in the lexicon and in speakers' grammars, so-called exceptional patterns also make a strong argument for a probabilistic grammar which is sensitive to lexical subtleties. Let us briefly revisit two such patterns, starting with XX́H words such as *fácil* 'easy', *jóvem* 'young', *projétil* 'projectile'.

We saw in Chapter 1 that XX́H words are traditionally classified as irregular in Portuguese, given that a heavy final syllable should result in final stress. However, we also observed that these words are commonly found in the language, which indicates that the most robust weight effect in the lexicon (final stress on a heavy final syllable) is also the least strict. Indeed, aside from words with antepenultimate stress, XX́H words comprise the largest 'irregular' subset in the Portuguese lexicon (11%; Table 1.1).

If the sub-regularity involved in XX́H words is reflected in speakers' grammars, we predicted that XX́H words should be a highly probable exceptional pattern to be learned and generalised in the language. Chapter 2 tested this prediction, and demonstrated that speakers' behaviour is consistent with the sub-regularity in question. We observed this consistency in two experiments

by inspecting the posterior distributions of word-final weight effects (H_1), which are wider relative to penultimate and antepenultimate weight effects, and thus reflect what we would expect based on the lexical patterns in question. On the one hand, the positive posterior distributions of H_1 in Chapter 2 mirrored the robust word-final weight effects found in the lexicon; on the other hand, because penultimate stress is relatively common in such words, the level of certainty of word-final weight effects was lowered, which resulted in wider distributions of credible parameter values in both experiments examined in Chapter 2. The probabilistic grammar assumed in this thesis accommodates these nuanced effects, given that weight distinctions are not understood as binary, and weight effects are not assumed to be categorical.

The second so-called exceptional pattern in Portuguese concerns antepenultimate stress. We saw that weight has an effect on stress in antepenultimate syllables, which contradicts the traditional assumption that this particular stress pattern is unpredictable. In the Portuguese lexicon, these effects are negative, as shown in Chapter 1. In other words, ĹLL words are more frequent than ĤLL words in the lexicon. Once we probe speakers' grammars, however, we find the opposite pattern: ĤLL words are statistically favoured over ĹLL words.

As discussed in Chapter 2, the negative effect in the Portuguese lexicon could be due to footing optimisation, given that ĤLL words yield more marked metrical configurations relative to ĹLL words. Indeed, this could indicate that footing in Portuguese constrains the effects of weight in order to avoid marked metrical structures. However, as shown in Chapter 2, speakers' grammars do not generalise the negative antepenultimate weight effects. Instead, speakers favour stress on heavy syllables across the entire stress domain in Portuguese. As a result, if footing can explain the negative weight effects found in the lexicon, it cannot account for the effects observed in speakers' grammars.

As we have seen, previous studies of Portuguese stress have employed feet not only to delimit the stress domain in the language, but also to determine where stress should fall. Clearly, however, the observation that weight effects are detected in all three syllables in the stress domain pose major challenges to a foot-based approach, given that the typology of weight-sensitive feet cannot account for weight effects in antepenultimate syllables. As we will see in the next chapter, however,

Portuguese poses additional challenges for the foot. Specifically, Chapter 3 examines other types of evidence that motivate footing across languages (e.g., truncation), and argues that there are no compelling reasons to assume that Portuguese builds feet. The chapter thus proposes an analysis of stress where feet play no role. Instead, weight is the main predictor of stress location in the language. To demarcate the stress domain, Chapter 3 employs an alternative theory of stress, namely, Accent-First Theory (van der Hulst 2012).

If footing can be challenged in a language like Portuguese, where regular stress is seemingly captured by binary left-headed weight-sensitive feet, an important question is whether other languages with similar stress patterns display behaviour that does indeed motivate feet. One such language is English. Like Portuguese, stress patterns in English nouns and adjectives can be captured with moraic trochees. As will be shown, however, the evidence for the foot in English is robust. The weight effects found in the English lexicon and in speakers' behaviour as well as truncation patterns observed in the language are compatible with what we would predict given the footing traditionally assumed in the literature. By investigating a language where the foot is well motivated, the chapter aims to strengthen the argument that Portuguese does not build feet: whereas in Portuguese stress is determined by weight alone, in English it is determined by weight and regulated by footing.

Chapter 3

Stress without feet: a parametric distinction between Portuguese and English

ABSTRACT

This paper argues that even though English and Portuguese present similar stress patterns on the surface, these two languages are fundamentally different: whereas English builds feet, Portuguese does not. To support this argument, we focus on weight effects on stress. We show that weight effects in the English lexicon as well as in native speakers' behaviour are consistent with an analysis of stress that employs feet. In contrast, weight effects in the Portuguese lexicon and in native speakers' behaviour cannot be accounted for by a foot-based analysis. Further evidence for the foot in English comes from word-minimality constraints, which are never violated in the language, unlike in Portuguese. To constrain the stress domain in Portuguese to a three-syllable window, we discuss an alternative to footing, namely, Accent-First Theory (van der Hulst 2012). Finally, we discuss how weight can be formally represented in gradient, rather than categorical, terms.

Keywords: stress, weight, probabilistic grammar, English, Portuguese, Accent-First Theory

3.1 Introduction

Prosodic Phonology assumes that syllables are organised into feet, which correspond to one of the domains where prominence is realised (Selkirk 1984, Nespor and Vogel 1986). One of the central

motivations for feet cross-linguistically stems from the observation that languages systematically constrain the window of syllables in which stress can fall (i.e., the stress domain): indeed, in the vast majority of languages, stress falls within a trisyllabic window—either at the left or the right edge of the word (see Gordon (2016) for a comprehensive review). Besides delimiting the domain of stress, feet also formalise *where* within this domain stress is expected to fall. Different studies, however, have questioned whether the foot is universal. Examples include French (Jun and Fougeron 2000), where the domain of obligatory prominence is the phonological phrase rather than the PWd, and Turkish (Özçelik 2013, 2014), where regular stress is characterised as PWd-final but the cues to prominence are often absent (Levi 2005). Examples are shown in (1).

(1) **French (a) and Turkish (b) regular stress**

a. *la petit natión*; *nation-ál*, *nation-al-ité*
'the small nation'; 'national', 'nationality'

b. *tabák*, *tabak-lár*, *tabak-lar-ím*
'plate', 'plates', 'my plates'

In the present paper, we question the universal status of the foot by arguing that Portuguese has no feet. Unlike French and Turkish, however, which have systems of prominence that are unusual from the perspective of footing, Portuguese has a very similar stress system compared to English, a language for which the presence of the foot has not been questioned. In both languages, regular stress seemingly can be captured by binary left-headed weight-sensitive feet, i.e., moraic trochees, as shown in (2): in English nouns and adjectives, stress falls on the penultimate syllable if that syllable is heavy, and on the antepenultimate syllable otherwise. In Portuguese non-verbs, stress falls on the final syllable if that syllable is heavy, and on the penultimate syllable otherwise (e.g., Bisol 1992). Indeed, the crucial difference between these two systems seems to be extrametricality, given that the final syllable is extrametrical in English (e.g., Hayes 1982), but not in Portuguese.

(2) **English (a) and Portuguese (b) regular stress**

a. *a*(*gén*)⟨*da*⟩, (*Cána*)⟨*da*⟩

b. *jor*(*nál*), *sa*(*páto*)
'newspaper', 'shoe'

Even though English and Portuguese stress look quite similar on the surface, we will show that they are fundamentally different with regard to footing. The weight effects found in the English lexicon (§3.3.2) and in experimental data (§3.4) are as predicted if we assume that the language builds moraic trochees. In Portuguese, on the other hand, weight effects are not consistent with any foot type, and thus pose a major challenge to foot-based approaches Garcia 2019; Chapter 2). Furthermore, existing words in English are at least one binary foot in length—the same is true of productive phenomena such as truncation and hypocorisation, which never generate monomoraic words. In other words, English never violates word-minimality (§3.2.4), which is indirectly imposed by the Prosodic Hierarchy and the Foot Binarity condition (e.g., McCarthy and Prince 1995). In Portuguese, however, we commonly find existing words which violate word-minimality, i.e., which are monomoraic and, therefore, smaller than a (binary moraic) foot. Sub-minimal words are also found in productive phenomena such as hypocorisation. In summary, we will see that English and Portuguese motivate distinct formal systems that regulate lexical stress despite having similar rhythmic patterns on the surface.

If Portuguese, unlike English, lacks motivation for the foot, two important questions are (i) how the stress window can be constrained in the language, and (ii) what regulates the location of prominence inside this window. To examine question (i), we discuss an alternative to feet, namely, Accent-First Theory (van der Hulst 2012), which accounts for more cross-linguistically observed patterns of word-level prominence than foot-based approaches do. Throughout the paper, we assume a probabilistic approach to weight and stress, along the lines of Garcia (2017b; Chapter 1). As will be shown, such an approach is empirically better supported than categorical analyses, given the weight effects on stress observed in both English and Portuguese.

The paper is organised in four parts. First, we review stress and footing in Portuguese (§3.2) and English (§3.3). Second, we statistically model experimental results on English stress (§3.4) using Bayesian regressions, and show that these results are consistent with the lexical patterns

found in the language, which in turn motivate footing in English. Third, in §3.6.1, we discuss how the stress domain can be constrained in Portuguese even if one assumes that feet do not exist in the language. Finally, in §3.6.2, we propose a gradient representation of weight, and discuss how weight and other aspects of the grammar interact to account for the patterns explored in the paper.

3.2 Stress and footing in Portuguese

Primary stress in Portuguese is constrained by a trisyllabic window: *marítimo* 'maritime', *martélo* 'hammer', *papél* 'paper'. As a result, pre-antepenultimate stress is not found in the language (**máritimo*). Even though both verbs and non-verbs (nouns and adjectives) respect this trisyllabic window, stress in these two classes of words is driven by different factors: stress in verbs is heavily influenced by morphological factors (see Wetzels 2007 for a review), while stress in non-verbs relies mostly on phonological factors, namely, weight (Bisol 1992, Lee 2007, Wetzels 2007, Garcia 2017b; Chapter 1). In this paper, we focus on stress in non-verbs, which is typically assigned *as per* (3), where H stands for a heavy syllable, L for a light syllable, and X for any syllable (H or L).

Traditionally, all stress patterns that deviate from (3) are considered to be irregular. These include all words with antepenultimate stress (13% of non-verbs; Garcia 2014), regardless of the weight profile involved (ẊXX): *fósforo* 'match (n)', *pénalti* 'penalty', *júpiter* 'Jupiter', *marítimo*. Irregular cases also include words with penultimate stress which have a heavy final syllable (XẊH; *fácil* 'easy', *nível* 'level'; 11% of non-verbs), and words with final stress which have a light final syllable (XXĹ; *jacaré* 'alligator', *tatú* 'armadillo'; 3% of non-verbs).

(3) **Regular stress in Portuguese: XXH́ else XX́L**

Assign final stress if the final syllable is heavy (H):	*papél* 'paper', *rapáz* 'boy'
Else, assign penultimate stress:	*martélo* 'hammer', *varánda* 'veranda'

The rule in (3) entails a categorical approach, where patterns are either regular or irregular, and where syllables are either heavy or light, as described above. Approximately 72% of the Portuguese

lexicon can be accounted for by (3) (Garcia 2014), which makes it a relatively robust rule.[1] However, as shown in Chapter 1, a probabilistic approach is more accurate at predicting stress patterns: it is able to capture some of the so-called irregular patterns. Such a probabilistic approach assumes that stress patterns are more or less likely, and, crucially, that weight is positionally defined and gradient, not categorical. In other words, the interpretation of a heavy syllable is determined relative to factors such as where in the word the syllable is located, and also to how many segments are found in said syllable (see §3.6.2).

The probabilistic approach in Chapter 1 focuses on the Portuguese lexicon, but a subsequent study (Garcia 2019; Chapter 2) has shown that speakers' grammars also display a gradient weight effect, whereby final stress is more strongly affected by weight than penultimate stress, which is in turn more strongly affected by weight than antepenultimate stress. Such gradient effects raise the question of how weight interacts with metrical structure in Portuguese. Indeed, weight gradience poses important challenges for foot-based approaches to stress which assume a categorical notion of weight such as mora count (Hyman 1985, Hayes 1989). In the following section, we examine these challenges in Portuguese, and argue that the existence of feet in this language is questionable at best.

3.2.1 Footing in Portuguese

Analyses of stress in Portuguese have traditionally relied on (or made reference to) metrical feet (Bisol 1992, Collischonn 1994, Lee 2007, Magalhães 2008). As we will see, foot-based studies have often assumed different types of feet, as well as extra machinery to account for the various irregular patterns found in the language.

Moraic and syllabic trochees

Bisol (1992) proposes that regular stress in Portuguese requires both moraic and syllabic trochees: moraic trochees capture XXH́ words, which contain a heavy final syllable and bear final stress; syllabic trochees capture XX́L words, which contain a light final syllable and bear penultimate

[1] XXH́ = 14.7%; XX́L = 57.27%.

stress. Bisol also assumes that irregular patterns are accounted for by syllabic and moraic trochees. She proposes that the final syllable in words with antepenultimate stress is exceptionally marked as extrametrical. As a result, these words have the same foot structure as regular penultimate stress (i.e., they form syllabic trochees)—except for the extrametrical final syllable.

Exceptional extrametricality also plays a role in XX́H words, where penultimate stress is found in spite of a heavy final syllable. In such cases, a syllabic trochee is built and the word-final coda is analysed as extrametrical, thus making the final syllable light.

Final stress in words with a light final syllable is accounted for by assuming that an underlying catalectic consonant is present word-finally. Importantly, this catalectic consonant is not phonetically realised but bears a mora, thus making the final syllable heavy.[2] Consequently, a moraic trochee is built in such words.

There are two important issues that arise in the context of the foot-based approach discussed here. First, there is a binary distinction between regular and irregular patterns. Consequently, this approach predicts that stress in novel words will always be regular, since so-called irregular cases are, by definition, unpredictable. However, as shown in Chapter 2, speakers accept novel words as well-formed which depart from the regular patterns in the language.

The second important issue arising from the analysis in question is directly related to the present paper: weight-sensitivity is only assumed to impact stress in word-final syllables. This is connected to the generalisation in (3). Consequently, if weight only matters word-finally, then antepenultimate stress should be equally likely in XLL and XHL words. However, Chapters 1 and 2 have shown otherwise.

Trochees and iambs

In trying to account for word-final stress on light syllables, some analyses of Portuguese assume that both trochees and iambs exist in the language (Bonilha 2004, Lee 2007). Bonilha (2004), for example, assumes that iambs are built in words such as *urubú* 'vulture' and *abacaxí* 'pineapple'

[2]The evidence for a word-final catalectic consonant comes from derived forms where the consonant surfaces. For example, *jacaré* + *-inho* (DIM) → *jacare-z-ínho* (cf. **jacare-ínho*). An extra consonant, however, is also found in words with non-final stress: *rómbo* → *rombo-z-ínho* ~ *rombínho* 'leak'. Indeed, Bachrach and Wagner (2007) argue that the consonant in question is inserted as a result of hiatus resolution, rather than being connected to final stress in CV-final words.

due to the word-final vowels in question (/i,u/): high vowels in word-final open syllables virtually always attract stress. According to Bonilha, /i,u/, 'when positioned at the end of the prosodic word, are considered good peak elements' (p. 41). This assumption, however, is inconsistent with the fact that high vowels have the lowest sonority (Ladefoged and Johnson 2011, de Lacy 2006, p. 286).[3] Indeed, in (Brazilian) Portuguese, /e,o/ in unstressed final syllables are reduced to [i,u], respectively.[4] It is not clear why word-final position would enhance the sonority level of vowels which are themselves low in sonority. If final stress on light syllables were driven by sonority, then XXL words ending in /a/ should be the best candidates for final stress—but that is not the case.

Lee (2007) also assumes that both iambic and trochaic feet play a role in the grammar of Portuguese. Lee proposes an optimality-theoretic (Prince and Smolensky 1993) account, and takes advantage of the view that constraints that strive for outputs that conform to both foot types will be present in every grammar: FtForm = Iamb and FtForm = Trochee (Lee 2007, p. 129). Iambic feet are built in XLĹ words such as *jacaré* 'alligator': ja(caré). One issue with this analysis is that having two foot types in a single language overgenerates the possible parsings available. For example, a XLH́ word could be parsed as a trochee, i.e., XL(H́), or as an iamb, i.e., X(LH́).

Morphologically-conditioned stress in non-verbs

Some analyses of Portuguese stress assume that theme vowels (TV) play a role in stress assignment (e.g., Pereira 1999, Lee 2007, Pereira 2007). TVs are always unstressed, and consist of {a, e, o} in Portuguese—in a word such as *gát-o* 'cat', the TV ('o') indicates gender (masc). In these analyses, stress is assumed to fall on the last vowel of the stem: *gát*]-*o*, *jornál*] 'newspaper', *café*], 'coffee'. As a result, feet are not necessary to account for regular stress patterns—which now include XXĹ in addition to XX́L words: if theme vowels are never stressed, then the final vowel in *café* cannot be thematic, and is therefore part of the stem. Naturally, the problem is that we can only know if a given vowel is thematic if it is not stressed. As the following pairs of words show, {a, e, o} can be thematic or not, depending on where stress falls: *fóm-e* 'hunger' *vs. café*; *cóbr-a* 'snake' *vs. sofá* 'sofa'; *gát-o vs. robó* 'robot' (see critique in Chapter 1). In addition, feet are presumably needed

[3] High vowels have the greatest degree of constriction in the oral cavity, given that their articulation requires that the tongue body be significantly raised towards the hard palate.

[4] The distribution of vowels in (Brazilian) Portuguese can be found in Appendix B.

in words with antepenultimate stress.

As we can see, foot-based approaches to stress in Portuguese are not consistent, and propose not only different types of trochees (Bisol 1992), but also different types of feet (Wetzels 1992, Bonilha 2004, Lee 2007). These inconsistencies across proposals stem from conflicting patterns in the language, which are more intricate than what is typically assumed. Indeed, this is what we observe once we examine the weight effects in the Portuguese lexicon, as we will see in the next section.

3.2.2 Gradient weight in the Portuguese lexicon

Chapter 1 proposes a probabilistic weight-based approach to Portuguese stress in non-verbs. The empirical motivation for such an approach lies in the lexical patterns found in Portuguese: weight-sensitivity affects *all* three syllables in the stress domain, and effects monotonically weaken as we move away from the right edge of the word. Indeed, weight effects are not only sensitive to the position of a given syllable, but also to the number of segments found in said syllable. For example, triphthongs are heavier than diphthongs, which are heavier than monophthongs. In other words, weight is gradient across and within syllables.

The approach in Chapter 1 entails that words are assigned stress probabilistically, and that stress remains stored in the lexicon once assigned. In other words, words are learned *with* stress. Naturally, different words will have different probabilities of bearing a particular stress pattern. In some cases, this probability will be virtually categorical, i.e., nearly 0 or 1, depending on the syllabic constituents in the word.

The weight gradience in the Portuguese Lexicon poses a major challenge to metrical approaches to stress in the language, which assume a categorical weight distinction. It also contradicts approaches which question or reject the effects of weight in the language (e.g., Lee 1994, Cantoni 2013). As a result, sub-regularities cannot be captured, given that irregular cases are considered to be unpredictable as a whole. Indeed, no foot type seems to accurately account for the patterns observed in the language. It seems, thus, that the weight gradience found in Portuguese requires

a set of rules or constraints which can modulate the likelihood of stress patterns *beyond* footing, given that weight is the crucial factor affecting the location of stress. Probabilistic frameworks such as MaxEnt (Goldwater and Johnson 2003, Wilson 2006, Hayes and Wilson 2008) could achieve the desired result by assuming weighted constraints in the grammar. The question, then, would be how to formally represent weight itself in a gradient manner (see §3.6.2).

Possible evidence for footing?

We have seen above that the gradient weight effects on stress in the Portuguese lexicon challenge the existence of feet in the language. Yet, it is possible to conceive that the language displays other evidence that motivates footing. One type of evidence may lie in the weight effects of antepenultimate syllables.

One important finding about the weight effects in the Portuguese lexicon is that the effect of a heavy syllable is not always positive: while final/penultimate stress is statistically more likely to occur in LLH/LHL words, respectively, antepenultimate stress is statistically more likely to occur in LLL words than in HLL words (Garcia 2017b; Chapter 1). In other words, weight effects in this particular position are *negative*, not positive. This could be interpreted as evidence for footing: if moraic trochees are built in the language, and if final syllables are extrametrical in words with antepenultimate stress (e.g., Bisol 1992), then, in a HLL word, a syllable would be left unparsed in the middle of the word (4a), which is cross-linguistically marked. Alternatively, uneven trochees could be built for HLL words (4b). These, however, are also more marked cross-linguistically relative to even trochees (Prince 1990): (L)L ≻ (HL)⟨L⟩. In contrast, (L)⟨L⟩ would leave no syllables unparsed, which could explain why LLL words are more frequent than HLL words: the latter are simply more marked with regard to metrical structure.

(4) **Antepenultimate weight effects result in more marked metrical structures**

LLL ≻ HLL due to metrical optimisation:

a. (H)L⟨L⟩ → Unparsed syllable

b. (HL)⟨L⟩ → Uneven trochee

As we will see in the next section, however, a recent study (Garcia 2019; Chapter 2) showed that speakers do not generalise this negative effect to novel words. Instead, they 'repair' the pattern, and generalise a *positive* weight effect, thus favouring ĹLL words over ĹLL words in a judgement task. Indeed, all three syllables in the stress domain exhibit a positive weight effect. We can conclude from this that, even though the Portuguese lexicon may present some evidence for vestigial feet, the current grammar of the language does not.

3.2.3 The productivity of weight patterns

The negative effect of antepenultimate heavy syllables found in Chapter 1 seems to contradict the typology of weight and stress. After all, in weight-sensitive languages, heavy syllables should not repel stress (Gordon 2006; but see Hayes (1995, p. 146) on trochaic shortening in languages such as Fijian, where Ĺ syllables are realised as L). The question, thus, is whether speakers' grammars generalise such a negative effect to novel words. Recent studies on phonological learning have shown that unnatural patterns are either harder to acquire or not acquired at all by speakers (Hayes and Londe 2006, Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012). The negative weight effect in question could also be a case where a given lexical pattern is not acquired by native speakers.

To investigate this question, Garcia 2019; Chapter 2) conducted an experiment involving nonce words in Portuguese. The results of the experiment, however, revealed that weight effects were *positive* and gradient across all syllables in the stress domain. Heavy syllables overall attract stress, regardless of the position they occupy in the stress domain. Therefore, speakers not only do not generalise the negative weight effect in the Portuguese lexicon, but they in fact *repair* such an effect, and learn the opposite pattern: heavy antepenultimate syllables favour antepenultimate stress (relative to light antepenultimate syllables). This pattern is exactly what one would expect from a weight-sensitive language where footing has no effect on stress.

3.2.4 Beyond stress: additional challenges

Thus far, we have seen that the weight effects found in speakers' grammars pose a major challenge to a foot-based approach to stress in Portuguese. However, stress is not the only aspect of Portuguese

that calls into question the status of the foot in the language. In this section, we show that additional challenges arise once we examine violations to word-minimality, hypocorisation, and truncation patterns.

Violations of word-minimality

We have seen that feet play a central role in metrical approaches to stress in Portuguese (whether they are trochaic or both trochaic and iambic in shape, as in Lee (2007)). Given that lexical words are prosodic words, that prosodic words must contain at least one foot, and that feet strive to be binary to be well-formed (McCarthy and Prince 1995), one would expect lexical words to minimally contain *two* syllables or *two* moras—a condition known as *word-minimality*. In English, for example, all lexical words respect word-minimality: bimoraic monosyllables such as *bee* ['biː] are common, but monomoraic monosyllables are not attested (e.g., *['bɪ]).

In Portuguese, however, sub-minimal prosodic words are not rare. In fact, the monosyllabic words listed in Table 3.1 are quite common in Portuguese, as can be inferred from their meanings. This poses a problem for metrical approaches to Portuguese stress, since prosodic words exist which do not dominate a binary foot (either syllabic or moraic).

Table 3.1: Common CV words (nouns and adjectives) in Portuguese

Word	**Gloss**	**Word**	**Gloss**
chá	'tea'	*pá*	'shovel'
dó	'pity'	*pé*	'foot'
fé	'faith'	*pó*	'dust'
má	'bad (f)'	*nú*	'nude'
nó	'knot'	*só*	'lonely'

Hypocoristics

That (C)CV words exist in the Portuguese lexicon may not necessarily be a substantial problem for metrical analyses which aim to build a grammar of Portuguese that employs feet. After all, it is possible that sub-minimality is not productive, and is therefore restricted to a handful of lexical items for etymological reasons, namely, the loss of final codas. Although this explanation could hold for some of the words (e.g., *só* from *sōlus* (lat.) 'lonely'), it predicts that speakers of Portuguese should not generalise the CV pattern to novel words. This, however, is not the case, as we can see from the formation of hypocoristics in Portuguese.

Gonçalves (2004) argues that the melodic material in names is mapped to a moraic trochaic 'template' to form the hypocoristic. Indeed, this seems to be the case for several hypocoristics (e.g., (a) in Table 3.2). However, outputs such as those in Table 3.2 (b), where hypocorisation results in a sub-minimal word, are very common in the language.[5]

Table 3.2: Hypocoristics in Portuguese

(a)	Name	Hypocoristic	(b)	Name	Hypocoristic
	Fabiána	Fábi		Felícia	Fé
	Isabél	Bél		Guilhérme	Guí (['gi])
	Rafaél	Ráfa		Luciána	Lú
	Robérto	Béto		Tiágo	Tí

In addition to the subminimal forms in Table 3.2 (b), other outputs that depart from bimoraic trochees are observed in hypocorisation in Portuguese. Specifically, when hypocoristics are reduplicated, both iambic and trochaic feet emerge: *Viviane* → *Vívi* ~ *Viví*; *Bibiana* → *Bíbi* ~ *Bibí*. However, there seems to be a preference for iambs, as all trochaic reduplicated hypocoristics can also be realised as iambs, but not the other way around: *Fátima* → *Fafá* but **Fáfa*; *Luciána, Luíza* → *Lulú* but **Lúlu*. In other words, trochees are more restricted than iambs in such cases. Note that this contradicts the main pattern observed elsewhere in Portuguese, where trochees are typically

[5]Note that some disyllabic hypocoristics can also alternate with monosyllabic forms: *Fabiána* → *Fábi* ~ *Fá*.

assumed to be the main or only foot type in the language.

Truncation patterns

The inconsistency of foot types in previous analyses of stress in Portuguese can be observed in truncation patterns, where both iambic (Table 3.3 (a)) and trochaic (Table 3.3 (b)) profiled outputs emerge in truncated forms.[6] In fact, minimal pairs can also be found: *professór* 'teacher' → *prófi*, but *profissionál* 'professional' → *profí*. Araújo (2002) proposes that the stress pattern in these truncated forms can be predicted by the position of secondary stress in the source word. In *pròfessór*, secondary stress is found word-initially, thus the truncated form bears penultimate stress: *prófi*. In this case, the resulting metrical structure corresponds to a trochaic foot. The stress pattern in the truncated form in *profìssionál*, on the other hand, is faithful to the secondary stress in the peninitial syllable of the source word: *profí*. In this case, the resulting metrical structure corresponds to[7] an iambic foot.

Table 3.3: Truncation patterns in Portuguese

	Word	**Truncated form**	**Footing**
(a)	*refrìgeránte* 'soda'	refrí	trochee → iamb
	dèpressa̋o 'depression'	depré	trochee → iamb
(b)	*cervéja* 'beer'	cé(r)va	trochee → trochee
	neuróse 'neurosis'	néura	trochee → trochee

It is important to note, however, that neither the pattern of primary nor secondary stress can accurately determine whether iambs or trochees will emerge in truncated forms. Indeed, some cases cannot be predicted at all: *dèpressa̋o* 'depression' → *depré* (cf. **dépre*). Here, Araújo (2002) argues that *depré* is a case of pseudo-truncation, since the source word cannot be unambiguously

[6] The word *profissionál* can also be truncated as *profíssa*, where a trochee is built.

[7] The impact of secondary stress on truncated forms seems to be robust in Portuguese, even though it cannot explain all patterns of truncation. One complication is that the location of secondary stress can vary. For example, in *profissionál*, as in *refrigeránte*, secondary stress varies: *pròfissionál* ~ *profìssionál*; *rèfrigeránte* ~ *refrìgeránte*. In this case, we need to assume *profissionál* as the source of *profí*.

determined. Unlike the cases above, where the source word is clear, in the case of *depré*, both *dépressão* and *dèprimída* 'depressed' can be the source word—but note that the resulting prosodic shape is not predictable from either possible source word in this case.

We can see that the inconsistency of foot types assumed in metrical approaches to stress in Portuguese can also be found in truncated forms in the language. In fact, this inconsistency may account for why footing is not actually discussed in Araújo (2002), or in Vilela et al. (2006), who provide an overview of previous studies on truncation in Portuguese.

To sum up, We have seen above that traditional metrical approaches to Portuguese stress assume inconsistent footing to account for the different patterns in the language. The same inconsistency is found in patterns of truncation, which result in iambs or trochees, depending on the word being analysed. Finally, sub-minimal words can be found in the Portuguese lexicon and, crucially, in hypocorisation, which indicates that derived words that violate word-minimality can be productive. These three observations entail that, if Portuguese does in fact build feet, the input to the language learner is far from unambiguous, and the metrical constraints that regulate footing must be excessively permissive. Indeed, this ambiguity in the language has led researchers to propose (i) moraic and syllabic trochees, (ii) trochees and iambs, as well as (iii) binary and degenerate feet. This has important implications for language acquisition, as learners attempting to construct a grammar for Portuguese will not be able to easily establish which foot type accurately characterises the language. In fact, Ferreira-Gonçalves (2010) discusses the metrical structure in Portuguese from the perspective of first language acquisition, and finds that words with both iambic and trochaic profiles are observed in children's productions, which is what we would predict if no particular foot shape emerges as optimal from the input to which learners are exposed.

No compelling evidence for footing in Portuguese

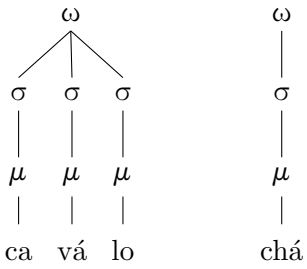
The discussion thus far has shown that no robust empirical evidence for a consistent metrical structure exists in Portuguese, which is reflected in previous studies on the phonology of the language. Crucially, we have seen that native speakers' grammars extend the gradient weight effects in Por-

tuguese to antepenultimate syllables, a finding which argues against the foot. As a result, if footing is what caused ĹLL words to be more frequent than ĤLL words in the Portuguese lexicon in the past, we have to conclude that this preference is not reflected in the synchronic grammar of the language.

In summary, no compelling evidence exists for footing in Portuguese. On the one hand, weight effects on stress make the existence of feet incompatible with the experimental patterns discussed above. On the other hand, other aspects of Portuguese where evidence for footing could be found offer strong evidence against it. We therefore conclude that Portuguese has no feet.

The idea that the foot does not exist in Portuguese is consistent with the hypothesis that the presence or absence of feet in the prosodic hierarchy is parametric (implied in McCarthy and Prince (1995) and assumed in Özçelik (2013, 2014)). Indeed, the conclusion that the foot is absent from Portuguese shows that even languages with seemingly ordinary patterns of prominence (i.e., unlike Turkish and French) cast doubt on a foot-based analysis of stress. Finally, since under a parametric account Portuguese has no feet, HEADEDNESS is vacuously satisfied at the foot level (Özçelik 2013, p. 55)—see Fig. 3.1.[8]

Figure 3.1: Prosodic structure of *caválo* 'horse' and *chá* 'tea' assuming FOOT = NO in Portuguese



The lack of feet in a language such as Portuguese raises the question of whether feet exist in other seemingly similar languages. One such language is English, where moraic trochees also appear to capture most stress patterns. However, as we will argue in the remainder of this paper, even though English and Portuguese have similar stress patterns on the surface, they are fundamentally

[8] We discuss how the window and location of stress can be constrained in a footless approach in §3.6.

different with regard to the formal system that regulates these patterns.

3.3 Stress and footing in English

As already mentioned, Portuguese and English share several characteristics with regard to stress. There has not been much dispute, however, as to which metrical patterns best characterise rhythm in the language (Chomsky and Halle 1968, Liberman and Prince 1977, Selkirk 1980, Hayes 1982, Halle and Vergnaud 1987b, Halle and Idsardi 1995, among others). In English nouns (and adjectives), which is our focus, stress falls on the penultimate syllable if that syllable is heavy, and on the antepenultimate syllable otherwise (5)—heavy penultimate syllables contain a coda consonant (*agénda*) or a long vowel (*Arizóna*). Final stress in nouns and adjectives tends to be avoided (Giegerich 2005, p. 185), but can be found in words ending in VVC syllables (Halle and Vergnaud 1987a)—see (5a).[9] Final stress can also be found in French borrowings, which tend to keep the source language's final stress pattern (e.g., *critíque*, *figuríne*, *souvenír*).

(5) **English stress in non-verbs**

a.	Final stress in VV(C)]$_\omega$ words	*políce, ballóon*
b.	Penultimate stress if the penultimate syllable is heavy: XX́X	*veránda, oppónent*
c.	Antepenultimate stress otherwise: X́LX	*Cánada, ártifice*

Given the behaviour of word-final syllables, extrametricality has played a central role in metrical accounts of English stress: Hayes (1982), for example, proposes that the final syllable in nouns is extrametrical, and is therefore 'invisible' during stress assignment: *a*(*gén*)⟨*da*⟩. For words with final stress such as those in (5a), Hayes proposes a rule (*Long Vowel Stressing*) which assigns a foot to such syllables. In summary, the stress algorithm in English non-verbs starts at the right edge of the word. If the final syllable contains a VV(C) rhyme, stress is likely final. Else, the penultimate syllable is checked. If a heavy syllable is found, stress is likely penultimate. Else, stress is antepenultimate. This algorithm can account for over 80% of the words in a subset of the

[9] Note that what counts as a heavy final syllable in this case is different from what counts as a heavy penultimate syllable.

CMU Dictionary (Weide 1993) containing 6,531 nouns and adjectives (excluding disyllables); see §3.3.2.

The generalisation discussed above implies that weight plays an important role in final and penultimate stress, but not in antepenultimate stress. After all, once the penultimate syllable in a noun is found to be light, the *default* stress is antepenultimate. As we will see in §3.3.2, this is one indication that weight effects on stress are regulated by footing in English.

3.3.1 Footing in English

Given that English is a weight-sensitive language, and that final stress is typically avoided in non-verbs, the foot type traditionally assumed in the language is the moraic trochee. This is the first relevant difference between Portuguese and English: there is wide agreement in the literature that moraic trochees, and not syllabic trochees (or iambs), capture stress in the language. Indeed, data from first language acquisition also point to trochees as the one consistent metrical pattern in children's productions (Kehoe 1998). Before we examine the productivity of stress patterns in English, however, let us first briefly look into other evidence for the foot in the language, namely, word-minimality constraints, truncation, and hypocorisation.

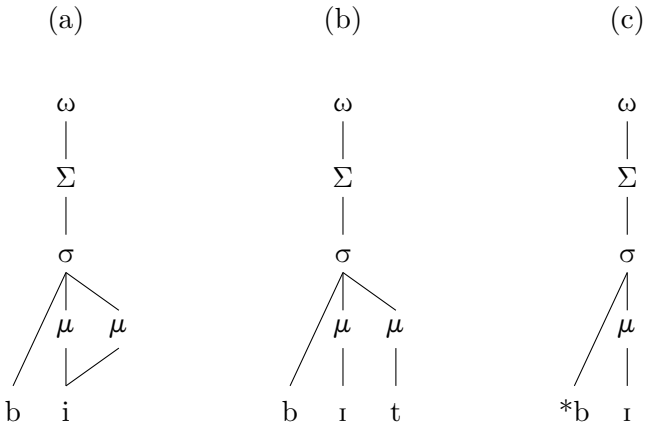
Word minimality, truncation, and hypocorisation in English

Unlike Portuguese, where we have lexical words that violate word-minimality constraints, English has no such words. Every lexical word in English must have at least two moras. This restriction bans CV words, but naturally allows CVV words. As a result, we can have *bee* ['biː] and *bit* ['bɪt], but not *['bɪ] (Fig. 3.2). These exceptionless word-minimality constraints in the language are nicely captured if English builds feet (moraic trochees), and if every foot is binary.

The word-minimality constraints mentioned above can also be observed in truncation patterns and hypocorisation in English. No truncation in the language results in a sub-minimal word: *bro(ther)*, *sis(ter)* and *doc(tor)* all have two moras: ['broʊ], ['sɪs], ['dɑk], respectively, but never *['brʌ], *['sɪ], *['dɑ].

The same is true for hypocoristics: *Nick*, *Bob*, *Sue* and *Joe*, but never *['nɪ], *['bɑ], *['sʊ] or

Figure 3.2: Word-minimality condition in English: lexical words require one foot, and feet require two moras



*['d͡ʒo]. These comparisons illustrate the interaction between vowel length, weight, and footing: feet in English need to be bimoraic to be licit, and this has consequences for the types of words we observe in the language.

The patterns discussed above motivate footing in English. Crucially, if we assume feet in the language, we predict the (near)[10] non-existence of sub-minimal words.[11] The question, however, is whether the interaction between stress and weight in English also supports a consistent metrical structure. This is what will be discussed next.

3.3.2 Lexical patterns in English

Recall that the generalisation discussed in (5c) implies that weight plays no role in antepenultimate stress in English: if a penultimate syllable is light, stress is antepenultimate regardless of the weight

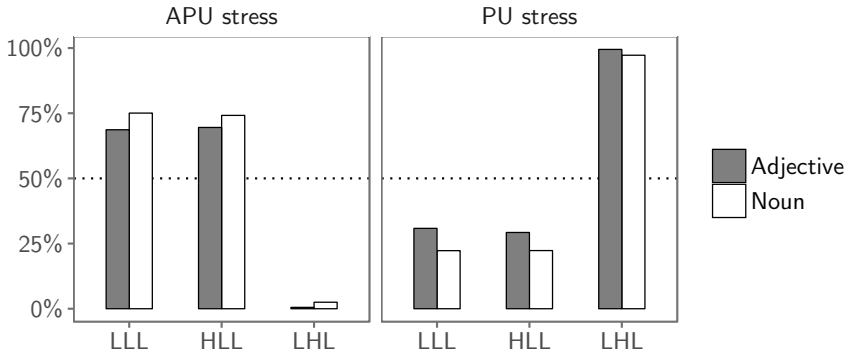
[10] CVCV words such as *city* may appear to be subminimal: (cí)⟨ty⟩. In such cases, building a binary foot conflicts with extrametricality: for a CVCV input, both (CV)⟨CV⟩ and (CVCV) are viable candidates. However, the existence of exceptional final stress indicates that extrametricality can be violated in English. In contrast, the fact that subminimal words are not attested in the language indicates that foot binarity must not be violated. In an optimality-theoretic account, where extrametricality is captured by NonFinality (No foot is final in ω; Kager (2011, p. 151)), these observations motivate the ranking Ft-Bin ≫ NonFinality (Goad 2016). As a result, the optimal candidate for *city* must be (CVCV).

[11] In some languages, feet are well-motivated even though sub-minimal words are attested. One example is Japanese, where word-minimality is violated by a handful of lexical items (e.g., *ya* 'arrow'; *ko* 'child'). Itô (1990) argues that word-minimality (bimoraicity) is enforced in the language as a *lexical* constraint (Kiparsky 1968), and therefore does not affect underived words. As a result, word-minimality is respected in truncated hypocoristics and shortened loanwords, unlike in Portuguese.

of the syllable itself. It seems, then, that unlike the gradient weight effects in Portuguese, weight does not play a role in all syllables in the stress domain in English. As we will see below, this is exactly what we find in the lexicon as well as in speakers' behaviour.

In this section, we use a subset of the CMU Pronouncing Dictionary (Weide 1993) to investigate how the generalisation in (5c) is reflected in the English lexicon. This will then provide a baseline for the experimental study presented in §3.4. The subset of CMU employed in this section is based on the filtered wordlist used in Moore-Cantwell (2016), a recent study which examined the English stress system in detail. However, to control for the possible conflicting effects of multiple heavy syllables, only words with one heavy syllable were selected (as coded in CMU).[12] Additionally, only trisyllabic nouns or adjectives were used, and words with final stress were removed. These conditions are in part motivated by the stimuli in the experimental study discussed in the next section. The resulting word list contained 4,573 words.

Figure 3.3: Stress and weight patterns in the CMU Dictionary ($n = 4,573$)



In Fig. 3.3, we can see that almost all LHL trisyllabic words plotted bear penultimate stress, a finding which is consistent with the generalisations in (5b). LLL and HLL words, on the other hand, have antepenultimate stress most of the time—note that these patterns are roughly the same for nouns and adjectives. Importantly, if we focus on antepenultimate stress, LLL and HLL words pattern together, which shows that having a heavy antepenultimate syllable does not impact antepenultimate stress. In other words, antepenultimate heavy syllables pattern with light syllables.

[12]The word list in question already codes word-final VC rhymes as light (e.g., *narcótic* is coded as HLL).

This is also consistent with the typical generalisations for English stress discussed above.

Like the absence of sub-minimal words discussed above, the patterns observed in Fig. 3.3 also motivate moraic trochees in English. If antepenultimate heavy syllables were indeed treated as heavy, as is the case in Portuguese (§3.2.2), antepenultimate stress would be more common in HLL words than in LLL words, which would lead to a more marked metrical structure (4).

Weight-sensitivity in final and penultimate syllables in English poses no problems for moraic trochees because no non-extrametrical syllables to the right of the stressed syllable remain unparsed in the stress domain. Weight in antepenultimate syllables, however, could lead to a marked metrical structure. In other words, the interaction between weight and footing predicts the presence of weight effects in final and penultimate syllables as well as a non-positive effect in antepenultimate syllables.

We have established that English, unlike Portuguese, offers compelling evidence for footing, summarised in Table 3.4. Crucially, weight effects on English stress are predicted if we assume moraic trochees in the language. Moraic trochees also predict why monomoraic words (i) do not exist in the English lexicon, and (ii) do not emerge in truncation or hypocorisation. The crucial question is to what extent such weight effects are actually generalised by native speakers of English. This is the topic of the next section.

Table 3.4: Weight effects and word-minimality in Portuguese and English

	Portuguese	**English (lexicon)**
Antepenultimate weight effects	*yes*	*no*
Violations of word-minimality		
Lexical words	*yes*	*no*
Hypocorisation	*yes*	*no*
Truncation	*yes*	*no*

3.4 Probing the productivity of weight patterns in English

Even though weight effects on English stress have been the focus of several studies, not many experiments have empirically probed such effects in native speakers' grammars. Guion et al. (2003), for example, examined weight effects by employing nonce words, but only bisyllabic words were used. As a result, nothing can be determined from this study about weight effects on antepenultimate stress. In a more recent study, Domahs et al. (2014) include longer words in a production task, which were orthographically presented to native speakers. Their results are overall consistent with the lexical patterns in the language, but not much is said about how weight affects antepenultimate stress. Furthermore, no minimal pairs for stress location were used, which creates an important confounding factor, namely, phonotactics: we cannot be certain that participants' preferences were guided solely by weight. It is possible that strings of segments with a lower phonotactic probability skew responses, for example. Indeed, several of the stimuli used in Domahs et al. (2014) are not phonotactically well-formed in English (e.g., *thimravas*, *posragols*, *domsanro*). As a result, some of the stimuli could have been analysed as compounds by speakers.

Most experimental studies examining weight effects in English employ orthographic stimuli. There are at least two reasons why this is a potential problem. First, the grapheme-phoneme correspondence is English is far from isomorphic. Second, the quality (and length) of unstressed vowels in English is very distinct from that of stressed vowels, which results in a considerably opaque orthography. As a result, if the same orthographic form yields different responses, it is not possible to know *a priori* if these different responses involve the same sequence of phonemes.

Instead of using orthographic forms, the present study employs an auditory judgement task involving minimal pairs for stress location. As we will see below, this task focuses specifically on the presence or absence of a heavy syllable in trisyllabic nonce words.

3.4.1 Methodology

In the present study, we developed a forced-choice judgement task which involved 180 trisyllabic nonce words. All words were generated by a script in R (R Core Team 2020), and were later

manually checked for phonotactic well-formedness. Stimuli containing attested but uncommon sequences were removed. Finally, because this is not a production task, vowel quality is constant for a given stimulus across participants.

The stimuli were developed on the basis of three main conditions: weight profile, coda type, and onset complexity, which has been shown to impact stress location (Davis 1988, Kelly 2004, Ryan 2014). Three weight profiles were used: LLL, HLL and LHL—LLL words served as the baseline. In HLL and LHL words, the coda in the heavy syllable included either an obstruent or a sonorant. To maximise naturalness, all final syllables in the stimuli were of the shape [CəC]—recall that this syllable profile is treated as extrametrical in nouns and adjectives in English, and therefore it patterns with a typical light syllable. As well, OCP effects were avoided by removing or adapting words which contained sequences of identical vowels or sequences of consonants which shared the same place of articulation. Finally, onset complexity was also varied. When a complex onset was present, it was located either in the antepenultimate syllable or in the penultimate syllable. Fig. 3.4 provides an overview of the experimental conditions involved, as well as the number of stimuli per condition.

The 180 stimuli (Appendix E) were recorded by a male native speaker of English with phonetic training. To ensure that vowel quality would remain as constant as possible, all stimuli were phonetically transcribed prior to recording: the set of antepenultimate and penultimate vowels consisted of /ɑ, ɪ, ɛ/, whereas final vowels were always schwas, as mentioned above. Each nonce word was then recorded with both antepenultimate and penultimate stress, which resulted in 180 minimal pairs that differed only in terms of stress location. Some examples are provided in Table 3.5.

Experiment

The forced-choice judgment task in the present study was developed using Praat (Boersma and Weenink 2020). Participants were auditorily presented with minimal pairs ($N = 180$), and were asked to choose which of the two pronunciations sounded more natural ('English-like') to them.

Figure 3.4: Experimental conditions. Stimuli ($N = 180$) were controlled for weight, coda type (obstruent or sonorant), and onset size (singleton or complex). Complex onsets are represented by 'CC' in the figure

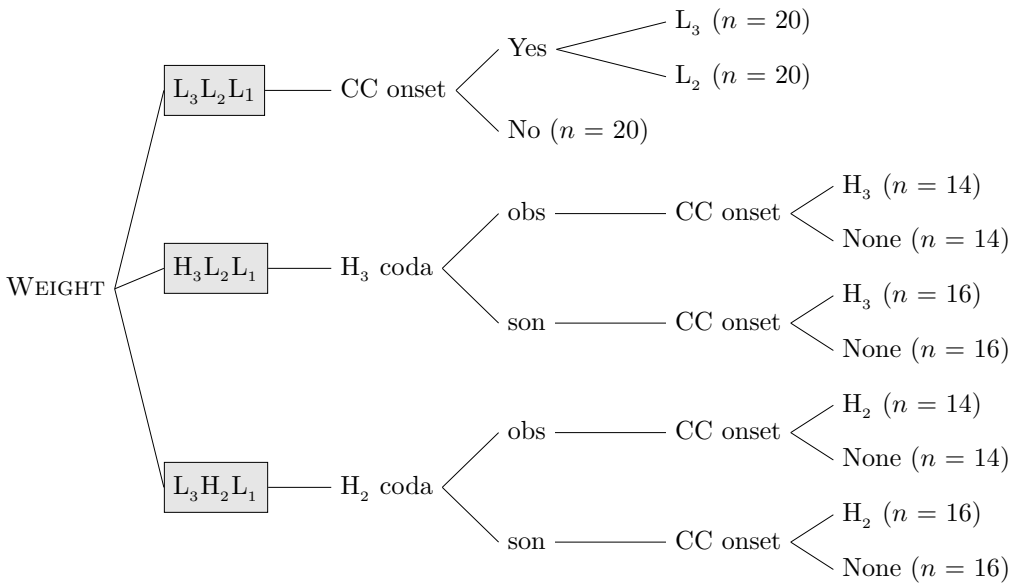
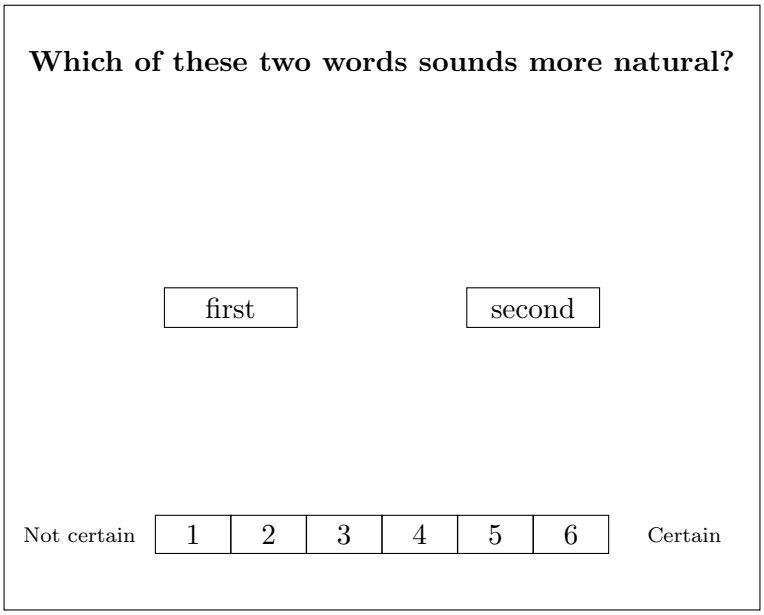


Table 3.5: Examples of stimuli used in the experiment

LLL	HLL	LHL
prɪ.tɑ.rək	nɑr.pɛ.lət	dɑ.sɛŋ.kəl
lɑ.prɛ.sən	prɑŋ.kɛ.mət	pɛ.trɑŋ.kəp
sɑ.pɪ.nər	krɪm.pɛ.dən	tɪ.prɛs.dəl

They were explicitly told that all the words were invented and represented objects, not actions or qualities. The stimuli were pseudo-randomised, as was the order in which the different stress patterns were presented. Participants were also asked to rate their level of certainty on a 6-point scale. This allowed them to modulate their otherwise binary responses. Finally, reaction times for each response were also recorded. Fig. 3.5 reproduces the screen participants saw during the experiment.

Figure 3.5: Experiment screen as presented to participants



Participants

The participants were native speakers of (North American) English living in Montreal (n = 13; 11 females). Most of them spoke other languages (e.g., French) at different proficiency levels, but none or them was bilingual from birth. Nearly all participants were students at McGill University: their level of education ranged from undergraduate to Master's/PhD, and their age ranged from 19 to 29. Overall, participants took 20-40 minutes to complete the experiment.

Predictions

If native speakers' grammars generalise the same weight effects observed in the English lexicon (Fig. 3.3), then we predict that heavy antepenultimate syllables will pattern with light antepenultimate syllables. In other words, we predict that weight will not play a detectable role in antepenultimate syllables. Heavy penultimate syllables, on the other hand, should strongly disfavour antepenultimate stress. Furthermore, we predict that more sonorous coda segments in penultimate syllables should pattern as heavier (Gordon 2006), and, as a result, should be more stress-attracting. Likewise, given the literature on onset effects on stress (e.g., Davis 1988, Topintzi 2010, Ryan 2011,

2014), we predict that onsets in penultimate syllables may also have an effect on speakers' preferences, insofar as syllables with onset clusters should be more stress-attracting.[13] Finally, we anticipate that speakers will be more certain and faster when choosing penultimate stress in LHL words, and antepenultimate stress in LLL and HLL words.

Given the consistent evidence for footing in English discussed thus far, we can assume that learners/native speakers have access to non-ambiguous input to build a grammar of stress, in contrast to learners/native speakers of Portuguese. As a result, participants are expected to generalise at least some of the robust weight effects in English without difficulty (e.g., penultimate stress in LHL words).

3.4.2 Results and analysis

Overall weight effects

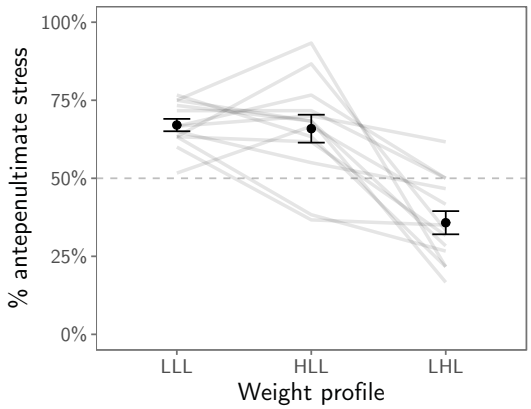
Let us start by analysing the main variable of interest, namely, weight. In Fig. 3.6, we can see the mean percentage of preference for antepenultimate stress across LLL, HLL and LHL words. Standard errors from the mean, as well as by-participant means (grey lines) are also provided. Preference for antepenultimate stress is above 50% for LLL and HLL words, but below 50% for LHL words. Crucially, even though their variance is distinct, LLL and HLL words do not differ in terms of their effect on speakers' stress preference.[14] This is exactly what we would predict given the lexical patterns shown in Fig. 3.3. Recall that this consistency between the lexicon and speakers' behaviour is remarkably different from what is found in Portuguese (§3.2), where speakers' behaviour goes against what is found in the lexicon vis-à-vis weight effects on antepenultimate stress.

To statistically model the effect of weight on speakers' responses, a Bayesian hierarchical logistic regression was run with by-speaker random slope (weight) and intercept, as well as a by-word random intercept. LLL was used as the reference level for weight, and is represented by the intercept of the model. As a result, we can interpret the effect size of HLL and LHL in relation

[13] Note, however, that because onset effects are predicted to be weaker than coda effects (Ryan 2011), capturing such effects requires more statistical power.

[14] It should be noted that LLL words (53%) are substantially more common in the English lexicon (§3.3.2) than HLL words (20%). This could explain why speakers' responses vary more in HLL words.

Figure 3.6: Overall weight effects on stress: percentages of preference for antepenultimate stress



to LLL (our baseline). The model was run using Stan (Carpenter et al. 2017) through the `brms` package (Bürkner 2016) in R.[15] As will be discussed below, the model's estimates confirm what is observed in Fig. 3.6, namely, that relative to LLL words, only LHL words affect speakers' responses, reducing the probability of preference for antepenultimate stress.

Table 3.6: Mean parameter estimates and associated standard deviations, credible intervals, $\hat{R}$, and Effective Sample Size (n_{eff})

Parameter	**Mean $\hat{\beta}$**	**SD**	**2.5%**	**50%**	**97.5%**	**$\hat{R}$**	**n_{eff}**
Intercept (LLL)	0.8	0.2	0.5	0.8	1.1	1.0	3600
HLL	0.0	0.3	-0.6	0.0	0.7	1.0	3600
LHL	-1.5	0.3	-2.1	-1.5	-0.9	1.0	3420
MODEL: `stress ~ weight + (weight | speaker) + (1 | word)`							

Table 3.6 lists the mean estimates (Mean $\hat{\beta}$) as well as the 50% and 95% credible intervals (CI) in the posterior distributions estimated. For example, the mean $\hat{\beta}$ for the intercept represents the mean of the posterior distribution for this parameter (in log-odds)—the positive intercept captures the preference for antepenultimate stress in LLL words, as observed in Fig. 3.6. The 95% CI in question goes from 0.5 (2.5%) to 1.1 (97.5%). $\hat{R}$ is used to inspect the convergence of the

[15] All models presented in this paper were run using four chains. For model specifications, see Appendix F.

model (an $\hat{R}$ of 1 indicates the model has converged). Finally, n_{eff} refers to the Effective Sample Size, i.e., the number of sampling steps assuming an uncorrelated chain. Fig. 3.7 plots the three posterior distributions in question. Unlike frequentist approaches, which provide a single parameter estimate and the probability of the data given such an estimate (given a null hypothesis; p-value), Bayesian approaches provide an entire distribution of credible parameter values given the data. Thus, even though the mean $\hat{\beta}$s in Table 3.6 are the most probable parameter values given the data, a distribution of parameter values can be inspected.[16]

As we can see, the mean estimate for HLL is zero. The 95% credible interval of the posterior distribution of HLL ranges from −0.6 to 0.7—i.e., it not only includes zero, but is centred around zero. Given that our reference level is LLL, this means that HLL and LLL words do not have a statistically different effect on participants' responses, i.e., there is no detectable difference between these two weight profiles given the data. On the other hand, the posterior distribution of LHL excludes zero, and has a mean of −1.5. This is the log-odds of antepenultimate stress given a LHL word (relative to a LLL word). Simply put, such a word lowers the odds of antepenultimate stress by a factor of 4.5.

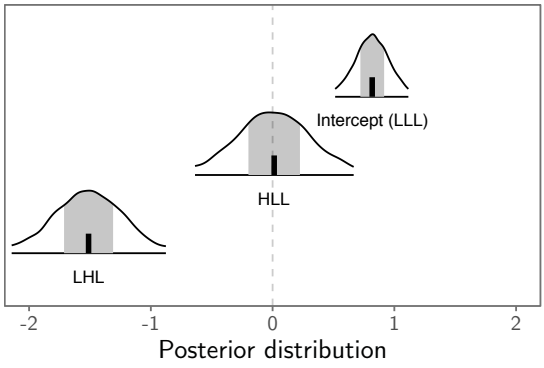
For the model in question, non-informative priors were used. We refer to this model as the 'naïve model', given that the model naïvely assumes that LLL, HLL and LHL words have the same probability of eliciting preference for antepenultimate stress. A model with mildly informative priors was also run to test whether the weight effects change or if the model itself has a better fit once we take into consideration speakers' (assumed) knowledge of English stress. In this model, the prior for the intercept was normally distributed around 1 with a standard deviation of 1—which can be represented as Intercept $\sim \mathcal{N}(1, 1)$. The prior distribution of HLL was assumed to be normally distributed around 0 (HLL $\sim \mathcal{N}(0, 1)$), and the prior distribution of LHL was assumed to be normally distributed around -1 (LHL $\sim \mathcal{N}(-1, 1)$). These priors are an attempt to approximate what we assume can characterise speakers' knowledge of English stress patterns, i.e., preference for antepenultimate stress in LLL words; weight effects in penultimate syllables; no weight effects in

[16] Note that credible intervals are not the same as confidence intervals, which are not a probability distribution, despite being frequently misinterpreted as such. For a comprehensive introduction to Bayesian data analysis, see Kruschke (2010) and McElreath (2016).

antepenultimate syllables.[17] Note, however, that all three prior distributions are sufficiently wide to be substantially affected by the experimental data being modelled if such data contradict the priors of the model. Both model specifications, with naïve and mildly informative priors, can be found in Appendices §F.1.1 and§F.1.2, respectively.

The model with mildly informative priors just described yielded a slightly lower WAIC,[18] compared to the naïve model, which indicates that a (slightly) better fit (2769.82 *vs.* 2770.96) was achieved by assuming some prior knowledge of English stress patterns. Another way to assess the fit of a model is to inspect the Leave-One-Out (LOO) cross-validation, which may be the preferred method to perform model comparisons according to Vehtari et al. (2017). This method also indicates that the model with mildly informative priors has a better fit than the naïve model (difference = 1.20, SE = 0.58).

Figure 3.7: Parameter estimates and associated posterior distributions: Mean (**|**), 50% (shaded grey) and 95% (solid line) credible intervals.



Participants' level of certainty and reaction times

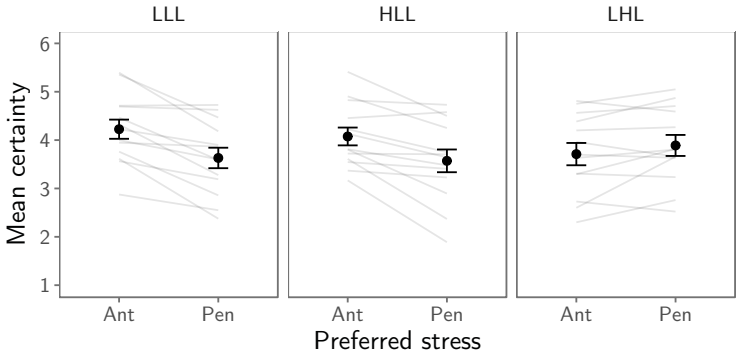
As previously mentioned, participants were also asked to rate their level of certainty for each response provided. As shown in Fig. 3.8, their levels of certainty mirror their response patterns: for

[17] Naturally, several other hypotheses (i.e., sets of prior values) could be entertained vis-à-vis speakers' knowledge. This paper considers one such hypothesis by using the mildly informative priors in question.

[18] WAIC, or *Watanabe-Akaike Information Criterion* (Watanabe 2010), is a method to assess the fit of a (Bayesian) model. It is calculated by taking averages of the log-likelihood over the posterior distribution taking into account individual data points. For more information on WAIC, see McElreath (2016, p. 191).

LLL and HLL words, participants' level of certainty is nearly identical: in both cases, antepenultimate stress yields higher certainty—each grey line represents the means for an individual subject. For LHL words, on the other hand, we observe a clearly different trend, as participants' certainty levels are slightly higher for words with penultimate stress. This effect is statistically credible, as confirmed in a (naïve) hierarchical ordinal regression, with by-speaker and by-word random intercepts. The model also added the interaction between stress and weight profile as a random effect.[19] The interaction of weight and stress in HLL words was not credibly different from the interaction of weight and stress in LLL words (Mean $\hat{\beta} = 0.21$, 95% CI = [-0.29, 0.7]). In contrast, the difference was credible between LHL and LLL words (Mean $\hat{\beta} = 1.12$, 95% CI = [0.65, 1.65].[20] Simply put, speakers' certainty is only affected by weight in penultimate syllables.

Figure 3.8: Participants' certainty on a 6-point scale by weight profile and preferred stress pattern



Similar results can be observed if we inspect participants' reaction times: whereas antepenultimate stress yields faster responses in LLL and HLL words, penultimate stress yields faster reaction times in LHL words. In other words, heavy antepenultimate syllables pattern with light antepenultimate syllables (Fig. 3.9).[21] A (naïve) mixed-effects linear regression was run with by-speaker and by-word random intercepts. As with the ordinal model discussed above, the interaction between

[19] `certainty ~ weight * stress + (weight * stress | speaker) + (1 | word)`.

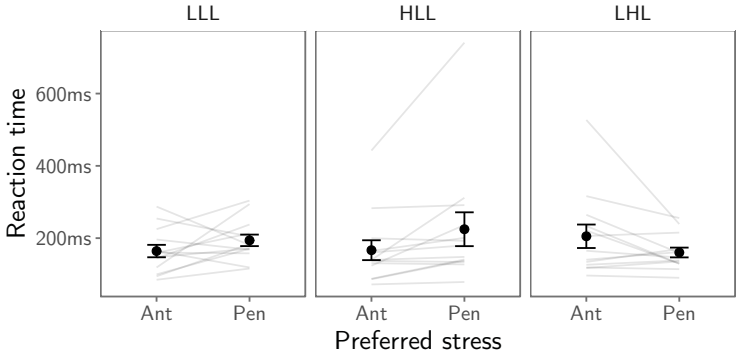
[20] The reference levels used were 'LLL' for `weight`, and 'antepenultimate' for `stress`.

[21] One of the subjects was particularly slow at responding (especially in HLL and LHL words), as shown by the grey lines in Fig. 3.9. The reaction time pattern in question, however, matches that of most participants, and this participant's response patterns were also consistent with the overall trends discussed thus far.

weight and stress was added to the model in question as a by-speaker random effect.[22] The interaction of weight and stress in HLL words was again not credibly different from the interaction of weight and stress in LLL words (Mean $\hat{\beta}$ = 0.11, 95% CI = [-0.16, 0.39]). In contrast, the difference was credible between LHL and LLL words (Mean $\hat{\beta}$ = -0.28, 95% CI = [-0.46, -0.10]).

Even though certainty levels and reaction times can be seen as secondary metrics in the present study, they are useful complements to the overall findings regarding weight and stress. The fact that both participants' certainty levels and reaction times are consistent with their response patterns increases the reliability of the results discussed thus far.

Figure 3.9: Participants' reaction times by weight profile and preferred stress pattern



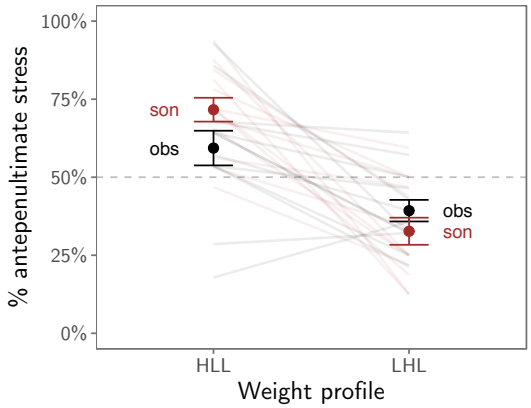
Sonority effects

Next, let us further examine weight effects by inspecting whether the quality of the coda in HLL and LHL words affected native speakers' stress preference. As predicted, more sonorous coda segments (i.e., sonorants) have a stronger effect on stress relative to less sonorous coda segments (i.e., obstruents)—especially in penultimate syllables, where weight effects are robustly observed. Fig. 3.10 plots the preference for antepenultimate stress (y-axis) by weight profile and coda type—means and standard errors are provided. As already seen in Fig. 3.6, antepenultimate stress is preferred in HLL words (> 50%), but dispreferred in LHL words (< 50%). Here, as predicted, we see that LHL words with sonorant codas disfavour antepenultimate stress more strongly relative to

[22] `log(reaction_time) ~ weight * stress + (weight * stress | speaker) + (1 | word)`.

LHL words with obstruent codas. Counter to our predictions, the same pattern is observed in HLL words, where sonorant antepenultimate codas favour antepenultimate stress more than obstruent codas. In other words, even though we do not observe an overall weight effect in antepenultimate heavy syllables (relative to light syllables), we observe a qualitative weight effect between sonorants and obstruents once we only examine HLL words.

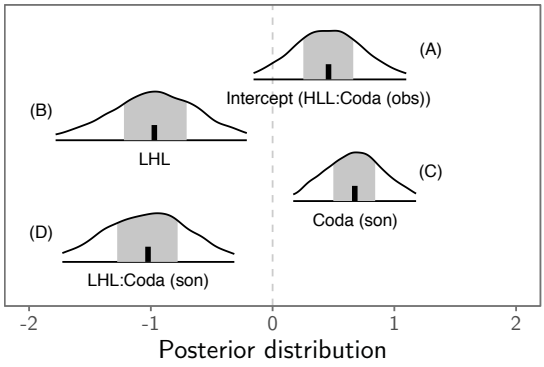
Figure 3.10: Effect of coda type on stress preference



To model the effect of coda type in HLL and LHL words, a (naïve) model was run with main effects as well as the interaction of weight and coda type. This additional model excludes LLL words, where no coda effect can be assessed. By-speaker random slopes for the interaction and random intercepts, as well as by-word random intercepts were added. The result is shown in Fig. 3.11. As we can see, the overall pattern still favours antepenultimate stress, as denoted by the almost entirely positive posterior distribution in (A), which represents the log-odds of antepenultimate stress in a HLL word with an obstruent consonant in coda position. The posterior distribution shown in (B) refers to LHL words with an obstruent coda. Unsurprisingly, the entire distribution is negative, given that LHL words disfavour antepenultimate stress. The distribution in (C) shows that having a sonorant coda in HLL words positively impacts the preference for antepenultimate stress (relative to an obstruent coda). This reflects the pattern we see in Fig. 3.10. Lastly, the entirely negative posterior distribution in (D) indicates that the interaction between weight and

coda type statistically impacts speakers' choices. However, this effect is clearly not robust enough to impact stress preference overall.

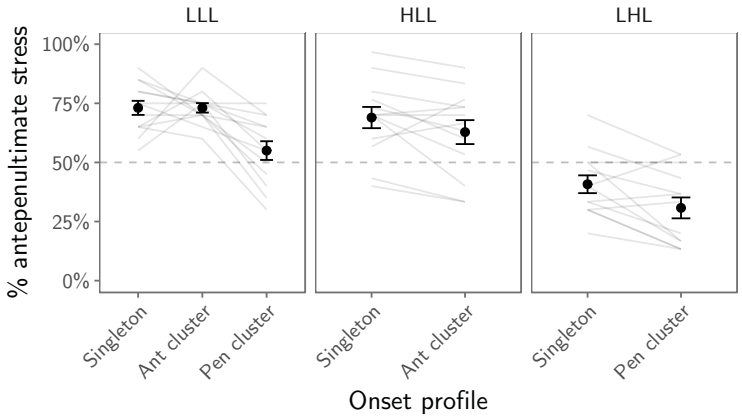
Figure 3.11: Parameter estimates and associated posterior distributions: Mean (**|**), 50% (shaded grey) and 95% (solid line) credible intervals. HLL and obstruent codas are used as reference (intercept)



Onset effects

Finally, let us briefly examine the effects of onset complexity on stress preference. Weight effects in English have been shown not to be restricted to syllable rhymes. Indeed, different studies have demonstrated that onsets also contribute to syllable weight (Davis 1988, Kelly 2004, Ryan 2014). In other words, syllables with more segments in onset position are more likely to attract stress, a tendency that is observed not only in the English lexicon (Ryan 2011, 2014), but also in experimental settings. The LLL words used in the present study are particularly useful here, given that no coda effects are possible and only onset complexity varies. In such words, three possible onset types were included in the stimuli: singleton in all syllables, complex onset in the antepenultimate syllable, and complex onset in the penultimate syllable.

Figure 3.12: Preference for antepenultimate stress as a function of onset and weight profiles. Complex onsets in penultimate position negatively affect speakers' preference for antepenultimate stress



As we can see in Fig. 3.12, onset complexity clearly has an effect on speakers' responses when located in penultimate syllables in LLL and LHL words, and in antepenultimate syllables in HLL words. Let us focus on LLL words, which overall favour antepenultimate stress—this is what we see in cases where no onset cluster is present ('Singleton') and in cases where an onset cluster is located in the antepenultimate syllable ('Ant cluster'). As predicted, having an onset cluster in the penultimate syllable, however, considerably affects speakers' responses, decreasing the preference for antepenultimate stress. This effect is statistically credible, as confirmed in a (naïve) hierarchical logistic regression with by-speaker random slope (onset complexity) and intercept, as well as a by-word random slope: Mean($\hat{\beta}$) = −0.91, 95% CI = [−1.68, −0.15].[23] Antepenultimate clusters do not show an effect relative to singleton antepenultimate onsets.

3.5 Discussion

The statistical models discussed in the previous section were not able to capture a difference between heavy and light antepenultimate syllables. This is consistent with the lexical patterns in Fig. 3.3. Crucially, these results also motivate moraic trochees in English. In other words, if English builds

[23]The regression in question modelled the effect of onset complexity on stress in LLL words, where all three levels of onset profile are can be compared.

moraic trochees, then weight effects should be constrained in antepenultimate syllables in order for all non-extrametrical syllables to be parsed, and that is what we empirically observe in the data analysed in the previous section. If footing constrains weight effects in English, then the observation that heavy syllables pattern as light in antepenultimate position is as expected.

Interestingly, whereas the weight of antepenultimate syllables does not impact the probability of antepenultimate stress, within HLL words, an effect of coda type was found. We can conclude that this effect, albeit present, is not strong enough to impact stress propensity once antepenultimate stress in LLL words is taken into account. This fine-grained weight effect can only be accounted for if we assume a gradient notion of weight: both heavy syllables with obstruent and sonorant codas are treated as light overall, but the former are 'lighter' than the latter—see Ryan (2016) for an overview of gradient weight.

We have seen that onsets also had an effect on participants' stress preference. When located in penultimate syllables, complex onsets were found to decrease speakers' preference for antepenultimate stress in LLL words. Unlike coda effects, however, no onset effects were captured in antepenultimate position. This is unsurprising given that onset effects are known to be weaker than coda effects. As a result, it is possible that onset effects are too weak to be statistically detectable in antepenultimate syllables, where weight effects are restricted by footing. In other words, it is expected that if some weight effect can be detected within antepenultimate syllables, codas are the most likely candidates.

The onset effects discussed above are consistent with previous studies, such as Kelly (2004), Topintzi (2010), Ryan (2011), Olejarczuk and Kapatsinski (2013), and Ryan (2014), who find that stress preference is sensitive to onset complexity. In such studies, word-initial onsets were found to affect stress in disyllables; i.e., to affect penultimate stress. These effects can only be captured if one assumes a gradient notion of weight.

In summary, we have argued thus far that whereas English provides strong evidence that motivates footing, Portuguese does not. Stress in English is the result of the interaction of footing and weight: heavy syllables attract stress as long as this does not result in a marked metrical struc-

ture, i.e., an unparsed syllable in the middle of the word or an uneven trochee. Otherwise, heavy syllables pattern with light syllables. Stress in Portuguese, on the other hand, does not require footing. Indeed, the gradient weight effects examined in Chapter 2 argue *against* the foot, as do violations of word-minimality, and the patterns of truncation and hypocorisation discussed above. As a result, if one attempts to pursue a foot-based approach, then the issues raised in this paper will present a significant challenge.

If feet are indeed parametric, as implied in McCarthy and Prince (1995) and proposed in Özçelik (2013, 2014), then Portuguese and English differ fundamentally in their prosodic structure. Indeed, once we assume that Portuguese has no foot projection, important questions arise. For example, since the domain of stress computation is typically defined metrically, we need to determine how the trisyllabic stress window is constrained in the language in the absence of feet. This is the topic of the next section, where we entertain an alternative approach to capturing stress in languages like Portuguese.

3.6 Constraining stress without feet

Even though main stress can logically fall on any given syllable in a word, the world's languages clearly favour positions closer to the edge of the word, a fact which facilitates word recognition (Cutler 2005, 2012). Indeed, Kager (2012) shows that stress is typically constrained to a trisyllabic window in bounded languages such as Portuguese—either at the left or right edge of the word. The relevant question in the present paper, however, is how to constrain stress in Portuguese without making use of the tools from existing foot typology, given that we have concluded above that no compelling evidence for the foot exists in the language. Below, we present an alternative to feet, namely, Accent-First Theory.

3.6.1 Accent-First Theory

Accent-First (AF) Theory, proposed by van der Hulst (2012), differs considerably from standard Metrical Theory (Liberman and Prince 1977, Hayes 1980, 1995, Idsardi 1992, 2009). In this theory,

stress[24] and rhythm are formalised as separate processes, which apply in different domains: lexical and post-lexical, respectively. Consequently, another important difference between AF and standard Metrical Theory is that, in AF, primary stress is not a result of rhythm. Rather, rhythm is only assigned once the location of primary stress has been determined.[25] Finally, unlike standard Metrical Theory, AF is neutral with respect to whether syllables are grouped into feet.

The main motivation behind the development of AF is the observation that no inventory of feet has ever been developed that is able to account for the typology of stress found across languages without additional machinery. As already discussed, the weight effects found in Portuguese, for example, necessarily require additional mechanisms, since moraic trochees and syllabic extrametricality (Bisol 1992, Lee 2007) alone cannot account for the interaction between weight and stress in antepenultimate syllables in the language.

AF relies on four word accent parameters, presented in Fig. 3.13, which are grouped into two categories, namely, **Domain** and **Accent**. Within **Domain**, Bounded determines which edge of the word is targeted by stress (i.e., L-eft or R-ight). In some languages, the stress domain is located at the left edge of the word (e.g., initial syllable in Czech, second syllable in Dakota). In other languages, stress targets the right edge of the word (e.g., final stress in Romani, penultimate in Polish (Goedemans and van der Hulst 2009, Dryer and Haspelmath 2013)). In AF, Bounded defines the stress domain. Unlike feet, however, Bounded does not always circumscribe the domain into which stress will fall (see below on Satellite). If Bounded-R, for example, a bisyllabic domain (demarcated by curly brackets) is formed on the right edge of the word: ...σ{(σσ)}].

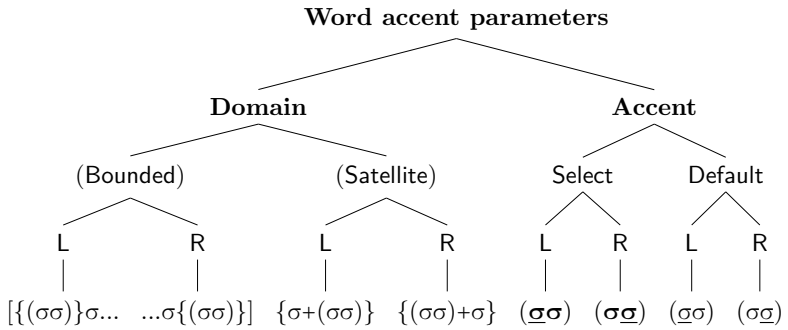
Stress is also commonly assigned to the third syllable from the right or left edge of a word, thus expanding the typology to include the trisyllabic window mentioned above. This is captured by Satellite, which adds one more syllable to the domain. For example, if Satellite-R and Bounded-R, an extra syllable is added to the right of the stress domain: ...{(σσ)+σ}. This external satellite is similar to an extrametrical syllable in metrical approaches. One crucial difference, however, is that satellites *can* be stressed, and are therefore not invisible. An internal satellite is achieved if

[24] In this paper, we do not adopt the term 'accent', used by van der Hulst (2012).

[25] A separate module (the *Rhythm module*, van der Hulst (2014)) is responsible for non-primary stress(es). This module, however, will not be examined in the present paper.

Satellite-L and Bounded-R, in which case an extra syllable is added to the left of the stress domain: ...{σ+(σσ)}]. Finally, note that both Bounded and Satellite may also be inactive in a language, in which case the domain is equivalent to the whole word, thus yielding an unbounded system—this is represented with parentheses in Fig. 3.13.

Figure 3.13: Accent parameters (adapted from van der Hulst (2012, p. 1499))



The two remaining parameters in Fig. 3.13 (within **Accent**) play a role in selecting the domain head in the case of domains that contain syllables of equal weight: Select in the case of two heavy syllables; Default in the case of two light syllables. For example, a LHH word in a weight-sensitive language such as Portuguese has two phonologically prominent positions. As a result, Select will define which of the two will be assigned primary stress. For Portuguese, the role of Select is redundant if the effect of a heavy syllable is probabilistically defined according to its position in the domain.

Finally, the possible sources for stress are also parameters in AF (not shown in Fig. 3.13). Syllable weight ('Stressed syllables are heavy') may be active (e.g., in weight-sensitive languages such as Portuguese and English) or inactive (e.g., in languages where stress is fixed or where stress is not affected by weight; such as Icelandic and Irish).[26] In summary, stress in Portuguese arises from a combination of parameter values under AF, given in (6).

[26]Likewise, Diacritic marking is also available for languages such as Russian, where stress is lexicalised. Note that even such languages may have partially predictable stress, e.g., onset size positively correlates with stress in Russian (Ryan 2011).

(6) **Portuguese stress**

Bounded-R, Satellite-L, (Select-R), Default-L $\{\sigma+(\sigma\sigma)\}]_{\text{PWd}}$

Bounded-R and Satellite-L categorically define the stress domain in Portuguese, and correctly predict the absence of pre-antepenultimate stress in the language. In other words, these two parameters appropriately capture the trisyllabic window in Portuguese.

As previously mentioned, if (binary) feet and extrametricality were employed to define the stress domain in Portuguese (i.e., $(\sigma\sigma)\langle\sigma\rangle]$), we would face problems selecting the type of foot needed to account for the effects of weight observed, which are gradient, not categorical. Indeed, for Portuguese, the crucial difference between (a) Bounded and Satellite, and (b) weight is that, whereas the former can be defined in categorical terms, the latter cannot. This distinction motivates a parametric view for (a), and a probabilistic view for (b). Naturally, both (a) and (b) can be accounted for in a probabilistic approach, since categoricity can be emulated with probabilities near 0 or 1—an important property, given that the stress window is not variable.

Although Bounded and Satellite could be treated as categorical in Portuguese, Default[27] cannot be. Otherwise, this would predict that all LHH words in the language would bear final stress, and that all LLL words would bear penult stress. Even though these predictions capture the general patterns in the Portuguese lexicon, neither generalisation completely holds: not all LHH words have final stress (e.g., *revólver* 'pistol'), and not all LLL words have penultimate stress (e.g., *patético, jacaré* 'pathetic, alligator'), as discussed earlier.

Assuming that English and Portuguese differ in terms of their prosodic representation (i.e., presence/absence of the foot), stress in these two languages is the result of weight and footing in English, and weight and domain parameters (AF) in Portuguese. In both languages, it seems that only a probabilistic grammar (e.g., MaxEnt Grammar; Goldwater and Johnson 2003, Wilson 2006, Hayes and Wilson 2008) can accurately account for the patterns observed, given the number of (sub-regular) exceptions found in these languages (see Ryan (2011, 2016) for English). In such a grammar, words that match the so-called 'regular' patterns are more likely to occur, but deviations

[27] As mentioned above, Select is redundant in Portuguese if the effect of a heavy syllable is probabilistically defined.

from these patterns are also possible, given the probabilistic nature of the grammar in question.

In a probabilistic grammar, however, we also have to emulate the stress windows in English (determined by footing constraints) and in Portuguese (determined by AF parameters),[28] which do not exhibit variation. Within a constraint-based framework, this could be achieved by assigning sufficiently large weights to constraints which are involved in defining the stress windows in these languages. As a result, an illicit stress pattern (i.e., which falls outside the stress domain) would be assigned probabilities near 0, and would virtually never surface. In (7), we provide possible constraint versions for the settings of Bounded and Satellite in Portuguese.[29]

(7) **Defining the stress domain in Portuguese with AF-based constraints**

Let $\mathcal{D}$ be the entire domain of stress in the language, represented by {...}.

Let d be a bisyllabic domain such that $d \subseteq \mathcal{D}$, represented by (...).

BOUNDED-L/Ⓡ:

Assign a mark to candidates where stress falls outside d at the L/Ⓡ edge of the word.

SATELLITE-Ⓛ/R:

Assign a mark to candidates where stress does not fall on a syllable to the Ⓛ/R of d.

Even though the probabilistic grammars of Portuguese and English yield similar patterns on the surface, we have seen that the formal representation of the stress domain differs across the two languages. Little has been said, however, about how weight can be represented in gradient terms. Given that the effect of heavy syllables seems to be modulated by a variety of factors, such as position (see Chapter 1) and segmental quality (Olejarczuk and Kapatsinski 2013; §3.10), for example, the traditional notion of weight is no longer sufficient. In the next section, we entertain one possible gradient representation of weight.

3.6.2 Representing weight gradience

We have seen that heavy syllables tend to attract stress in weight-sensitive languages such as English and Portuguese. This effect, however, is not categorical, which means that stress is not

[28] We leave it for future research to translate the AF parameters in question into a constraint-based approach.
[29] The circle indicates the active value in the language.

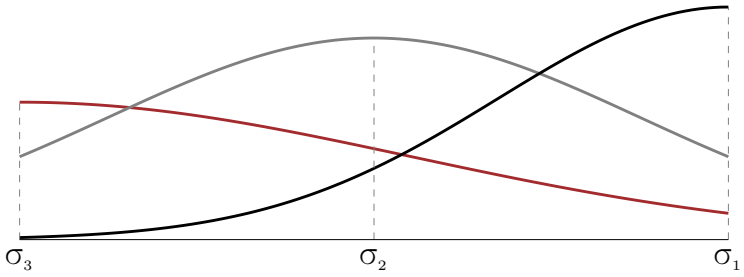
always assigned to a heavy syllable; for example, both English and Portuguese have exceptions that deviate from what is expected vis-à-vis the relationship that holds between weight and stress. In other words, we need to account for the overall tendency towards weight-sensitivity, but also for the fact that deviations are found in the lexicon—and, importantly, in experimental data. For instance, as already discussed, (L)ĹH words, which deviate from the expected final stress, albeit less common than (L)LḢ words, are not rare in the Portuguese lexicon—a fact which is empirically replicated in experimental data (Chapter 2).

A second important fact regarding weight effects is that, in Portuguese, each position in the stress domain is affected differently by a heavy syllable, hence the ranking $H_3 < H_2 < H_1$—where the number represents the position of a given heavy syllable in the stress domain (1 = final syllable). However, note that because weight is a local property of a syllable, a heavy final syllable cannot directly inform us about the probability of antepenultimate *vs*. penultimate stress. As a result, even though this hierarchy of relative weight captures gradience in the stress domain, it does not capture the fact that penultimate stress is much more common than antepenultimate stress in the presence of a heavy final syllable. In other words, the ranking $H_3 < H_2 < H_1$ does not predict that LĹH words are more common than ĹLH words. As these patterns do not seem to be random, it would be ideal if we could represent the effects of weight such that the facts described above are accounted for in a principled way.

We will assume that while weight is a property of the syllable, its impact on stress goes beyond the syllable—given that, as a suprasegmental phenomenon, stress can only be understood in relational terms. A heavy syllable generates a suprasegmental peak of phonological prominence. As a result, in the absence of other heavy syllables in the domain, this syllable is the ideal candidate to bear stress. The effect of such a peak, however, is not constrained by the syllable boundary, given that the effect of heaviness holds on the suprasegmental level. Rather, it ripples away from the heavy syllable following a distribution: other syllables in the domain are thus also possible (albeit less likely) candidates to bear stress. For illustrative purposes only, we will assume that this distribution follows a Gaussian curve (we return to this point below), graphically shown in Fig. 3.14.

The weight effects of σ_1 in Fig. 3.14 peak over the final position in the stress domain. The distribution of such effects, however, also affects other positions in the stress domain. This representation predicts that, if we only take the weight dimension into consideration, the probability of penultimate stress in LLH words will be higher than the probability of antepenultimate stress, because of the weight effects generated by the distribution of σ_1 (the black line is non-zero over σ_2, and virtually zero over σ_3). Regardless of how wide the distribution of σ_1 is, its constrictive effect on the antepenultimate syllable will always supersede its constrictive effect on the penultimate syllable—assuming a unimodal distribution such as the one in question.

Figure 3.14: A probabilistic representation of weight.
More robust weight effects yield a narrower distribution



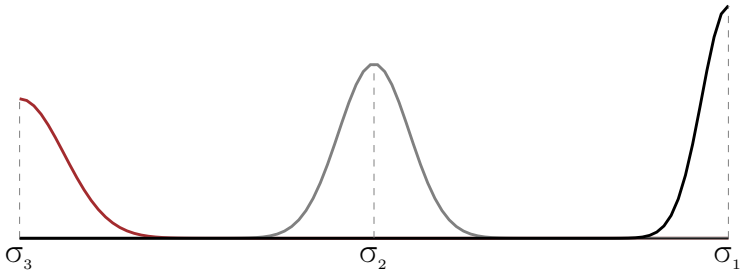
In Fig. 3.14, the more robust the weight effects in position n, the narrower the distribution peaked over n, and the higher the probability of stress on σ_n. Conversely, when the effects of σ_n are less robust, the distribution is wider, which increases the probability of stress elsewhere in the domain. This relationship mirrors the experimental results in Chapter 2, where the posterior distributions of σ_3 contain the lowest estimates in the domain. In Fig. 3.14, this is represented with the widest distribution in the domain: both HX́X and HXX́ words are relative common in Portuguese.

In a hypothetical language that is identical to Portuguese but in which weight categorically affects stress, the distributions of weight effects in the stress domain would be sufficiently narrow to avoid any overlap, as shown in Fig. 3.15.[30] In other words, the different weight effects and (patterned) exceptionalities in a given language can be captured by overlapping distributions in

[30]Note, however, that the different heights of the distributions in this illustrative example imply that final stress is more likely in words with multiple heavy syllables in such a language.

the grammar, which in turn impact the probabilities of different stress patterns.[31]

Figure 3.15: The representation of weight in a language with no exceptional stress



Because all distributions in Fig. 3.14 are Gaussian, they are all symmetrical. As a result, in the presence of a heavy syllable in σ_2, ĹHL and LHĹ are equally likely outcomes on the basis of their weight profiles. In other words, the effect of σ_2 ripples away in both directions with equal strength. At first, this may appear to be an undesirable prediction, given that the Portuguese lexicon has virtually no LHL words with antepenult stress. This lexical gap could be accidental or systematic: if it is systematic, then native speakers are expected to judge ĹHL words as illicit or unnatural. This, however, is not the case. As shown in Chapter 2, speakers' judgments are in fact consistent with the representational assumptions in Fig. 3.14, insofar as antepenultimate stress is accepted in the presence of a heavy penultimate syllable.

Naturally, this fine-grained probabilistic representation of weight sensitivity is illustrative, in that the actual distributions of the weight effects of a heavy syllable are not known exactly. However, given the distributions used, this type of representation makes predictions regarding stress contrasts not tested in the literature on Portuguese. For example, it predicts that if the weight-stress pattern in ĹHL words is accepted as natural by native speakers of the language, so should the weight-stress pattern in LHĹ words, given that weight effects are assumed to be symmetrical in the gradient representations in Figs. 3.14 and 3.15.[32] Crucially, the probability of stress on a light syllable is lowered as a function of the distance between this light syllable and the closest heavy syllable in the domain.

[31] A similar representation is employed in models such as the Gradual Learning Algorithm (Boersma 1998).

[32] Other factors, such as the sonority profile of codas, could certainly play a role in stress assignment, as previously mentioned.

If weight is represented in gradient terms, then each of the effects of a given heavy syllable can be modulated accordingly. In English (as in Portuguese), weight effects are positionally defined. First, heavy syllables are overall not treated as heavy in antepenultimate position in the language. Second, we have seen that final stress in English is more likely when a VVC rhyme is present word-finally, and less likely otherwise. This implies that word-final weight effects are weaker than those found in penultimate syllables, and therefore heavier (final) syllables are needed for stress to be attracted to the right edge of the word. In gradient representations such as those shown in Figs. 3.14 and 3.15, VVC rhymes in σ_1 would be represented with a narrower distribution than other rhyme shapes in σ_1, following the assumption that the heavier a syllable becomes, the narrower its distribution is expected to be. Representing weight in non-categorical terms can not only better capture the empirical effects discussed thus far, but it can also help us better understand how stress and weight interact in the grammar of a given language.

In constraint-based approaches, weight effects have typically been captured by the principle in (8) (Prince 1990), which states that if a given syllable is heavy, it should be stressed. If WSP is positionally defined, then a probabilistic approach where constraints are weighted (e.g., Hayes and Wilson 2008) could assign different weights to different positions in the stress domain. This would entail that WSP is a family of constraints, WSP_n, where n is a given syllable in the stress domain. Foot-based constraints (for English) and accent-based constraints (for Portuguese) would restrict n such that no weight effects outside the stress domain are active in the grammar.

(8) WEIGHT-TO-STRESS PRINCIPLE (WSP) (Prince 1990)
If heavy, then stressed

If WSP_n captures the weight effects in Portuguese and English, it is the interaction between WSP_n and other constraints in the grammar that captures deviations from what is expected based solely on weight. For example, the constraint weight of NONFINALITY in English will trump WSP_1 most of the time, but once the weight of a given final syllable reaches a certain threshold (i.e., VVC), violating NON-FINALITY will be preferred to violating WSP_1. This implies that WSP_1 is more seriously violated when the final syllable contains a VVC rhyme relative to a VC rhyme in

English. In other words, to assign violations to candidates on the basis of WSP_n, we first need to compute how heavy σ_n is, *as per* Fig. 3.14.

What the gradient formalisation of weight proposed above provides, then, is a representation of heavy syllables which is consistent with positionally-defined WSP constraints (e.g., Ryan 2011). In other words, we move from a categorical view where syllables are, most of the time, either heavy or light, and where weight effects are captured by a single WSP constraint, to a gradient view where WSP_n is positionally modulated, and weight is represented in gradient terms (9).

(9) POSITIONAL WEIGHT-TO-STRESS PRINCIPLE (WSP_n)

If σ_n is heavy, then σ_n is stressed

Where the left or right edge of the word is represented by $n = 1$

Ranking: $\text{WSP}_3 < \text{WSP}_2 < \text{WSP}_1$

In summary, English stress can be formalised as the interaction between weight-based constraints and metrical constraints—which are also responsible for restricting the stress domain in the language. In Portuguese, on the other hand, stress is essentially accounted for by weight-based constraints, leaving AF-based constraints with the role of delimiting the trisyllabic stress window.

3.7 Final remarks

In this paper, we have argued that even though English and Portuguese have similar stress patterns, these two languages are fundamentally different. On the one hand, Portuguese does not offer compelling evidence for footing, given that weight effects are found across all syllables in the stress domain. On the other hand, English offers robust evidence for the foot, as the weight effects observed in the language are not detectable in antepenultimate syllables, an observation that follows from the assumption that the language builds moraic trochees. Footing also predicts the truncation and hypocorisation patterns observed in English, as well as the absence of sub-minimal words.

In contrast to English, Portuguese presents multiple violations of word-minimality, and truncation patterns yield non-uniform metrical patterns. Indeed, feet have little added value in the

language if their only function is to restrict the stress window to trisyllabic, given that alternatives such as Accent-First Theory accomplish this goal without assuming the foot. English and Portuguese therefore present different formal systems that regulate stress: whereas English stress is the result of the interaction between feet and weight, Portuguese stress is the result of weight. In summary, the evidence *for* the foot in English is as robust as the evidence *against* the foot in Portuguese.

Given that these Portuguese and English present different formal systems that regulate the stress patterns in these languages, future research is needed to investigate whether these two types of systems (metrical structure *vs.* **Domain** and **Accent** parameters) can co-exist and not overgenerate the typology of stress patterns attested across languages.

Conclusion

This thesis has proposed a probabilistic approach to stress. In Chapter 1 (Garcia 2017b), I examined the Portuguese lexicon, and showed that weight effects on stress are far from categorical. Instead, these effects are gradient, and can be found in all syllables in the stress domain. Even onset size was shown to impact the location of stress in the Portuguese lexicon. These findings contradict not only studies which question (or reject) weight-sensitivity in the language (e.g., Lee 1994, Santos 2001, Cantoni 2013), but also standardly-held assumptions regarding weight effects in the literature of Portuguese stress, namely, (i) that weight only directly impacts word-final syllables, (ii) that the weight distinction in the language is binary, i.e., syllables are either heavy or light, and (iii) that only rhymes contribute to weight. I have also shown that a probabilistic approach based solely on weight is not only empirically better motivated (given what we find in the Portuguese lexicon), but is also more accurate at predicting the location of stress in existing words.

The grammar alluded to in Chapter 1 (Garcia 2017b) assumes that words are assigned stress probabilistically on the basis of (weight) patterns which are already present in the lexicon. Once a given word enters the lexicon, stress remains stored. This is an important difference between traditional analyses of stress in Portuguese and the probabilistic proposal argued for in Chapter 1 (Garcia 2017b): once assigned (probabilistically), stress is stored for *all* words. In contrast, traditional analyses assume that stress is only stored for irregular cases, and derived for regular patterns (cf. Cantoni 2013). The probabilistic approach in Chapter 1 is consistent with a Bayesian approach, given that the probability of a given stress pattern in a novel word is determined not only by the weight profile of said word, but also by the distribution of patterns already present in the lexicon.

Whether the weight gradience in the lexicon is learned and generalised by speakers was the

object of investigation in Chapter 2 (published: Garcia 2019). Of particular interest was the negative weight effect found in antepenultimate syllables. The chapter first assessed the reliability of such an effect, and then empirically tested whether speakers' generalised it to novel words. Because the lexicon modelled in Chapter 1 contained virtually all non-verbs in Portuguese, subtle patterns such as the negative weight effect in antepenultimate syllables might not be realistically present in the input to which learners are exposed when acquiring the grammar of Portuguese. In addition, the negative weight effect in question might be absent from the lexica of individual speakers, which are typically assumed to be smaller than the set of all the existing words in the language. To ensure the effect in question was robust, the chapter modelled antepenultimate weight effects in simulated smaller lexica and in a lexicon containing only frequent words in Portuguese. All the simulations suggested that the negative weight effect found in the lexicon is reliably present in learners' input as well as in speakers' lexica.

Even though the lexical simulations in Chapter 2 statistically confirm the negative weight effect in antepenultimate syllables, the experimental data discussed in the chapter show that speakers do *not* generalise this effect. Instead, speakers favour antepenultimate stress more often in HLL words than in LLL words, i.e., the opposite effect is generalised to novel words. In contrast, the overall weight gradience found in the language was shown to be reflected in speakers' generalisations.

The finding that speakers favour stress on heavy syllables throughout the stress domain in Portuguese poses a major challenge for foot-based analyses. Indeed, as discussed in Chapter 3, the problems of a foot-based approach go beyond weight effects, and include violations of word-minimality not only in existing words, but also in hypocoristics. Furthermore, metrical patterns in Portuguese are often inconsistent, as existing words as well as truncated forms support different foot types (both trochees and iambs). As was discussed, it is unsurprising that previous foot-based analyses have assumed a range of foot types and additional mechanisms (such as extrametricality and catalexis) to account for the stress patterns in the language. Taking all the observations together, Chapter 3 argued that the foot cannot be motivated for Portuguese.

To strengthen the argument that Portuguese has no feet, Chapter 3 also examined English, a language where, at first glance, stress patterns look strikingly similar to the patterns found

in Portuguese. It is argued, however, that these two languages are fundamentally different with regard to the formal system that regulates stress, given that, unlike Portuguese, English offers robust evidence for footing: existing lexical words never violate word-minimality; truncation and hypocorisation can be captured by a consistent foot type (moraic trochees); and, crucially, weight effects are predicted exactly if we assume footing. The interaction between footing and weight in English means that weight plays a major role in stress assignment, but is regulated by footing. This, in turn, explains why, unlike in Portuguese, heavy antepenultimate syllables pattern with light syllables, i.e., no statistical difference regarding speakers' preferences was captured between ĤLL and ĹLL words.

Even though the main objective in Chapter 3 was to argue that Portuguese has no feet, it also provided experimental data on weight effects on English stress. The findings included not only the crucial lack of a weight effect in antepenultimate syllables, but also the effect of onset size in penultimate syllables, as well as the effect of sonority in coda segments. As was shown, even though sonorant codas are more stress-attracting than obstruent codas, this effect is not strong enough to override the overall lack of a weight effect in antepenultimate syllables. The same is true of onset size, which positively correlates with stress preference in both antepenultimate and penultimate syllables.

Like the weight effects in Portuguese discussed in Chapter 1, the weight effects in English discussed in Chapter 3 are also consistent with a probabilistic approach where weight is understood as gradient. Furthermore, weight effects in Portuguese and English are positionally defined (e.g., coda effects are only sufficiently strong in penultimate syllables in English; onsets are weaker than codas in both languages). Onset effects in English, however, were found to be more consistent than those detected in Portuguese (Chapter 1), given that a positive effect was confirmed in both antepenultimate and penultimate syllables—an effect which is consistent with previous studies (Davis 1988, Ryan 2011, 2014). Crucially, we cannot accurately account for these effects if we assume that weight effects are categorical.

Contributions

The statistical approach to stress presented in this thesis allows us to objectively assess standardly-held theoretical assumptions regarding weight. Chapter 1 is the first study to statistically model weight effects on stress in a comprehensive lexicon of Portuguese, and provides not only evidence for weight effects in the entire stress domain, but also new evidence for the contribution of onsets to stress location. Importantly, whereas previous studies examining onset effects on stress have focused mostly on word-initial position, Chapter 1 has shown that the effect of onsets on stress is detectable in *all* positions in the trisyllabic window in Portuguese. The probabilistic approach in Chapter 1 also departs from traditional analyses insofar as stress is assumed to remain stored in the lexicon—indeed, it is *how* stress is assigned in the first place that is the focus of the chapter. Finally, the lexicon examined in Chapter 1 was elaborated as part of this thesis, and is now publicly available to researchers who wish to investigate other phonological aspects of Portuguese stress (Garcia 2014).

Chapter 2 is the first study to empirically test weight effects on stress in Portuguese. Crucially, the study has shown that the gradient weight effects found in the Portuguese lexicon are learned and generalised to novel words by native speakers. Chapter 2 has also shown that speakers can actually repair an unnatural pattern present in their lexicon. Unlike previous studies on phonological (under)learning, which have shown that unnatural patterns are harder to acquire and generalise (e.g., Hayes et al. 2009, Becker et al. 2011, Becker et al. 2012), these results show that speakers can in fact go one step further, and learn the opposite pattern—a similar effect is reported for the acquisition of Polish complex onsets in Jarosz (2017).

Chapter 3 is the first study to argue against the foot in Portuguese with experimental evidence based on lexical patterns (Chapter 1) as well as native speakers' behaviour (Chapter 2). Unlike French and Turkish, the focus of previous studies that have questioned the universal status of the foot, Portuguese is a language with seemingly ordinary patterns of prominence. As a result, footing has been assumed to play an important role in most approaches to stress.

Chapter 3 also investigated English, a language which shares most of its stress patterns in non-verbs with Portuguese. To my knowledge, this was the first experimental study to focus specifically on antepenultimate weight effects on stress in English, and to confirm that such effects follow from

what a foot-based account would predict.

Chapters 2 and 3 both modelled weight effects using Bayesian models with mildly informative priors—which were shown to have a better fit than their naïve counterparts. This approach attempted not only to provide a more explanatory and meaningful statistical examination than traditional frequentist methods, but also to show that models which encompass certain (justified) expectations can statistically outperform models which assume a neutral baseline. In other words, these two studies show that our expectations regarding speakers' grammars can take active part in the hypothesis being tested, which in turn allows us to bridge the traditional gap between statistical analysis and theoretical assumptions. To my knowledge, this is the first study to have employed sensitivity analysis to explore how different expectations regarding speakers' grammars can affect the fit of Bayesian models examining weight effects on stress.

Future directions

As previously mentioned, among the topics which will benefit from further examination is the role of onsets in Portuguese. We have seen that onsets have variable effects in the lexicon (positive effect in antepenultimate syllables; negative effect in penultimate and final syllables). No domain of weight computation can account for such effects (e.g., Garcia 2016). The first question that needs to be investigated is therefore whether these seemingly contradictory effects are indeed reflected in speakers' grammars. If they are, then an explanation for the asymmetric effects of onsets need to be explored in more detail.

Given that most studies investigating weight-sensitivity have focused on primary stress, another question arising from this thesis is whether (and how) weight affects secondary stress. The location of secondary stress in Portuguese can vary in sufficiently long words, as long as no clashes or three-syllable lapses are created: /desmataˈmento/ 'deforestation' → [ˌdezmataˈmẽntʊ] ~ [dezˌmataˈmẽntʊ]; /falesiˈmento/ 'decease' *n* → [ˌfalesiˈmẽntʊ] ~ [faˌlesiˈmẽntʊ]. The question is whether an initial heavy syllable in such words tends to attract secondary stress in production data more often than a light initial syllable.

In the probabilistic approach to weight and stress proposed in this thesis, representational

assumptions still play a central role. In Chapter 3, binary feet were employed along with a richer representation of weight. One relevant question is how exactly gradient weight is mapped onto feet, which are, by definition, categorical constituents with fixed edges. In other words, future research should explore how a 'moraic' foot can be formally represented given that it has been argued that moras should be replaced by a weight continuum.

In spite of the weight continuum assumed in this thesis, it is clear that some phenomena are nicely captured by binary distinctions. The condition on word-minimality, for example, only requires a '0 or 1' approach to be evaluated. The representation of gradient weight needs to allow for such binarity. Indeed, when we analysed the posterior distributions of weight effects in Chapter 3, one of the first questions of interest was whether the distributions included zero for a given credible interval. Thus, a non-categorical approach does not preclude binary generalisations.

Finally, the findings of this thesis have consequences for second language acquisition. Given that Portuguese has no evidence for the foot, future research needs to explore how this affects the acquisition of Portuguese as a second language by learners whose native language has feet, as well as the acquisition of other languages that motivate the foot by speakers of Portuguese, as there is significant evidence that second language learners transfer the grammar of their first language into their second language (Schwartz and Sprouse 1994, 1996). One relevant empirical question is whether Portuguese-speaking learners of English would differentiate ĤLL and ĹLL words in the experiment discussed in Chapter 3. If a heavy antepenultimate syllable is favoured in such cases, this would indicate that Portuguese speakers are merely transferring their weight-only grammar from Portuguese, where feet play no role. The prediction is that, if learners do differentiate ĤLL and ĹLL words in English, they should also judge sub-minimal words in English as possible words, given that the presence or absence of feet in a language impacts not only stress patterns, but also the shapes of possible words in said language.

∘ ∘

In summary, this thesis has proposed a probabilistic approach to weight and stress, in which patterns are more or less likely to be assigned to a novel word on the basis of their lexical distribution. In the case of Portuguese, I have shown that a weight-based approach is sufficiently accurate

to outperform previous categorical approaches. Finally, the present proposal argues *against* the foot in Portuguese, and, at the same time, strengthens the argument *for* the foot in English.

References

Agresti, A. (2010). *Analysis of ordinal categorical data*. John Wiley & Sons, New Jersey.

Albright, A. and Hayes, B. (2006). Modeling productivity with the Gradual Learning Algorithm: the problem of accidentally exceptionless generalizations. In Fanselow, G., Féry, C., Vogel, R., and Schlesewsky, M., editors, *Gradience in grammar: Generative perspectives*, chapter 10, pages 185–204. Oxford University Press, Oxford.

Amaral, M. P. d. (2000). *As proparoxítonas: teoria e variação*. PhD thesis, Pontifícia Universidade Católica do Rio Grande do Sul.

Araújo, G. A. (2002). Truncamento e reduplicação no português brasileiro. *Revista de Estudos da Linguagem*, 10(1):61–90.

Araújo, G. A. (2007). *O acento em português: abordagens fonológicas*. Parábola, São Paulo.

Araújo, G. A., Guimarães-Filho, Z. O., Oliveira, L., and Viaro, M. (2007). As proparoxítonas e o sistema acentual do português. In Araújo, G. A., editor, *O acento em português: abordagens fonológicas*, pages 37–60. Parábola, São Paulo.

Araújo, G. A., Guimarães-Filho, Z. O., Oliveira, L., and Viaro, M. E. (2011). Algumas observações sobre as proparoxítonas e o sistema acentual do português. *Cadernos de Estudos Linguísticos*, 50(1):69–90.

Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge University Press, New York.

Bachrach, A. and Wagner, M. (2007). Syntactically driven cyclicity vs. output-output correspondence: the case of adjunction in diminutive morphology. Ms. available on LingBuzz at http://ling.auf.net/lingbuzz/000383.

Barker, M. A. (1964). *Klamath grammar*. University of California Press, Los Angeles.

Battisti, E. (1997). *A nasalização no português brasileiro e a redução dos ditongos nasais átonos: uma abordagem baseada em restrições*. PhD thesis, Pontifícia Universidade Católica do Rio Grande do Sul.

Becker, M., Ketrez, N., and Nevins, A. (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Language*, 87(1):84–125.

Becker, M., Nevins, A., and Levine, J. (2012). Asymmetries in generalizing alternations to and from initial syllables. *Language*, 88(2):231–268.

Beckman, J. (1997). Positional faithfulness, positional neutralization, and Shona vowel harmony. *Phonology*, 14(1):1–46.

Belsley, D. A., Kuh, E., and Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. John Wiley & Sons, New York.

Benevides, A. d. L. (2017). O acento primário em pseudopalavras: uma abordagem experimental. Master's thesis, Universidade de São Paulo.

Bisol, L. (1992). O acento e o pé métrico binário. *Cadernos de Estudos Linguísticos*, 22:69–80.

Bisol, L. (2000). O troqueu silábico no sistema fonológico (um adendo ao artigo de Plínio Barbosa). *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 16(2):403–413.

Bisol, L. (2013). O acento: duas alternativas de análise. *Organon*, 28(54):281–321.

Boersma, P. (1998). *Functional Phonology: formalizing the interactions between articulatory and perceptual drives*. PhD thesis, University of Amsterdam.

Boersma, P. and Pater, J. (2016). Convergence properties of a gradual learning algorithm for harmonic grammar. In McCarty, J. and Pater, J., editors, *Harmonic Grammar and Harmonic Serialism*, chapter 13, pages 389–434. Equinox, London.

Boersma, P. and Weenink, D. (2020). Praat: doing phonetics by computer [Computer program].

Bonilha, G. F. G. (2004). A influência da qualidade da vogal na determinação do peso silábico e formação dos pés: o acento primário do português. *Organon*, 18(36):41–56.

Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, Boca Raton.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Bürkner, P.-C. (2016). `'brms'`: Bayesian regression models using Stan. R package version 1.5.0.

Câmara Jr., J. M. (1970). *Estrutura da língua portuguesa*. Editora Vozes, Petrópolis.

Cantoni, M. M. (2013). *O acento no português brasileiro: uma abordagem experimental*. PhD thesis, Universidade Federal de Minas Gerais.

Carpenter, A. C. (2010). A naturalness bias in learning stress. *Phonology*, 27(3):345–392.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: a probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. Harper & Row, New York.

Christensen, R. H. B. (2013). Analysis of ordinal data with cumulative link models—estimation with the R-package ordinal. Vignette available at http://cran.rproject.org/web/packages/ordinal/vignettes/clm_intro.pdf.

Collischonn, G. (1994). Acento secundário em português. *Letras de Hoje*, 29(4):43–55.

Collischonn, G. (2010). O acento em português. In *Introdução a estudos de fonologia do português brasileiro*, chapter 4, pages 132–165. EDIPUCRS, Porto Alegre, 5th edition.

Cristófaro-Silva, T. (2005). Fonologia probabilística: estudos de caso do português brasileiro. *Lingua(gem)*, 2(2):223–248.

Cutler, A. (2005). Lexical stress. In Pisoni, D. B. and Remez, R. E., editors, *The handbook of speech perception*, pages 264–289. Blackwell, Oxford.

Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*. MIT Press, Cambridge, MA.

d'Andrade, E. (1994). *Temas de fonologia*. Edições Colibri, Lisboa.

Dave, H. (2012). *Frequency word lists: Brazilian Portuguese*. Frequency corpus publicly available at https://invokeit.wordpress.com/frequency-word-lists/.

Davis, S. (1988). Syllable onsets as a factor in stress rules. *Phonology*, 5(1):1–19.

de Freitas, M. A. and Barbosa, M. F. M. (2013). A alternância do diminutivo -inho/-zinho no português brasileiro: um enfoque variacionista. *ALFA: Revista de Linguística*, 57(2):577–605.

de Lacy, P. (2006). *Markedness: reduction and preservation in phonology*. Cambridge University Press, New York.

Domahs, U., Plag, I., and Carroll, R. (2014). Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *The Journal of Comparative Germanic Linguistics*, 17(1):59–96.

Dryer, M. S. and Haspelmath, M., editors (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Ferreira-Gonçalves, G. (2010). Aquisição prosódica do português: o acento em suas formas marcadas. *Revista Virtual de Estudos da Linguagem (ReVEL)*, 8(15):61–81.

Frota, S. and Vigário, M. (2001). On the correlates of rhythmic distinctions: the European/Brazilian Portuguese case. *Probus*, 13(2):247–275.

Frota, S., Vigário, M., Martins, F., and Cruz, M. (2010). Frepop–Frequency of Phonological Objects in Portuguese (version 1.0). *Laboratório de Fonética da Faculdade de Letras de Lisboa*.

Fudge, E. (1987). Branching structure within the syllable. *Journal of Linguistics*, 23(02):359–377.

Fudge, E. C. (1969). Syllables. *Journal of Linguistics*, 5(2):253–286.

Garcia, G. D. (2014). *Portuguese Stress Lexicon*. Comprehensive list of non-verbs in Portuguese. Available at http://guilhermegarcia.github.io/psl.html.

Garcia, G. D. (2015). *Word generator: an R script for generating nonce words.* Commented code available at http://guilhermegarcia.github.io/resources.html.

Garcia, G. D. (2016). The computation of weight in Portuguese: syllables and intervals. In Kim, K.-m., Umbal, P., Block, T., Chan, Q., Cheng, T., Finney, K., Katz, M., Nickel-Thompson, S., and Shorten, L., editors, *Proceedings of the 33rd West Coast Conference on Formal Linguistics*, pages 137–145. Cascadilla Proceedings Project.

Garcia, G. D. (2017a). Grammar trumps lexicon: Typologically inconsistent weight effects are not generalized. In Lamont, A. and Tezloff, K., editors, *Proceedings of the 47th Annual Meeting of the North East Linguistic Society* (NELS), volume 2, pages 25–34. Amherst, MA: Graduate Linguistics Student Association (GLSA), University of Massachusetts.

Garcia, G. D. (2017b). Weight gradience and stress in Portuguese. *Phonology*, 34(1):41–79.

Garcia, G. D. (2019). When lexical statistics and the grammar conflict: learning and repairing weight effects on stress. *Language*, 95(4):612–641.

Garcia, G. D. and Goad, H. (2020). Stress without feet: a parametric distinction between English and Portuguese. In preparation.

Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis*, 3(3):445–449.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014a). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC, Boca Raton, 3rd edition.

Gelman, A., Hwang, J., and Vehtari, A. (2014b). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Giegerich, H. J. (2005). *English phonology: an introduction*. Cambridge University Press, New York, 8th edition. First published in 1992.

Goad, H. (2016). Phonological processes in children's productions: convergence with and divergence from adult grammars. In Lidz, J., Snyder, W., and Pater, J., editors, *The Oxford Handbook of Developmental Linguistics*, pages 43–67. Oxford University Press, Oxford.

Goedemans, R. and van der Hulst, H. (2009). StressTyp: a database for word accentual patterns in the world's languages. In Everaert, M. and Musgrave, S., editors, *The Use of Databases in Cross-linguistics Research*, pages 235–282. Mouton de Gruyter, Berlin.

Goedemans, R. and van der Hulst, H. (2013). *Weight factors in weight-sensitive stress systems*. Leipzig. Available at http://wals.info/chapter/16.

Goldsmith, J. (1976). *Autosegmental Phonology*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA. Published by Garland Press, New York , 1979.

Goldwater, S. and Johnson, M. (2003). Learning OT constraint rankings using a Maximum Entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, pages 111–120.

Gonçalves, C. A. (2004). A morfologia prosódica e o comportamento transderivacional da hipocorização no português brasileiro. *Revista de Estudos da Linguagem*, 12(1):7–38.

Gordon, M. (2004). Positional weight constraints in OT. *Linguistic Inquiry*, 35(4):692–703.

Gordon, M. (2005). A perceptually-driven account of onset-sensitive stress. *Natural Language & Linguistic Theory*, 23(3):595–653.

Gordon, M. (2006). *Syllable weight: phonetics, phonology, typology*. Routledge, New York.

Gordon, M. (2011). Stress systems. In Goldsmith, J. A., Riggle, J., and Alan, C. L., editors, *The handbook of phonological theory*, pages 141–163. John Wiley & Sons, Hoboken.

Gordon, M. (2016). *Phonological typology*. Oxford University Press, Oxford.

Griffiths, T. L. and Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17(9):767–773.

Guion, S. G., Clark, J. J., Harada, T., and Wayland, R. P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(4):403–426.

Halle, M. and Idsardi, W. (1995). General properties of stress and metrical structure. In Goldsmith, J. A., editor, *The Handbook of Phonological Theory*, pages 403–443. Blackwell, Cambridge, MA, and Oxford, UK.

Halle, M. and Kenstowicz, M. (1991). The free element condition and cyclic versus noncyclic stress. *Linguistic Inquiry*, 22(3):457–501.

Halle, M. and Vergnaud, J.-R. (1980). Three-dimensional phonology. *Journal of Linguistic Research*, 1(1):83–105.

Halle, M. and Vergnaud, J.-R. (1987a). *An essay on stress*. MIT Press, Cambridge, MA.

Halle, M. and Vergnaud, J.-R. (1987b). Stress and the cycle. *Linguistic Inquiry*, 18(1):45–84.

Harris, J. (1997). Licensing inheritance: an integrated theory of neutralisation. *Phonology*, 14(3):315–370.

Harris, J. (2011). Deletion. In Oostendorp, M. v., Ewen, C., Hume, E., and Rice, K., editors, *The Blackwell Companion to Phonology*, pages 1597–1621. Wiley-Blackwell, Oxford.

Harris, J. W. (1983). *Syllable structure and stress in Spanish: a non-linear analysis*. MIT Press, Linguistic Inquiry Monographs, Cambridge, MA.

Hayes, B. (1980). *A metrical theory of stress rules*. PhD thesis, Massachusetts Institute of Technology.

Hayes, B. (1982). Extrametricality and English stress. *Linguistic Inquiry*, 13(2):227–276.

Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry*, 20:253–306.

Hayes, B. (1995). *Metrical stress theory: principles and case studies*. University of Chicago Press, Chicago.

Hayes, B. and Londe, Z. (2006). Stochastic phonological knowledge: the case of Hungarian vowel harmony. *Phonology*, 23(1):59–104.

Hayes, B., Siptár, P., Zuraw, K., and Londe, Z. (2009). Natural and unnatural constraints in Hungarian vowel harmony. *Language*, 85(4):822–863.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440.

Hermans, B. and Wetzels, L. (2012). Productive and unproductive stress patterns in Brazilian Portuguese. *Revista Letras & Letras*, 28:77–114.

Houaiss, A., Villar, M., and de Mello Franco, F. M. (2001). *Dicionário eletrônico Houaiss da língua portuguesa*. Objetiva, Rio de Janeiro.

Hyman, L. (1985). *A theory of phonological weight*. Foris Publications, Dordrecht.

Idsardi, W. (1992). *The computation of prosody*. PhD thesis, Massachusetts Institute of Technology.

Idsardi, W. (2009). Calculating metrical structure. *Contemporary Views on Architecture and Representations in Phonology*, 48:191–212.

Itô, J. (1990). Prosodic minimality in Japanese. *CLS*, 26(2):213–239.

Jarosz, G. (2017). Defying the stimulus: acquisition of complex onsets in Polish. *Phonology*, 34(2):269–298.

Jun, S.-A. and Fougeron, C. (2000). A phonological model of French intonation. In Botinis, A., editor, *Intonation*, pages 209–242. Springer, Dordrecht.

Kager, R. (2011). *Optimality Theory*. Cambridge University Press, Cambridge, 11th edition. First published in 1999.

Kager, R. (2012). Stress in windows: language typology and factorial typology. *Lingua*, 122(13):1454–1493.

Kass, R. E., Carlin, B. P., Gelman, A., and Neal, R. M. (1998). Markov Chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, 52(2):93–100.

Kehoe, M. (1998). Support for metrical stress theory in stress acquisition. *Clinical Linguistics & Phonetics*, 12(1):1–23.

Kelly, M. H. (2004). Word onset patterns and lexical stress in English. *Journal of Memory and Language*, 50(3):231–244.

Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Blackwell, Oxford.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E., Harms, R., Bach, E., and Harms, R., editors, *Universals in Linguistic Theory*, pages 170–202. Holt, Rinehart and Winston, New York.

Klautau, A. (2013). *UFPADic 3.0*. Retrieved from http://www.laps.ufpa.br/falabrasil on 14 Sep, 2013.

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(5):658–676.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573–603.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*. Academic Press, London, 2nd edition.

Kruschke, J. K., Aguinis, H., and Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4):722–752.

Ladefoged, P. and Johnson, K. (2011). *A course in Phonetics*. Wadsworth, Boston, 6th edition.

Lee, M. D. and Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: a practical course*. Cambridge University Press, Cambridge.

Lee, S.-H. (1994). A regra de acento do português: outra alternativa. *Letras de Hoje*, 24(4):37–42.

Lee, S.-H. (1995). *Morfologia e fonologia lexical do português do Brasil*. PhD thesis, Universidade Estadual de Campinas.

Lee, S.-H. (2007). O acento primário no português: uma análise unificada na Teoria da Otimalidade. In Araújo, G. A., editor, *O acento em português: abordagens fonológicas*, pages 120–143. Parábola, São Paulo.

Levi, S. V. (2005). Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association*, 35(1):73–97.

Levin, J. (1985). *A metrical theory of syllabicity*. PhD thesis, Massachusetts Institute of Technology.

Liberman, M. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8(2):249–336.

Lowenstamm, J. and Kaye, J. (1986). Compensatory lengthening in Tiberian Hebrew. In Wetzels, L., Sezer, E., Wetzels, L., and Sezer, E., editors, *Studies in Compensatory Lengthening*, pages 97–146. Foris, Dordrecht.

Magalhães, J. S. (2008). O acento dos não verbos no português brasileiro no plano multidimensional. *ALFA: Revista de Linguística*, 52(2):405–430.

Major, R. C. (1985). Stress and rhythm in Brazilian Portuguese. *Language*, 61(2):259–282.

Massini-Cagliari, G. (1999). *Do poético ao linguístico no ritmo dos trovadores: três momentos da história do acento*. Faculdade de Ciências e Letras, Laboratório Editorial, UNESP, Araraquara.

Mateus, M. H. and d'Andrade, E. (1998). The syllable structure in European Portuguese. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 14(1):13–32.

Mateus, M. H. and d'Andrade, E. (2000). *The phonology of Portuguese*. Oxford University Press, Oxford.

Mateus, M. H. M. (1983). O acento da palavra em português: uma nova proposta. *Boletim de Filologia*, 28:211–229.

McCarthy, J. J. (1979). *Formal problems in Semitic phonology and morphology*. PhD thesis, Massachusetts Institute of Technology. Published by Garland Press, New York, 1985.

McCarthy, J. J. and Prince, A. (1995). Prosodic Morphology. In Goldsmith, J. A., editor, *The Handbook of Phonological Theory*, pages 318–366. Blackwell, Cambridge, MA.

McElreath, R. (2016). *Statistical rethinking: a Bayesian course with examples in R and Stan*, volume 122. Chapman & Hall/CRC, Boca Raton.

Mitchell, T. F. (1960). Prominence and syllabication in Arabic. *Bulletin of the School of Oriental and African Studies*, 23(2):369–389.

Moore-Cantwell, C. (2016). *The representation of probabilistic phonological patterns: neurological, behavioral, and computational evidence from the English stress system*. PhD thesis, University of Massachusetts.

Moore-Cantwell, C. and Pater, J. (2016). Gradient exceptionality in Maximum Entropy Grammar with lexically specific constraints. *Catalan Journal of Linguistics*, 15:53–66.

Morén, B. (2000). The puzzle of Kashmiri stress: implications for weight theory. *Phonology*, 17(3):365–396.

Munro, P. and Willmond, C. (1994). *Chickasaw: an analytical dictionary*. University of Oklahoma Press, Norman.

Nagy, W. E. and Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 19(3):304–330.

Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Foris, Dordrecht.

Neto, N., Rocha, W., and Sousa, G. (2015). An open-source rule-based syllabification tool for Brazilian Portuguese. *Journal of the Brazilian Computer Society*, 21(1):1–10.

Olejarczuk, P. and Kapatsinski, V. (2013). The syllabification of medial clusters: evidence from stress assignment. Poster presented at the 87th Annual Meeting of the Linguistic Society of America, Boston.

Özçelik, Ö. (2013). *Representation and acquisition of stress: the case of Turkish*. PhD thesis, McGill University.

Özçelik, Ö. (2014). Prosodic faithfulness to foot edges: the case of Turkish stress. *Phonology*, 31(2):229–269.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science*, 33(6):999–1035.

Pereira, M. I. (1999). *O acento de palavra em português—uma análise métrica*. PhD thesis, Universidade de Coimbra.

Pereira, M. I. (2007). Acento latino e acento em português: que parentesco? In Araújo, G. A., editor, *O acento em português: abordagens fonológicas*, pages 61–83. Parábola, São Paulo.

Pinker, S. (2012). *The better angels of our nature: why violence has declined*. Penguin Books, New York City, NY.

Prince, A. (1990). Quantitative consequences of rhythmic organization. *CLS*, 26(2):355–398.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: constraint interaction in Generative Grammar*. Blackwell, Malden, MA & Oxford. A revision of the 1993 technical report was published in 2004 (Rutgers University Center for Cognitive Science). Available on Rutgers Optimality Archive, ROA-537.

R Core Team (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Roca, I. M. (1999). Stress in the Romance languages. In van der Hulst, H., editor, *Word Prosodic Systems in the Languages of Europe*, pages 672–811. Mouton de Gruyter, Berlin.

Ryan, K. M. (2011). Gradient syllable weight and weight universals in quantitative metrics. *Phonology*, 28(3):413–454.

Ryan, K. M. (2014). Onsets contribute to syllable weight: statistical evidence from stress and meter. *Language*, 90(2):309–341.

Ryan, K. M. (2016). Phonological weight. *Language and Linguistics Compass*, 10(12):720–733.

Santos, R. S. (2001). *A aquisição do acento primário no português brasileiro*. PhD thesis, Universidade Estadual de Campinas.

Schwartz, B. D. and Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. In Hoekstra, T. and Schwartz, B. D., editors, *Language acquisition studies in generative grammar*, pages 317–368. John Benjamins, Amsterdam.

Schwartz, B. D. and Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12(1):40–72.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, 11(3):563–605.

Selkirk, E. (1984). *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, Cambridge, MA.

Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. PhD thesis, Massachusetts Institute of Technology.

Steriade, D. (1994). Positional neutralization. Unpublished manuscript. University of California, Los Angeles.

Steriade, D. (2012). Intervals vs. syllables as units of linguistic rhythm. Handouts, EALING, Paris.

Tang, K. (2012). A 61 million word corpus of Brazilian Portuguese film subtitles as a resource for linguistic research. *UCL Working Papers in Linguistics 24: 208–214*.

Thomas, E. W. (1974). *A grammar of spoken Brazilian Portuguese*. Vanderbilt University Press, Nashville.

Topintzi, N. (2010). *Onsets: suprasegmental and prosodic behaviour*. Cambridge University Press, New York.

van der Hulst, H. (2012). Deconstructing stress. *Lingua*, 122(13):1494–1521.

van der Hulst, H. (2014). Representing rhythm. In van der Hulst, H., editor, *Word stress: theoretical and typological issues*, pages 325–365. Cambridge University Press, Cambridge.

van Oostendorp, M. (2012). Quantity and the three-syllable window in Dutch word stress. *Language and Linguistics Compass*, 6(6):343–358.

Vehtari, A., Gelman, A., and Gabry, J. (2016). LOO: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian models. R package version 1.1.0.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27(5):1413–1432.

Vilela, A. C., Godoy, L., and Silva, T. C. (2006). Truncamento no português brasileiro: para uma melhor compreensão do fenômeno. *Revista de Estudos da Linguagem*, 14(1):149–174.

Vogel, I. (2008). The morphology-phonology interface: isolating to polysynthetic languages. *Acta Linguistica Hungarica*, 55(1):205–226.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and Widely Applicable Information Criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Weide, R. (1993). Carnegie Mellon Pronouncing Dictionary.

Wetzels, W. L. (1992). Mid vowel neutralization in Brazilian Portuguese. *Cadernos de Estudos Linguísticos*, 23:19–55.

Wetzels, W. L. (1997). The lexical representation of nasality in Brazilian Portuguese. *Probus*, 9(2):203–232.

Wetzels, W. L. (2007). Primary word stress in Brazilian Portuguese and the weight parameter. *Journal of Portuguese Linguistics*, 5:9–58.

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, 30(5):945–982.

Yun, S. (2010). The typology of compensatory lengthening: a phonetically-based Optimality-Theoretic approach. Master's thesis, Seoul National University.

Zuraw, K. (2000). *Patterned exceptions in phonology*. PhD thesis, University of California at Los Angeles.

Appendix A

Statistical terms

$\hat{\mathbf{R}}$ Also known as the Gelman-Rubin convergence diagnostic. A metric used to monitor the convergence of the chains of a given model. $\hat{\mathrm{R}}$ is a measure that compares the variance of the simulations from each chain with the variance of all the chains mixed together. If all chains are at equilibrium, $\hat{\mathrm{R}}$ should equal 1. See Brooks et al. (2011).

chain A 'random walk' in the parameter space during the sampling using Monte Carlo simulation. At least two chains are used, which are later checked for convergence regarding the posterior distribution of parameter values. All the chains in a given model should converge to the same parameter space.

credible interval Interval containing the probability distribution of the most credible values in the posterior. Arbitrary ranges can be established (e.g., the 95% Credible Interval) as a decision tool. Note that credible intervals differ considerably from **confidence intervals**, which are based on hypothetical future samplings, and which are not a probability distribution.

Effective Sample Size The number of sampling steps assuming a completely uncorrelated chain, where each step provides independent information about the posterior distribution. See Kass et al. (1998).

LOO *Leave-One-Out.* Information criterion commonly used to assess a model's fit (mainly in relation to other models). The idea is to capture the out-of-sample prediction error by (i) training a given model on a data set that contains all but one item (or set), and (ii) evaluating

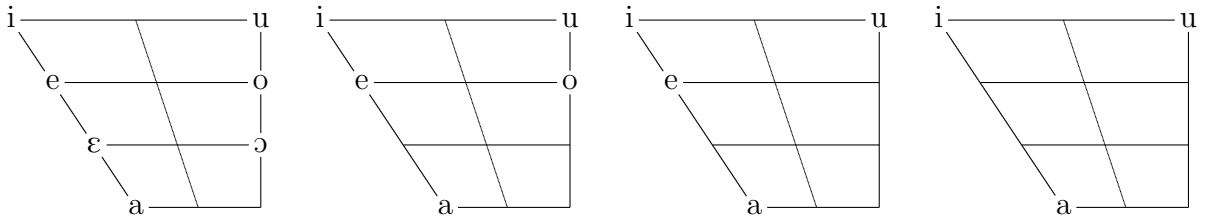
its accuracy at predicting said item (or set). To calculate the LOO for the models fit in Chapter 3, `brms` (Bürkner 2016) uses the `loo` package in R (Vehtari et al. 2016).

WAIC *Widely Applicable Information Criterion* (Watanabe 2010), also known as *Watanabe-Akaike Information Criterion*. A cross-validation method to assess the fit of a (Bayesian) model. It is calculated by averaging the log-likelihood over the posterior distribution taking into account individual data points. See McElreath (2016, p. 191) for a discussion on the differences between WAIC, AIC, BIC and DIC. For advantages of WAIC over DIC, see Gelman et al. (2014b).

Appendix B

Brazilian Portuguese vowels

The vowel distributions shown in the figures below represent the phonemic vowels in standard Brazilian Portuguese (Câmara Jr. 1970). The lower-mid vowels /ɛ, ɔ/ are only contrastive in stressed positions: *café* /ka.ˈfɛ/ → *cafeteira* /ka.fe.ˈtej.ɾa/ 'coffee', 'coffee maker'. Word-finally, mid-vowels are neutralised: /ɛ, e/ → /i/; /ɔ, o/ → /u/. Finally, /a/ is often phonetically realised as [ɐ] in unstressed position. A comprehensive analysis of vowel neutralisation in Brazilian Portuguese can be found in Wetzels (1992).



Stressed position | Pre-tonic position | Post-tonic position | Post-tonic (final) position

Appendix C

Stimuli (Chapter 2)

babedɾal	balafo	bamezil	bampɾere	baprabor	baronal	batasil	batɾakoŋ
batɾatoŋ	beroko	bestedo	bibanko	bikɾodar	binogal	blikanfo	bogɾenda
bomblina	botablor	braboŋko	bralino	branfora	brapesme	brasmare	brelero
brondale	brospeno	kabrervo	kabriza	kadrida	kaprifo	karipo	katrater
katrenir	semitur	siblanda	sikraba	sigroko	siminer	sinfrate	sirmika
sitredir	klasfika	klikarfe	klikumbo	kokuro	kogidar	kogotril	koltale
komadriŋ	kondito	konvade	koprobil	kortemo	kosavir	krafomo	krikombo
krikone	krizero	kritina	krobira	krolerpa	kuladil	kumbrosa	darnido
dekade	denoro	denzito	depebrir	detinsa	detubral	dikriba	dipigar
dipramal	dirana	ditradal	ditruspa	doroto	doteflar	dotorto	drangipa
drapeze	dundito	falegor	faŋkrela	fedado	figredo	fitilbo	fitrofa
fladeze	flandovo	flizonta	fontena	frakko	framera	frangile	freminso
frimpelo	frospato	fugrosto	fuliste	fuvosta	gadalo	ganomo	gapospe
ʒeninta	ʒerimul	ʒesprila	ʒestika	gimaro	glindebo	gofridor	grakolo
grozista	guplaro	ʒakliŋko	ʒakomar	ʒapemba	ʒobarto	ʒoteror	makobaŋ
maglimbo	malgrodo	mampedo	marobrar	merotrer	meskriva	metanta	metribul
mikamiŋ	mipleska	momemir	moŋkriko	mopropir	mornola	mulopraŋ	muploste
pakrurol	padirto	padrenga	pafrizo	paleso	panvata	patrikoŋ	pekogo
pedenso	pempano	pengata	petritol	pibidral	pisipal	pifreno	pinalbo
piprogur	pizaprar	plaberme	plabunta	plarolo	plikurvo	plirame	podrido
popranva	porebros	potoplir	prabamo	prefanto	prempedo	preŋkofa	primodo
prinisto	prinore	prisbade	prolurvo	putrenso	xadota	xavompa	xelaliŋ
xepreste	xidolo	xitelvo	xobitrir	xontruse	xumbraro	xuriko	sampodo
semalo	semprozo	senvide	sertrolo	setroko	sikresol	simbrime	siritral
sokrondo	sostrole	taklanda	tagrane	tarala	tarbita	taromil	tatremer
tergrame	tetrito	tetruko	tikrona	tifriro	tinalko	tiŋkrika	toblonso
tokronto	tompreda	toprizo	toputril	totrense	traduka	traduno	tredolto
tredonsa	trentode	trezime	tringabo	trizanga	tristuda	trolarto	trombafe
trostizo	truskome	tubradal	tunobral	vadronsa	vanispe	vasteko	velordo
venfrado	veralo	vesplako	vidatrir	volitriŋ	zagrente	zitrado	zorasto

N = 240

Appendix D

Statistical models (Chapter 2)

D.1 Lexicon

D.1.1 Lexical simulation (Fig. 2.3)

D.1.1.1 R script

```
# Load Portuguese Stress Lexicon first ('psl' variable)
nIter = 10000 # Number of iterations
lexSample = 10000 # Sample size of each lexicon

antPen = psl[psl$stressLoc != "final" &
              psl$weightProfile %in% c("LLL", "HLL", "LHL"), ]
antPen = droplevels(antPen)
antPen$weightProfile = relevel(antPen$weightProfile, ref = "LLL")
output = data.frame(matrix(nrow=nIter, ncol=4,
                 dimnames=list(c(seq(1:nIter)),
                 c("iter", "intercept", "HLL", "LHL"))))
output$iter = 1:nrow(output)
for(i in 1:nIter){
  tempData = antPen[sample(1:nrow(antPen), lexSample, replace = TRUE), ]
  tempModel = glm(stressLoc ~ weightProfile, tempData, family = "binomial")
  output$intercept[i] = tempModel$coef[[1]]
  output$HLL[i] = tempModel$coef[[2]]
  output$LHL[i] = tempModel$coef[[3]]}
```

D.1.2 APU *vs.* PU (Fig. 2.4)

D.1.2.1 Model specification in Stan

All the models below were run using four chains in parallel.

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  int<lower=1> K;  // number of population-level effects
  matrix[N, K] X;  // population-level design matrix
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
  int Kc;
  matrix[N, K - 1] Xc;  // centered version of X
  vector[K - 1] means_X;  // column means of X before centering
  Kc = K - 1;  // the intercept is removed from the design matrix
  for (i in 2:K) {
    means_X[i - 1] = mean(X[, i]);
    Xc[, i - 1] = X[, i] - means_X[i - 1];
  }
}
parameters {
  vector[Kc] b;  // population-level effects
  real temp_Intercept;  // temporary intercept
}
transformed parameters {
}
model {
  vector[N] mu;
  mu = Xc * b + temp_Intercept;
```

```
  // prior specifications
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  b_Intercept = temp_Intercept - dot_product(means_X, b);
}
```

D.2 Experimental data: Version A

D.2.1 APU *vs.* PU (Fig. 2.7a)

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  int<lower=1> K;  // number of population-level effects
  matrix[N, K] X;  // population-level design matrix
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  vector[N] Z_1_2;
  vector[N] Z_1_3;
  int<lower=1> NC_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
  int Kc;
  matrix[N, K - 1] Xc;  // centered version of X
  vector[K - 1] means_X;  // column means of X before centering
  Kc = K - 1;  // the intercept is removed from the design matrix
  for (i in 2:K) {
    means_X[i - 1] = mean(X[, i]);
    Xc[, i - 1] = X[, i] - means_X[i - 1];
```

```
  }
}
parameters {
  vector[Kc] b;  // population-level effects
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  matrix[M_1, N_1] z_1;  // unscaled group-level effects
  // cholesky factor of correlation matrix
  cholesky_factor_corr[M_1] L_1;
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  matrix[N_1, M_1] r_1;
  vector[N_1] r_1_1;
  vector[N_1] r_1_2;
  vector[N_1] r_1_3;
  // group-level effects
  vector[N_2] r_2_1;
  r_1 = (diag_pre_multiply(sd_1, L_1) * z_1)';
  r_1_1 = r_1[, 1];
  r_1_2 = r_1[, 2];
  r_1_3 = r_1[, 3];
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = Xc * b + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_1_2[J_1[n]]) * Z_1_2[n] + (
        r_1_3[J_1[n]]) * Z_1_3[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  b[1] ~ normal(1, 0.5);
```

```
  b[2] ~ normal(-1, 0.5);
  temp_Intercept ~ normal(-1, 0.5);
  sd_1 ~ student_t(3, 0, 10);
  L_1 ~ lkj_corr_cholesky(1);
  to_vector(z_1) ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  corr_matrix[M_1] Cor_1;
  vector<lower=-1,upper=1>[NC_1] cor_1;
  b_Intercept = temp_Intercept - dot_product(means_X, b);
  // take only relevant parts of correlation matrix
  Cor_1 = multiply_lower_tri_self_transpose(L_1);
  cor_1[1] = Cor_1[1,2];
  cor_1[2] = Cor_1[1,3];
  cor_1[3] = Cor_1[2,3];
}
```

D.2.2 U *vs.* PU (Fig. 2.7b)

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
}
parameters {
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  vector[N_1] z_1[M_1];  // unscaled group-level effects
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  vector[N_1] r_1_1;
  // group-level effects
  vector[N_2] r_2_1;
  r_1_1 = sd_1[1] * (z_1[1]);
```

```
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = rep_vector(0, N) + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  temp_Intercept ~ normal(1, 1);
  sd_1 ~ student_t(3, 0, 10);
  z_1[1] ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  b_Intercept = temp_Intercept;
}
```

D.3 Experimental data: Version B

D.3.1 APU *vs.* PU (Fig. 2.9a)

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  int<lower=1> K;  // number of population-level effects
  matrix[N, K] X;  // population-level design matrix
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  vector[N] Z_1_2;
  vector[N] Z_1_3;
  int<lower=1> NC_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
  int Kc;
  matrix[N, K - 1] Xc;  // centered version of X
  vector[K - 1] means_X;  // column means of X before centering
  Kc = K - 1;  // the intercept is removed from the design matrix
  for (i in 2:K) {
    means_X[i - 1] = mean(X[, i]);
    Xc[, i - 1] = X[, i] - means_X[i - 1];
```

```
  }
}
parameters {
  vector[Kc] b;  // population-level effects
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  matrix[M_1, N_1] z_1;  // unscaled group-level effects
  // cholesky factor of correlation matrix
  cholesky_factor_corr[M_1] L_1;
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  matrix[N_1, M_1] r_1;
  vector[N_1] r_1_1;
  vector[N_1] r_1_2;
  vector[N_1] r_1_3;
  // group-level effects
  vector[N_2] r_2_1;
  r_1 = (diag_pre_multiply(sd_1, L_1) * z_1)';
  r_1_1 = r_1[, 1];
  r_1_2 = r_1[, 2];
  r_1_3 = r_1[, 3];
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = Xc * b + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_1_2[J_1[n]]) * Z_1_2[n] + (
        r_1_3[J_1[n]]) * Z_1_3[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  b[1] ~ normal(1, 1);
```

```
  b[2] ~ normal(-1, 1);
  temp_Intercept ~ normal(-1, 1);
  sd_1 ~ student_t(3, 0, 10);
  L_1 ~ lkj_corr_cholesky(1);
  to_vector(z_1) ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  corr_matrix[M_1] Cor_1;
  vector<lower=-1,upper=1>[NC_1] cor_1;
  b_Intercept = temp_Intercept - dot_product(means_X, b);
  // take only relevant parts of correlation matrix
  Cor_1 = multiply_lower_tri_self_transpose(L_1);
  cor_1[1] = Cor_1[1,2];
  cor_1[2] = Cor_1[1,3];
  cor_1[3] = Cor_1[2,3];
}
```

D.3.2 U *vs.* PU (Fig. 2.9b)

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
}
parameters {
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  vector[N_1] z_1[M_1];  // unscaled group-level effects
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  vector[N_1] r_1_1;
  // group-level effects
  vector[N_2] r_2_1;
  r_1_1 = sd_1[1] * (z_1[1]);
```

```
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = rep_vector(0, N) + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  temp_Intercept ~ normal(1, 1);
  sd_1 ~ student_t(3, 0, 10);
  z_1[1] ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  b_Intercept = temp_Intercept;
}
```

Appendix E

Stimuli (Chapter 3)

bɑkrɛstən	dɑkɪmpəl	dɑsɛŋkəl	dɑlkɛpər	dɑskɛrəs	dɛfɪdəl
dɛlɑrpək	dɛlɪspən	dɛprɪstək	dɪkrɑlpət	dɪtrɛŋkəl	drɑlɛtəl
fɛtrɪdəl	frɑlsɛləp	frɑskɛlər	kɑdɪmpəl	kɑpɛməl	kɑprɛməl
kɑprɛndəl	kɑprɛstən	kɑprɪlən	kɑprɪlər	kɑprɪntəl	kɑtrɛpən
kɑtrɛlpət	kɑtrɛspət	kɑrsɪlən	kɛprɑntəs	kɛtrɑlək	kɪlɑspər
kɪmɛsər	kɪnɑrlən	kɪprɑdər	kɪtɑspən	kɪtrɑpəl	kɪtrɑsəl
kɪtrɑlpən	kɪstɛnəp	klɪspɛdən	krɑntɛnər	krɛbɑtən	krɛpɑtən
krɛstɪnəl	krɛstɪnəp	krɛstɪrəl	krɪmpɛdən	krɪmpɛtəl	krɪndɛnəf
lɑdɛsən	lɑdɛtəl	lɑprɛsən	lɑrsɛkən	lɑrsɛtəp	lɑstɛpəl
lɑstɛtər	lɛtɑrpəs	lɛstɪrəf	lɪkɑspər	lɪmɛdəl	lɪpɛtən
lɪnsɛkəf	lɪstɛmər	mɑsɛnəl	mɑtɛstəl	mɑrsɛkən	mɑstɛtər
mɛdɑpər	mɛtɑrpəs	mɛtrɑŋkəp	mɛŋkɪpəl	mɛstɪnəp	mɪkɑstər
mɪkrɛsəl	mɪlɛdəl	nɑpɪstən	nɑprɛtən	nɑrpɛlət	nɑrpɪlət
nɑstɛpəl	nɛbɑkət	nɛdɑrək	nɛprɑkət	nɛtrɑlər	nɛstɪrəf
nɪkɑrlən	nɪtrɑlpən	pɑkɛlən	pɑkrɛndəl	pɑkrɛntəl	pɑkrɛstəp
pɑlɛstər	pɑnɪstən	pɑtrɛspək	pɑtrɛspət	pɑrfɛməs	pɑrsɛtəl
pɑzmɛrət	pɛkrɑntəs	pɛlɑrkət	pɛlɑrpək	pɛlɪstən	pɛtrɑŋkəp
pɪmɑstər	pɪtrɑrək	pɪtrɛstəl	pɪtrɛzmət	plɪstɛdən	prɑkɪdər
prɑnɛtən	prɑsɪrət	prɑŋkɛmət	prɑskɛləs	prɑstɛrət	prɛndɪməf
prɛstɪkəf	prɪkɑlər	prɪlɛtən	prɪsɑnər	prɪtɑrək	prɪndɛləs
prɪstɛkəf	rɑtɛŋkəl	rɛŋkɪpəl	rɪmsɛkəf	rɪstɛmər	sɑdɛŋkəl
sɑmɛnəl	sɑpɪnər	sɑprɪrət	sɑprɪskər	sɑtrɪskər	sɪkɑdər
sɪkɑrək	sɪklɑrək	sɪkrɛnəl	sɪnɑrkəp	sɪprɑndək	sɪtɑlən
sɪtɛnər	tɑkɪsən	tɑklɪrət	tɑkrɛstəp	tɑmɛstəl	tɑpɛstər
tɑprɪlən	tɑrɛŋkəl	tɑlkɛpər	tɑrfɛməs	tɑrsɛtəl	tɑskɛrəs
tɑzmɛrət	tɛkrɑzmət	tɛlɑrkət	tɛprɪŋkəl	tɪkɑspən	tɪkrɑndək
tɪnɑrkəp	tɪprɛstəl	trɑkɛpən	trɑkɪlər	trɑkɪrət	trɑmɛrət
trɑlsɛləp	trɑŋkɛmət	trɑŋkɛnər	trɑskɛlər	trɑskɛləs	trɑstɛrət
trɛkɑlət	trɛlɑpər	trɛnɑlər	trɛnɑrək	trɛmpɪdən	trɛspɪrəl
trɪmɛrət	trɪsɑlən	trɪmpɛtəl	trɪndɛləs	trɪskɛnəl	trɪstɛnəp

N = 180

Appendix F

Statistical models (Chapter 3)

F.1 Stan models

All the models below were run using four chains in parallel.

F.1.1 Naïve model

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  int<lower=1> K;  // number of population-level effects
  matrix[N, K] X;  // population-level design matrix
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  vector[N] Z_1_2;
  vector[N] Z_1_3;
  int<lower=1> NC_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
```

```
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
  int Kc;
  matrix[N, K - 1] Xc;  // centered version of X
  vector[K - 1] means_X;  // column means of X before centering
  Kc = K - 1;  // the intercept is removed from the design matrix
  for (i in 2:K) {
    means_X[i - 1] = mean(X[, i]);
    Xc[, i - 1] = X[, i] - means_X[i - 1];
  }
}
parameters {
  vector[Kc] b;  // population-level effects
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  matrix[M_1, N_1] z_1;  // unscaled group-level effects
  // cholesky factor of correlation matrix
  cholesky_factor_corr[M_1] L_1;
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  matrix[N_1, M_1] r_1;
  vector[N_1] r_1_1;
  vector[N_1] r_1_2;
  vector[N_1] r_1_3;
  // group-level effects
  vector[N_2] r_2_1;
  r_1 = (diag_pre_multiply(sd_1, L_1) * z_1)';
  r_1_1 = r_1[, 1];
  r_1_2 = r_1[, 2];
```

```
  r_1_3 = r_1[, 3];
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = Xc * b + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_1_2[J_1[n]]) * Z_1_2[n] + (
        r_1_3[J_1[n]]) * Z_1_3[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  b[1] ~ normal(0, 1e+10);
  b[2] ~ normal(0, 1e+10);
  temp_Intercept ~ normal(0, 1e+10);
  sd_1 ~ student_t(3, 0, 10);
  L_1 ~ lkj_corr_cholesky(1);
  to_vector(z_1) ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  corr_matrix[M_1] Cor_1;
  vector<lower=-1,upper=1>[NC_1] cor_1;
  b_Intercept = temp_Intercept - dot_product(means_X, b);
  // take only relevant parts of correlation matrix
  Cor_1 = multiply_lower_tri_self_transpose(L_1);
  cor_1[1] = Cor_1[1,2];
  cor_1[2] = Cor_1[1,3];
  cor_1[3] = Cor_1[2,3];
}
```

F.1.2 Informed model

```
// generated with brms 1.5.0
functions {
}
data {
  int<lower=1> N;  // total number of observations
  int Y[N];  // response variable
  int<lower=1> K;  // number of population-level effects
  matrix[N, K] X;  // population-level design matrix
  // data for group-level effects of ID 1
  int<lower=1> J_1[N];
  int<lower=1> N_1;
  int<lower=1> M_1;
  vector[N] Z_1_1;
  vector[N] Z_1_2;
  vector[N] Z_1_3;
  int<lower=1> NC_1;
  // data for group-level effects of ID 2
  int<lower=1> J_2[N];
  int<lower=1> N_2;
  int<lower=1> M_2;
  vector[N] Z_2_1;
  int prior_only;  // should the likelihood be ignored?
}
transformed data {
  int Kc;
  matrix[N, K - 1] Xc;  // centered version of X
  vector[K - 1] means_X;  // column means of X before centering
  Kc = K - 1;  // the intercept is removed from the design matrix
  for (i in 2:K) {
    means_X[i - 1] = mean(X[, i]);
    Xc[, i - 1] = X[, i] - means_X[i - 1];
  }
}
```

```
parameters {
  vector[Kc] b;  // population-level effects
  real temp_Intercept;  // temporary intercept
  vector<lower=0>[M_1] sd_1;  // group-level standard deviations
  matrix[M_1, N_1] z_1;  // unscaled group-level effects
  // cholesky factor of correlation matrix
  cholesky_factor_corr[M_1] L_1;
  vector<lower=0>[M_2] sd_2;  // group-level standard deviations
  vector[N_2] z_2[M_2];  // unscaled group-level effects
}
transformed parameters {
  // group-level effects
  matrix[N_1, M_1] r_1;
  vector[N_1] r_1_1;
  vector[N_1] r_1_2;
  vector[N_1] r_1_3;
  // group-level effects
  vector[N_2] r_2_1;
  r_1 = (diag_pre_multiply(sd_1, L_1) * z_1)';
  r_1_1 = r_1[, 1];
  r_1_2 = r_1[, 2];
  r_1_3 = r_1[, 3];
  r_2_1 = sd_2[1] * (z_2[1]);
}
model {
  vector[N] mu;
  mu = Xc * b + temp_Intercept;
  for (n in 1:N) {
    mu[n] = mu[n] + (r_1_1[J_1[n]]) * Z_1_1[n] + (r_1_2[J_1[n]]) * Z_1_2[n] + (
        r_1_3[J_1[n]]) * Z_1_3[n] + (r_2_1[J_2[n]]) * Z_2_1[n];
  }
  // prior specifications
  b[1] ~ normal(0, 1);
  b[2] ~ normal(-1, 1);
  temp_Intercept ~ normal(1, 1);
```

```
  sd_1 ~ student_t(3, 0, 10);
  L_1 ~ lkj_corr_cholesky(1);
  to_vector(z_1) ~ normal(0, 1);
  sd_2 ~ student_t(3, 0, 10);
  z_2[1] ~ normal(0, 1);
  // likelihood contribution
  if (!prior_only) {
    Y ~ bernoulli_logit(mu);
  }
}
generated quantities {
  real b_Intercept;  // population-level intercept
  corr_matrix[M_1] Cor_1;
  vector<lower=-1,upper=1>[NC_1] cor_1;
  b_Intercept = temp_Intercept - dot_product(means_X, b);
  // take only relevant parts of correlation matrix
  Cor_1 = multiply_lower_tri_self_transpose(L_1);
  cor_1[1] = Cor_1[1,2];
  cor_1[2] = Cor_1[1,3];
  cor_1[3] = Cor_1[2,3];
}
```