

# DINÂMICA SINCRÔNICA E DIACRÔNICA DA DERIVA DE ABERTURA DAS VOGAIS MÉDIAS TÔNICAS DO PORTUGUÊS<sup>1</sup>

Eleonora C. ALBANO

Lafape-IEL/Dinafon-CNPq<sup>2</sup>

## RESUMO

*A distinção portuguesa entre as vogais médias abertas e fechadas é em grande parte inovadora. Com a perda do acento quantitativo latino, alguns contrastes desapareceram e outros emergiram via analogia, não só mórfica como fônica. A cadeia milenar de mudanças aí envolvidas pode ser entendida como uma deriva de abertura na posição tônica. Este estudo reconstrói a trajetória dessa deriva através da análise estatística de um corpus do português brasileiro falado atual. As distribuições de segmentos fônicos nos tipos e ocorrências do corpus são comparadas em diferentes ambientes prosódicos lexicais para mostrar que a deriva evolui de acordo com previsões de Zipf, mas responde a desequilíbrios do sistema racionalizáveis pelas noções de marca e carga funcional de Trubetzkoy. A sua atuação sobre neologismos, empréstimos e demais inovações é o lado zipfiano do quadro. A sua ação para restaurar a carga funcional dos contrastes na posição acentual e no tipo silábico não-marcados é o lado trubetzkoyano do quadro. A análise se completa com uma reconstrução do papel do detalhe fonético em atrelar a deriva à posição tônica durante a transição do acento quantitativo para o acento intensivo que vai do romance ibérico ao galego-português. A síntese teórica e empírica por fim alcançada deve a sua inspiração aos modelos fonológicos dinâmicos atuais.*

---

<sup>1</sup> Agradeço a Didier Demolin a leitura e discussão do texto. Todos os erros são de minha responsabilidade.

<sup>2</sup> Processo no. 311154/2009-3.

## ABSTRACT

*The Portuguese distinction between open and close mid vowels is to a great extent innovative. With the loss of Latin quantitative stress, some contrasts have disappeared and others have emerged through analogy, both morphological and phonetic. The millennial change chain involved therein can be understood as a drift for vowels to open under stress. This study reconstructs the trajectory of such a drift through a statistical analysis of a corpus of contemporary spoken Brazilian Portuguese. Distributions of segments in the types and tokens of the corpus are compared in different lexical prosodic environments to show that the drift evolves according to Zipf's predictions, but responds to imbalances explainable by Trubetzkoy's notions of mark and functional load. Its operation on neologisms, loans and other innovations is the zipfian side of the picture. Its action to restore the functional load of the contrasts in the unmarked stress position and syllable type is the trubetzkoyan side of the picture. The analysis is completed with a reconstruction of the role of phonetic detail in tying the drift to stressed position during the transition from quantitative to intensive stress extending from Iberian romance through Galician-Portuguese. The theoretical and empirical synthesis finally achieved owes its inspiration to current dynamical phonological models.*

## PALAVRAS-CHAVE

*Abertura. Acento. Deriva. Estatística fônica. História do português. Vogais médias.*

## KEY WORDS:

*Drift. Mid Vowels. Opening. Stress. Portuguese Language History. Segment Statistics.*

## Introdução

Embora o método experimental venha sendo advogado há quase quarenta anos por estudiosos da relação entre a fonética e a fonologia (OHALA, 1974), a sua disseminação pela última é recente e não consensual. Além disso, a formação em ambas as áreas ainda é rara, visto que, tradicionalmente, a fonética se afilia às ciências naturais, e a fonologia, às ciências humanas. Assim, o interesse crescente dos fonólogos pelo método requer atenção<sup>3</sup>.

Essa tendência atual deve-se em parte ao sucesso do movimento do software livre, que vem disponibilizando aos pesquisadores ferramentas que prometem apagar as fronteiras entre os gabinetes e os laboratórios. Tal conquista é tanto mais oportuna quanto maior é a promessa de que também as fronteiras entre a fonética e a fonologia venham a se apagar. Mas, para abraçá-la com fôlego, ainda há muito a refletir sobre o que seria de fato uma fonologia experimental.

Este ensaio focaliza o papel da estatística fônica nesse cenário, com base no estudo de um dos aspectos mais típicos da fonologia do português: o contraste entre as quatro vogais médias, /e, ε, o, ɔ/, e a problemática da sua manutenção e evolução histórica em ambientes tônicos e neutralização em ambientes átonos.

A estatística fônica – *i.e.*, o levantamento da frequência de ocorrência e o tratamento estatístico das unidades fônicas e suas combinações – foi objeto de debate entre dois dos pioneiros da fonologia. O primeiro é não menos que o seu aclamado pai: Nicolai Trubetzkoy. O segundo é um dos malditos do campo: George Kingsley Zipf, cujos esforços em prol da quantificação das unidades linguísticas são hoje, ainda, mais reconhecidos nas

---

<sup>3</sup> V., p.ex., COETZEE, KAGER & PATER (2009).

ciências naturais e sociais do que na linguística.

Assim como Trubetzkoy, Zipf adquiriu influência crescente depois de uma morte prematura. A sua defesa do uso da estatística para formular leis linguísticas afastou-o do estruturalismo e levou-o a explorar outras áreas do saber, onde, curiosamente, encontrou fenômenos semelhantes. É digno de nota, portanto, que o fundador da fonologia não o tenha ignorado.

O referido debate é aqui mobilizado para respaldar uma análise das distribuições das vogais /e, ε, o, ɔ/ e suas combinações no português brasileiro falado, calculadas de duas formas já clássicas: sobre a lista do vocabulário – os chamados tipos, reveladores de vieses fônicos lexicais, objeto da reflexão de Trubetzkoy –; ou sobre o total do corpus, incluindo repetições – as chamadas ocorrências, reveladoras de vieses fônicos discursivos, objeto da reflexão de Zipf.

Esta análise tenta integrar aspectos complementares das duas abordagens a fim de iluminar a diacronia do vocalismo português através da sua estatística fônica sincrônica. Cabe destacar que iniciativas como esta têm raros precedentes na literatura devido à novidade e à dispersão do conhecimento necessário à integração buscada. Basta lembrar que, mal Trubetzkoy instituir o primado da sincronia no estudo dos fatos fônicos, Zipf tentou resgatar a ótica diacrônica com a sua abordagem quantitativa.

O conhecimento que permite aproximar ideias aparentemente tão díspares foi construído nos últimos 40 anos e só em parte floresceu na linguística. Quatro linhas de pensamento contribuíram para sedimentá-lo. A primeira nasceu do fértil cruzamento entre a linguística histórica e a dialetologia; a segunda, da influência da neurofisiologia sobre a nascente ciência da computação; a terceira, do esforço de alguns

foneticistas experimentais interessados em fonologia; e a quarta, de um movimento transdisciplinar originado na física.

A primeira linha de pensamento originou-se na sociolinguística laboviana, cujos minuciosos estudos de mudanças fônicas em curso (LABOV, 1972) não só resgataram e ampliaram a visão, devida aos neogramáticos, da mudança como um processo gradual (OSTHOFF & BRUGMANN, 1967 [1878]), mas também tentaram coaduná-la com uma visão nova e aparentemente contraditória que, com base na história do chinês, reivindica a existência de mudanças foneticamente abruptas e lexicalmente graduais – fenômeno denominado difusão lexical (WANG, 1969, CHEN & WANG, 1975). A obra de Labov inspirou fortemente as atuais fonologias baseadas no uso, que incluem não só o trabalho pioneiro de BYBEE (2001), mas também a teoria dos exemplares (PIERREHUMBERT, 2001) e a sociofonética inglesa (FOULKES *et al.*, 2010).

A segunda linha de pensamento reside no connexionismo (RUMELHART *et al.*, 1986, ELMAN *et al.*, 1996), cujo nascimento foi bastante discutido em ALBANO (2009). Para os presentes fins, basta esclarecer que se trata de um tipo de computação, inspirado no funcionamento dos neurônios, que se tem revelado capaz de simular vários processos de aprendizagem, inclusive linguísticos. Particularmente relevante é o seu sucesso em extrair regularidades de paradigmas flexionais (p.ex., HARE *et al.*, 1995) – o que o habilita a operacionalizar a noção de analogia, cuja definição clássica carece de precisão.

A terceira linha de pensamento teve como foro a série bianual de colóquios denominada “*Conference on Laboratory Phonology*”, deslanchada por John Kingston e Mary Beckman em 1985. Foi nesse ambiente, constituído por foneticistas

empenhados em “naturalizar”<sup>4</sup> a Fonologia, que começou a ser sistematicamente posta em questão a hipótese, largamente aceita pela literatura de ambas as áreas, da existência de um mecanismo universal, puramente biológico, de “realização fonética”. Assim, passou-se a cultivar a hipótese alternativa de que a fonologia – *i.e.*, a lógica das propriedades fônicas que operam distinções lexicais em uma língua – atua “da gramática à física da fala” (KINGSTON & BECKMAN, 1990).

A quarta linha de pensamento deu sustentação teórica à mudança de paradigma assim delineada<sup>5</sup>. Trata-se da teoria dos sistemas dinâmicos, que, surgida na matemática para racionalizar observações da física, alastrou-se pelas ciências naturais, e delas para as biológicas e sociais (PORT & VAN GELDER, 1995). Essa teoria, que define como dinâmico todo sistema cuja variável independente é o tempo, prevê que a evolução temporal de parâmetros contínuos possa dar saltos discretos – previsão que promete iluminar a discussão, deslanchada por LABOV (1981), sobre as relações possíveis entre os dois contextos internos da mudança fônica: o fonético e o lexical.

Ora, a relevância do debate entre Trubetzkoy e Zipf para o novo paradigma reside em apontar a estatística como ferramenta capaz de investigar as proporções em que as distinções fônicas são utilizadas ou combinadas numa língua – seja no léxico, enfatizado pelo primeiro; seja no discurso oral ou escrito, enfatizados pelo segundo. Tais proporções revelam preferências motoras e/ou perceptivas que fornecem pistas sobre como cada língua lida, sincrônica e diacronicamente, com restrições de ordem biomecânica ou cognitiva tendentes a universais.

Cabe, a esse respeito, ressaltar a clarividência de ambos os

---

<sup>4</sup> Em resposta a uma longa discussão, iniciada por KIPARSKY (1973 [1968]), sobre a arbitrariedade do formalismo do campo.

<sup>5</sup> Nos termos de KUHN (1970).

pioneiros. Para Trubetzkoy, o caráter marcado ou não-marcado dos termos de uma distinção fônica não era necessariamente universal – ideia que foi, mais tarde, obscurecida pela leitura da sua obra (TRUBETZKOY, 1957 [1939]) por CHOMSKY & HALLE (1968). Para Zipf, as leis quantitativas a interligar fenômenos humanos e naturais eram claramente não-determinísticas, *i.e.*, estocásticas – ideia que foi, analogamente, obscurecida pela leitura da sua obra (ZIPF, 1935, 1949) por MILLER & CHOMSKY (1963). Um retorno, hoje, a ambos os autores revela uma complementaridade da qual emergem surpreendentes pontos de contato.

Um primeiro achado deste trabalho é que as distribuições de tipos e ocorrências tornam-se semelhantes em amostras grandes, com exceções raras, aparentemente isoladas, e de alto valor heurístico – pois identificam segmentos propensos à mudança.

Essa heurística é ilustrada com o estudo da tendência milenar das vogais médias do português a se abrirem na sílaba tônica. Trata-se, na verdade, de uma cadeia de mudanças foneticamente abruptas que se disseminaram, gradual e intermitentemente, por vários setores do léxico desde a formação da língua (ALBANO, 2010). A sua persistência e coerência sob diferentes formas qualifica-as como um caso de deriva (SAPIR, 1921).

A exposição prossegue mostrando que uma maior compreensão da deriva de abertura das vogais médias do português exige a integração de ideias trubetzkoyanas e zipfinas.

A história do vocalismo do português e seus vestígios na estatística fônica sincrônica revelam uma tendência, ainda viva, a aumentar a carga funcional<sup>6</sup> da distinção entre as vogais médias

---

<sup>6</sup> TRUBETZKOY (1939 [1957]: 288) define a carga ou rendimento funcional como “o grau de utilização das diversas oposições distintivas” (tradução minha). Embora tenha havido, depois, várias tentativas de quantificá-la - p.ex., KING (1967) -, nenhuma delas

abertas e fechadas não pela livre combinação, mas pela extensão de vieses fonotáticos fônica e/ou morficamente condicionados.

Esse modo de restabelecer o equilíbrio do sistema tem um aspecto trubetzkoyano e outro zipfiano: a marca de abertura prefere a acentuação e o tipo silábico não-marcados, mas, longe de ser neutra, indica a nativização das palavras portadoras, atestando a sua incorporação ao vernáculo e datando-as como posteriores ao romance ibérico, *i.e.*, como da era iniciada com o galego-português.

A análise mostra que tal marca de identidade está intimamente relacionada a outra, que ajudou a diferenciar, na sequência, o romance ibérico e o galego-português: o surgimento de sílabas finais pesadas, com consequente acentuação oxítona – devido, primeiro, à queda do /e/ final, e, em seguida, à perda da consoante intervocálica pré-final.

A conclusão especula sobre os efeitos da expansão do acento intensivo românico sobre a deriva de abertura, para reconstruir um quadro em que a inovação fônica e a inovação léxico-gramatical teriam moldado, *pari passu*, esse importante aspecto da identidade galego-portuguesa.

## 1. Metodologia

A maior parte dos dados foi extraída do corpus de internet conhecido como “LAEL fala”<sup>7</sup> (doravante LAEL), disponibilizado para pesquisa por BERBER-SARDINHA (2000).

---

será aqui utilizada.

<sup>7</sup> A sigla significa *Linguística Aplicada ao Ensino de Línguas*, programa de pós-graduação da PUC-SP em que o autor atua. O corpus está disponível em: <http://www2.lael.pucsp.br/corpora/>.



Trata-se de uma lista de frequências de ocorrência de palavras obtida pela transcrição ortográfica de gravações de aulas, conversas e entrevistas. Contém aproximadamente 46.000 tipos e três milhões de ocorrências.

Esse material foi corrigido, a fim de eliminar distorções devidas a erros ortográficos ou problemas de notação (p.ex., abreviaturas, códigos de transcrição), e submetido a uma conversão ortográfico-fônica com o programa ORTOFON<sup>8</sup> (ALBANO & MOREIRA, 1996), que tem uma alta taxa de acerto nas vogais médias (96%). Os erros do conversor foram eliminados automaticamente (p.ex., palavras estrangeiras transcritas como siglas) ou corrigidos manualmente. O corpus fonetizado resultante contém cerca de 45.000 tipos e 2.700.000 ocorrências.

A conversão distingue 19 consoantes e 7 vogais, totalizando 26 segmentos fônicos. São marcados também o acento lexical, a divisão silábica e as posições de neutralização típicas do dialeto paulistano<sup>9</sup>. As vogais pós-tônicas e as semivogais são transcritas como arquifonemas vocálicos, *i.e.*, /I, U/. As líquidas dos grupos consonantais de ataque e as codas também são transcritas com maiúsculas, inclusive a nasal, para a qual se adota a solução mattosiana (CAMARA JR., 1970).

Todos os cálculos são realizados com o auxílio de fórmulas montadas com funções do programa Excel.

Dado o tamanho da amostra de palavras, o número de segmentos fônicos é suficientemente grande para permitir a comparação entre os tipos e as ocorrências. A composição fônica geral do corpus é a que segue.

A lista de tipos tem 175.920 consoantes, 162.038 vogais –

<sup>8</sup> A saída do Ortofon é “arquifonêmica”, *i.e.*, uma transcrição fônica larga que distingue ambientes propensos à neutralização (p.ex., coda, posição pós-tônica, etc.).

<sup>9</sup> A saber: na palavra, a pós-tônica; e, na sílaba, a coda.

portanto, 337.958 segmentos fônicos –, além de 152.257 sílabas e 44.518 acentos. A lista de ocorrências tem, por sua vez, 5.156.382 consoantes, 5.557.320 vogais – portanto, 10.713.702 segmentos fônicos –, além de 5.135.409 sílabas e 2.054.829 acentos.

Isso caracteriza o português brasileiro (doravante PB) como uma língua tendente à estrutura silábica CV, pois a taxa de consoantes é menos de 10% superior à de vogais no léxico, representado pelos tipos. Na fala corrente, representada pelas ocorrências, a mesma pequena vantagem inverte-se a favor das vogais, o que indica que há muitas palavras de alta frequência com sílabas abertas, ditongos ou hiatos.

Também fica caracterizada uma preferência por dissílabos e trissílabos, pois o número médio de sílabas por palavra é de 3,4 no léxico e 1,8 na fala corrente. No léxico, cerca de 30% dessas sílabas são acentuadas; e, na fala corrente, 40% – o que condiz com as respectivas médias por palavra.

As análises que requeriam etiquetagem gramatical do corpus, indisponível no LAEL, lançaram mão de duas outras bases de dados, ambas de língua escrita: a versão do Minidicionário Aurélio preparada por SILVA *et al.* (1993), com cerca de 27.000 palavras (doravante Mini-Aurélio); e o CETEN-Folha (doravante CETEN), disponibilizado pelo *Núcleo Interinstitucional de Linguística Computacional* (NILC), da USP de São Carlos, na *Linguateca*<sup>10</sup>, com cerca de 130.000 tipos e 34.380.045 ocorrências.

Os procedimentos de conversão e contagem são idênticos aos aplicados ao LAEL e as estatísticas são proporcionais às acima reportadas, de modo que não seria pertinente insistir aqui sobre elas.

---

<sup>10</sup> Centro de recursos *online* para o estudo computacional do português construído por um consórcio de universidades portuguesas. Ver: <http://www.linguateca.pt/>.

A análise estatística das distribuições de frequência e dos seus vieses utiliza as seguintes técnicas não-paramétricas: para detectar e analisar as associações entre categorias, os testes de  $X^2$  de Pearson e razão de verossimilhança; para medir a força de associação, o  $\varphi$  de Pearson; para testar a significância das células, os desvios padronizados; e, para avaliar o papel de cada fator e respectivas interações, a estimação de máxima verossimilhança.

## **2. A deriva de abertura das vogais médias tônicas do português**

### **2.1. Semelhanças e diferenças entre as distribuições de tipos e ocorrências**

Vale destacar uma das passagens mais esclarecedoras do capítulo de Sapir sobre a deriva:

The drift of a language is constituted by the unconscious selection on the part of its speakers of those individual variations that are cumulative in some special direction. This direction may be inferred, in the main, from the past history of the language. In the long run any new feature of the drift becomes part and parcel of the common, accepted speech, but for a long time it may exist as a mere tendency in the speech of a few, perhaps of a despised few. As we look about us and observe current usage, it is not likely to occur to us that our language has a “slope,” that the changes of the next few centuries are in a sense prefigured in certain obscure tendencies of the present and that these changes, when consummated, will be seen to be but continuations of changes that have been already

effected (SAPIR, E. 1921: 166)

A tendência à abertura das vogais médias tônicas, observável hoje no português brasileiro, assim como no europeu (doravante PE; MATEUS & PARDAL, 2000), encaixa-se perfeitamente nessa descrição. Trata-se de uma cadeia de mudanças que gradualmente fez proliferar em ambientes não-etimológicos, especialmente em posição tônica, o contraste entre vogais médias abertas e fechadas, derivado da distinção latina de quantidade, perdida no português assim como nas demais línguas românicas.

Os fatos etimológicos são bastante simples, embora a sua evolução histórica tenha sido complexa: as longas latinas tinham uma qualidade fechada, assim como as breves tinham uma qualidade aberta (CAMARA JR., 1976; SILVA NETO, 1979). Tais distinções preservaram-se no português, ainda que de forma parcial, inovadora e extremamente variável.

O PE perdeu-as quase completamente em posição átona, através de um processo radical de fechamento que as neutralizou em favor de /i, u/, salvo em casos ocasionais de crase, (p.ex., ‘p[ɛ]gada’). O PB do sul e do sudeste também as perdeu na mesma posição, através de uma neutralização análoga, porém mais branda, em favor de /e, o/. Já outros dialetos, em especial os do nordeste, preservaram algumas vogais abertas etimológicas em posição átona (p.ex., ‘qu[ɛ]rer’, ‘p[ɔ]der’), o que desencadeou processos variáveis de neutralização no sentido oposto, *i.e.*, em favor de /ɛ, ɔ/.

A conhecida tendência à abertura na tônica data de antes da chegada dos portugueses ao Brasil e tem manifestações variáveis, intermitentes e extremamente complexas.

Essa deriva, apesar de ter base fonética no abaixamento da mandíbula, que ocorre tanto na abertura vocálica como na acentuação, é foneticamente abrupta, ou seja, não se manifesta

em termos de detalhe fonético – contrariamente ao que se poderia esperar à primeira vista. Isso significa que não há indício acústico de que vogais médio-fechadas tônicas sejam ligeiramente mais abertas que as suas contrapartes átonas. Ao contrário, segundo ROCES-RODRIGUES (2010), o  $F_1$  de /e, o/ apresenta diferenças significativas entre as duas posições, mas no sentido de maior fechamento na tônica.

Os fatos históricos também apontam para surtos intermitentes de expansão da tendência à abertura na tônica, por analogia e difusão lexical. Como as informações diacrônicas a esse respeito são escassas, é prudente começar a investigação pelo lado mais visível da linha do tempo, ou seja: o momento atual.

O primeiro fato contemporâneo que chama atenção é a baixa frequência de /ɛ, ɔ/. A Figura 1 mostra a distribuição da frequência de tipos dos segmentos fônicos do PB no LAEL.

A fim de evitar mal-entendidos, cabe esclarecer que o alfabeto fonético utilizado nas Figuras, o *Speech Assessment Methods Phonetic Alphabet*, conhecido como SAMPA<sup>11</sup>, trata as maiúsculas não como arquifonemas, mas como segmentos fônicos distintos das minúsculas correspondentes – portanto, fonemas<sup>12</sup>.

Cabe também notar que a saída arquifonêmica do ORTOFON não cria ambiguidade nas Figuras, pois, nelas, os arquifonemas foram somados ao membro majoritário da oposição neutralizada. Assim, /i/ representa /i, I/; /n/ representa /n, N/; /s/ representa /s, S/; e assim sucessivamente<sup>13</sup>. No caso dos róticos, o arquifonema de coda

<sup>11</sup> Versão do IPA para uso em ambientes que não aceitam fontes especiais. Disponível em: <http://www.phon.ucl.ac.uk/home/sampa/portug.htm>.

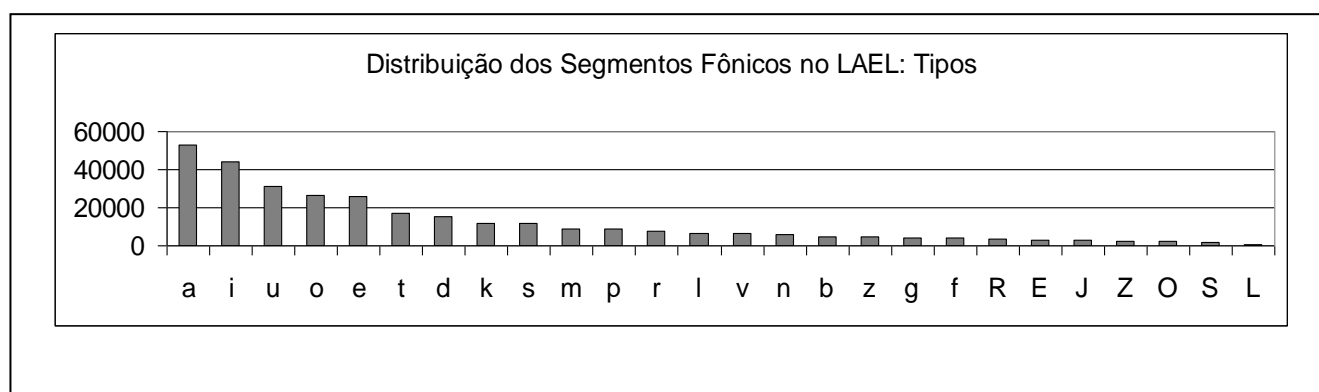
<sup>12</sup> ɛ=E, ɔ=O, ɾ=r, ʒ=Z, ʃ=S, ɲ=J, λ=L. R representa o “r forte”, que, no PB, tem realizações fonéticas vibrantes e fricativas, passando por vários pontos de articulação.

<sup>13</sup> Isto é, os arquifonemas vocálicos, da posição pós-tônica, e os arquifonemas

foi somado ao “r brando”, transcrito como /r/, conforme a convenção SAMPA para o português. Pela mesma convenção, o “r forte” é transcrito como /R/.

Salta aos olhos que a distribuição tem um aspecto zipfiano, a saber: a frequência cai exponencialmente do primeiro ao último elemento da ordem decrescente. Além disso, a posição de /ɛ, ɔ/, na cauda, em meio às consoantes palatais, está de acordo com a previsão de ZIPF (1949) de que as aquisições mais recentes de uma língua são também as menos frequentes.

FIGURA 1 - Distribuição de frequência dos segmentos fônicos do PB nos tipos do LAEL

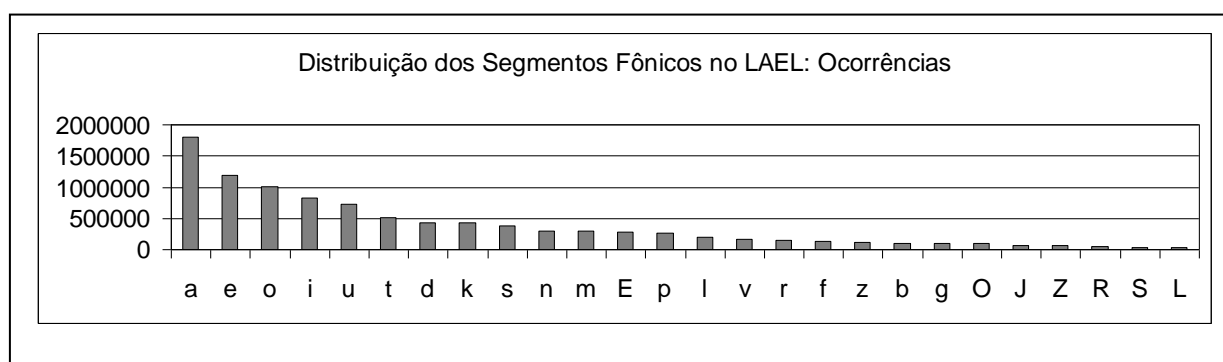


consonantais, da coda silábica, são somados ao fonema majoritário do par neutralizado.

Não obstante, o método usado para levantar essa distribuição não é o de Zipf, mas o de Trubetzkoy: as frequências foram computadas na lista de tipos. Vejamos agora, na Figura 2, a distribuição dos mesmos segmentos contados à maneira de Zipf, *i.e.*, na lista de ocorrências.

O que salta aos olhos aqui é que os segmentos que sofrem as maiores alterações de ordem da Figura 1 para a Figura 2 são as vogais /ε, ɔ/ (NB: grafadas /E, O/). Visto que as duas Figuras são muito semelhantes e as respectivas escalas de frequências têm um alto coeficiente de correlação de Spearman ( $R=0,95$ ), é razoável supor que tais discrepâncias, por seu caráter minoritário, apontam para desequilíbrios no sistema, *i.e.*, mudanças em curso.

FIGURA 2 - Distribuição de frequência dos segmentos fônicos do PB nas ocorrências do LAEL



A hipótese é corroborada por uma inspeção caso a caso: poucos segmentos se movem mais de dois pontos para cima ou para baixo na escala de frequência.

As exceções cuja ordem é mais baixa nas ocorrências que

nos tipos são /e/ e /R/ – com descidas de 3 pontos. Sugestivamente, ambos participam de processos de enfraquecimento no PB atual<sup>14</sup>.

As exceções cuja ordem é mais alta nas ocorrências que nos tipos são, por sua vez, /ε/, com uma subida de 7 pontos, e /ɔ/, com uma subida de 3 pontos. Sugestivamente, a sua incidência na posição tônica primária, em oposição à secundária<sup>15</sup>, chega a 98% nas ocorrências, com uma vantagem de 7% sobre os tipos. Essa preferência por co-ocorrer com o acento primário na fala corrente indica que a abertura é um processo de fortalecimento.

Assim, mesmo sem fazer justiça a processos variáveis de enfraquecimento<sup>16</sup>, devido à natureza ortográfica da transcrição original, a comparação entre as distribuições de tipos e ocorrências parece identificar segmentos fônicos sujeitos a mudança. Isso pode significar que Trubetzkoy e Zipf estavam ambos certos, apesar das grandes diferenças entre os seus pontos-de-vista. Vejamos como é possível racionalizar essa aparente contradição.

As Figuras 3 e 4 apresentam os mesmos dados das Figuras 1 e 2 em escala logarítmica. Essa escala acentua a semelhança entre as distribuições, evidenciando que as maiores diferenças recaem na cauda, ou seja, nos segmentos de baixa frequência. As curvas resultantes são boas aproximações da reta, exceto no trecho final, onde se encontram: /ε, ɲ, ʒ, ɔ, ʃ, λ/, nos tipos; e /ɔ, ɲ, ʒ, R, ʃ, λ/, nas ocorrências. A “filologia dinâmica” de Zipf explica dois importantes aspectos dessas distribuições, a saber: o fato de as curvas terem o aspecto típico de uma lei de potência, *i.e.*, o valor de cada ponto ser inversamente proporcional à sua

<sup>14</sup> V., p.ex., BISOL (1981), CALLOU *et al.* (2002).

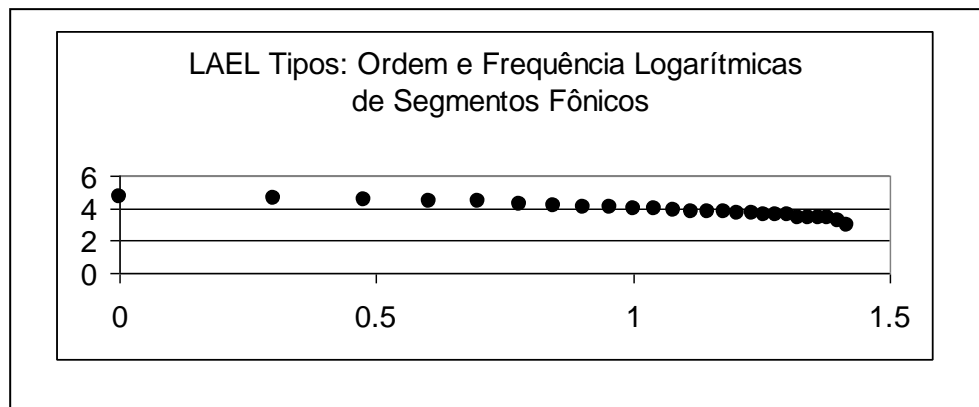
<sup>15</sup> O ORTOFON dá conta da manutenção da abertura da vogal em diminutivos, aumentativos e advérbios em ‘-mente’.

<sup>16</sup> Por exemplo, a perda de consoantes de coda, o fechamento de vogais médias, etc.



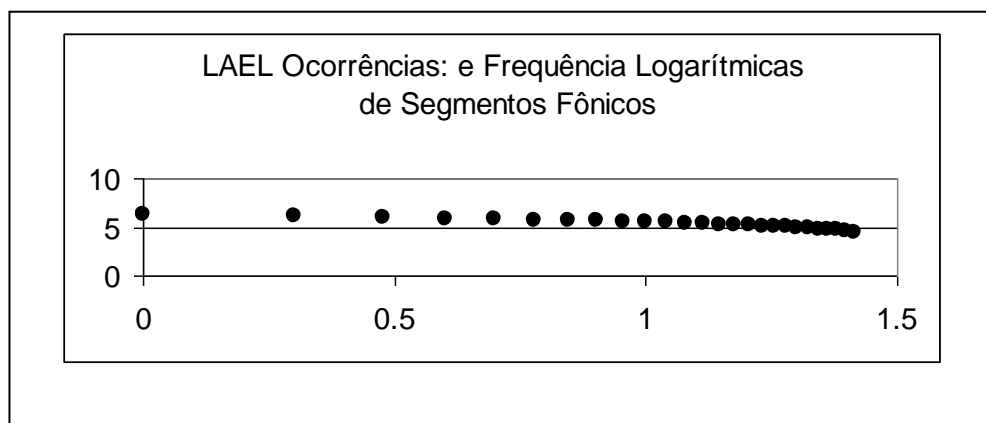
posição na ordem decrescente de frequência<sup>17</sup>; e o fato, já mencionado, de as aquisições recentes da língua, tais como as consoantes palatais e as vogais médias abertas, estarem nas posições mais baixas dessa ordem.

FIGURA 3 - Ordem e frequência logarítmicas dos segmentos fônicos do PB nos tipos do LAEL



<sup>17</sup> Nos casos mais conhecidos, o expoente envolvido é aproximadamente 1, o que produz uma série harmônica: 1/2, 1/3, etc.

FIGURA 4 - Ordem e frequência logarítmicas dos segmentos fônicos do PB nas ocorrências do LAEL



Quando, porém, se examinam os segmentos envolvidos, percebe-se claramente que a noção de complexidade de Zipf, que supõe uma mecânica articulatória universal, enfrenta dificuldades. Basta, por exemplo, inspecionar as Figuras 1 e 2 para constatar que as líquidas /l, r/, cuja aquisição tem sido descrita como tardia (MEZZOMO & RIBAS, 2004), estão à frente de muitas obstruintes geralmente adquiridas primeiro.

Para explicar exceções, Zipf contava com o ponto de vista estocástico e com o princípio do menor esforço, ao qual atribuía um caráter compartilhado – *i.e.*, com distribuição do ônus da comunicação entre o falante e o ouvinte. Mesmo assim, este caso não tem solução clara, já que não há evidência, na literatura, de que a percepção das líquidas seja mais simples que a sua produção (AOYOMA *et al.*, 2004).

Já a noção de marca de Trubetzkoy sai-se melhor aqui, por rechaçar o naturalismo, enfatizando as relações entre os termos de uma oposição e não o seu conteúdo fonético. Assim, por exemplo, ao associar marca a frequência, o autor ressalta a

possível variabilidade interlinguística do termo não-marcado (TRUBETZKOY, 1957: 282-284), ilustrando-a com as aspiradas do lesghé (*ibidem*, p. 282) – mais frequentes que as não-aspiradas, ao contrário do que ocorre em outras línguas. Aparentemente, Trubetzkoy estava mais à vontade que Zipf ou mesmo que seus próprios seguidores (p.ex., CHOMSKY & HALLE, 1968) com o fato de que a complexidade pode variar de uma língua para outra.

Quanto ao PB, as Figuras 1 e 2 parecem sugerir fortemente que os termos não-marcados das oposições consonantais são: apical, para ponto de articulação; oclusiva, para modo de articulação; e surda, para sonoridade.

Não obstante, as frequências de alguns desses termos estão longe de indicar marcas inequívocas. Assim, a baixa posição da velar /g/ é assimétrica à da sua contraparte /k/, que está à frente de todas as labiais, perdendo apenas para as apicais. Outra assimetria é a inversão nas ordens de /v/ e /f/ e de /ʒ/ e /ʃ/ com relação aos demais contrastes de sonoridade, fato que se observa tanto nos tipos como nas ocorrências.

Analogamente, os subsistemas vocálicos átonos, de três e cinco termos, ainda que obviamente preferidos ao tônico, de sete, não se ordenam inequivocamente, pois nem mesmo o caráter conservador da transcrição ortográfica exime as vogais /e, o/ e /i, u/ de mudarem de posição dos tipos para as ocorrências. Outra ambiguidade envolve os graus de abertura: embora /a/ seja o segmento fônico menos marcado do PB, as demais vogais abertas, *i.e.*, as médias /ɛ, ɔ/, são fortemente marcadas, o que obscurece o funcionamento dessa marca.

Todas essas assimetrias e ambiguidades põem em pauta uma questão com a qual nem Trubetzkoy nem Zipf estavam preparados para lidar: a não-linearidade dos efeitos conjuntos

dos componentes mínimos da fala<sup>18</sup>. O termo “linear” é aqui usado no sentido matemático: um sistema é linear se a sua saída é diretamente proporcional à sua entrada. Num sistema não linear, a relação entre a entrada e a saída é mais complexa.

O problema é que os conceitos explicativos de ambos os autores – respectivamente, a marca e o menor esforço – supõem que aqueles efeitos sejam aditivos, o que hoje se sabe falso. Ambos acertaram em tentar decompor a complexidade dos sistemas fônicos, mas erraram nas simplificações escolhidas para tanto.

Evidentemente, não se pode responsabilizá-los por isso, pois o conhecimento sobre a não-linearidade desses sistemas era inexistente durante as suas vidas e permanece incipiente ainda hoje.

De qualquer modo, é possível, agora, rever à luz do exposto a obscura história da deriva de abertura das vogais médias tônicas em português. A evidência histórica, apesar de fragmentária, aponta para uma relação não-linear da abertura com o seu contexto fônico e lexical.

## **2.2. A abertura das vogais médias na flexão galego-portuguesa**

Houve duas correntes de disseminação analógica da abertura na história do português. A primeira, tratada nesta seção e nas duas próximas, é a analogia propriamente dita, que tem fundamento mórfico e pode se manifestar de duas maneiras: entre diferentes paradigmas, a chamada “extensão”; ou no mesmo paradigma, o chamado “nivelamento”. A segunda,

---

<sup>18</sup> Cabe observar que a teoria dos traços distintivos supõe essa linearidade, mesmo quando se apresenta como “não-linear”, com o que quer dizer apenas não-sequencial.

tratada na seção 2.5, é a “analogia fonética” – termo devido a SCHUCHARDT (1885) –, cuja disseminação ocorre por semelhança fônica no léxico, independentemente de filiação paradigmática.

A primeira e mais antiga expansão analógica da abertura remonta ao galego-português e é, certamente, mais recente que a metafoia que fechou /ɛ, ɔ/ etimológicos em substantivos e adjetivos tais como ‘*sōcrum*’ e ‘*nōvum*’ no latim vulgar (SILVA NETO, *op. cit.*). Trata-se da tendência a incorporar o contraste entre vogais médias abertas e fechadas à morfologia flexional, associando-o, em substantivos e adjetivos, à flexão de gênero e número e, em verbos, à flexão de tempo, modo, e pessoa-número (CAMARA JR., *op. cit.*).

Como, ao norte de Portugal e na Galícia, há, na verdade, uma situação linguística complexa, em que cada dialeto apresenta, em maior ou menor grau, características tidas como divisores das duas línguas, as observações que seguem referem-se às versões oficiais de cada uma (MAIA, 1986). Ressalve-se, porém, que é possível encontrar em território português traços atribuídos abaixo ao galego, e vice-versa.

Na flexão nominal, as duas línguas convergem em apresentar casos de /ɛ, ɔ/ etimológicos a sinalizar o feminino e o plural, tal como em ‘porco, porca, porcos’. Divergem, contudo, no modo como a analogia estendeu essa alternância a étimos fechados. Um exemplo é o sufixo ‘*ōsum*’, que, no português, por extensão analógica das alternâncias criadas por metafoia, tem vogal fechada no masculino singular, e aberta no feminino, singular e plural, bem como no masculino plural. No galego a mesma mudança foi seguida de um nivelamento analógico que generalizou a abertura a todo o paradigma.

Da mesma forma, na flexão verbal, as duas línguas convergem em apresentar /ɛ, ɔ/ em todas as formas rizotônicas

da primeira conjugação, por extensão analógica de verbos frequentes com étimos abertos tais como ‘levar’ e ‘olhar’<sup>19</sup> (galego ‘ollar’). Divergem, porém, quanto à preservação desses étimos nas formas arrizotônicas do mesmo paradigma. No galego, mais sistematicamente que nos dialetos do PB em que /ε, ɔ/ ocorrem nas pré-tônicas, verbos derivados de bases com vogal aberta tendem a preservá-la em formas arrizotônicas, p.ex., ‘v[ɔ]tamos’.

As divergências maiores, no entanto, residem no tratamento das formas rizotônicas do presente do indicativo e do imperativo na segunda e terceira conjugações.

Aqui as duas línguas convergem em haver estendido a abertura à segunda pessoa do singular e à terceira pessoa do singular e do plural do presente do indicativo, por pressão analógica de formas frequentes tais como ‘p[ε]rde’, ‘m[ɔ]ve’, ‘p[ε]de’ e ‘d[ɔ]rme’. Divergem, porém, quanto à extensão dessa mudança ao imperativo: diferentemente do português, o galego possui o contraste entre imperativos fechados, cf. ‘b[e]be’ e ‘g[o]sta’<sup>20</sup>, e abertos, cf. ‘p[ε]rde’ e ‘m[ɔ]ve’. Também os sufixos ‘ecer’ e ‘escer’, arrastados pela deriva de abertura em português, permaneceram fechados em galego.

Assim, a pressão analógica que, independentemente de conjugação, generalizou a abertura aos temas paroxítonos não-truncados em vogal média (*i.e.*, com vogal média na tônica e vogal temática na átona final), acabou por ter um alcance maior no português do que no galego.

Coerentemente com isso, o galego favoreceu uma tendência contrária, desaparecida do português moderno depois de haver competido com a abertura até o fim do período medieval. Trata-

<sup>19</sup> As frequências de ocorrência de ‘leva’ e ‘olha’ no LAEL são, respectivamente, 717 e 2387.

<sup>20</sup> Derivados, respectivamente, de /ĩ, ũ/ latinos.

se do nivelamento analógico das vogais altas decorrentes da morfologização da metafoia desencadeada pelo iota latino na primeira pessoa do singular do presente do indicativo da quarta conjugação, *i.e.*, ‘sirvo’ < *s[ɛ]rvio* (DUBERT-GARCIA, 2010). Assim, usam-se até hoje na Galícia formas tais como ‘sirve’ e ‘fuge’ – esta uma simples reversão do ‘f[ɔ]ge’ analógico do galego-português.

O estado em que o português chegou ao Brasil no século XVI indica que a corrente da deriva de abertura por extensão e nivelamento no paradigma verbal já tinha se completado em grande parte então. Além das grandes semelhanças entre o PB e do PE, há exceções comuns à abertura – p.ex., ‘fechar’, ou os verbos em ‘ejar’ e ‘gredir’. Tais exceções sofrem pressões analógicas semelhantes em diferentes dialetos portugueses e brasileiros, criando formas tais como ‘f[ɛ]cha’, ‘ar[ɛ]ja’ e, até, o estigmatizado ‘progr[ɛ]de’.

Já a segunda corrente da deriva de abertura das vogais médias tônicas, que recorre sempre que há um aumento de vocabulário novo ou estrangeiro, permanece ativa até hoje sob diferentes formas. Trata-se da tendência, manifesta em substantivos e adjetivos, a tomar /ɛ, ɔ/ como o *default* das vogais médias em posição tônica.

Antes de prosseguir, convém, todavia, aprofundar o estudo do verbo, a fim de enfrentar uma questão comum a ambas as correntes da deriva de abertura, a saber: o papel dos estratos lexicais<sup>21</sup> na disseminação de uma mudança fônica. Esse é, aliás, um terreno fértil para tentar integrar as ideias de Zipf e Trubetzkoy sobre a estatística fônica.

---

<sup>21</sup> Termo proposto por ITO & MESTER (1999) para designar um subconjunto do léxico que parece seguir regras próprias.

### 2.3. O sistema verbal

ZIPF (1949) sustenta, com base em análises do inglês, que a frequência de uma palavra é inversamente proporcional à sua idade, assim como ao seu tamanho e ao número das suas diferentes acepções. Assim, palavras frequentes costumam ser curtas e ter muitos sentidos, em geral básicos, abstratos e genéricos, enquanto palavras infrequentes costumam ser mais longas e ter poucos sentidos, em geral acessórios, concretos e específicos ou mesmo especializados.

Essas ideias têm sido retomadas nas últimas décadas, principalmente na literatura linguística ligada à gramaticalização e à ciência cognitiva<sup>22</sup>. Mais recentemente, porém, BYBEE (2002) retomou também a discussão levantada por Zipf da relação entre frequência e enfraquecimento fonológico. A autora e seus colaboradores mostram que processos de redução como a queda de *schwa* são, de fato, mais comuns em palavras de alta frequência no inglês. Isso corrobora a tese zipfiana de que as palavras mais velhas e frequentes são também mais curtas.

Um corolário ainda inexplorado das teses de Zipf sobre as relações entre esses vários aspectos das entradas lexicais é que uma mudança fônica pode nem sempre independender do sentido – o que abala pressupostos já muito enraizados na Fonologia sobre a sua autonomia frente a outros componentes da gramática.

Pode-se prever, por exemplo, que formas etimológicas de alta frequência do fone-alvo de uma mudança analógica em curso possam exercer forte pressão para a sua extensão em determinado paradigma. Ora, isso é admitir a existência de possíveis associações indiretas entre a propensão à analogia e os significados genéricos e/ou gramaticais.

---

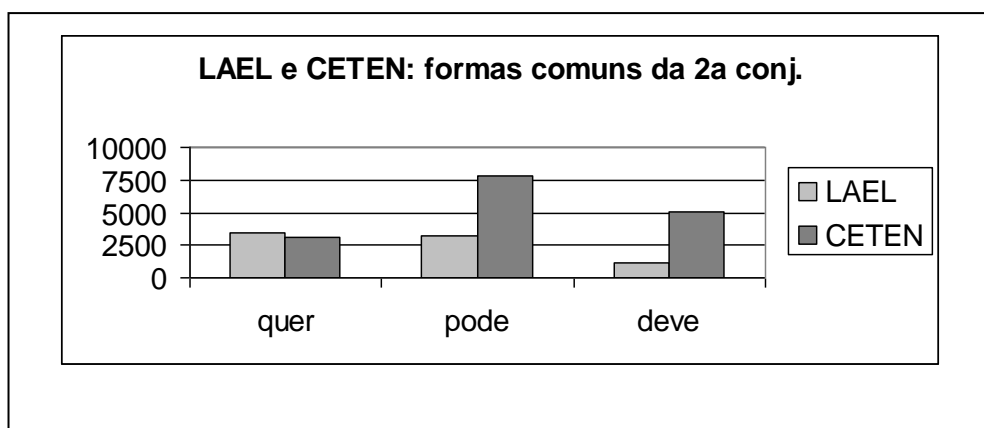
<sup>22</sup> V. a esclarecedora resenha de PUSTET (2004).



Tal parece ter sido o papel da segunda conjugação na extensão da abertura aos temas paroxítonos não-truncados em vogal média, a saber: a segunda e a terceira pessoas do singular, e a terceira do plural, do presente do indicativo.

Embora a abertura esteja hoje nivelada na conjugação produtiva – *i.e.*, a primeira –, é na segunda que se concentra a maioria dos étimos abertos da terceira pessoa do singular do presente do indicativo. Por exemplo, as formas ‘quer’, ‘pode’ e ‘deve’ devem ter tido, como têm hoje, um grande impacto sobre a frequência de ocorrência das vogais /ɛ, ɔ/. A Figura 5 exhibe as suas ocorrências no LAEL e no CETEN.

FIGURA 5 - Frequências de ocorrência de ‘quer’, ‘pode’ e ‘deve’ no LAEL e no CETEN



É interessante notar que esses três verbos têm significados básicos, abstratos e genéricos, o que propicia o seu uso como auxiliares modais. A sua faixa de frequência não deve, portanto, ter mudado muito desde a Idade Média. Isso, porém, por si só,

não teria bastado para deslancar a analogia. Outro fator importante é a sua semelhança formal com a maioria dos outros verbos da mesma conjugação.

Aqui o dado relevante são as percentagens dos radicais em vogal média nos verbos do CETEN, exibidas na Tabela 1 abaixo. Já que, na segunda conjugação, a vogal temática também é média, o alto percentual de tipos com essa abertura vocálica no radical indica a presença, no léxico, de uma forte tendência à harmonia vocálica, provavelmente herdada do latim. Tais radicais redundantes podem ter atraído a extensão analógica da abertura e disseminado a apofonia.

Cabe notar que o ambiente da mudança é troubetzkoyano, *i.e.*, lexical, e não zipfiano, *i.e.*, discursivo. A abertura atingiu formas duplamente classificáveis como não-marcadas, *i.e.*: morfologicamente, por serem imunes ao truncamento; e fonologicamente, por serem majoritárias e parcialmente redundantes quanto à vogal do radical.

A apofonia de abertura teve, então, possivelmente, o papel de regularizar a flexão dos radicais em vogal média das três conjugações, unificando os temas paroxítonos não-truncados do presente do indicativo pelo modelo de verbos tais como ‘leva’, ‘pede’, ‘segue’. O presumível contexto analógico é em parte mórfico e em parte fônico. É mórfica a restrição a temas verbais não-truncados. É fônica a restrição a paroxítonas – a qual adquiriu autonomia ao longo dos séculos, como se verá nas próximas seções.

Os percentuais da Tabela 1 corroboram a hipótese de que a apofonia de abertura tenha se propagado primeiro aos temas não-truncados da segunda e terceira conjugações, dada a sua semelhança, e só depois à primeira – na qual viria a nivelar-se na primeira pessoa do singular do presente do indicativo e nas formas rizotônicas do presente do subjuntivo.

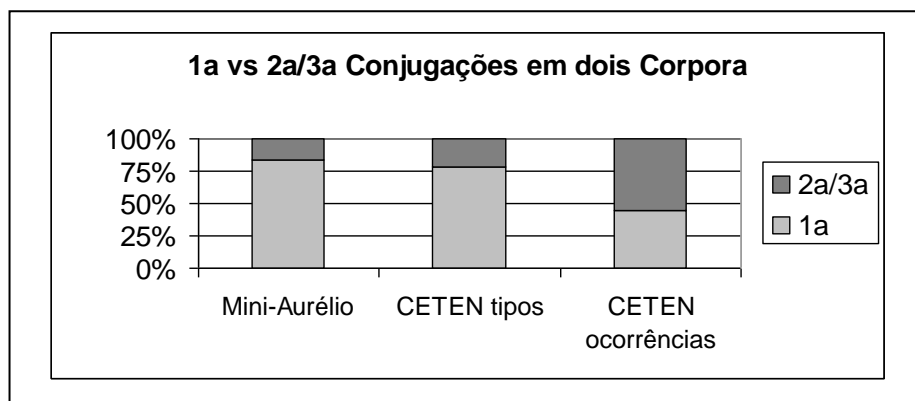
TABELA 1 - Percentuais de radicais em vogal média nos verbos no CETEN

Radicais em V média no CETEN	% tipos	% ocorrências
1ª. conjugação	26	23
2ª. conjugação	75	28
3ª. conjugação	23	17

88254863

Embora carecendo ainda de respaldo diacrônico, essa hipótese coaduna-se com o fato de a segunda e a terceira conjugações, somadas, terem, hoje, uma alta frequência de ocorrência – o que, segundo Zipf, indica a sua antiguidade. A Figura 6 compara a sua frequência relativa (em %) à da conjugação produtiva no Mini-Aurélio e nos tipos e ocorrências do CETEN. Esses números juntam-se à perda histórica de produtividade para respaldar a interpretação zipfiana.

FIGURA 6 - Percentuais das 3 conjugações no MiniAurélio e no CETEN



Note-se que, no CETEN tipos, a proporção de formas da segunda e terceira conjugações (21%) não é muito maior que a do dicionário (16%), onde só há infinitivos. Já no CETEN ocorrências, ela é até maior que a da primeira conjugação (55% *vs.* 45%).

Outro índice que corrobora a interpretação zipfiana é a razão tipo/ocorrência, calculada com base no CETEN. Tal razão, expressa em termos percentuais, indica grosseiramente a produtividade ou novidade de uma forma lexical. Assim, enquanto, na segunda e terceira conjugações, os tipos perfazem apenas 0,30% das ocorrências, na primeira chegam a 1,38%, ou seja, quase cinco vezes mais.

Os dados da Tabela 1 e da Figura 6 coadunam-se também com o fato de a maioria dos verbos de alta frequência da primeira conjugação ter étimos fechados; assim, p.ex., ‘pegar’ < *picare*, ‘gostar’ < *gustare*, etc. A terceira pessoa do singular do presente do indicativo, por seu caráter fônica e morficamente não-marcado, pode, portanto, ter levado verbos de conjugações distintas a responder à pressão analógica que embutiu a flexão de presente na prosódia lexical por meio da apofonia de abertura.

Para concluir, cabe rever uma velha controvérsia sobre a interpretação sincrônica do processo de abertura. HARRIS (1974) propõe uma só regra de apofonia aplicável a todas as formas rizotônicas do presente, como caso omissivo ou *default*<sup>23</sup>, complementar à harmonia vocálica abstrata dos temas truncados da segunda e terceira conjugações. Já MATEUS (1975) sustenta que a abertura da primeira conjugação é, na verdade, uma

---

<sup>23</sup> De acordo com a “*elsewhere condition*” ou “*elsewhere principle*” (KIPARSKY, 1973). Por exemplo, em ‘m[ɔ]v+e+o’ > ‘m[o]vo’, a harmonia com a vogal temática truncada, caso específico, teria precedência sobre a abertura, caso geral.

extensão dessa harmonia; e que a apofonia se restringe às formas abertas do presente do indicativo das demais conjugações. A maior parte da literatura subsequente limitou-se a amear argumentos em favor de uma ou outra posição (p.ex., QUICOLI, 1990; WETZELS, 1995).

Para a visão dinâmica aqui defendida, basta apontar que qualquer fonologia que relacione regularidade a uso verá os dados acima como favoráveis à análise de HARRIS. Veremos, abaixo, que há mais razões para tratar todas as tendências à abertura do léxico português como parte de uma única deriva, ainda em curso.

## 2.4. O sistema nominal

Como já vimos, as exceções à abertura no paradigma verbal estão perdendo força, o que é evidenciado por variantes tais como ‘f[e]cha’/ ‘f[ɛ]cha’. A pressão das alternâncias majoritárias é tal que afeta até, por hipercorreção, um verbo de alta frequência cujo radical não contém vogal média: ‘viver’; haja vista a sua variante estigmatizada da fala popular: ‘v[ɛ]ve’. Tudo isso indica que a deriva de abertura permanece ativa no sistema verbal (CAMPOS & CRISTÓFARO-SILVA, 2008), embora seu ímpeto tenha, naturalmente, diminuído com o próprio crescimento.

Também na flexão do feminino e do plural dos substantivos e adjetivos essa deriva permanece ativa. Autores brasileiros (p.ex., TOMAZ, 2006) e portugueses (p.ex., CAVACAS, 1920) reportam a emergência de femininos tais como ‘abs[ɔ]rta’ ou plurais tais como ‘abs[ɔ]rtos’, ‘b[ɔ]lsos’ e ‘pesc[ɔ]ços’.

Outro mecanismo em que a abertura tem agido é a derivação regressiva. Aqui ela recorre à morfologização da

metafonia para a extensão das alternâncias. Substantivos derivados de verbos com vogal média no radical costumam ter a tônica fechada, se masculinos (p.ex., ‘ap[e]go’, ‘enr[e]do’, ‘tr[o]co’, ‘suf[o]co’), e aberta, se femininos (p.ex., ‘l[ε]va’, ‘r[ε]ga’, ‘tr[ɔ]ca’, ‘p[ɔ]da’).

Já no adjetivo a derivação regressiva tende ao fechamento. Assim, participios reduzidos costumam apresentar vogal fechada, independentemente de gênero ou número, p.ex., ‘qu[e]do/a’, ‘p[e]go/a’. De qualquer modo, essa regra é às vezes contrariada por formas inovadoras tais como ‘p[ε]go/a’, encontradas no dialeto carioca, entre outros.

Ora, inovações como essa são manifestações do já mencionado uso das vogais abertas como *default* na sílaba tônica. Trata-se, portanto, de um caso de analogia fonética em que o fator preponderante é, obviamente, a tonicidade. Na seção 2.5, serão examinados os demais fatores que intervêm no processo.

Aparentemente, essa versão mais genérica da deriva de abertura se iniciou com o aportuguesamento das palavras estrangeiras desde a ocupação árabe da Península Ibérica, e deslanchou quando o comércio exterior, estimulado pela navegação, trouxe uma nova onda de empréstimos.

É oportuno observar que, apesar de não possuir vogais médias, o árabe deu origem a muitas palavras portuguesas que as contêm. Algumas tiveram origem nos ditongos /aj/ e /aw/, p.ex. ‘xeque’ < *xayh* e ‘ceroulas’ < *sarawil*. Outras, porém, originaram-se nos monotongos /i, a, u/, p.ex., ‘alecrim’ < *al-iklil*, ‘alferes’ < *al-faris*, ‘alforge’ < *al-kburj*.

Isso leva à suspeita de que as vogais médias do galego-português tenham cedo adquirido o duplo papel exercido hoje: aproximar o vocabulário estrangeiro do vernáculo, e, ao mesmo tempo, marcá-lo como novidade. A marca de inovação é, justamente, a abertura na tônica.

Essa hipótese é corroborada por estatísticas da etimologia do vocabulário atual. A Tabela 2 exibe as frequências das vogais médias abertas e fechadas com étimos latinos e não-latinos nos substantivos e adjetivos com radical em vogal média do Mini-Aurélio (N=2039). As formas derivadas foram excluídas em virtude dos vieses introduzidos por alguns sufixos. A classificação etimológica segue o Dicionário Houaiss (HOUAISS & VILLAR, 2001).

São reportadas, além das frequências observadas, as frequências esperadas (com base no aleatório), e os desvios padronizados, que expressam a significância da célula em termos da curva normal. Desvios positivos maiores que 1,96 indicam uma associação entre as categorias cruzadas; desvios negativos menores que -1,96 indicam uma dissociação entre as categorias cruzadas.

As tabelas 3 e 4 exibem os valores de  $X^2$  e de  $\varphi$ , medida da força da associação, que, levando em conta o tamanho da amostra e o número de graus de liberdade, indica o percentual da variância dos dados explicado pelos cruzamentos da tabela de contingência.

TABELA 2 - Tabela de contingência para abertura *vs.* etimologia nas vogais médias portuguesas

Abertura vs. Etimologia					
		Etimologia		Total	
		Latina	Outra		
Abertura	Fechada	Observado	1457	119	1576
		Esperado	1236,7	339,3	1576,0
		Desvio padronizado	6,3	-12,0	
	Aberta	Observado	143	320	463
		Esperado	363,3	99,7	463,0
		Desvio padronizado	-11,6	22,1	
Total	Observado	1600	439	2039	
	Esperado	1600,0	439,0	2039,0	

TABELA 3 -  $\chi^2$  calculado pelos métodos de Pearson e razão de verossimilhança

Testes de $\chi^2$			
	Valor	Graus de liberdade	p
Pearson	803	1	0,00
Razão de verossimilhança	708	1	0,00
No. de casos válidos	2039		



TABELA 4 -  $\varphi$  para os valores de  $X^2$  da Tabela 3

Medida Simétrica de Força de Associação			
		Valor	p
2 x 2 Nominal	$\Phi$	0,63	0,00
No. de casos válidos		2039	

É instrutivo observar que a maioria dos 320 substantivos e adjetivos que apresentam vogal aberta na Tabela 2 data do português moderno, no qual o influxo de palavras estrangeiras aumentou consideravelmente devido ao avanço da língua escrita. Há, porém, palavras mais antigas cujas vogais fechadas foram arrastadas pela deriva de abertura, p.ex., ‘pele’ (séc. X), ‘forma’ (séc. XIII). É oportuno notar que essas têm, em geral, frequências altas e acepções múltiplas. A propósito, o latim *fōrma* deu dois substantivos em português: um abstrato de alta frequência, ‘f[o]rma’, afetado pela abertura, e outro concreto, de baixa frequência, ‘f[o]rma’, que preserva o étimo fechado.

Tudo isso faz suspeitar que, desde cedo, a deriva de abertura tenha sido uma poderosa estratégia para incorporar termos estrangeiros ou latinos cultos ao vernáculo. Assim, étimos fechados são preservados onde já estão consagrados pelo uso e/ou fazem parte da morfologia produtiva, como, p.ex., o sufixo ‘-dor’. Em contrapartida, quando a derivação concatena sufixos átonos a radicais em vogal média, o *default* é a abertura. Aí se incluem até formas cultas, tais como o sufixo adjetivo ‘-ico’, p.ex., ‘ético’, ‘ótico’, etc.; ou os radicais gregos átonos, p.ex., ‘telégrafo’, ‘hipódromo’, etc.

No século XX, o aumento do número de siglas e de palavras estrangeiras provenientes de múltiplas línguas provocou novos

surtos desse processo de nativização. São exemplos: ‘FAPESP’, ‘IBOPE’, etc.; e ‘coquet[ɛ]l’, ‘M[ɔ]tors’, etc. Note-se que os últimos são derivados de palavras inglesas com vogais longas e fechadas. Há também muitos nomes próprios cuja vogal fechada na língua de origem abre-se em português, p.ex.: ‘Scherer’ (alemão), ‘Lois’ (inglês), etc.

Outra manifestação atual da deriva de abertura é o seu uso como *default* em caso de dúvida. Assim, palavras raras, tais como ‘cepa’, ‘terso’, ‘absorto’, ‘torpe’, tendem a ser pronunciadas com vogais abertas quando encontradas pela primeira vez por leitores que as desconhecem.

Em suma, a suspeita levantada pela comparação das distribuições de tipos e ocorrências está, agora, amplamente corroborada: a deriva de abertura continua viva e conspira, em todas as classes gramaticais, para manter a carga funcional do contraste entre vogais médias abertas e fechadas na posição tônica. Cabe, agora, perguntar que fatores subjazem a essa conspiração.

## 2.5. Abertura, contexto fônico e identidade

Como já se viu, a abertura é previsível em vários contextos fônicos. Além dos casos já discutidos, cabe lembrar, ainda: o fechamento categórico em ambiente nasal<sup>24</sup>, e quase categórico em ditongos<sup>25</sup>; e, também, a existência de sufixos derivacionais produtivos, seja em vogal aberta, p.ex., ‘-eco’, ‘-el’, ‘-ório’, ‘-ol’, etc., seja em vogal fechada, p.ex., ‘-ete’, ‘-ês’, ‘-oto’, ‘-dor’, etc.

<sup>24</sup> Em alguns dialetos, o fechamento da vogal média tônica diante de uma consoante nasal heterossilábica tem exceções, tais como ‘[ɔ]ma’.

<sup>25</sup> Os ditongos ‘éi’ e ‘ói’ só ocorrem antes de vogal ou /s/ de plural, p.ex., ‘ideia’, ‘boia’, ‘papéis’, ‘sóis’.

TESCHNER (2006), seguindo PIZZINI (1982), investiga as percentagens de ocorrência da abertura e do fechamento diante das consoantes pós-vocálicas no corpus de FERREIRA (2005). Embora os resultados sejam sugestivos, essa análise tem dois problemas. Um é que as frequências brutas da maioria das classes de co-ocorrência são muito pequenas, o que compromete o cálculo da percentagem ou qualquer outra expressão da frequência relativa. O outro é que a previsibilidade devida à morfologia derivacional não é descontada. Fatores fonéticos e morfológicos estão, portanto, confundidos.

Mais importante que os fatores segmentais na determinação da abertura é a prosódia lexical, *i.e.*, a acentuação e a estrutura silábica.

Apesar de estar atuando há mais de um milênio, a deriva de abertura ainda está longe de haver equilibrado a distribuição das vogais médias, mesmo na posição tônica. As Figuras 7 e 8 exibem as quatro médias tônicas nos tipos e ocorrências no LAEL. Em ambos os casos, as fechadas são maioria absoluta.

A Figura 7 é consistente com a ideia de que a carga funcional da oposição entre as médias fechadas e abertas ainda é baixa no PB. Isso se deve, provavelmente, ao fato de o galego-português tê-la preservado à revelia da tendência à extinção no romance ibérico, consumada, depois, no espanhol. Em contrapartida, a Figura 8 mostra que a oposição entre as duas vogais médias anteriores vem sendo usada na língua a ponto de alterar significativamente a posição de /ɛ/ na escala de frequência. Esse uso constante favorece o seu recrutamento em processos de criação vocabular (BYBEE, 2001).

FIGURA 7 - Distribuição das vogais médias tônicas no LAEL tipos

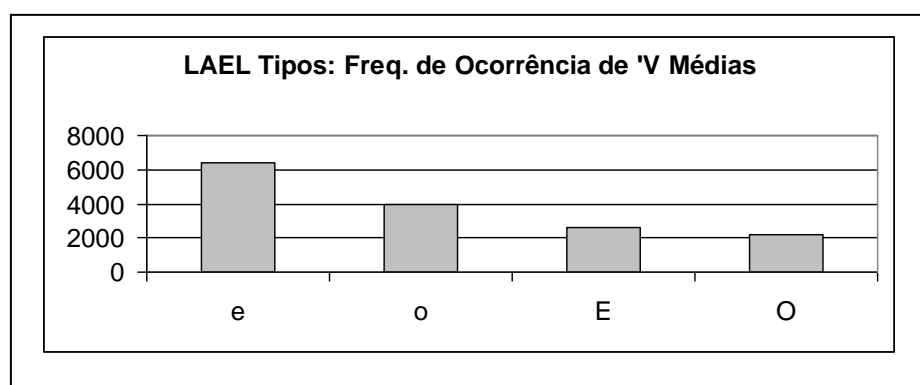
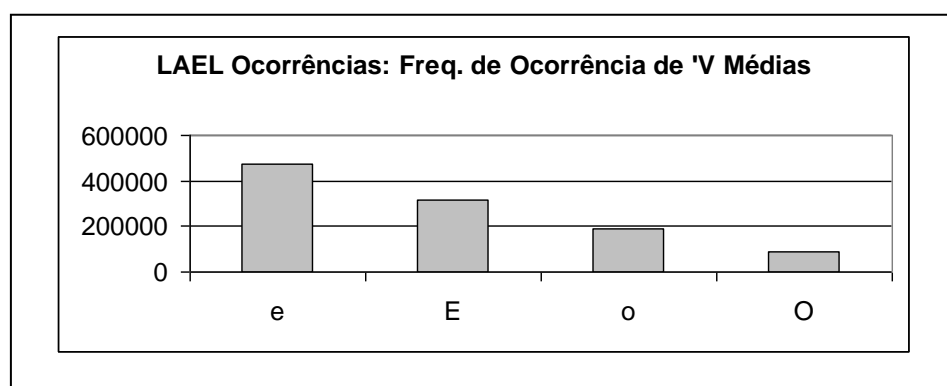


FIGURA 8 - Distribuição das vogais médias tônicas no LAEL ocorrências

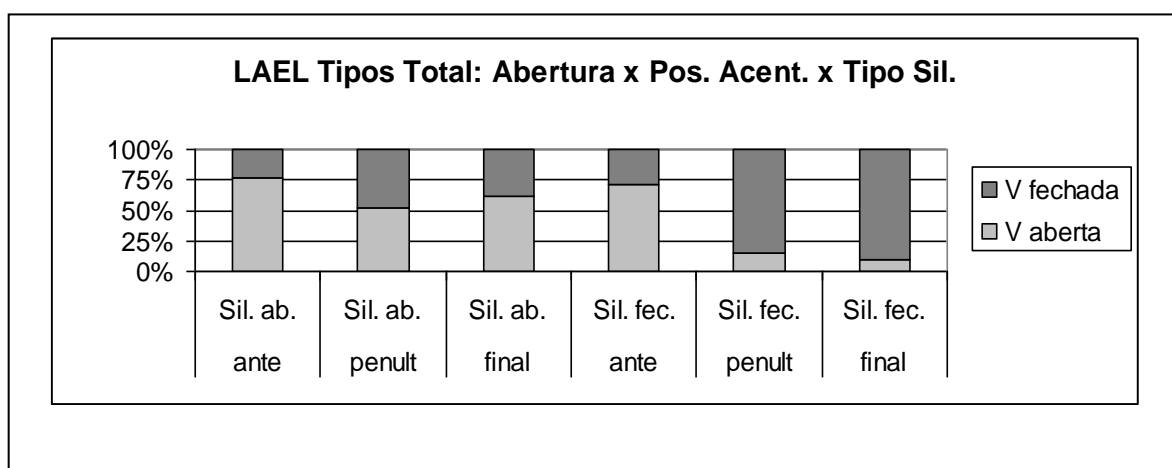


Aspectos mais específicos da prosódia lexical afetam a mesma deriva além da tonicidade. O primeiro a saltar aos olhos é a posição acentual.

A preferência das proparoxítonas pela vogal aberta é óbvia; e encontra-se, como já foi dito, morfologizada em certos processos derivacionais produtivos. Já as paroxítonas e oxítonas apresentam uma situação bem mais complexa, em que intervêm não só a morfologização, mas também o tipo silábico da tônica.

Isso fica claro na Figura 9, que exhibe os percentuais de tônicas médias abertas e fechadas por posição acentual e tipo silábico no LAEL tipos. As oxítonas e paroxítonas claramente preferem a vogal fechada em sílabas fechadas. Em sílabas abertas, por outro lado, aproximam-se do aleatório.

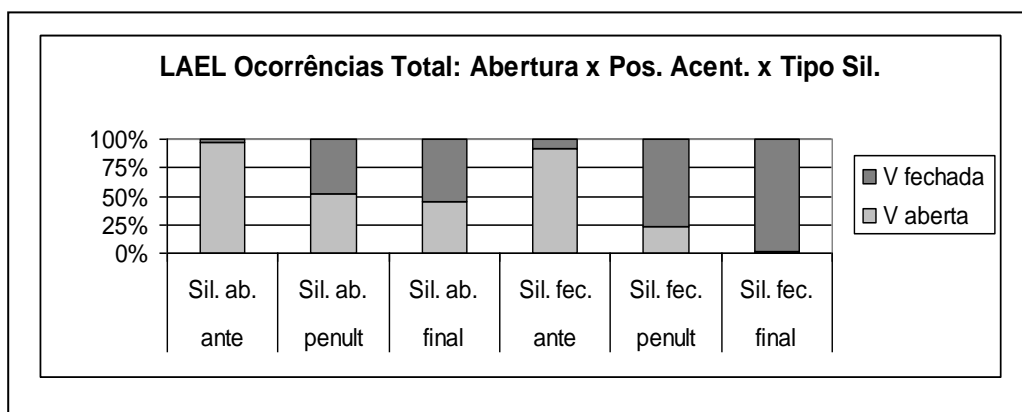
FIGURA 9 - Abertura *vs.* posição acentual e tipo silábico no LAEL tipos



Esse padrão é confirmado pelo LAEL ocorrências, exibido na Figura 10. A preferência das proparoxítonas pela vogal aberta é quase categórica. Contrasta com ela a preferência das paroxítonas e oxítonas pela vogal fechada em sílaba fechada, que é muito forte nas primeiras e praticamente categórica nas segundas. Já em sílaba aberta, a ausência de preferência de

ambas fica ainda mais clara, com conseqüente redução da marca da vogal aberta. A carga funcional da distinção de abertura é, portanto, aparentemente maior nessas posições, embora, como já vimos, esteja, também aí, em grande parte morfologizada.

FIGURA 10 - Abertura *vs.* posição acentual e tipo silábico no LAEL ocorrências



Esses fatos indicam que a deriva de abertura está de fato associada à incorporação de inovações ao vernáculo e não a uma possível tendência à perda da distinção. Essa interpretação é respaldada pelo caráter majoritário das paroxítonas (75% nos tipos e 73% nas ocorrências) e das sílabas abertas (59% dos tipos e 53% das ocorrências) no total do corpus. As paroxítonas abertas respondem, assim, por 35% dos tipos e 36% das ocorrências no subconjunto das vogais médias, o que é mais que o dobro do esperado sob a hipótese da equiprobabilidade das categorias prosódicas lexicais (17%).

Nesse ambiente, processos de derivação com sufixos tais como ‘-ete’, ‘-oso’, etc. contribuem para manter a proporção de vogais fechadas, ao mesmo tempo em que processos análogos

com sufixos tais como ‘-eco’, ‘-ote’, etc. somam-se aos novos radicais introduzidos pela deriva de abertura para manter a proporção de vogais abertas. Essa proporção é alimentada, ainda, por verbos derivados em ‘-ar’ e ‘-ecer’, sendo a contribuição dos últimos dividida entre a abertura e o fechamento<sup>26</sup>.

Parece, pois, que a carga funcional da distinção de abertura tem um baixo teor estritamente fonológico, manifesto em pares raros tais como ‘s[e]de/s[ɛ]de’ ou ‘m[o]ça’/‘m[ɔ]ssa’, e um alto teor morfofonológico, manifesto não só na morfologia flexional, mas também na morfologia derivacional. Em ambos os casos os processos implicados estão, como acabamos de ver, intimamente ligados à prosódia lexical.

A fim de aprofundar o último ponto, investiguemos um pouco mais a complexa interação entre as distribuições de frequência da abertura, da posição acentual e do tipo silábico no LAEL. Para isso a ferramenta apropriada é o modelo log-linear.

Trata-se de um método logarítmico de estimação de  $X^2$  que permite considerar as variáveis de uma tabela de contingência como fatores cruzados. As variáveis em questão e suas interações são reescritas como termos de uma soma; e estima-se o quanto cada termo e seus totais marginais contribuem para o valor total de  $X^2$ . Chama-se ‘parcial’ a estimativa baseada nos termos; e ‘marginal’, a baseada nos totais.

As Tabelas 5 e 6 exibem os resultados da análise log-linear dos fatores abertura, posição acentual e tipo silábico para o LAEL tipos e ocorrências. Um indício da complexidade das relações envolvidas é que, em ambas as tabelas, todos os fatores, bem como os seus cruzamentos, alcançam significância em ambas as estimativas<sup>27</sup>.

<sup>26</sup> Através da flexão de presente, do indicativo e do subjuntivo, respectivamente.

<sup>27</sup> O modelo log-linear exclui automaticamente os fatores cuja contribuição é

As estimativas parciais e marginais permitem comparar os tamanhos do efeito para graus de liberdade idênticos. Quando esses diferem, a interpretação é menos segura, principalmente se há discrepância entre os valores de  $X^2$  parcial e marginal. Abaixo os resultados estão ordenados decrescentemente pela estimativa parcial.

TABELA 5 - Modelo log-linear de 3 fatores aplicado ao LAEL tipos

LAEL Tipos	Graus de Liberdade	Estimativa Parcial	p Parcial	Estimativa Marginal	p Marginal
Pos. Ac.	2	12995.73	0.00	12995.73	0.00
Abert.	1	2471.17	0.00	2471.17	0.00
Abert. x Tipo Sil.	1	2282.36	0.00	2999.59	0.00
Pos. Ac. x Tipo Sil.	2	1395.79	0.00	2113.02	0.00
Tipo Sil.	1	578.84	0.00	578.84	0.00
Abert. x Pos. Ac.	2	320.88	0.00	1038.10	0.00

São dignas de nota, a propósito, as mudanças de ordem dos fatores entre os tipos e as ocorrências, à exceção da posição acentual. De fato, como já vimos, o caráter majoritário das paroxítonas permanece estável, respondendo, portanto, pela alta previsibilidade do fator.

Já a queda da abertura e de sua interação com tipo o silábico, e a consequente subida da interação entre o último fator e a posição acentual, devem-se ao ganho das tendências categóricas nas oxítonas e proparoxítonas, evidente no confronto das Figuras 9 e 10. Isso também explica a subida da interação entre a

---

insignificante para o  $X^2$  total.



abertura e a posição acentual.

TABELA 6 - Modelo log-linear de 3 fatores aplicado ao LAEL  
ocorrências

LAEL Ocorrências	Graus de Liberdade	Estimativa Parcial	p Parcial	Estimativa Marginal	p Marginal
<b>Pos. Ac.</b>	<b>2</b>	<b>1094647</b>	<b>0.00</b>	<b>1094647</b>	<b>0.00</b>
<b>Pos. Ac. x Tipo Sil.</b>	<b>2</b>	<b>205478</b>	<b>0.00</b>	<b>388463</b>	<b>0.00</b>
<b>Abert. X Pos. Ac.</b>	<b>2</b>	<b>166835</b>	<b>0.00</b>	<b>349820</b>	<b>0.00</b>
<b>Abert.</b>	<b>1</b>	<b>150271</b>	<b>0.00</b>	<b>150271</b>	<b>0.00</b>
<b>Abert. X Tipo Sil.</b>	<b>1</b>	<b>110329</b>	<b>0.00</b>	<b>293314</b>	<b>0.00</b>
<b>Tipo Sil.</b>	<b>1</b>	<b>24993</b>	<b>0.00</b>	<b>24993</b>	<b>0.00</b>

Note-se, contudo, que nenhuma mudança de ordem compromete a significância dos fatores ou interações, e que todas são semelhantes, embora não idênticas, para ambas as estimativas.

Pode-se dizer, portanto, que a disseminação da deriva de abertura na tônica está longe de ser aleatória no PB. Além das restrições categóricas quanto à acentuação, à ditongação e à nasalidade, há restrições probabilísticas quanto ao tipo silábico e à posição acentual – com algumas, como a das proparoxítonas, tendendo a categóricas. A mais importante é o predomínio de vogais fechadas nas oxítonas e paroxítonas em sílabas fechadas. Esse subconjunto perfaz 59% do corpus das vogais médias tônicas e fornece uma pista para o possível papel do acento intensivo na deriva de abertura.

O que nele chama atenção é a sensibilidade do acento ao peso silábico, maior que no resto do corpus, onde as tônicas fechadas perfazem apenas 40%. Essa diferença é posterior ao latim, já que as responsáveis por quase todo o excedente são as

oxítonas (18,5%). Lembrando que as codas eram proibidas nas penúltimas átonas do latim, é possível imaginar um cenário em que o acento quantitativo tenha convivido com o acento intensivo, dada a sua compatibilidade fonética. A fragilização da quantidade seria aí indicada por um aumento do número de sílabas tônicas com codas consonantais.

Ora, essa direção, cujos rastros estão no corpus, é respaldada pelo fato de certas consoantes terem sido atraídas para a coda precedente na época da queda de /e/ final no romance ibérico. Além disso, por volta do séc. X, o galego-português parece ter enfraquecido a pós-tônica final ainda mais, com a perda de consoantes intervocálicas e a preservação da vogal final em ditongo ou hiato, p.ex., ‘grau’, ‘boi’, ‘irmão’, ‘sabões’, ‘boa’, ‘voo’, etc., – o que ocasionou novas sílabas finais pesadas (WILLIAMS, 1962; MATTOS E SILVA, 1991). É, portanto, plausível que, durante a consolidação das distinções de abertura das vogais médias, o acento intensivo já possuísse uma força considerável – tendo como correlato fonético uma duração alofônica sensível ao tipo silábico e à posição acentual, como a do PB atual.

Cabe, a propósito, lembrar que, ao menos no PB sudeste, onde há poucas vogais médias abertas em posição átona, as relações entre duração e abertura são diametralmente opostas às das do latim: as mais longas são as abertas, ainda que não categoricamente (AQUINO, 1997). Trata-se, na verdade, de uma abertura gradiente, sensível ao tipo silábico e à posição acentual. Assim, as fechadas estão em geral ligadas à atonicidade, ou, em menor grau, à tonicidade em sílaba fechada, onde dividem a duração total com a consoante de coda. A esse respeito, vale notar que a razão átona/tônica nos tipos do LAEL é de 3,5 para as vogais /e, o/, contra 2,5 para o total do corpus.

Tudo isso sugere fortemente que o detalhe fonético

relacionado ao acento intensivo tenha tido um importante papel em disseminar a deriva de abertura. Como a documentação existente não permite abordar diretamente essa questão, é prudente examiná-la à luz das demais discussões, retomando-a apenas no final. Antes, porém, é preciso fechar o estudo da prosódia lexical, discutindo as suas implicações para a problemática da incorporação de inovações ao vernáculo e sua datação.

Nas proparoxítonas, pode-se dizer que a abertura é categórica, exceto nos contextos que exigem a vogal fechada. São eles: a nasalidade concomitante ou imediata, tal como em ‘ênfase’ ou ‘ônibus; e a presença mediata de vogal média na penúltima, tal como em ‘pêssego’ e ‘fôlego’<sup>28</sup>. A aparente exceção ‘bêbedo’ (séc. XV) confirma, na verdade, a regra, pois a variante ‘bêbedo’ é mais antiga (séc. XIII). De qualquer maneira, essas formas “harmônicas” são escassas e não parecem pesar muito atualmente. Haja vista o deslocamento acentual sofrido por ‘lêvedo’ (séc. XVI) no século XX, com a consagração de ‘levedo’.

Nas oxítonas, por outro lado, a abertura e o fechamento competem genuinamente. Nos empréstimos, por exemplo, a qualidade vocálica da língua de origem pode ser respeitada ou alterada por analogia fonética, conforme a “ganguê vocabular”<sup>29</sup> evocada. Assim, tem-se de um lado, ‘buquê’, ‘suflê’, ‘mantô’, ‘bistrô’; e, de outro, ‘café’, ‘balé’, ‘faraó’, ‘enxó’. Vale notar que uma mesma terminação, p.ex., o francês ‘-et’ [ɛ], pode gerar duplicidade, tal como em ‘croché’ e ‘crochê’.

Essa situação complexa contrasta com a das paroxítonas, nas quais a perda da marca da vogal aberta em sílaba aberta

<sup>28</sup> Observe-se que não há proparoxítonas com tônicas ditongadas.

<sup>29</sup> Termo proposto por BYBEE (2001: 28) para designar um conjunto de palavras foneticamente análogas.

contribuiu para a sua sedimentação como *default*, salvo em caso de interferência de alguma gangue antagônica (p.ex., Ban[e]SPA, Bov[e]SPA<sup>30</sup>). O ambiente prosódico lexical não-marcado parece, assim, ter atraído a analogia fonética<sup>31</sup> – presente, conforme visto no caso mais restrito do verbo, desde o português medieval. O atual *default* de abertura na paroxítona seria, portanto, apenas a generalização de uma tendência antes morfologizada. Vale lembrar, a propósito, que generalizações do tipo *default* podem ser modeladas por redes conexionistas dinâmicas (HARE *et al.*, *op. cit.*).

O que parece estar, de fato, em jogo aqui é a tendência, já documentada, a marcar o neologismo ou empréstimo como tal. Essa pressão se exerce mais naturalmente nas paroxítonas porque o fechamento poderia mascarar a nativização por estar associado a um contingente majoritariamente latino, conforme indicado na Tabela 2. As oxítonas são menos afetadas porque pertencem a um estrato lexical de datação mais recente, surgido no romance ibérico<sup>32</sup> e caracterizado por aportes estrangeiros significativos, mesmo se descontados os brasileirismos de origem indígena ou africana. Neste caso a própria acentuação já marca a inovação.

Vale notar que essa interpretação pode ser estendida às proparoxítonas, cuja maioria advém da língua escrita, constituindo, portanto, um estrato ainda mais especial e recente – sujeito, portanto, a dupla marcação.

Um argumento zipfiano a favor da abertura como marca de inovação incorporada ao vernáculo é a taxa de vogais médias

---

<sup>30</sup> Apesar de a gangue das fechadas ser mínima e infrequente ('vespa', 'crespa'), as duas palavras em '[e]spa' da língua, 'encrespa' e 'desencrespa', pertencem ao paradigma verbal da primeira conjugação.

<sup>31</sup> Esta noção, com que SCHUCHARDT (*op. cit.*) combateu os neogramáticos, vem sendo operacionalizada pela fonologia de laboratório (DE SCHRYVER *et al.*, 2008).

<sup>32</sup> Cabe lembrar que o latim não possuía oxítonas.

fechadas e abertas nas palavras com frequência superior a 1000 no LAEL. Nos tipos, ela é de 2,5, valor superior ao do total do corpus, que é de 2.1. Isso implica uma maior concentração de fechadas entre as palavras de alta frequência, que, segundo Zipf, são as mais antigas. Já nas ocorrências, como é de esperar, dada a alta frequência de algumas palavras em vogal aberta (p.ex., 'é', 'pode'), os valores são muito próximos, a saber: 1,5 e 1,6, respectivamente.

Assim, parece que a deriva de abertura atuou intermitentemente em quatro tempos. O primeiro data do galego-português, e inovou principalmente a flexão verbal. O segundo data do renascimento, e nativizou o léxico trazido pela navegação e demais avanços da nova era. O terceiro data do início da Idade Moderna, e incorporou ao vernáculo os empréstimos devidos à expansão da escrita. O quarto data do séc. XX, e vem nativizando o léxico trazido pela sociedade pós-industrial e a globalização.

É inescapável a conclusão de que as vogais médias do português têm uma relação complexa, mas altamente produtiva, com o modo pelo qual o uso renova o sentido gramatical e lexical. Aí reside, provavelmente, a sua intratabilidade na literatura estruturalista e gerativista. Os fatos são contundentes em abalar o pressuposto da autonomia da fonologia.

Com isso em mente, passemos, por fim, à discussão do papel do acento intensivo e do detalhe fonético associado na deriva de abertura.

### À guisa de conclusão: sobre a possível motivação da deriva de abertura

Não caberia aqui resenhar a vasta literatura sobre o ritmo do português, dada a distância entre o PB e o PE<sup>33</sup>. Ademais, a controvérsia sobre os ritmos silábico ou acentual é, na verdade, marginal à principal questão aqui envolvida, que é a da relação entre o acento intensivo e a altura da mandíbula.

Embora essa relação ainda não tenha sido investigada no português, há suficiente evidência, com base no inglês, de que o acento intensivo e a ênfase estão correlacionados ao abaixamento da mandíbula (FUJIMURA, 1986; ERICKSON, 1998).

Sabe-se, por outro lado, que esse gesto articulatório está envolvido, ainda que não necessariamente, na distinção entre as vogais médias abertas e fechadas. A outra possibilidade é retraindo a raiz da língua em direção à faringe – gesto cuja recorrência em processos de harmonia vocálica nas línguas do mundo levou à proposição do traço [+/-ATR] – “raiz avançada da língua” (*advanced tongue root*). Assim, é igualmente natural realizar a diferença entre [ɛ, ɔ] e [e, o] apenas pelo avanço ou recuo da raiz da língua, embora, por economia de esforço, as línguas tendam a preferir o abaixamento da mandíbula em ambientes tônicos.

Vieses estatísticos afins são atestados no LAEL: as tônicas preferem /ɛ, a, ɔ/, assim como as átonas preferem /i, e, o, u/<sup>34</sup>. Não obstante, como é de esperar, já que um viés maior comprometeria a carga funcional do sistema vocálico, essa associação é fraca, exibindo valores de  $\varphi$  inferiores a 0,20 e explicando, portanto, menos de 20% da variância dos dados.

Com o aumento de duração na tônica devido ao acento intensivo, o uso alternativo do abaixamento da mandíbula e da

<sup>33</sup> V., p.ex., FROTA & VIGÁRIO (2001), MEIRELES *et al.* (2010).

<sup>34</sup>  $X^2$  tipos = 4874, e  $X^2$  ocorrências = 136519; respectivamente,  $\varphi$  = -0,15, e  $\varphi$  = -0,17.

retração da raiz da língua pode ter-se consagrado no romance ibérico na época do surgimento das primeiras oxítonas. O desenvolvimento subsequente desse uso – com declínio no espanhol, e cultivo no galego-português – pode estar relacionado à diferenciação rítmica das duas línguas. Deve-se, entretanto, lembrar que a deriva de abertura prosperou no galego e no PB, não dependendo, portanto, da exacerbação do ritmo acentual que se observa hoje no PE.

O que parece ter alimentado a preservação da distinção entre as vogais médias no galego-português é, portanto, uma conjunção de fatores fonéticos e léxico-gramaticais que desafia qualquer noção de sistema fonológico abstrato e autônomo. De um lado, havia uma tendência a associar o gradiente fisiológico da altura mandibular aos gradientes acústicos de intensidade, duração e abertura que veiculam a distinção sob acento intensivo. De outro, havia uma tendência não só a lexicalizar tal associação – marcando novos estratos de vocabulário –, mas também a morfologizá-la – recorrendo à apofonia para nivelar ou estender paradigmas em tônica aberta, o que foi uma forma inovadora de compensar a perda da flexão latina.

Com o passar dos séculos, a associação entre os fatores fonéticos e léxico-gramaticais subjacentes à deriva de abertura tornou-se natural e mesmo necessária. Hoje o uso das distinções entre as vogais médias do português exige o comando de todos eles, constituindo não só um índice de proficiência (TESCHNER, *op. cit.*), mas também uma marca de identidade.

É lícito, portanto, propor que as mais marcadas dessas vogais, *i.e.*, as abertas, tenham, hoje, o estatuto de fonestemas<sup>35</sup> difusos, *i.e.*, que a sua presença numa palavra evoque uma gama

---

<sup>35</sup> Um fonestema é um fragmento fônico sem estatuto mórfico que, entretanto, se associa frequentemente a algum significado. O exemplo mais conhecido é o grupo inicial ‘gl’ em inglês, que evoca luminosidade: ‘glee’, ‘gleam’, ‘gloss’, etc.

de sentidos gramaticais ou lexicais aos quais se pode agregar informação estratal, identitária, ou mesmo, como se verá, certo simbolismo fonético.

Há sentidos tradicionais bem estabelecidos, como o de “presente” e “modo”, na flexão verbal, ou os de “gênero” e “número”, na flexão nominal. Há também os sentidos “vernacular” e “nativizado”, sobre os quais já muito se discorreu. E há, ainda, sentidos menos óbvios, como os dos diminutivos pejorativos em ‘-ote’ e ‘-eco’, – que parecem evocar, de maneira contraditória e humorística, o gesto de abertura mandibular. Sem contar com a forte presença, no PB, dos estratos lexicais indígenas e africanos, manifestos em palavras tais como ‘pajé’, ‘cipó’, ‘axé’, ‘bobó’, etc. Esse padrão oxítono é, aliás, um solo fértil para a criatividade popular: vejam-se certos termos coloquiais de origem duvidosa, tais como ‘picolé’ e ‘xodó’.

Cabe, finalmente, retornar às considerações teóricas da introdução. A compreensão da deriva de abertura alcançada pela tentativa de coadunar as noções de marca e carga funcional de Trubetzkoy com a visão dinâmica do léxico inaugurada por Zipf coloca muitas questões para as próprias linhas de pensamento que a tornaram possível.

Para os estudos da disseminação da mudança fônica, tal como entendida por LABOV (1981), surge uma categoria ainda não reportada: a mudança foneticamente abrupta, e lexicalmente, nem abrupta, nem gradual, mas intermitente. Essa intermitência levanta, aliás, questões sobre os limites entre a analogia fônica e a analogia mórfica que os avanços do conexionismo talvez possam um dia iluminar.

Para os estudos da motivação da mudança fônica, tal como entendida pelas fonologias do uso, surge o desafio de encontrar um papel não-circular para a noção de frequência. Talvez as



discrepâncias aqui encontradas entre os tipos e as ocorrências nas distribuições segmentais do LAEL sejam um primeiro passo nessa direção.

É perfeitamente plausível que uma intensificação do uso de um gesto articulatório possa induzir a sua extensão a novos tipos. Essa extensão tem um motivo zipfiano: o poder do vocabulário básico de não apenas se conservar, mas também precipitar possíveis inovações em palavras menos comuns. Aí subjaz, todavia, um motivo trubetzkoyano: tal poder só pode se exercer graças a uma pressão por equilíbrio decorrente da baixa carga funcional de alguma distinção vigente.

Para a ciência contemporânea, essa pressão nada tem de teleológico: indica apenas a emergência de um patamar de estabilidade durante a trajetória temporal do sistema fônico, entendido como sistema dinâmico, conforme sugerido ao longo desta análise.

Finalmente, para as fonologias de laboratório, surge a questão de explicar a associação entre a abertura e o acento. Trata-se de um problema abordável para visões que contemplam representações articulatórias; e algo desconcertante para visões que contemplam apenas representações auditivas. A fonologia gestual, em particular, tem ferramentas que podem fomentar o entendimento da interação prosódia/segmentos de modo a esclarecer o papel que aí desempenham os gestos alternativos e/ou compensatórios (SALTZMAN *et al.*, 2008).

Por isso, num futuro não muito remoto, talvez seja possível fazer medidas fonéticas e/ou psicolinguísticas que sirvam de base a um modelo capaz de simular a não-linearidade e a intermitência da trajetória da deriva de abertura no português. A teoria dos sistemas dinâmicos estará à mão, então, para inserir tal esforço num quadro de referência transdisciplinar que faça jus ao caráter ao mesmo tempo físico e sócio-cultural das línguas

naturais humanas.

## Referências

ALBANO, E. C. **O “Tear Encantado”, a Complexidade, a Tecnologia Digital e o Pensamento Interdisciplinar sobre a Linguagem.** Campinas: Remate de Males, 2009. vol. 29, no.1, p. 41-57.

ALBANO, E. C. **Some challenging facts of Brazilian Portuguese phonology.** Conferência de abertura da São Paulo School of Advanced Studies in Speech Dynamics, Universidade de São Paulo, Instituto de Matemática, 7 de junho de 2010. Disponível em:

[http://www.dinafon.iel.unicamp.br/spsassd\\_files/Some-challenging-facts-of-BP-phonology.pdf](http://www.dinafon.iel.unicamp.br/spsassd_files/Some-challenging-facts-of-BP-phonology.pdf)

ALBANO, E. C. & MOREIRA, A. A. **Archisegment-based letter-to-phone conversion for concatenative speech synthesis in Portuguese.** *Proceedings: ICSLP'96.* 1996. v.3, p. 1708-1711.

AOYAMA, K.; FLEGE, J. E.; GUION, S.; AKAHANE-YAMADA, R. & YAMADA, T. **Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/.** *Journal of Phonetics*, 2004. vol. 32, 233-250.

AQUINO, P. A. **O Papel das Vogais Reduzidas Pós-Tônicas na Construção de um Sistema de Síntese Concatenativa para o Português do Brasil.** Campinas: Universidade Estadual de Campinas, 1997. Dissertação

(Mestrado).

BERBER-SARDINHA, T. **Linguística de Corpus: Histórico e problemática.** DELTA. Documentação de Estudos em Linguística Teórica e Aplicada. São Paulo: DELTA, 2000. v. 16, n. 2, p. 323-367.

BISOL, L. **Harmonização vocálica: uma regra variável.** Rio de Janeiro: Universidade Federal do Rio de Janeiro, 1981. Tese (Doutorado).

BYBEE, J. **Phonology and Language Use.** Cambridge: Cambridge University Press, 2001.

BYBEE, J. **Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change.** Language Variation and Change. 2002. vol. 14, p. 261-290.

CALLOU, D.; LEITE, Y. & MORAES, J. **Processos de enfraquecimento consonantal no português do Brasil.** In: ABAURRE, M. B; RODRIGUES, A. (Orgs.) *Gramática do português falado*. Vol. VIII: novos estudos descritivos. Campinas, UNICAMP/FAPESP, 2002. p. 537-555.

CAMARA JÚNIOR, J. M. **Estrutura da língua portuguesa.** 3ª ed. Petrópolis: Vozes, 1970.

CAMARA JR., J. M. **História e estrutura da língua portuguesa.** 2ª ed. Rio de Janeiro, Padrão, 1976.

CAMPOS, C. S. O. & CRISTÓFARO-SILVA, T. **Abertura vocálica em verbos irregulares do português.** Alfa: Revista de Linguística, 2008. vol. 52, no. 2, p. 371-390.

CAVACAS, A. A. **A língua portuguesa e sua metafoia.** Coimbra: Imprensa da Universidade, 1920.

CHEN, M. & WANG, W. S-Y. **Sound change:** actuation and implementation. *Language*, 1975. vol. 51, p. 255-81.

CHOMSKY, N. & HALLE, M. **The sound pattern of English.** Nova Iorque: Harper & Row, 1968.

COETZEE, A.; KAGER, R. & PATER, J. **Introduction:** phonological models and experimental data. *Phonology*, 2009. vol. 26, p. 1-8.

DE SCHRYVER, J.; NEIJT, A.; GHESQUIÈRE, P. & ERNESTUS, M. **Analogy, frequency and sound change, the case of Dutch devoicing.** *Journal of Germanic Linguistics*, 2008. vol. 20, no. 1, 59-195.

DE JONG, K.; BECKMAN, M. & EDWARDS, J. **The interplay between prosodic structure and coarticulation.** *Language & Speech*, 1993. 36, p. 197-212.

DUBERT-GARCIA, F. **Sobre as origens da vogal radical /i/ em sigo-siga no verbo galego-português:** um fenómeno de contacto linguístico? Instituto da Língua Galega-USC, 2010.

Disponível em:

<http://webspersoais.usc.es/export/sites/default/persoais/francisco.dubert/arquivos/origensalternancia.pdf>

Acesso em 5/11/2010.

ELMAN, J. L.; BATES, E. A.; JOHNSON, M. H.; KARMILOFF-SMITH, A.; PARISI, D. & PLUNKETT, K. **Rethinking Innateness:** A connectionist perspective on

development. Cambridge MA: MIT Press, 1996.

ERICKSON, D. **Effects of contrastive emphasis on jaw opening.** *Phonetica*, 1998, vol. 55, p.147-69.

ERICKSON, D.; FUJIMURA, O. & PARDO, B. **Articulatory correlates of prosodic control:** Emotion and emphasis. *Language & Speech*, 1998. vol. 41, p. 395-413.

FERREIRA, A. B. H. **Novo Dicionário da língua portuguesa.** 4ª ed. Rio de Janeiro: Nova Fronteira, 2005.

FOULKES, P.; SCOBIE, J. M. & WATT, D. **Sociophonetics.** In: HARDCASTLE W. J.; LAVER, J.; GIBBON, F. E. (Orgs.) *The Handbook of Phonetic Sciences*. 2ª ed. Oxford: Blackwell Publishing Ltd., 2010, cap.19, p. 703-754.

FROTA, S. & VIGÁRIO, M. **On the correlates of rhythmic distinctions:** the European/Brazilian Portuguese case. *Probus*, 2001. 13.2, p. 247-275.

FUJIMURA, O. **Relative Invariance of Articulatory Movements.** In: PERKELL, J. S.; KLATT, D. H. (Orgs.) *Invariance and Variability in Speech Processes*. Hillsdale: Lawrence Erlbaum, 1986. p. 226-42.

HARE, M.; ELMAN, J. L. & DAUGHERTY, K. G. **Default Generalization in Connectionist Networks.** *Language and Cognitive Processes*. 1995. v.10, 6, p. 601-630.

HARRIS, J. **Evidence from Portuguese for the ‘Elsewhere Condition’ in phonology.** *Linguistic Inquiry*. 1974. vol. 5, p. 61–80.

HOUAISS, A. & VILLAR, M. S. **Dicionário Houaiss da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2001.

ITO, J. & MESTER, A. **The structure of the phonological lexicon**. In: TSUJIMURA, N. (Org.) *The Handbook of Japanese Linguistics*. Malden, MA, & Oxford: Blackwell Publishers, 1999. p. 62-100.

KING, R. D. **Functional load and sound change**. *Language*. 1967. vol. 43, p. 831-852.

KINGSTON, J. & BECKMAN, M. **Papers in Laboratory Phonology I**. Cambridge: Cambridge University Press, 1990.

KIPARSKY, P. **How abstract is phonology?** In: FUJIMURA, O. (Org.). *Three Dimensions of Linguistic Theory*. Tóquio: TEC, 1973 [1968]. p. 1-136.

KIPARSKY, P. **“Elsewhere” in phonology**. In: Anderson, S. R.; Kiparsky, P. (Orgs.). *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston, 1973. p. 93-106.

KUHN, T. S. *The Structure of Scientific Revolutions*. 2<sup>a</sup> ed. Chicago: Univ. of Chicago Press, 1970.

LABOV, W. **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press, 1972.

LABOV, W. **Resolving the Neogrammarian Controversy**. *Language*. 1981. vol. 57, no. 2, p. 267-308.

MAIA, C. A. **História do Galego-Português**. Coimbra: Fundação Calouste Gulbenkian & JNICT, 1986.

MATEUS, M. H. M. **Aspectos da fonologia portuguesa.** Lisboa: Instituto Nacional de Investigação Científica, 1975.

MATEUS, M. H. M. & PARDAL, E. A. **The Phonology of Portuguese.** Oxford: Oxford University Press, 2000.

MATTOS E SILVA, R. V. **O português arcaico:** fonologia. São Paulo/Bahia: Contexto/Universidade Federal da Bahia, 1991.

MEIRELES, A. R.; TOZETTI, J. P. & BORGES, R. R. **Speech rate and rhythmic variation in Brazilian Portuguese.** Speech Prosody 2010. Disponível em: <http://aune.lpl.univ-aix.fr/~sprosog/sp2010/papers/100875.pdf>  
Acesso em 15/12/2010.

MENEZES, C.; PARDO, B.; ERICKSON, D. & FUJIMURA, O. **Changes in syllable magnitude and timing due to repeated correction.** Speech Communication. 2003. vol. 40, 71-85.

MEZZOMO, C; RIBAS, L. **Sobre a aquisição das líquidas.** In: LAMPRECHT, R.; BONILHA, G.; FREITAS, G.; MATZENAUER, C. L.; MEZZOMO, C.; OLIVEIRA, C.; RIBAS, L. *Aquisição fonológica do português: perfil de desenvolvimento e subsídios para terapia.* Porto Alegre: Artmed, 2004.

MILLER, G. & CHOMSKY, N. **Finitary models of language users.** In: LUCE, R.D.; BUSH, R.R.; GALANTER E. (Orgs.). *Handbook of Mathematical Psychology.* Nova Iorque: Wiley, 1963. Vol. II. p. 419-493.

OHALA, J. **Experimental historical phonology.** In: ANDERSON, J. M.; JONES, C. (Orgs.). *Historical linguistics II. Theory and description in phonology.* Amsterdã: North Holland, 1974. pp. 353-389.

OSTHOFF, H. & BRUGMANN, K. **Preface to Morphological Investigations in the Sphere of the Indo-European Languages.** In: LEHMANN, W. P. (Org.). *A Reader in Nineteenth-Century Historical Indo-European Linguistics. Indiana University Studies in the History and Theory of Linguistics.* Bloomington & Londres: Indiana University Press, 1967 [1878], Vol. I. p. 197-209.

PAUL, H. **Principles of the history of language.** Londres: Swan Sonnenschein, 1888.

PIERREHUMBERT, J. **Exemplar dynamics:** Word frequency, lenition, and contrast. In: J. BYBEE and P. HOPPER (Orgs.). *Frequency effects and the emergence of lexical structure.* Amsterdam: John Benjamins, 2001. p. 137-157.

PIZZINI, Q. **Portuguese mid-vowels.** *Papers in Romance*, 1982. vol. 4, no. 3, 197-209.

PORT, R. & VAN GELDER, T. **Mind as Motion:** Explorations in the Dynamics of Cognition. Cambridge: MIT Press, 1995.

PUSTET, R. **Zipf and his heirs.** *Language Sciences*. 2004. vol. 26, p. 1-25.

QUÍCOLI, A. C. **Harmony, lowering and nasalization in Brazilian Portuguese.** *Lingua*. 1990. vol. 80, p. 295–331.



ROCES-RODRIGUES, L. **Relações gradientes V\_V em sequências CVC no português brasileiro.** Campinas: Universidade Estadual de Campinas, 2010. Tese (Doutorado).

RUMELHART, D. E.; MCCLELLAND, J. L. & THE PDP RESEARCH GROUP. **Parallel Distributed Processing: Explorations in the Microstructure of Cognition.** Foundations, Cambridge: MIT Press, 1986. vol 1.

SALTZMAN, E.; NAM, H.; KRIVOKAPIC, J. & GOLDSTEIN, L. **A task-dynamic toolkit for modeling the effects of prosodic structure on articulation.** In: BARBOSA, P. A.; MADUREIRA, S.; REIS, C. (Orgs.). *Proceedings of the 4th International Conference on Speech Prosody* (Speech Prosody 2008). Campinas, 2008. p. 175-184.

SAPIR, E. **Language:** An introduction to the study of speech. Nova Iorque: Harcourt, Brace & Co., 1921.

SCHUCHARDT, H. **On sound laws:** against the Neogrammarians. In: VENNEMANN, T.; WILBUR, T. H. (Orgs.). *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change.* Frankfurt: Athenäun, 1972 [1885]. p. 39-72.

SILVA, A. H. P.; MOREIRA, A. A.; VILLA, M. H. L. & AQUINO, P. A. **Codificação Fonológica Informatizada do Minidicionário Aurélio:** um banco de dados para o estudo da fonologia portuguesa. XLI Seminário do Grupo de Estudos Linguísticos do Estado de São Paulo (GEL), 1994, Ribeirão Preto. Anais do XLI GEL, v. 1, p. 1321-1327, 1993.

SILVA NETO, S. **História da língua portuguesa.** 3.a ed. Rio

de Janeiro & Brasília: Presença, Instituto Nacional do Livro, 1979.

TESCHNER, R. V. **Graphotactics of Portuguese Tonic and Oral Mid-Vowels: Open, Closed and Offglided.** *Hispania*. 2006. vol. 89, no. 4, 950-968.

TOMAZ, K. S. **Análise de vogais médias posteriores em formas nominais de plural do português brasileiro.** Belo Horizonte: Universidade Federal de Minas Gerais, 2006. Dissertação (Mestrado).

TRUBETZKOY, N. S. **Principes de Phonologie.** Paris: Klincksieck, 1957 [1939].

WANG, W. S-Y. **Competing sound changes as a cause of residue.** *Language*. 1969. vol. 45, p. 9-25.

WETZELS, W. L. **Mid-vowel neutralization in the Brazilian Portuguese verb.** *Phonology*. 1995. vol. 12, p. 281–304.

WILLIAMS, E. **From Latin to Portuguese: Historical phonology and morphology of the Portuguese language.** Filadélfia: University of Pennsylvania Press, 1962.

ZIPF, G. K. **The Psychobiology of Language.** Boston: Houghton-Mifflin, 1935.

\_\_\_\_\_. **Human Behavior and the Principle of Least Effort.** Cambridge: Addison-Wesley, 1949.