

Apagamento vocálico e binariedade no português: uma investigação baseada em preditivas Bayesianas*

Vowel deletion and binarity in portuguese: a study based on Bayesian predictive inference

Maria Bernadete Marques Abaurre
(Unicamp, IEL, DL)

Filomena Sandalo
(Unicamp, IEL, DL)

Verónica González-López
(Unicamp, IMECC, DE)

RESUMO

As aplicações de modelos estatísticos em fonologia de laboratório têm se mostrado produtivas na avaliações do poder de predição de hipóteses. Neste trabalho, a força da restrição FootBin foi testada estatisticamente, no português brasileiro e no português europeu, a partir de preditivas Bayesianas. Partimos do pressuposto de que, se a restrição FootBin for ativa, há uma preferência, na língua observada, por palavras com um número par de sílabas. Com base neste pressuposto, o estudo foi realizado a partir de dados experimentais (apagamento/redução vocálica) analisados acusticamente e tratados estatisticamente. Nossos resultados estatísticos apontam para o fato de FootBin estar ativa nas duas variedades de português, embora a força para binariedade se manifeste de maneira mais forte em PB. O fato de FootBin estar ativa nas duas variedades da língua com forças diferentes suporta uma análise otimalista baseada em restrições.

Palavras-chave: *Fonologia de laboratório, Inferência preditiva Bayesiana, FootBin, Apagamento / redução vocálica em português.*

* Esta pesquisa foi financiada pela FAPESP e pelo CNPq. Uma versão preliminar deste trabalho foi apresentada no PaPI'07, Phonetics and Phonology in Iberia, Braga, Portugal. Agradecemos os comentários recebidos nesse evento.

ABSTRACT

The applications of statistical models in laboratory phonology have been shown to be productive in evaluating the predictive power of hypotheses. In this paper, the strength of the restriction FootBin was tested statistically in Brazilian and European Portuguese, on the basis of a Bayesian predictive inference model. We assumed that if the restriction FootBin is active there is a preference, in a language, for words with an even number of syllables. This study, based on such assumption, was conducted on the basis of acoustic and statistical analysis of data from an experiment on vowel deletion / reduction. Our statistical results show that FootBin is active in both PB and EP, although the strength of binarity is stronger in PB. The fact that FootBin is active in both varieties with a different strength in BP and in EP supports an optimality constraint-based analysis.

Key-words: *Laboratory phonology, Bayesian predictive inference, FootBin, Vowel deletion / reduction in Portuguese*

1. Introdução

A literatura sobre padrões rítmicos do português brasileiro (PB) aponta para o fato de esta língua apresentar um padrão binário, tanto em termos de atribuição de acentos primários (cf. Wetzels 1997, Bisol 1992, 2000, Massini-Cagliari 1995, entre outros), como de secundários (cf. Carvalho 1988, Collischonn 1993, Abaurre & Galves 1998, entre outros). Já a literatura sobre o português europeu (PE) é controversa. D'Andrade & Laks (1991) afirmam que os acentos secundários do PE são atribuídos via pés binários. Carvalho (1988, 1989) afirma que são atribuídos via pés ternários. Mais recentemente, Frota (1998) e Vigário (1998) argumentam a favor de pés ilimitados para acentos secundários no PE.

Vários autores apontam para uma possível correlação entre apagamento vocálico e organização rítmica (cf. Kager 1997). Em Sandalo et alii (2006), chama-se a atenção para o fato de que a vogal /i/ e outras vogais não baixas podem ser elididas no português do Brasil sob condicionamento rítmico, como discutido a seguir. Em Matos & Sandalo (2006), são apresentados resultados de um estudo exploratório da relação entre binariedade e apagamento vocálico.

Bisol (1991) observou que /i/ é apagado no contexto [t_s], e Bisol & Hora (1993) notaram ainda o apagamento de vogais não baixas entre consoantes coronais homorgânicas. O apagamento discutido por esses autores ocorre independentemente da estrutura métrica e do número de sílabas (cf. *satisfatória* e *satisfeito*, onde o /i/ pode sempre ser apagado entre o /t/ e o /s/). Sandalo & alii (2006) observam, no entanto, que há outros contextos que permitem o apagamento de vogal apenas se o número de sílabas da palavra for ímpar. Considerando-se, por exemplo, o par *modernização* (cinco sílabas) e *modernizaria* (seis sílabas), com acento secundário na sílaba inicial, observa-se que o apagamento de /i/ tende a ocorrer apenas no primeiro caso. Em Sandalo et alii (2006), rotulam-se estes últimos casos de apagamento rítmico e formula-se a hipótese de que a redução vocálica otimiza uma tendência para binariedade.

Este trabalho testa estatisticamente a afirmação de que o português é uma língua de padrão rítmico binário. Partindo da Teoria da Otimidade, que sustenta que as línguas são geradas a partir de restrições hierarquicamente ordenadas, assumimos que há uma restrição rotulada de *FootBin* que milita por binariedade. Se esta restrição é alta na hierarquia, esperaríamos que palavras com número ímpar de sílabas fossem evitadas, como sugerido em Sandalo et alii (2006).

Com base em uma busca, em *corpora* processados acusticamente, por ocorrências de apagamento/redução drástica de vogais em PB e PE, buscaremos responder especificamente às seguintes perguntas a partir de uma análise estatística preditiva Bayesiana (cf. Gupta & Nadarajah 2004):

1. Essa pressão que milita a favor de binariedade existe de fato no português?
2. Caso exista, ela se manifesta apenas em PB ou em ambas as variedades?
3. Se a pressão existe em ambas as variedades, manifesta-se da mesma maneira ou de forma distinta?

A análise é baseada na investigação acústica de 426 sentenças (2502 palavras lexicais), lidas por falantes nativos de PB e de PE. O software PRAAT foi usado na identificação de ocorrências de apaga-

mentos/redução de vogais. Em relação à metodologia estatística preditiva escolhida para o tratamento destes dados, consideramos como mais apropriada a abordagem através de preditivas construídas pela distribuição beta-binomial. As distribuições beta-binomiais permitem definir intervalos de confiança de natureza Bayesiana que levam em conta o tamanho reduzido do banco de dados. A construção da preditiva é feita utilizando o banco de dados do PE para fazer predição do resultado observado no PB. Em seguida, é feito o caminho inverso. Construímos a partir dos dados do PB a preditiva do PE e comparamos a predição com o valor observado.

Este texto está organizado da seguinte maneira: na seção 2, apresentamos um resumo dos trabalhos que fazem referência à relação entre acento secundário e apagamento vocálico. Na seção 3, apresentamos o *corpus* e, na seção 4, a análise estatística baseada nas preditivas Bayesianas. Finalmente, na seção 5, procuramos mostrar que os resultados obtidos a partir da análise estatística favorecem um modelo fonológico representacional baseado em restrições.

2. Apagamento vocálico, acento secundário e ritmo em português e espanhol

As análises do ritmo do português em geral propõem que o acento secundário no português do Brasil segue um padrão binário (pés de duas sílabas), raramente violado. As exceções a esse sistema binário são constituídas principalmente pelos chamados casos de dátilos iniciais (Prince 1983, Collischonn, 1993). O dátilo inicial (pé trissilábico com acento na primeira sílaba) não é, no entanto, obrigatório. Uma palavra como *abacaxi*, por exemplo, pode ser acentuada como em *abacaxi*, apresentando um dátilo inicial, ou como *abacaxi*.¹

Sabe-se que o espanhol apresenta o mesmo fenômeno (Harris 1983, 1989, Roca 1986). De acordo com Harris (1983), assumindo uma análise derivacional, o acento secundário em espanhol é atribuído pela construção de pés troqueus da direita para a esquerda, na cadeia das sílabas que precedem a sílaba portadora do acento primário. Se for

1. Nos exemplos, negrito indica acento primário e sublinhado indica acento secundário.

permitida a ocorrência de pés degenerados em um estágio intermediário da derivação, ter-se-á, como resultado, o tipo de choque acentual ilustrado em (1), abaixo. Segundo Harris, dátilos iniciais podem então ser derivados pela aplicação de uma regra de desacentuação para a direita e por um novo parseamento, como se vê em (2), onde a sílaba medial (ti) não é parseada. A outra opção, em uma análise derivacional, é resolver o choque acentual com desacentuação para a esquerda, como mostrado em (3).

(1)

(x)
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ
<u>cons</u>	tan	ti	<u>no</u>	po	<u>li</u>	ta	nis	mo		

(2)

(x)
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ
<u>cons</u>	tan	ti	<u>no</u>	po	<u>li</u>	ta	nis	mo			

(x)
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ
<u>cons</u>	tan	ti	<u>no</u>	po	<u>li</u>	ta	nis	mo			

(3)

(x)
(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)	(x)
σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ	σ
cons	<u>tan</u>	ti	<u>no</u>	po	<u>li</u>	ta	nis	mo			

Com relação a esta análise, Hayes (1995: 97) observa que “the crucial point of Harris’s analysis is that it relies on a temporary degenerate foot, set up in the middle of the derivation (1), that either is expanded into a proper foot by destressing and reparsing, or is itself deleted.” Note-se que, em nenhum dos dois casos, o pé degenerado se superfi-

cializa, o que, segundo Hayes, mostra que o ponto crucial da fonologia do espanhol é a presença de uma restrição que bane pés degenerados, ou seja, uma restrição que milita a favor de binariedade.

Pode-se argumentar que a análise de Harris poderia dar conta dos fatos do português. Sandalo et alii (2006), no entanto, mostram que essa análise enfrenta problemas empíricos. Uma análise acústica dos fatos do PB realizada por Sandalo et alii indica que muitas palavras que contêm um número ímpar de sílabas sofrem um processo de apagamento de vogal. O fato de que a vogal /i/ e outras vogais não baixas podem ser elididas em posição não acentuada no contexto t_s foi observado por Bisol (1991) e o fato de que vogais não baixas podem ser apagadas entre consoantes (coronais) homorgânicas foi apontado por Bisol & Hora (1993). O apagamento discutido por esses autores ocorre independentemente da estrutura métrica (cf. *satisfatória* e *satisfeito*, onde o /i/ pode sempre ser apagado entre o /t/ e o /s/). Sandalo et alii observam, no entanto, que há outros contextos que permitem o apagamento de vogal se o número de sílabas da palavra for ímpar. Note que a sílaba que na análise de Harris é deixada não parseada é aquela que, no PB, em alguns contextos, não se realiza porque sofre apagamento. Assim, uma das possíveis realizações de uma palavra como *constantinopolitano* é *constan[tn]opolitano*, com apagamento da vogal /i/ entre as consoantes coronais homorgânicas /t/ e /n/. Desse apagamento resulta uma estrutura binária perfeita, em termos de pés:

(4) ((constant)Σ (nopo)Σ (lita)Σ (nismo)Σ)ω.

Alguém poderia argumentar que a estratégia empregada pelo português brasileiro para evitar pés degenerados é o apagamento de vogal não parseada (em vez do simples reparseamento). Assim, uma análise como a de Harris poderia ser proposta, contanto que incorporasse uma regra de apagamento de /i/.

O fenômeno do apagamento de vogais em PB, no entanto, indica que os fatos são mais complexos do que uma análise métrica poderia prever. As palavras contendo um número ímpar de sílabas podem ser o alvo para o apagamento rítmico de vogal, o que sugere que de fato

estamos diante de uma língua que prefere evitar pés degenerados, como sugere Harris. Ou seja, que sofre uma pressão por binariedade.

Considerando-se, por exemplo, o par *modernização* (cinco sílabas) e *modernizaria* (seis sílabas), com acento secundário na sílaba inicial, observa-se que o apagamento de /i/ tende a ocorrer apenas no primeiro caso, ainda que, dentro de uma análise derivacional como a de Harris apresentada acima, o apagamento desta vogal não seja previsto uma vez que a sílaba que tem como núcleo esta vogal está em posição forte (i.e., é portadora de acento secundário).

(5)

(x)
(x)	(x)
σ	σ	σ	σ	σ	σ
mo der ni za ção					

(6)

(x)
(x)	(x)	(
σ	σ	σ	σ	σ	σ	σ
mo der ni za ri a						

Portanto, a realização registrada por Sandalo et alii (2006) para a palavra *modernização*, retomada em (7) abaixo, sentença retirada do corpus utilizado por esses autores, é problemática para a Teoria Métrica porque, se o acento secundário resulta de uma alternância na ocorrência de acento e não-acento da direita para a esquerda nas sílabas que precedem a sílaba portadora de acento primário, não haveria razão para o apagamento de vogal, uma vez que existem quatro sílabas precedendo a sílaba com o acento primário em *modernização* e, portanto, uma alternância binária perfeita resultaria de um parseamento métrico das sílabas pretônicas.

(7) A modernização foi satisfatória
a mo der[n z]a ção foi sa[ts] fa tó ria

Em (7), acima, o primeiro apagamento vocálico parece ser ritmicamente induzido, e somente faz sentido se assumirmos que há uma restrição que força a ocorrência de pés binários na palavra como um todo (i.e. (mo dern) Σ (za ção) Σ) ω), não sendo necessário introduzir direcionalidade (i.e., contagem da direita para a esquerda a partir do acento primário) para obter binariedade por meio da alternância perfeita entre sílabas fortes e fracas, como previsto por uma análise derivacional baseada na Teoria Métrica, como a análise de Harris exemplificada em (5). Basta uma restrição, como a prevista pela Teoria da Otimalidade (*FootBin*), que milita por binariedade.

Uma evidência adicional para uma análise não métrica vem de palavras com número ímpar de sílabas que não apresentam redução vocálica no que, pela análise de Harris seria a vogal não parseada do possível pé ternário inicial, mas apagam a vogal em outra posição desde que ela seja alta:

(8) (gra ma) Σ (ti ca) Σ (lidad) Σ) ω

A análise da Harris prevê apagamento do /i/ da sílaba *ti*, uma vez que a análise daquele autor prevê que esta sílaba possa ficar não parseada em uma dada derivação, como em (2). No entanto, no corpus aqui analisado (cf. 8), o apagamento ocorre na posição final, o que não é previsto pela análise de Harris.

Sandalo et alii (2006) afirmam ainda que palavras com número par de sílabas parecem resistir ao apagamento de vogais. Interpretamos este fato como indicativo de que o apagamento rítmico não se faz necessário, nesses casos, pelo fato de essas palavras já terem uma estrutura binária (e.g., (mo der) (ni za) (ri a)) ω).

Portanto, o fato de que o apagamento rítmico pode ocorrer quando o número total de sílabas da palavra é ímpar pode ser tomado como evidência contra uma análise derivacional do acento, segundo a qual o acento primário é gerado pela construção de pés no léxico e somente as vogais pretônicas constituem o domínio de aplicação do algoritmo de acento secundário no nível pós-lexical, conforme proposto por Harris (1989) para o espanhol e por Collischonn (1993) para o PB.

Neste trabalho, assumimos que a restrição *FootBin*, que milita a favor da binariedade, toma inteiras palavras como domínios e não apenas sílabas pretônicas, como nas análises métricas derivacionais.

No português europeu não há um consenso sobre a implementação de um padrão rítmico binário. D'Andrade & Laks (1991) propõem que o acento secundário seja atribuído, em PE, através da construção de pés binários. Carvalho (1988, 1989) sustenta, por outro lado, que o acento secundário é atribuído através de pés ternários. Frota (1998) e Vigário (1998) assumem que a acentuação secundária não é obrigatória, em PE e, se ocorre, tende a manifestar-se em pés ilimitados. Vigário (2003) afirma que a posição inicial de palavras fonológicas porta acento não-primário em PE.

Em suma, vimos que trabalhos anteriores apontam para a existência no PB de um fenômeno de apagamento de vogais ritmicamente motivado. Sabe-se que o PE também apresenta apagamento vocálico (Mateus 1975, Mateus et alii 1983, Mateus & D'Andrade 2000). Acreditamos que o apagamento de vogais também é influenciado pelo ritmo no PE porque, de acordo com Vigário (2003), o apagamento de [i] e [u] é muito produtivo em PE, mas, segundo essa autora, “the presence of a prosodic word initial (non-primary) stress disfavors the deletion of [i] and [u] in prosodic word initial position”. Embora a afirmação de Vigário indique que o apagamento de vogal e o acento são fenômenos relacionados no PE, não é claro se tais apagamentos também podem ser atribuídos a uma pressão da língua em direção a um ritmo binário.

3. O corpus

Para este estudo, analisamos dois *corpora* de português, um contendo 20 sentenças e o outro com 54 sentenças. O primeiro *corpus*, inicialmente utilizado pelo projeto temático “Padrões Rítmicos, Fixação de Parâmetros e Mudança Lingüística” (Projeto FAPESP 98/3382-0), foi elaborado por Marina Vigário (Universidade de Lisboa), Sónia Frota (Universidade de Lisboa) e Charlotte Galves (Universidade Estadual de Campinas). É composto por 20 frases lidas três vezes por dois informantes do PB e do PE de mesma faixa etária, mesmo grau de escolaridade e mesmo sexo (feminino). O segundo *corpus*, traduzido

e adaptado para o português com base no trabalho de Ramus et alii (1999) por Sónia Frota e Flaviane Fernandes-Svartman (Universidade de São Paulo), é composto por 54 frases lidas duas vezes pelos mesmos informantes do primeiro *corpus*.

Todas as sentenças foram lidas em taxa de elocução normal, e são declarativas em foco largo, ou seja, respostas à pergunta “O que aconteceu?”. Também o número de sílabas é controlado, para que haja um número equivalente de palavras com número ímpar e par de sílabas.

O programa PRAAT (<http://www.fon.hum.uva.nl/praat/>) foi usado para a análise acústica e identificação de ocorrências de apagamento ou redução drástica de vogais. Na nossa investigação, qualquer ocorrência de apagamento/redução de vogais foi considerada, independentemente da posição da sílaba na palavra. Foram analisadas apenas palavras lexicais.

Com estes dados, totalizamos 2502 palavras lexicais que foram analisadas acusticamente. 220 palavras lexicais do BP sofreram apagamento, sendo que 132 palavras ficaram com número par de sílabas após sofrerem apagamento/redução drástica de vogais, e 88 ficaram com número ímpar de sílabas.

Um número muito maior de palavras do PE sofreu o processo de apagamento (622), sendo que 356 apresentaram um número par de sílabas depois de sofrer apagamento, e 266 apresentaram um número ímpar.

Abaixo apresentamos, à guisa de exemplo, duas tabelas (cf. Matos & Sandalo 2006) representativas do corpus de 20 sentenças com indicação dos apagamentos ocorridos em PB e PE. Uma primeira observação dos dados permite notar que PE sofre mais frequentemente o processo de apagamento de vogais e que PB parece apresentar uma tendência para apagamento de vogais em palavras com número ímpar de sílabas. As Tabelas 1 e 2 estão organizadas da seguinte maneira: na primeira coluna apresentamos algumas sentenças do corpus com a sílaba que teve uma vogal apagada indicada em *itálico*. Na segunda coluna indicamos o número de sílabas da palavra antes e depois do processo de apagamento, considerando que o apagamento do núcleo da sílaba leva ao apagamento daquela sílaba como um todo.

Tabela 1: PB, falante 1, terceira leitura

A modernização foi <i>satisfatória</i>	5 → 4
O professor também o descreveu aos <i>alunos</i>	3 → 2
O governador aceitou a <i>modernização</i>	5 → 4
A falta de <i>autoridade</i> foi alarmante	5 → 4
O investigador já lhe devolveu o <i>dinheiro</i>	3 → 2
A professora discutiu a <i>gramaticalidade</i>	7 → 6
A procura da <i>gramaticalidade</i> é nosso objetivo	7 → 6
O professor também <i>descreveu</i> os <i>alunos</i>	3 → 2 3 → 2
A <i>autoridade</i> cabe ao governador	5 → 4
A <i>gramaticalidade</i> das frases foi conseguida	7 → 6 4 → 3

Tabela 2: PE, falante 2, segunda leitura

A modernização foi <i>satisfatória</i>	5 → 4
A <i>autoridade</i> do governador diminuiu	5 → 4
O investigador já <i>devolveu</i> o <i>dinheiro</i>	5 → 4 3 → 2 3 → 1
O organizador <i>apresentou</i> a <i>catalogadora</i>	4 → 3 6 → 5
O <i>trabalho</i> da <i>pesquisadora</i> foi <i>publicado</i>	3 → 2 5 → 4 4 → 3
O professor também o <i>descreveu</i> aos <i>alunos</i>	3 → 2 (2x)
O governador aceitou a <i>modernização</i>	4 → 3 5 → 4
A falta de <i>autoridade</i> foi alarmante	5 → 3 4 → 3
O investigador já lhe <i>devolveu</i> o <i>dinheiro</i>	5 → 4 3 → 2 3 → 2
A <i>catalogadora</i> compreendeu o <i>trabalho</i> da <i>pesquisadora</i>	6 → 5 3 → 2 5 → 4
A <i>professora</i> <i>discutiu</i> a <i>gramaticalidade</i>	4 → 2 3 → 1 7 → 6
A <i>inteligência</i> da <i>catalogadora</i> foi <i>determinante</i>	5 → 4 5 → 2
O investigador já <i>ofereceu</i> <i>dinheiro</i>	5 → 4 4 → 2 3 → 1
A procura da <i>gramaticalidade</i> é o nosso objetivo	2 → 1 4 → 3
A <i>pesquisadora</i> perdeu a <i>autoridade</i>	5 → 4 (2x)
O professor também <i>descreveu</i> os <i>alunos</i>	3 → 2 (3x)
A <i>autoridade</i> cabe ao governador	5 → 3 4 → 2
O investigador já me <i>ofereceu</i> <i>dinheiro</i>	5 → 4 4 → 2 3 → 1
A <i>gramaticalidade</i> das frases foi conseguida	7 → 6 2 → 1 4 → 3

Considerando-se a natureza dos dados disponíveis nos *corpora* tomados para análise, uma forma de extrair informações estatísticas seria mediante a construção de intervalos de confiança das proporções de apagamentos que resultam em palavras com um número par de sílabas, buscando observar, assim, a força da restrição *FootBin* em ambas

as modalidades. Abaixo, na Tabela 3, mostram-se os resultados das estimativas das proporções de apagamentos que resultam em palavras com um número par de sílabas. Além disso, apresentam-se os intervalos de confiança para tais proporções considerando-se níveis tradicionais de confiança 0.99, 0.95 e 0.90.

Tabela 3: Intervalos de confiança frequentistas para proporções de apagamentos que resultam em palavras com número par de sílabas

Confiança	Variedade	Estimativa da Proporção	Limite Inferior	Limite superior
0,9	PB	0,6	0,55	0,65
	PE	0,57	0,54	0,60
0,95	PB	0,6	0,54	0,66
	PE	0,57	0,53	0,61
0,99	PB	0,6	0,51	0,69
	PE	0,57	0,52	0,62

Conforme se pode observar, os intervalos de confiança construídos incluem ambas as estimativas de proporção, o que demonstra que esta ferramenta é insensível às possíveis diferenças existentes entre as proporções estimadas no que diz respeito à força de *FootBin*.

Na seção seguinte, propõe-se uma ferramenta alternativa que permite uma leitura alternativa das proporções obtidas. Com base na problemática Bayesiana, que assume a existência de uma conjectura ou distribuição a priori sobre o fenômeno, construímos regiões de credibilidade² com o propósito de identificar possíveis diferenças entre as proporções de ocorrência.

4. Análise estatística

A distribuição preditiva foi construída para cada banco de dados (PB e PE) e baseou-se no número de sucessos (apagamentos que

2. A inferência frequentista clássica apela para a utilização de intervalos de confiança. Esta noção é redefinida e reformulada na inferência Bayesiana, que adota o conceito “região ou intervalo de credibilidade”.

resultam em palavras com um número par de sílabas), os quais serão denotados por Y e Y' para PB e PE, respectivamente. Os valores efetivamente observados no banco de dados serão denotados por y e y' . O total de realizações (apagamentos) nos bancos de dados foi $n=220$ para o PB e $n'=622$ para o PE.

Conhecido o número de realizações n' , a preditiva construída com base nos dados do PB (que depende de n , y) será denotada por $\text{Pred}(Y'/n, y, n')$ (leia-se preditiva para a variável Y' num total de n' apagamentos, partindo da informação do PB: n e y).

A definição analítica da preditiva é baseada na informação de duas fontes: a) observação dos dados, b) conjectura linguística. A seguir listamos os elementos que compõem a preditiva:

1. A verossimilhança (no caso, denominada distribuição Binomial), ou seja, a distribuição do número de sucessos em n realizações independentes de um mesmo experimento que pode fornecer apenas dois resultados: sim ou não (sucesso ou fracasso), com a mesma probabilidade de sucesso p : $\text{Binomial}(n, p)$;
2. A distribuição a priori sobre p , que representa a conjectura linguística sobre as chances de sucesso, denotada por $\text{Beta}(\alpha, \beta)$;
3. A informação geral do banco de dados do PE: número de realizações (apagamentos), n' , e o número de sucessos observados (apagamentos que resultam em palavras com um número par de sílabas), y' ;
4. A preditiva, $\text{Pred}(Y'/n, y, n')$, é uma média entre a verossimilhança e a distribuição a priori Beta, sobre as chances de sucesso, p . Na literatura, esta preditiva recebe o nome de distribuição Beta-Binomial. Rigorosamente, a Beta Binomial depende dos valores dos hiperparâmetros α e β , ou seja, $\text{Pred}(Y'/n, y, n', \alpha, \beta)$.

4.1. Formulação (elicitación) da distribuição a priori

Mediante a especificação dos hiperparâmetros α e β da distribuição a priori, podem ser representadas diversas conjecturas linguísticas.

Por exemplo, assumindo $\alpha = \beta = 1$, indica-se que todos os valores de p têm a mesma chance, ou seja, em média, p é $1/2$. Essa especificação é conhecida como não informativa e a distribuição a priori Beta(1,1) coincide com a distribuição Uniforme sobre p , no intervalo $[0,1]$. Já para situações nas quais os hiperparâmetros são escolhidos segundo $\alpha \gg \beta > 1$, indica-se que é mais provável que $p = \text{Prob}(\text{sucesso})$ assuma valores próximos de 1 ou, equivalentemente, que a chance de fracasso $1-p = \text{Prob}(\text{fracasso})$ assume valores próximos de zero. Em oposição, se assumimos $1 < \alpha \ll \beta$, indicamos que $p = \text{Prob}(\text{sucesso})$ é baixa e $1-p = \text{Prob}(\text{fracasso})$ é alta.

4.2. Formulação matemática e regra de decisão

Partindo da informação do PB, digamos n e y , e sabendo que no banco de dados do PE o número de apagamentos foi n' , pode-se construir a predição da variável Y' (apagamentos que resultam em palavras com um número par de sílabas). Segundo a formulação da Beta-Binomial, o valor esperado, denotado por E , e a variância, denotada por V , de Y' , são dados por

$$\begin{aligned} E &= E(Y' / n, y, n', \alpha, \beta) \\ &= n'(\alpha + y) / (\alpha + \beta + n). \end{aligned}$$

$$\begin{aligned} V &= \text{Var}(Y' / n, y, n', \alpha, \beta) \\ &= n'(\alpha + y)(\beta + n - y)(\alpha + \beta + n + n') / ((\alpha + \beta + n)^2(\alpha + \beta + n + 1)). \end{aligned}$$

Se o número n' for suficientemente grande, $(Y' - E) / V^{1/2}$ tem distribuição aproximadamente Normal padrão.

4.3. Regra de decisão

Dado o nível de significância γ , o valor observado y' de Y' é considerado previsível a partir de Y (ou seja, de PB) se $(y' - E) / V^{1/2}$ encontra-se no intervalo de predição $[-z_{\gamma/2}, z_{\gamma/2}]$, onde $z_{\gamma/2}$ é o $1 - \gamma/2$ quantil da distribuição Normal padrão. Neste trabalho, ilustraremos os resulta-

dos utilizando três valores para γ . Os valores foram $\gamma=0.1$ ($z_{0.1}= 1.64$), $\gamma=0.05$ ($z_{0.05}= 1.96$), $\gamma=0.01$ ($z_{0.01}= 2.58$).

4.4. Resultados

Olharemos as preditivas do PB em relação ao PE, e de PE em relação a PB, segundo os valores indicados na Tabela 4 abaixo. O Gráfico 1 representa palavras cujo número de sílabas após processos de apagamento/redução drástica é par. Utilizamos as informações contidas na tabela 4 para construir as preditivas representadas no Gráfico 1.

Tabela 4: Informações de valores para as preditivas

	Gráfico 1: Resultados pares	
	PE	PB
n	622	220
y	356	132
n'	220	622
y'	132	356

No Gráfico 1 abaixo, nas linhas horizontais rotuladas PB e PE, as regiões preditivas, delimitadas pelos três diferentes níveis de confiança $1-\gamma$ (99%, 95% e 90%), estão representadas pelos cruzamentos de linhas verticais (rosa, roxo e verde, respectivamente). Cada uma dessas linhas delimitadas representa a predição que uma modalidade faz em relação à outra em termos da probabilidade de produzir apagamentos que resultam em palavras com número par de sílabas.

O zero indicado pelo eixo das ordenadas no Gráfico 1 é tomado como ponto de referência para a leitura das predições, isto é, ele é tomado como ponto ideal do fenômeno em estudo. À direita do zero, indicam-se valores referentes a uma tendência maior de ocorrência do fenômeno. À esquerda do zero, indicam-se valores referentes a uma tendência menor de ocorrência do fenômeno.

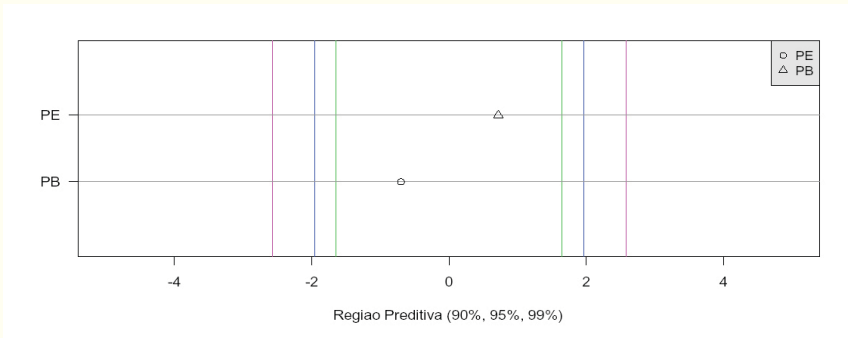


Gráfico 1: Preditivas feitas por PE e PB a partir de apagamentos que resultam em palavras com número par de sílabas.

A primeira observação é que ambas as modalidades têm uma tendência a realizar apagamentos de vogais para criar palavras com número par de sílabas, uma vez que ambas estão dentro da mesma região, para todos os três níveis de confiança. Esse resultado já era evidente a partir da Tabela 3, porém as diferenças entre as proporções não resultaram estatisticamente significantes, analisando-se apenas os intervalos frequentistas.

Observando os intervalos preditivos, podemos inferir uma leitura mais detalhada sobre as tendências das modalidades com relação à binariedade. Na linha horizontal PB, o zero no eixo horizontal das ordenadas representa o ponto de referência de apagamentos em direção a um número par de sílabas, para o PB. O círculo representa o número de apagamentos preditos por PB como resultando em palavras com um número par de sílabas no PE, neste caso localizado à esquerda do zero.

A predição feita a partir do PB (linha horizontal PB) é de que o PE efetua apagamentos que levam a um número par de sílabas em menor proporção do que os efetuados por PB.

Já na linha horizontal PE, o zero no eixo horizontal das ordenadas representa o ponto de referência de apagamentos em direção a par para o PE. O triângulo representa o número de apagamentos preditos por PE como resultando em palavras com um número par de sílabas no PB, neste caso localizado à direita do zero.

A predição feita a partir do PE (linha horizontal PE) é de que o PB efetua apagamentos que levam a um número par de sílabas em maior proporção do que os efetuados por PE.

Nossos resultados indicam, assim, que há de fato uma tendência maior do PB, com relação ao PE, em realizar apagamentos em direção a número par de sílabas.

5. Considerações linguísticas

Na análise otimalista de Sandalo et alii (2006) e Sandalo & Abaurre (2007), a força de *FootBin* foi assumida como bastante alta para o PB, mas como baixa em PE, embora ativa. Abaixo apresentamos as hierarquias de restrições para o PB e o PE, conforme propostas pelos autores:³

PB: DEPST : RIGHTMOST : CLASHINT >> INTLEX >> **BINGRAD** : PARSE :
NOLAPSE >> CLASHEXT >> TROCHEE >> ALIGN

PE: DEPST : RIGHTMOST : CLASHINT >> TROCHEE >> ALIGN : INTLEX >>
PARSE >> CLASHEXT >> **FOOTBIN** : NOLAPSE

A análise estatística tem como resultado mais significativo as predições feitas a partir do PB com relação ao PE, e a partir do PE com relação ao PB, no tocante à tendência para a realização de apagamentos de vogais que levam a palavras com número par de sílabas. Repetimos abaixo estas previsões, e a seguir elaboramos uma interpretação linguística a partir de tais previsões.

A predição feita a partir do PB (linha horizontal PB, no Gráfico 1) é de que o PE efetua apagamentos que levam a um número par de sílabas em menor proporção do que os efetuados por PB. Já a predição feita a partir do PE (linha horizontal PE, no Gráfico 1) é de que o PB

3. Embora as restrições tenham rótulos distintos em PB e PE (*BinGrad* e *FootBin*, respectivamente), ambas militam a favor de binariedade. *BinGrad* recebeu um nome distinto para explicitar que o cálculo de processamento de violações é gradiente em PB, isto é, tem uma força maior. A definição de cada uma destas restrições, em português, está em Sandalo & Abaurre (2007).

efetua apagamentos que levam a um número par de sílabas em maior proporção do que os efetuados por PE.

Uma interpretação possível é a de que tais previsões, obtidas a partir da inferência Bayesiana, corroboram as hierarquias propostas em Sandalo et alii (2006) e Sandalo & Abaurre (2007). Isto é, em ambas as variedades do português, há uma pressão para binariedade, o que se traduz na análise otimalista proposta pela presença ativa, nas duas hierarquias, da restrição *FootBin/BinGrad*. No entanto, pode-se concluir que há uma força maior desta pressão no PB.

Na análise de Sandalo et alii, a força maior da pressão para binariedade em PB é prevista através do lugar na hierarquia (posição mais alta) e através de um cálculo gradiente de violações. Cabe observar que esta análise otimalista foi feita a partir de julgamentos de gramaticalidade. Os resultados estatísticos obtidos neste trabalho corroboram esta análise no que diz respeito, especificamente, à posição de *FootBin/BinGrad* na hierarquia de restrições de cada uma das variedades.

6. Conclusão

Neste trabalho, procuramos avaliar a força de uma pressão a favor de um ritmo binário no português, como apontado em vários trabalhos sobre organização rítmica nesta língua. Optamos por verificar esta afirmação a partir da observação da ocorrência de apagamentos/redução drástica de vogais em dados de PB e de PE, uma vez que a literatura aponta para uma correlação entre apagamento vocálico e implementação de um determinado padrão rítmico. Partimos do pressuposto de que, se a restrição *FootBin* for ativa nesta língua, há uma preferência por palavras com um número par de sílabas, ou seja, constituídas de pés binários.

A força de *FootBin* foi testada estatisticamente a partir de preditivas Bayesianas em PB e em PE. Nossos resultados apontam para o fato de esta restrição estar ativa nas duas variedades do português, o que corrobora uma análise otimalista baseada em restrições já apresentada em Sandalo et alii (2006) e em Sandalo & Abaurre (2007). A análise estatística realizada, entretanto, revelou uma diferença entre PB e PE: a força para binariedade parece se manifestar de maneira mais forte

em PB, reforçando novamente a posição diferente da restrição *FootBin* nas hierarquias propostas na análise de Sandalo et alii e de Sandalo & Abaurre.

O estudo realizado por Sandalo et alii (2006) foi desenvolvido a partir da metodologia clássica nos estudos gerativistas, que parte da avaliação de gramaticalidade feita a partir de julgamentos de falantes nativos. Este trabalho agora realizado traz evidência estatística para essa análise linguística, demonstrando a importância da fonologia de laboratório (Pierrehumbert, Beckman & Ladd 2000) para o teste de modelos de análise. Assim, nossa investigação procura trazer evidência adicional para corroborar conjecturas fonológicas propostas em trabalhos anteriores, baseadas na metodologia hipotético-dedutiva. Neste sentido, vale lembrar aqui as palavras de Pierrehumbert, Beckman & Ladd (2000:273):

Research activities within laboratory phonology involve the cooperation of people who may disagree about phonological theory, but who share a concern for strengthening the scientific foundations of phonology through improved methodology, explicit modeling, and cumulation of results. These goals, we would argue, all reflect the belief that phonology is one of the natural sciences, and that all of language, including language specific characteristics and sociolinguistic variation, is part of the natural world. [...] We explore the ramifications of this position for the relationship of data and methods to phonological theory; for the denotations of entities in that theory; and for our understanding of Universal Grammar and linguistic competence.

Recebido em novembro de 2011

Aprovado em novembro de 2013

E-mail: bernadete.abaurre@gmail.com

Referências bibliográficas

- ABAURRE, M.B. & C. Galves. 1998. Rhythmic differences between European and Brazilian Portuguese: an optimalist and minimalist approach. *DELTA* 14: 377-403.
- BISOL, L. 1991. Palatalization and its variable restriction. *International Journal of the Sociology of Language*, 89: 107-124.

- _____. 1992. O acento e o pé métrico binário. *Cadernos de Estudos Lingüísticos*, 22: 69-80. Universidade Estadual de Campinas .
- _____. 2000. O troqueu silábico no sistema fonológico. *DELTA* 16: 403-413.
- BISOL, L. & D. Hora. 1993. Palatalização da oclusiva dental e fonologia lexical. *Letras* 5:25-40.
- CARVALHO, J. B. 1988. Réduction vocalique, quantité et accentuation: pour une explication structurale de la divergence entre portugais lusitanien et portugais brésilien. *Boletim de Filologia* 32:5-26.
- _____. 1989. Phonological conditions on portuguese clitic placement: on syntactic evidence for stress and rhythmical patterns. *Linguistics* 27: 405-436.
- COLLISCHONN, G. 1993. *Um estudo do acento secundário em português*. Universidade Federal do Rio Grande do Sul. Dissertação de mestrado.
- D'ANDRADE, E. & S.B. Laks. 1991. Na crista da onda: o acento de palavra em português. *Actas do 7º Encontro da Associação Portuguesa de Linguística*: 15-26.
- FROTA, S. 1998. *Prosody and Focus in European Portuguese*. Universidade de Lisboa. Tese de doutorado.
- GUPTA, A. & S. Nadarajah (orgs). 2004. *Handbook of Beta Distribution and its Applications*. New York: Marcel Dekker, Inc.
- HARRIS, J. W. 1983. *Syllable structure and stress in Spanish: a nonlinear analysis*. Cambridge, MA: MIT Press.
- _____. 1989. A podiatric note on secondary stress in Spanish. MIT ms.
- HAYES, B.. 1995. *Metrical stress theory: Principles and Case Studies*. Chicago: University of Chicago Press.
- KAGER, R. 1997. Rhythmic vowel deletion in Optimality Theory. In I. Roca (ed.), *Derivations and Constraints in Phonology*, 463-499. Oxford: Oxford University Press.
- MASSINI-CAGLIARI, G. 1995. *Cantigas de amigo; do ritmo poético ao lingüístico – um percurso histórico da acentuação em português*. Universidade Estadual de Campinas. Tese de doutorado.
- MATEUS, M. H. M. 1975. *Aspectos da Fonologia do Português*. Lisboa: Instituto Nacional de Investigação Científica.
- MATEUS, M. H. M. et alii. 1983. *Gramática da Língua Portuguesa*. Coimbra: Almedina.
- MATEUS, M.H. M. & D'Andrade, E. 2000. *The Phonology of Portuguese*. Oxford: Oxford University Press.
- MATOS, M.P. & F. Sandalo, 2006. Síncope vocálica no Português Brasileiro. In Anais do VI Encontro do CELSUL, Círculo de Estudos Linguísticos do Sul. Florianópolis: UFSC.

- PRINCE, A. 1983. Relating to the Grid. *Linguistic Inquiry* 14: 19-100.
- PIRREHUMBERT, J, M Beckman & D.R. Ladd. 2000. Conceptual Foundations of Phonology as a Laboratory Science: Conceptual and Empirical Issues. In: N. Burton-Roberts, P. Carr & G. Docherty (orgs), *Phonological Knowledge*, Oxford: Oxford University Press, pp. 273-303.
- RAMUS, F., Nespore, M. & Mehler, J. 1999. Correlates of Linguistic Rhythm in the Speech Signal, *Cognition* 73, 265–292.
- ROCA, I. M. 1986. Secondary Stress and Metrical Rhythm. *Phonology Yearbook* 3: 341- 370.
- SANDALO, F., Abaurre M.B., Mandel, A. & Galves, C. 2006. Secondary stress in two varieties of Portuguese and the Sotaq optimality based computer program. *Probus - International Journal of Latin and Romance Linguistics* 8.1: 97-125.
- SANDALO, F. & Abaurre, M.B. 2007. Acento secundário em duas variedades de português: uma análise baseada na OT. In: Araújo, G. A. (org) *O acento em português: abordagens fonológicas*. São Paulo: Parábola.
- VIGÁRIO, M. 1998. Cliticização no português europeu: uma operação pós-lexical. *Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística*, vol. I. Aveiro, Portugal, pp557-598.
- _____. 2003. *The prosodic word in European Portuguese*. Interface Explorations 6. Berlin: Mouton de Gruyter.
- WETZELS, L. 1997. The lexical representation of nasality in Brazilian Portuguese. *Probus* 9.2.