

O papel do léxico no alçamento sem motivação aparente das vogais médias pretônicas do português brasileiro

The role of the lexicon in pretonic mid-vowel raising without apparent motivation in Brazilian Portuguese

Susiele Machry da Silva

UTFPR

susiele.machry@gmail.com

Ana Paula Correa da Silva Biasibetti

PUCRS

biasibetti.ana@gmail.com

Resumo: Este artigo tem por objetivo discutir o papel da frequência lexical no processo variável de alçamento sem motivação aparente das vogais médias pretônicas na variedade de português brasileiro falada em Porto Alegre/RS. O estudo, realizado a partir de dados dos trabalhos desenvolvidos por Bisol (1981), Klunck (2007), Cruz (2010) e Silva (2014), investiga a atuação do léxico a partir dos preceitos da Fonologia de Uso (BYBEE, 2001) e da Teoria de Exemplares (JOHNSON, 1997; PIERREHUMBERT, 2001, 2002). Parte-se da hipótese de que o alçamento se propaga prioritariamente pelos paradigmas frequentes cujas redes de associação apresentam estrutura mais densa, isto é, quanto maior o número de itens lexicais associados ao paradigma, maior a probabilidade de o alçamento se propagar. Os resultados obtidos, em conformidade com essa hipótese, sugerem que o processo atinge, inicialmente, os paradigmas mais frequentes e, gradualmente, se espalha para outras palavras com estrutura igual ou semelhante.

Palavras-chave: vogais médias pretônicas; alçamento sem motivação aparente; variação; léxico; frequência.

Abstract: This paper aims to discuss the role of lexical frequency in pretonic mid-vowel raising without apparent motivation in the Brazilian Portuguese variety spoken in Porto Alegre/RS. This study reexamines the data analyzed by Bisol (1981), Klunck (2007), Cruz (2010) and Silva (2014) and investigates the role of the lexicon under the assumptions of Usage-Based Phonology (BYBEE, 2001) and Exemplar Theory (JOHNSON, 1997; PIERREHUMBERT, 2001, 2002). It is hypothesized that the raising propagates preferentially through the frequent paradigms whose networks of association have a denser structure, that is, the larger the number of lexical items associated with the paradigm, the higher the chance of the raising propagates. The results confirmed the hypothesis, suggesting that the process initially occurs in the most frequent paradigms and spreads gradually to other words with the same or similar structure. **Keywords:** pretonic mid-vowels; raising without apparent motivation; variation; lexicon; frequency.

Recebido em 21 de maio de 2015.

Aprovado em 15 de outubro de 2015.

Introdução

A análise sobre como os processos de mudança sonora se propagam pela língua pode ser conduzida sob duas diferentes perspectivas, uma centrada no som e a outra na palavra. As abordagens remetem à controvérsia teórica dos modelos neogramático (PAUL, 1983) e difusionista (WANG, 1969; CHEN; WANG, 1975): o primeiro enfatiza o papel do ambiente fonético/fonológico para a ocorrência da mudança sonora, enquanto o segundo enfatiza o papel do léxico na propagação da mudança.

Segundo a proposta neogramática, a mudança linguística opera de forma lexicalmente abrupta e foneticamente gradual, com tendência a propagar-se por todos os ambientes que possuem contexto fonético propício para a aplicação da regra. Em outras palavras, a mudança ocorre a partir de um condicionamento estrutural interno ao sistema linguístico.

Ainda que a perspectiva neogramática encontre embasamento teórico em muitos fenômenos variáveis observados, há casos em

que a mudança sonora não se propaga com regularidade mesmo em contexto favorável para a sua aplicação. Ao que parece, o potencial de implementação da mudança gerado por um condicionamento fonético supostamente favorecedor não é, necessariamente, a única força capaz de impulsionar a aplicação de um processo variável. Diante desse argumento, a pressuposição, corroborada por Labov (1981), é de que nem todos os processos de mudança sonora observados nas línguas são de natureza neogramática e, por isso, a escolha do modelo teórico deve pautar-se pelo tipo de fenômeno analisado.

O alçamento variável das vogais /e/ e /o/ em posição pretônica no português brasileiro (doravante PB) pode acontecer por harmonização vocálica (*menino~minino; coruja~curuja*), ou, por motivo não aparente (*senhor~sinhor; governo~guverno*). De acordo com Bisol (2009, p. 78), o alçamento por harmonização vocálica (doravante HV) e o alçamento sem motivação aparente (doravante AL) são processos formalmente diferentes. A HV é entendida como um caso de assimilação em que a vogal média pretônica assimila o traço de altura de uma vogal alta localizada em sílaba subsequente, tratando-se, portanto, de um processo fonético. O AL, por sua vez, representa um caso de neutralização entre as vogais médias altas e as vogais altas, ainda não categórico, cujo modo de implementação sugere um processo de cunho lexical, haja vista a ausência de um gatilho fonético explícito para a ocorrência do alçamento.

Em relação à variedade gaúcha, o processo de HV é o representante típico de uma regra neogramática, pois apresenta aplicação moderada e condicionamento fonético explícito (BISOL, 1981; SCHWINDT, 1995, 2002; CASAGRANDE, 2003). O processo de AL, por sua vez, apresenta baixo índice de aplicação e, quando ocorre, tende a afetar somente determinados itens lexicais, com maior propensão a atuar sobre paradigmas específicos (*piqueno, piquena, piquenos, piquenas*, por exemplo), o que compromete a força do papel condicionador dos grupos de fatores linguísticos contextuais (KLUNCK, 2007; CRUZ, 2010; SILVA, 2014). Nesse sentido, a natureza do AL afasta-se de uma regra neogramática e apresenta fortes indícios de que se trata de um processo difusionista.

A propagação do AL, pela hipótese de que parte este estudo, é controlada pelo léxico e afeta, majoritariamente, os itens lexicais conectados a paradigmas, ou seja, aqueles que apresentam estruturas sonora e semântica semelhantes. Trabalha-se, portanto, com a premissa

de que a morfologia tende a influenciar no processo de propagação do alçamento, no sentido de que propicia a formação de redes de associação com base nas relações paradigmáticas entre os itens lexicais (BYBEE, 2001, 2002, 2010). Não obstante, espera-se que essa propagação seja favorecida por paradigmas que se apresentem como mais frequentes na língua e que possuam ambiente fonético propício.

Ante o exposto, este artigo se propõe a investigar a atuação do léxico na aplicação variável do AL na variedade do PB falada em Porto Alegre/RS a partir dos dados analisados por Bisol (1981), Klunck (2007), Cruz (2010) e Silva (2014). Assume-se um possível efeito de frequência e a formação de redes de palavras a partir de paradigmas como fatores viabilizadores do AL. Tomam-se por base os pressupostos da Teoria de Exemplares (JOHNSON, 1997; PIERREHUMBERT, 2001, 2002, 2003) e da Fonologia de Uso (BYBEE, 2001, 2002, 2010), uma vez que, pela hipótese de que parte este estudo, os referidos modelos podem oferecer o aparato teórico necessário para explicar a forma como o AL se propaga na variedade em exame.

Fonologia de Uso e Teoria de Exemplares

O modelo de Difusão Lexical (WANG, 1969; CHEN; WANG, 1975) prevê que a mudança sonora se propaga gradativamente através do léxico, atingindo palavras individuais ou grupos de palavras que possuem estrutura sonora semelhante. Por admitir exceções, a proposta difusionista mostra-se convincente para explicar diferentes tipos de mudança sonora, incluindo até mesmo aquelas que são entendidas como regulares por serem foneticamente motivadas (caso da HV no PB, por exemplo). Entretanto, conforme questiona Oliveira (1991), uma vez que a mudança não atinge todas as palavras ao mesmo tempo, torna-se necessário investigar por que algumas palavras sofrem alterações mais rapidamente do que outras.

Uma possível explicação para a aplicação irregular de processos variáveis advém da Fonologia de Uso (BYBEE, 2001, 2002, 2010), para a qual a seleção lexical é determinada pela frequência de uso das palavras. De acordo com a proposta desse modelo, as mudanças podem ser simultaneamente fonética e lexicalmente graduais, porém é a frequência de uso o fator que determina a ordem e a extensão em que as palavras são afetadas.

Ainda segundo a Fonologia de Uso, há dois tipos de frequência: a frequência de ocorrência (*token*), que se refere à ocorrência de determinado item lexical em um corpus, e a frequência de tipo (*type*), que diz respeito à frequência de uma estrutura no léxico, ou seja, indica o quanto determinado padrão é usado na língua (BYBEE, 2001, p. 10). Neste trabalho, toma-se a frequência de tipo como uma forma de medir a frequência dos paradigmas.

Entende-se que, em relação à análise do AL aqui proposta, o exame da frequência de tipo mostra-se importante para explicar o comportamento de determinados paradigmas nos quais o alçamento se concentra, pois pressupõe-se que a recorrência de alguns itens lexicais alçados gera a automatização de um padrão que se propaga gradualmente a outros itens lexicais que, por similaridade estrutural e semântica, possam estar conectados à mesma rede. Segundo Bybee (2001, p. 12, tradução nossa), “parece que a produtividade de um padrão, expressa em um esquema, é amplamente, mas não inteiramente, determinada pela sua frequência de tipo: quanto mais itens estão contidos em um esquema, mais forte o esquema fica e mais disponível se torna para a aplicação em novos itens”¹.

Por conseguinte, todas as palavras com estruturas sonora e semântica semelhantes são suscetíveis de serem afetadas pelo processo: a análise de formas recorrentes como *começar* e *cumecei*, por exemplo, permite a apreensão do padrão *cumeç-* que, por sua vez, confere embasamento para a produção de outras formas, como *cumeçou* e *cumeçando*.

Para que isso seja possível, a teorização anteriormente defendida pressupõe que haja detalhamento fonético associado às palavras no léxico. Assim sendo, faz-se necessário adotar um modelo que assumo o armazenamento desse detalhe na representação cognitiva das palavras, reivindicação formalmente atendida pela Teoria de Exemplares (JOHNSON, 1997; PIERREHUMBERT, 2001, 2002, 2012), modelo que permite multirrepresentações (BYBEE, 2001; CRISTÓFARO-SILVA; GOMES, 2004) e que, por sua flexibilidade, admite diferentes níveis de representação para as informações que o falante armazena nas suas experiências, incluindo nessas representações a palavra individual associada a seus detalhes de pronúncia.

¹No original: “it appears that the productivity of a pattern, expressed in a schema, is largely, though not entirely, determined by its type frequency: the more items encompassed by a schema, the stronger it is, and the more available it is for application to new items”.

Por esse entendimento, as variantes fonéticas de uma palavra são armazenadas na memória do falante e organizadas em grupos com base na similaridade, ou seja, instâncias (ou exemplares) semelhantes agrupam-se próximas umas das outras, enquanto instâncias divergentes ficam separadas. Conforme Pierrehumbert (2001, p. 140), há diferentes níveis de ativação para os exemplares, de maneira que os mais recentes e de experiências mais frequentes apresentam níveis de ativação mais altos em relação aos exemplares menos frequentes ou pouco utilizados. É a frequência de uso, portanto, que determina o fortalecimento dos exemplares.

No que tange aos casos de variação fonético-fonológica, pressupõe-se que o falante possui, para um mesmo fonema, múltiplas possíveis realizações fonéticas e que sua escolha por uma variante em detrimento de outra depende do quanto aquela categoria se encontra fortalecida. Nesse caso, quando o falante entra em contato com novas ocorrências, essas são comparadas e classificadas de acordo com a similaridade com outras categorias já armazenadas em sua memória (PIERREHUMBERT, 2003, p. 131). Entende-se, por conseguinte, que há uma constante atualização dos sons e itens lexicais no mapa cognitivo do falante, e isso lhe permite fazer escolhas de acordo com as representações mais recentes e mais fortalecidas.

Diante disso, assume-se, em concordância com a Fonologia de Uso e a Teoria de Exemplares, que os processos de variação podem difundir-se gradualmente pelo léxico, atingindo primeiramente as palavras mais frequentes na língua e com estrutura sonora semelhante. Assim, em relação ao processo de AL aqui investigado, tem-se a hipótese de que a variação se propaga por entre os itens lexicais que compartilham o mesmo paradigma (**governo**, **governa**, **governar**) e que a gradualidade e a velocidade com que o processo de variação se propaga pelo léxico depende tanto da frequência dos itens lexicais, quanto da frequência dos paradigmas na língua.

Não se exclui, com isso, o papel que os fatores contextuais possam exercer para mais rapidamente desencadear o processo de alçamento. Isto é, embora possa não ser a força primária da propagação da mudança, entende-se que a presença de um contexto fonético favorável ao alçamento (o contexto seguinte palatal ou velar, por exemplo, como se verifica nas formas *senhor* e *pequeno*) aumenta a propensão para o processo se manifestar.

Metodologia

O corpus utilizado na discussão desenvolvida neste estudo é constituído pelas ocorrências das vogais /e/ e /o/ pretônicas em palavras que não apresentam uma vogal alta subsequente, ou seja, palavras que não são o alvo da harmonização vocálica. As palavras incluídas na análise foram selecionadas a partir das amostras investigadas por Bisol (1981), Klunck (2007), Cruz (2010) e Silva (2014) sobre o alçamento das vogais médias pretônicas. Todas as amostras integram pesquisas de cunho variacionista e seguem pressupostos metodológicos semelhantes.

Primeiramente, realizou-se o levantamento de todas as ocorrências de palavras que apresentavam vogal média pretônica, sem vogal alta subsequente, nos quatro estudos (cinco amostras) selecionados para a constituição do corpus. Em seguida, desconsideraram-se as palavras funcionais, os nomes próprios e os numerais, bem como as palavras que apresentavam contextos em que o alçamento tende a ser categórico: vogal-alvo no início de palavra seguida por /N/ ou /S/ (*enxada, escola*) e vogal-alvo em casos de prefixo *des-* (*descartar, desconectar*). Além dessas, excluíram-se as palavras suscetíveis à formação de ditongo ou hiato (*toalha, teatro*), assim como todas as palavras que apresentavam vogal alta fonológica em posição pretônica (*futebol*). Por fim, delimitou-se como escopo de análise deste trabalho apenas os itens lexicais que apresentaram cinco ou mais ocorrências em cada amostra consultada.

Concomitantemente ao trabalho de levantamento dos dados, procedeu-se à tabulação dos índices de alçamento e de preservação das vogais médias apresentados em cada uma das amostras analisadas. A Tabela 1, exposta a seguir, apresenta o resumo do levantamento de dados utilizados neste estudo. Inclui-se nessa tabela, além do total de ocorrências de cada amostra, o número de ocorrências que sofreram a aplicação do AL.

Bisol (1981) analisou dados de HV e de AL coletados em 1979 nas comunidades de Veranópolis, Livramento, Taquara e Porto Alegre/região metropolitana. Das amostras apresentadas pela autora, foram consideradas, para a análise neste estudo, apenas os casos de vogal média pretônica suscetíveis de sofrer o processo de AL nas variedades culta e popular faladas na região metropolitana de Porto Alegre. Nessa seleção foram obtidas 685 ocorrências totais para /e/ e 667 para /o/.

Tabela 1 – Levantamento das ocorrências das vogais /e/ e /o/ pretônicas em contexto de AL

| Vogal /e/ | | | |
|-----------------------|--------------------|------------------|--------------|
| AMOSTRA | PRESERVAÇÃO | APLICAÇÃO | TOTAL |
| Bisol (1981) | 594 | 91 | 685 |
| Klunck (2007) | 604 | 49 | 653 |
| Cruz (2010) 1988-1989 | 592 | 80 | 672 |
| Cruz (2010) 2007-2009 | 1.310 | 64 | 1.374 |
| Silva (2014) | 1.527 | 26 | 1.553 |
| TOTAL | 4.627 | 310 | 4.937 |
| Vogal /o/ | | | |
| AMOSTRA | PRESERVAÇÃO | APLICAÇÃO | TOTAL |
| Bisol (1981) | 544 | 123 | 667 |
| Klunck (2007) | 535 | 136 | 671 |
| Cruz (2010) 1988-1989 | 764 | 158 | 922 |
| Cruz (2010) 2007-2009 | 931 | 148 | 1.079 |
| Silva (2014) | 1.216 | 37 | 1.253 |
| TOTAL | 3.990 | 602 | 4.592 |

Fonte: elaborado pelas autoras

Klunck (2007), Cruz (2010) e Silva (2014), por sua vez, realizaram estudos exclusivamente sobre a manifestação do AL em Porto Alegre. Aplicados os critérios de exclusão de ocorrências definidos anteriormente, a amostra investigada por Klunck (2007), coletada em 1996, resultou em 653 ocorrências de /e/ e 671 ocorrências de /o/ em contexto pretônico. Cruz (2010) trabalhou com duas amostras coletadas em períodos distintos, a saber, em 1988-1989 e 2007-2009, das quais foram selecionadas 672 ocorrências de /e/ e 922 ocorrências de /o/ na amostra mais antiga, e 1.374 ocorrências de /e/ e 1.079 ocorrências de /o/ na amostra mais recente. Silva (2014), por fim, analisou uma amostra coletada em 2004 da qual foram selecionadas 1.553 ocorrências de /e/ e 1.253 ocorrências de /o/.

Em resumo, observa-se na Tabela 1 que os índices de aplicação do AL por item lexical são baixos, sendo o alçamento de /e/ menor que o alçamento de /o/. Após a computação das ocorrências por item lexical, procedeu-se ao levantamento dos paradigmas mais frequentes nas

amostras, registrando-se tanto o número de ocorrências totais quanto o número de ocorrências que sofreram alçamento sob o mesmo paradigma. Para essa seleção, estipulou-se, para todos os paradigmas, o mínimo de dez ocorrências em cada amostra, independentemente do número de itens lexicais associados. Seguindo a mesma estratégia utilizada na seleção dos itens lexicais, não foram considerados paradigmas com vogal alta em sílaba pretônica ou em sílaba subsequente (*futebol*, *perigo*), paradigmas com sequência *des-* (*despertar*, *destrói*) e paradigmas de numerais (*dezesesseis*, *dezoito*).

A consulta sobre a frequência de ocorrência dos itens lexicais e dos paradigmas selecionados nas cinco amostras investigadas foi realizada no Corpus Brasileiro (SARDINHA; MOREIRA FILHO; ALAMBERT, 2010) e no banco de dados do projeto Avaliação Sonora do Português Atual (ASPA) (CRISTÓFARO-SILVA, 2004; CRISTÓFARO-SILVA; ALMEIDA, 2005; ALMEIDA; CRISTÓFARO-SILVA, 2006). O Corpus Brasileiro (doravante CB) possui um total de 1.133.416.757 *tokens* e 871.117.178 *types*. O ASPA, por sua vez, possui um total de 220.509.454 *tokens* e 611.480 *types*. Optou-se por utilizar os dados do CB para a verificação do comportamento do item lexical, haja vista o grande número de *tokens* incluídos no referido banco, enquanto o ASPA foi utilizado na verificação da frequência de tipo, devido a sua versatilidade na extração dessa informação, a qual é necessária para a discussão aqui desenvolvida.

Bybee (2006) revela que não há parâmetros bem definidos para a mensuração da frequência das palavras em um corpus. A medida de frequência mais recorrentemente utilizada tem sido a frequência por milhão de palavras, pautada em Kucera e Francis (1967) a partir da investigação sobre as frequências lexicais do inglês americano no *The Brown Corpus* (1.014.000 ocorrências totais). Com base nesse estudo, uma ocorrência por milhão de palavras seria a frequência mais baixa possível, sendo que as palavras frequentes apresentariam, em média, 32,6 ocorrências por milhão e as palavras não frequentes apresentariam, em média, 5,7 ocorrências por milhão (CHAN; YBARRA; SCHWARTZ, 2006, p. 368).

No entanto, dois problemas devem ser considerados ao assumir essas medidas como parâmetros na mensuração da frequência de palavras. Primeiramente, conforme discutem Van Heuven *et al.* (2014, p. 1179), bancos de dados mais amplos revelam que muitas palavras ocorrem com uma frequência menor do que uma vez por milhão, a exemplo do banco

SUBTLEX-UK (201.712.237 ocorrências totais) de inglês britânico, em que 80% das palavras do corpus apresentam frequência menor do que uma ocorrência por milhão. Além disso, as normas para a delimitação do ponto de corte para palavras mais frequentes e menos frequentes em uma língua necessitam ser estabelecidas de acordo com o corpus da língua em estudo, ou seja, os mesmos parâmetros de um corpus da língua inglesa não podem ser transpostos para outras línguas, pois estas apresentam composições distintas daquela.

Diante disso, os parâmetros de frequências estabelecidos através do trabalho de Kucera e Francis (1967) representam dois problemas metodológicos ao estudo aqui proposto: inicialmente, o parâmetro de corte entre frequente e não-frequente não é fixo, haja vista que o tamanho do banco interfere nesse valor; em segundo lugar, o parâmetro baseia-se exclusivamente em dados do inglês americano.

Com o intuito de resolver problemas relacionados à mensuração da frequência lexical em bancos de dados, Van Heuven *et al.* (2014) recentemente propuseram a utilização da escala Zipf: a aplicação do logaritmo $\text{LOG}_{10}(fpm)+3$, em que *fpm* significa frequência por milhão, indica a frequência das palavras em um corpus por meio de uma escala que vai de 1 a 6. As palavras que apresentam valores entre 1 e 3 são consideradas palavras pouco frequentes, enquanto as palavras com valores entre 4 e 6 são interpretadas como palavras frequentes.

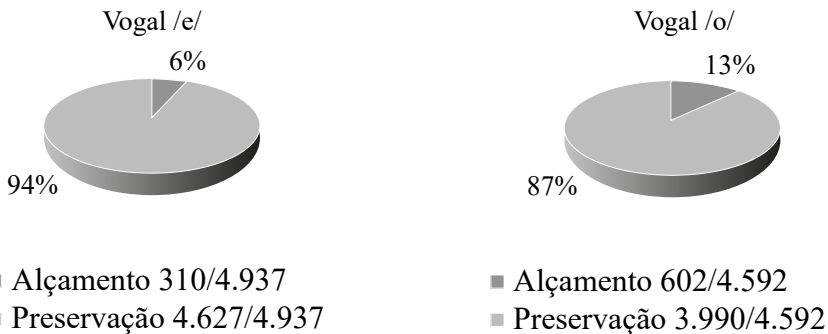
A escala proposta pelos autores atende às demandas metodológicas deste estudo, uma vez que permite realizar projeções a partir de uma língua específica e, também, porque define claramente os parâmetros a serem considerados como ponto de corte para estipular quão frequentes são as palavras. Na análise de resultados, discutida na próxima seção, a escala Zipf é apresentada para os itens lexicais selecionados nas amostras investigadas com o intuito de apontar se há um possível papel da frequência na manifestação de AL.

A atuação do léxico no processo de alçamento sem motivação aparente

O processo de AL tende a ser pouco recorrente na variedade falada na região metropolitana de Porto Alegre e, quando ocorre, atinge mais a vogal /o/ do que a vogal /e/. Esse comportamento pode ser observado nos totais de aplicação apresentados para cada uma das amostras (cf.

Metodologia) e corroborado pela taxa de aplicação do AL em relação à soma dos dados selecionados nas cinco amostras consideradas por esta pesquisa, conforme o Gráfico 1, a seguir.

Gráfico 1 – Taxas de aplicação do AL nas amostras investigadas



Fonte: elaborado pelas autoras

Observa-se no gráfico anterior que a vogal /e/ obteve um percentual de 94% de preservação e 6% de alçamento (N=4.937). A vogal /o/, por sua vez, apresentou 87% de preservação e 13% de alçamento (N=4.592). Confirma-se, portanto, a baixa aplicação do alçamento por AL e a tendência a maior probabilidade de alçamento para a vogal /o/. Ainda que as duas vogais tenham apresentado comportamentos diferenciados no que tange à taxa de aplicação do processo de alçamento, ambas parecem compartilhar o mesmo mecanismo de implementação, em que, ao que parece, o léxico representa ser atuante. Essa observação é fortalecida tanto pela forma como o processo se manifesta, a saber, afetando apenas determinados itens lexicais, bem como pela conformidade encontrada nas cinco amostras investigadas quanto às palavras que tendem a sofrer o alçamento. Além disso, o ambiente fonético favorável, quando presente, não parece ser a força que viabiliza o alçamento das vogais médias pretônicas.

A forma como o AL tem se manifestado na variedade em exame, atingindo itens lexicais específicos, indica que o léxico é o principal motivador do processo, tal como sugerido por Klunck (2007) e Cruz (2010). Além disso, segundo os autores, há o entendimento de que o alçamento é verificado mais frequentemente em paradigmas do que em palavras isoladas. A diferença, entretanto, está no tipo de paradigma que

favorece a propagação do alçamento, que para /o/ é verbal, atingindo em maior extensão formas como *c[u]meçar*, *c[u]meço*, *c[u]mecei*, *c[u]meçando*, e para /e/ é nominal, afetando formas como *s[i]nhor*, *s[i]nhora*, *s[i]nhores*, *s[i]nhoras* (Anexo I).

Por fim, embora indique ser uma força favorecedora ao alçamento, o contexto adjacente, por si só, não parece dar conta da forma como o processo de AL se propaga na língua. Diante disso, a hipótese que se propõe é de que o alçamento inicia em algumas palavras isoladas e gradualmente se espalha para outras palavras assemelhadas. O fator que determina quais itens lexicais são primeiramente atingidos pelo AL e a forma como o processo se propaga pode estar relacionado tanto com a frequência de palavra quanto com a frequência de paradigmas.

Essa hipótese foi testada a partir das análises dos itens lexicais mais frequentes nas amostras investigadas (Anexos I e II²). A partir desse levantamento e da verificação da frequência dessas palavras no CB, com a subsequente conversão dos valores pela escala Zipf, foi possível observar que praticamente todas as palavras selecionadas nas amostras são frequentes no corpus e, por tese, na língua. Assim sendo, a hipótese de que o alçamento se propagaria majoritariamente pelas palavras frequentes na língua, de certa forma, torna-se sem efeito. A observação de que itens lexicais de alta frequência na língua, tais como *problema*, *momento*, *pessoas*, *semana*, *gostava*, entre outros, não apresentam ocorrências alçadas em nenhuma das cinco amostras investigadas, reforça o papel pouco (ou nada) significativo do fator frequência para a aplicação do AL. Por outro lado, formas pouco frequentes, a exemplo de *boneca* (Zipf 3,38; 5 ocorrências alçadas em 8 ocorrências totais), *bolacha* (Zipf 2,6; 1 alçamento em 7) e *costela* (Zipf 3,11; 8 alçamentos em 9), apresentam ocorrências de alçamento.

Em outras palavras, embora o AL se manifeste em itens lexicais específicos e apresente um padrão uniforme quanto à aplicação nas amostras investigadas, no sentido de que as mesmas palavras sofrem alçamento nas diferentes amostras, não parece ser a frequência o fator que crucialmente determina a propagação do processo. Diante do papel pouco significativo da frequência lexical, optou-se por realizar um levantamento dos paradigmas e dos itens lexicais conectados aos

²Reproduziram-se nos Anexos I e II todos os itens lexicais que apresentaram 50 ocorrências ou mais no somatório das amostras.

paradigmas que sofreram alçamento nas amostras examinadas. A Tabela 2 a seguir apresenta os itens lexicais associados a cada paradigma (*types*), a recorrência com que esses itens se manifestaram nas amostras (*tokens*), a frequência no corpus (de acordo com a escala Zipf) e os totais de aplicação e preservação.

Tabela 2 – Paradigmas alçados – vogal /o/

(Continua)

| PARADIGMA | ITENS ASSOCIADOS (<i>types</i>) | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (n° de <i>tokens</i>) | PRE-SERVAÇÃO | APLICAÇÃO |
|----------------|-----------------------------------|--------------------------------|--|--------------|-----------|
| acontec | acontece | 4,75 | 84 | 65 | 19 |
| | acontecendo | 4,20 | 23 | 21 | 2 |
| | acontecer | 4,43 | 14 | 6 | 8 |
| | aconteceu | 4,63 | 46 | 26 | 20 |
| convers | conversa | 4,36 | 53 | 16 | 37 |
| | conversando | 3,64 | 55 | 29 | 26 |
| | conversar | 4,13 | 92 | 51 | 41 |
| conhec | conhece | 4,30 | 55 | 18 | 37 |
| | conhecem | 3,90 | 7 | 2 | 5 |
| | conhecer | 4,69 | 88 | 61 | 27 |
| | conheço | 3,75 | 62 | 27 | 35 |
| começ | começa | 4,75 | 43 | 17 | 26 |
| | começando | 3,90 | 10 | 6 | 4 |
| | começar | 4,49 | 70 | 43 | 27 |
| | comecei | 3,82 | 133 | 79 | 54 |
| | começo | 4,42 | 30 | 25 | 5 |
| | começou | 4,85 | 52 | 24 | 28 |

(Conclusão)

| | | | | | |
|-----------------|------------|------|-----|-----|----|
| com | comer | 4,18 | 91 | 27 | 64 |
| | comendo | 3,34 | 10 | 2 | 8 |
| comerç | comércio | 4,78 | 52 | 34 | 18 |
| pod | poder | 5,40 | 18 | 17 | 1 |
| | podendo | 4,67 | 1 | 0 | 1 |
| | poderem | 3,63 | 1 | 0 | 1 |
| conseg | consegue | 4,53 | 58 | 55 | 3 |
| fog | fogão | 3,34 | 10 | 7 | 3 |
| | fogueira | 3,32 | 8 | 2 | 6 |
| acompanh | acompanhar | 4,20 | 6 | 5 | 1 |
| coleg | colegas | 4,60 | 75 | 74 | 1 |
| | colega | 4,18 | 54 | 54 | 0 |
| | colégio | 4,09 | 272 | 261 | 11 |
| bot | botar | 3,49 | 21 | 20 | 1 |
| | botava | 2,84 | 12 | 11 | 1 |
| bonec | boneca | 3,38 | 8 | 3 | 5 |
| govern | governador | 4,75 | 5 | 0 | 5 |
| | governo | 5,62 | 61 | 19 | 42 |

Fonte: elaborado pelas autoras

A observação dos dados apresentados na Tabela 2 permite depreender que o açamento da vogal /o/ se concentra, conforme hipótese prevista, em determinados itens lexicais, associados, em sua maioria, a paradigmas verbais, a exemplo do que acontece em formas como *c[u]nheçe*, *c[u]nhecem*, *c[u]nhecem* e *c[u]nheço*. Deve-se atentar, todavia, para o fato de que não há um padrão uniforme na propagação do açamento entre os itens lexicais vinculados ao paradigma.

Sendo assim, de acordo com as considerações pontuadas, o AL, no que tange à vogal /o/, propaga-se gradualmente por itens lexicais, preferencialmente entre aqueles itens associados a paradigmas verbais. Verifica-se, dessa forma, a relevância do efeito morfológico para a propagação do AL. A propagação, entretanto, não acontece de maneira

uniforme, uma vez que se manifesta com maior ou menor propensão a depender do item lexical.

O mesmo levantamento de ocorrências de alçamento em paradigmas realizado para a vogal /o/ foi também realizado para a vogal /e/. Entretanto, nesse caso as ocorrências de vogal alçada são em menor número e concentram-se em poucos itens lexicais, associados, majoritariamente, a paradigmas nominais. O resumo das informações correspondentes aos paradigmas, itens lexicais associados (*types*), frequência no corpus, bem como o número de produções nas amostras (*tokens*) podem ser visualizados na Tabela 3, apresentada a seguir.

Tabela 3 – Paradigmas alçados – vogal /e/

| PARADIGMA | ITENS ASSOCIADOS (types) | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRESEÇÃO | APLICAÇÃO |
|------------------|--------------------------|--------------------------------|--|----------|-----------|
| melhor | melhor | 5,45 | 121 | 114 | 7 |
| | melhores | 4,85 | 12 | 12 | 0 |
| | melhorando | 3,74 | 6 | 6 | 0 |
| pequen | pequena | 4,8 | 74 | 1 | 73 |
| | pequenas | 4,74 | 7 | 0 | 7 |
| | pequeno | 4,79 | 122 | 12 | 110 |
| seman | semana | 5,18 | 127 | 125 | 2 |
| | semanas | 4,71 | 5 | 5 | 0 |
| pedaç | pedaço | 3,09 | 18 | 17 | 1 |
| professor | professor | 5,20 | 155 | 154 | 1 |
| | professores | 5,26 | 120 | 120 | 0 |
| | professora | 4,65 | 142 | 142 | 0 |
| | professoras | 4,07 | 6 | 6 | 0 |
| senhor | senhor | 4,5 | 14 | 1 | 13 |
| | senhora | 4 | 73 | 7 | 66 |

Fonte: elaborado pelas autoras

A manifestação do AL em palavras com vogal /e/ pretônica, de acordo com a Tabela 3, concentra-se em poucos itens lexicais, escolhendo mais precisamente as formas associadas aos paradigmas nominais *senhor-* e *pequen-*. Outros paradigmas, a exemplo de *melhor-*, *professor-*

e *seman-*, mesmo sendo frequentes na língua, não possuem a mesma força, apresentando baixa aplicação do AL: as formas *melhorar*, *melhores* e *melhorando*, por exemplo, estão associadas ao paradigma *melhor-*, porém não mostram nenhuma ocorrência de alçamento nas amostras investigadas. Reitera-se, assim, a falta de um efeito de frequência diretamente observável. Em outras palavras, de forma semelhante ao observado para a vogal /o/, não há, portanto, como depreender o papel da frequência na ocorrência do alçamento de /e/.

Um possível papel da frequência poderia ser depreendido no sentido de serem as formas mais frequentes na língua, quando associadas a paradigmas, as mais suscetíveis a serem atingidas pelo AL. No entanto, essa frequência, ao que se tem por hipótese, está mais relacionada aos paradigmas, ou melhor, à densidade da rede formada por itens lexicais em torno de um paradigma, do que a itens lexicais isolados.

Ante o exposto, os achados sugerem que a concentração do alçamento em determinados paradigmas – nominais para /e/ e verbais para /o/ – está relacionada com a frequência desses paradigmas na língua e, conseqüentemente, o maior reforço e densidade da rede. Paradigmas como *convers*, *conhec*, *começ* são frequentes na língua e, por essa razão, são potencialmente suscetíveis a primeiro sofrer o processo de variação. É essa força que determina, portanto, a difusão do alçamento, posteriormente reforçada pelo ambiente fonético condicionador. Isso explicaria o fato de que nem todas as palavras possíveis de sofrer a regra são, de fato, atingidas.

No caso de /o/, a propagação do AL tende a ser ainda maior em decorrência de, nesse caso, haver maior incidência de paradigmas verbais que formam densas redes com suas formas derivadas. Uma vez presente o alçamento no paradigma, como em *c[u]nhece*, *c[u]nvers*, por exemplo, há uma projeção do padrão alçado para as formas derivadas *c[u]nheço*, *c[u]nhece*, *c[u]nhecia*; *c[u]nverso*, *c[u]nversar* e *c[u]nversa*.

O fato de que o alçamento se espalha por outros itens lexicais associados a um paradigma que, em muitos casos, é detentor de ambiente fonético favorável (consoante dorsal precedente, consoante nasal seguinte, consoante dorsal seguinte entre outros) conduz ao entendimento de que o ambiente fonético atua secundariamente no processo de implementação do AL. Em outras palavras, o contexto não constitui o fator principal que determina a propagação do AL.

Diante da hipótese de que a propagação do AL está relacionada com a frequência dos paradigmas na língua, sendo a densidade das redes

paradigmáticas que determina as formas a serem primeiramente atingidas pelo processo, realizou-se um levantamento dos paradigmas mais recorrentes nas amostras em exame. A partir dessa relação, verificou-se a frequência desses paradigmas no ASPA, o qual permitiu fazer a busca específica por paradigma. A Tabela 4, apresentada a seguir, mostra os dados de frequência no corpus³ e nas amostras, bem como computa as formas alçadas e preservadas em cada paradigma.

Tabela 4 – Paradigmas mais frequentes

| PARADIGMAS | Nº ITENS LEXICAIS ASSOCIADOS (nº de types) | 1 FREQUÊNCIA NO CORPUS (nº tokens) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRE-SERVAÇÃO | APLICAÇÃO |
|------------------|--|------------------------------------|--|--------------|-----------|
| Vogal /e/ | | | | | |
| melhor- | 3 | 180.174 | 139 | 132 | 7 |
| senhor- | 2 | 28.096 | 87 | 8 | 79 |
| pequen- | 3 | 70.472 | 203 | 13 | 190 |
| seman- | 2 | 167.613 | 132 | 130 | 2 |
| pessoa- | 2 | 226.778 | 764 | 764 | 0 |
| Vogal /o/ | | | | | |
| govern- | 2 | 471.410 | 66 | 19 | 47 |
| conhec- | 4 | 57.185 | 212 | 108 | 104 |
| acontec- | 4 | 96.327 | 167 | 118 | 49 |
| comec- | 6 | 176.021 | 338 | 194 | 144 |
| com- | 3 | 9.930 | 153 | 63 | 90 |
| coleg- | 3 | 27.258 | 347 | 335 | 12 |
| convers- | 3 | 3.345.274 | 200 | 96 | 104 |

Fonte: elaborado pelas autoras

No que concerne à vogal média /e/, os mais altos totais de alçamento concentram-se nos paradigmas *senhor-* e *pequen-*, com aplicação de 79 (em 87 casos) e 190 (em 203 casos), respectivamente. Esse resultado não se aplica para os paradigmas *seman-*, *melhor-* e *pessoa-*, os quais apresentam baixa aplicação do alçamento: *seman-* 2 alçamentos (em 132 casos), *melhor-* 7 alçamentos (em 139 casos) e

³Valores extraídos do banco de dados do Projeto ASPA.

peessoa- nenhuma aplicação (em 764 casos). Essa diferença de ocorrência do alçamento verificada nos paradigmas *senhor-* e *pequen-* em relação aos demais poderia ser explicada pelo fato de o contexto fonético presente nos paradigmas *senhor-* e *pequen-* possuir mais força para alçar a vogal em relação aos paradigmas *peessoa-* e *seman-*. A partir desses dados, poder-se-ia inferir, então, que os paradigmas *senhor-* e *pequen-*, por apresentarem ambiente fonético propício ao alçamento, tendem a projetar aos seus itens lexicais associados a vogal alçada. Isto é, para esses casos, é mais provável o falante associar a ocorrência de vogal alta no paradigma.

Esse mesmo processo de análise não pode, todavia, ser atribuído ao paradigma *melhor*, o qual, por possuir uma consoante palatal seguinte, por sua articulação alta, também seria contexto favorável à realização de vogal alta. No caso de *melhor*, nem mesmo o paradigma e o contexto adjacente favorável representam forças determinantes para a manifestação do AL, verificado em apenas oito ocorrências, todas referentes ao dado *melhor*. Além disso, observa-se o comportamento diferenciado de alguns itens lexicais independentemente do comportamento do grupo (*melhor* não reproduz o comportamento de *melhores* e *melhorando* de acordo com os dados levantados).

Uma possível explicação sobre a aplicação do AL e a forma como o alçamento se propaga dentro de um mesmo paradigma não atingindo de maneira uniforme todos os itens conectados provém da frequência e do quanto cada estrutura se encontra fortalecida na rede. Estruturas semelhantes, segundo Bybee (2001, 2010), tendem a ser conectadas em uma única rede e tornam-se fortalecidas pelo uso. Contudo, mesmo que conectadas à mesma rede, nem todas as formas derivadas são necessariamente suscetíveis de alçamento. Assim, pode-se encontrar, por exemplo, a produção *m[i]lhor* e não, necessariamente, a produção **m[i]lhorei*.

Nesse caso, em consonância com Phillips (1984), pode-se sugerir que a diferença de comportamento de *melhor-* em relação aos paradigmas *senhor-* e *pequen-* pode estar relacionada à função assumida por essas formas. Vocábulos como *senhor*, *senhora* e *pequeno* assumem funções de adjetivo, substantivo e pronome de tratamento, ou seja, são padrões recorrentes na língua. O paradigma *melhor-*, por sua vez, alterna suas funções entre adjetivo, verbo e advérbio. Quando *melhor-* assume a função de advérbio, por exemplo, é uma palavra funcional invariável, isto

é, não possui paradigma e, com isso, está menos sujeita à variação. Nas amostras analisadas, esse paradigma se estende a palavras como *melhor*, *melhorar*, *melhoraremos*, *melhoramento*, *melhorou*. Em formas verbais, a exemplo de *melhorar* e *melhoraremos*, recorrentes nas amostras, a vogal é raramente produzida com o alçamento, o que explica o baixo índice de aplicação do alçamento desse paradigma.

Com relação aos paradigmas correspondentes à vogal /o/, observa-se, de forma semelhante à vogal /e/, que a ocorrência de vogal alta depende do tipo de paradigma e de suas derivações. A propagação, nesse caso centrada, em maior parte, nos paradigmas verbais *começ-*, *conheç-* *acontec-* *convers*, tende a apresentar alta aplicação dentro desses paradigmas, mas distribui-se de maneira não uniforme. A exemplo, no paradigma *começ-* é muito mais recorrente a ocorrência da forma *c[u]meçar* (27 aplicações em 70 ocorrências) em relação às formas *c[u]meço* (5 aplicações em 30 ocorrências) e *c[u]meçando* (4 aplicações em 10 ocorrências) (Tabela 3). Esses dados reforçam a sugestão de que a propagação do AL é difusa e não chega a atingir na mesma proporção todos os itens lexicais associados ao paradigma.

Essa propagação gradual e difusa, a qual a análise por item lexical muitas vezes ignora, pode ser reconhecida através da análise por paradigmas. São essas estruturas da língua que possibilitam uma visão mais transparente do processo e que captam melhor o mecanismo de propagação da variação. Diante desse fato, assume-se que, no caso do AL, o uso mais recorrente da variante com vogal alta está associado a determinados paradigmas, conforme já sugerido por Cruz (2010), e que sua propagação é potencializada por um ambiente fonético favorável e pela frequência que assumem essas estruturas na língua (BYBEE, 2001, 2010). Há, nesse sentido, uma atuação de forças conjuntas, em que o alçamento se propaga de diferentes formas constituindo redes, com estruturas mais ou menos densas dos paradigmas e suas formas derivadas.

Assim, na medida em que as formas *pequen-* e *govern-*, por exemplo, são frequentemente produzidas com a vogal alçada, o mesmo padrão tende a ser aplicado a outras estruturas que envolvem o mesmo paradigma, embora, conforme referido anteriormente, isso não aconteça na mesma proporção para todas as formas envolvidas. Pressupõe-se, assim, conforme Bybee (2002, p. 271), que o falante cria, a partir de um determinado paradigma, uma rede de associações entre palavras similares, tal como exemplificado na figura 1:

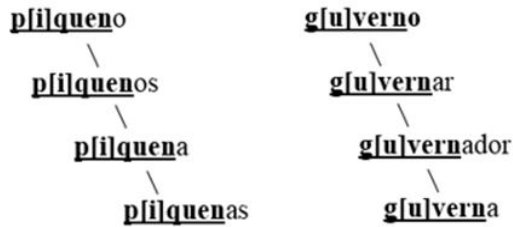


Figura 1 – Rede de associações dos paradigmas *pequen-* e *govern-*
 Fonte: elaborado pelas autoras com base em Bybee (2001) e Huback (2007)

Os exemplos anteriores revelam que a semelhança fonética e semântica entre as formas permite que o falante associe todas as palavras que possuem o mesmo paradigma a uma mesma rede (BYBEE, 2010). Pressupõe-se que, na medida em que esses paradigmas são frequentemente produzidos com a vogal pretônica alçada, a contínua exposição a esse alçamento afeta o mapa de exemplares do falante, que passa a ter a representação de vogal alta para esse contexto.

O fato de que determinada estrutura é replicada paradigmaticamente em diversos itens lexicais, conectados em termos de similaridade fonética e semântica, na sua forma alçada ou não-alçada, sustenta a ideia de que o falante pode possuir diferentes representações fonéticas de uma mesma forma fonológica e que a sua escolha, por uma ou outra forma, depende do fortalecimento dessa forma através do uso.

Assim, pressupõe-se um nivelamento ou padronização que só ocorre a partir da existência de uma forma ou representação básica que possa ser utilizada como parâmetro de comparação. Nesse caso, ao produzir um dado como *governador*, o falante automaticamente busca, por semelhança fonética e semântica, realizar a associação dessa forma com os itens já armazenados, tais como *governar* e *governo*. Se essas palavras com as quais o vocábulo é comparado se encontram representadas na forma elevada, no caso *guverno* e *guvernar*, há grande probabilidade de ocorrência do alçamento também em *guvernador*.

Com isso, explica-se a tendência de processos, tais como o AL, propagarem-se gradualmente pelo léxico, atingindo com maior recorrência os itens conectados a determinados paradigmas. Nesse sentido, não há como prever em que grau a mudança vai se propagar pelo léxico e quanto novas formas podem ou não ser atingidas, já que isso também depende de fatores outros, tais como o próprio indivíduo.

A representação pode não ser a mesma para todos os indivíduos, uma vez que alguns optam pela forma canônica (*governo*), enquanto outros optam pela forma inovadora (*guverno*). Dessa forma, a propagação da mudança não estaria somente associada à estrutura da língua, mas também a fatores individuais e socioculturais. Não é possível, portanto, estabelecer uma representação cognitiva única da vogal média pretônica, isto é, se média ou alta, para todos os falantes, mesmo em ambientes de condicionamento fonético potencial.

Dessa forma, é confirmada a hipótese de que o alçamento de vogais médias pretônicas sem motivação aparente se manifesta, majoritariamente, em itens lexicais que possuem relações paradigmáticas mais densas com outros exemplares. Sugere-se que um ambiente fonético específico, embora não seja o responsável por determinar a forma como a variação se propaga, potencialize o alçamento. Explica-se, portanto, a razão pelas quais formas como *governo* e *senhor* são mais suscetíveis ao alçamento do que formas isoladas, ou seja, formas que não constituem uma rede paradigmática, como *cebola*, por exemplo.

Por fim, uma vez que a mudança se difunde potencialmente através de uma determinada estrutura, os demais vocábulos conectados à rede em termos de semelhança tendem a se nivelar, gradualmente, a esse padrão, justificando dados com altos índices de alçamento, porém pouco frequentes, tais como *comércio* e *sotaque*.

Considerações finais

Este estudo corresponde a uma tentativa de depreender o papel do léxico no processo de AL na variedade falada na região metropolitana de Porto Alegre. Para tanto, fez-se a leitura dos dados analisados nos estudos de Bisol (1981), Klunck (2007), Cruz (2010) e Silva (2014), a partir do levantamento das palavras e paradigmas mais frequentes nessas amostras. O objetivo foi buscar possíveis generalizações e padrões linguísticos capazes de explicar, sob a perspectiva dos modelos multirepresentacionais (BYBEE, 2001; CRISTÓFARO-SILVA; GOMES, 2004), a maior incidência do AL em determinados itens lexicais e paradigmas.

Considerando a variedade analisada, a discussão, sustentada por preceitos teóricos da Fonologia de Uso (BYBEE, 2001) e da Teoria de Exemplares (JOHNSON, 1997; PIERREHUMBERT, 2001, 2002), permite inferir que o processo do AL se propaga pelo léxico por meio

de paradigmas, responsáveis por estabelecer as relações entre os itens lexicais. Uma vez ativada, a propagação do AL depende de quão densa encontra-se a rede formada pelos itens lexicais associados ao mesmo paradigma: quanto mais densa a rede, maior tende a ser a aplicação e a propagação do alçamento. A motivação parece ser, portanto, de natureza morfofonológica, uma vez que se propaga, gradualmente, por entre palavras que apresentam base morfológica comum.

Agradecimento

Ana Paula Correa da Silva Biasibetti agradece ao CNPq pela bolsa de doutorado.

Referências

- ALMEIDA, L. S.; CRISTÓFARO SILVA, T. Projeto ASPA: um mapeamento dos sons do português brasileiro. In: Congresso Internacional de Fonética e Fonologia, 3, 2006, Belo Horizonte. *Caderno de Resumos*. v. 1. Belo Horizonte: Faculdade de Letras da UFMG, 2006.
- SARDINHA, A. P. B.; MOREIRA FILHO, J. L.; ALAMBERT, E. A. *Corpus Brasileiro*. Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, 2010. Disponível em: <<http://www.sketchengine.co.uk/>>. Acesso em: 20 mar. 2015.
- BISOL, L. *Harmonia vocálica: uma regra variável*. 1981. 333f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.
- _____. O alçamento da pretônica sem motivação aparente. In: BISOL, L.; COLLISCHONN, G. (Orgs.). *Português no sul do Brasil: variação fonológica*. Porto Alegre: EdUPUCRS, 2009, p. 73-92.
- BYBEE, J. *Phonology and language use*. New York: Cambridge University, 2001.
- _____. Word frequency and context use in the lexical diffusion of phonetically conditioned sound change. *Language, Variation and Change*, v. 14, n. 3, p. 261-290, 2002.
- _____. From usage to grammar: the mind's response to repetition. *Language*, v. 82, n. 4, p. 711-733, 2006.

_____. *Language, usage and cognition*. New York: Cambridge University, 2010.

CASAGRANDE, G. *Harmonização vocálica: análise variacionista em tempo real*. 2003. 112f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2003.

CHAN, E.; YBARRA, O.; SCHWARTZ, N. Reversing the affective congruency effect: the role of target word frequency of occurrence. *Journal of Experimental Social Psychology*, v. 42, n. 3, p. 365-372, 2006.

CHEN, M.; WANG, W. Sound change: actuation and implementation. *Language*, v. 51, n. 2, p. 255-282, 1975.

CRISTÓFARO-SILVA, T. ASPA (Avaliação Sonora do Português Atual): uma ferramenta de apoio aos estudos em fonética e fonologia. In: ENANPOLL, 19., 2004, Maceió. *Caderno de Resumos*. Maceió: ANPOLL, 2004.

CRISTÓFARO-SILVA, T.; ALMEIDA, L. S. ASPA: a formulação de um banco de dados de referência da estrutura sonora do português contemporâneo. In: Congresso da Sociedade Brasileira de Computação, 25, 2005, São Leopoldo. *Anais*. São Leopoldo: Sociedade Brasileira de Computação, v. 1, p. 2268-2277, 2005

CRISTÓFARO-SILVA, T.; GOMES, C. A. Representações múltiplas e organização do componente linguístico. *Fórum Linguístico*, Florianópolis, v. 4, n. 1, p. 147-177, jul. 2004.

CRUZ, M. C. *As vogais médias pretônicas em Porto Alegre/RS: um estudo sobre o alçamento sem motivação aparente*. 2010. 203f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2010.

HUBACK, A. P. da S. Efeitos de frequência nas representações mentais. 2007. 318 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

JOHNSON, K. Speech perception without speaker normalization: an exemplar model. In: JOHNSON, K.; MULLENNIX, J. *Talker variability in speech processing*. San Diego: Academic, 1997. p. 145-165.

KLUNCK, P. *Alçamento das vogais médias pretônicas sem motivação aparente*. 2007. 112f. Dissertação (Mestrado em Letras) – Faculdade de

Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2007.

KUCERA, H.; FRANCIS, W. *Computational analysis of presentday American English*. Providence, RI: Brown University, 1967.

LABOV, W. Resolving the neogrammarian controversy. *Language*, v. 57, n. 2, p. 267-308, 1981.

OLIVEIRA, M. A. de. The neogrammarian controversy revisited. *International Journal of the Sociology of Language*, Berlin, v. 89, n. 1, p. 93-105, 1991.

PAUL, H. *Princípios fundamentais da história da língua*. Lisboa: Gulbenkian, 1983.

PHILLIPS, B. Word frequency and the actuation of sound change. *Language*, v. 60, n. 2, p. 320-342, 1984.

PIERREHUMBERT, J. Exemplar dynamics: word frequency, lenition and contrast. In: BYBEE, J.; HOPPER, P. *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins, 2001. p. 137-157.

_____. Word specific phonetics. In: GUSSENHOVEN, C.; WARNER, N. (Ed.). *Laboratory Phonology 7*. The Hague: Mouton de Gruyter, 2002. p. 101-139.

_____. Phonetic diversity, statistical, learning, and acquisition of phonology. *Language and Speech*, n. 46, p. 115-154, 2003.

_____. The dynamic lexicon. In: COHN, A.; FOUGERON, C.; HUFFMAN, M. *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University, 2012. p. 173-183.

SCHWINDT, L. C. *Harmonia vocálica em dialetos do Sul do país: uma análise variacionista*. 1995. 76f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 1995.

_____. A regra variável de harmonização vocálica no RS. In: BISOL, L.; BRESCANCINI, C. (Orgs.). *Fonologia e variação: recortes do português brasileiro*. Porto Alegre: EdiPUCRS, 2002. p. 161-182.

SILVA, A. P. C. da. *Elevação sem motivação aparente das vogais médias pretônicas entre os jovens porto-alegrenses*. 2014. 172f. Dissertação (Mestrado em Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014.

VAN HEUVEN, W. J. B. *et al.* SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, v. 67, n. 6, p. 1176-1190, 2014.

WANG, W. Competing changes as a cause of residue. *Language*, v. 45, n. 1, p. 9-24, 1969.

ANEXOS

Anexo I – Itens lexicais mais frequentes nas amostras – vogal /o/
(Continua)

| ITEM LEXICAL | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRESE-VAÇÃO | APLICAÇÃO |
|--------------|--------------------------------|--|-------------|-----------|
| colégio | 4,09 | 272 | 261 | 11 |
| problema | 5,32 | 198 | 198 | 0 |
| professor | 5,20 | 186 | 186 | 0 |
| comecei | 3,82 | 133 | 79 | 54 |
| comprar | 4,53 | 125 | 122 | 3 |
| professores | 5,26 | 120 | 120 | 0 |
| professora | 4,65 | 114 | 114 | 0 |
| gostava | 3,79 | 99 | 99 | 0 |
| conversar | 4,13 | 92 | 51 | 41 |
| comer | 4,18 | 91 | 27 | 64 |
| concreto | 4,41 | 91 | 91 | 0 |
| novela | 4,23 | 91 | 91 | 0 |
| conhecer | 4,69 | 88 | 61 | 27 |
| acontece | 4,75 | 84 | 65 | 19 |
| vontade | 4,79 | 82 | 82 | 0 |
| morava | 3,79 | 78 | 78 | 0 |
| momento | 5,39 | 76 | 76 | 0 |
| colegas | 4,60 | 75 | 74 | 1 |
| começar | 4,49 | 70 | 43 | 27 |
| problemas | 5,36 | 63 | 0 | 63 |
| conheço | 3,75 | 62 | 27 | 35 |
| governo | 5,62 | 61 | 19 | 42 |
| consegue | 4,53 | 58 | 55 | 3 |
| conhece | 4,30 | 55 | 18 | 37 |

(Conclusão)

| ITEM LEXICAL | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRESE-VAÇÃO | APLICAÇÃO |
|--------------|--------------------------------|--|-------------|-----------|
| conversando | 3,64 | 55 | 29 | 26 |
| programa | 5,32 | 55 | 55 | 0 |
| colega | 4,18 | 54 | 54 | 0 |
| conversa | 4,36 | 53 | 16 | 37 |
| começou | 4,85 | 52 | 24 | 28 |
| comércio | 4,78 | 52 | 34 | 18 |
| gostei | 3,50 | 50 | 50 | 0 |
| tomate | 3,80 | 10 | 9 | 1 |
| bolacha | 2,60 | 7 | 6 | 1 |
| costela | 3,11 | 9 | 1 | 8 |
| TOTAL | | 2.835 | 2.362 | 473 |

Anexo II – Itens lexicais mais frequentes nas amostras – vogal /e/

(continua)

| ITEM LEXICAL | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRESE-VAÇÃO | APLICAÇÃO |
|--------------|--------------------------------|--|-------------|-----------|
| pessoas | 5,61 | 538 | 538 | 0 |
| pessoa | 5,12 | 226 | 226 | 0 |
| semana | 5,18 | 177 | 175 | 2 |
| melhor | 5,45 | 168 | 161 | 7 |
| professor | 5,20 | 124 | 123 | 1 |
| pequeno | 4,79 | 122 | 12 | 110 |
| comecei | 3,83 | 116 | 116 | 0 |
| legal | 4,63 | 115 | 115 | 0 |
| professores | 5,26 | 114 | 114 | 0 |
| verdade | 5,11 | 111 | 111 | 0 |
| senhora | 4 | 103 | 96 | 7 |

(Conclusão)

| ITEM LEXICAL | FREQUÊNCIA NO CORPUS (em Zipf) | FREQUÊNCIA NAS AMOSTRAS (nº de tokens) | PRESERVAÇÃO | APLICAÇÃO |
|--------------|--------------------------------|--|-------------|-----------|
| pequena | 4,80 | 74 | 73 | 1 |
| conhecer | 4,69 | 71 | 0 | 71 |
| conversando | 3,64 | 64 | 9 | 55 |
| começar | 4,49 | 61 | 0 | 61 |
| conversar | 4,13 | 61 | 61 | 0 |
| negócio | 4,60 | 61 | 61 | 0 |
| mercado | 5,43 | 55 | 55 | 0 |
| chegar | 4,85 | 54 | 54 | 0 |
| depende | 4,69 | 53 | 53 | 0 |
| professora | 4,65 | 53 | 53 | 0 |
| matemática | 4,36 | 52 | 52 | 0 |
| certeza | 4,65 | 50 | 50 | 0 |
| TOTAL | | 2.691 | 2.393 | 298 |