

A EPÊNTESE NA PRODUÇÃO DE PLOSIVAS EM CODAS MEDIAIS DO PORTUGUÊS BRASILEIRO POR COLOMBIANOS: UMA ABORDAGEM COM BASE EM RESTRIÇÕES

Carmen Lúcia MATZENAUER

Doutor. Universidade Católica de Pelotas/UCPEL
E-mail: carmenluc@terra.com.br

Ubiratã Kickhöfel ALVES

Doutor. Universidade Federal do Rio Grande do SUL/UFRGS
E-mail: ukalves@gmail.com

Roberta Quintanilha AZEVEDO

Doutoranda. Universidade Católica de Pelotas/UCPEL
E-mail: robertaquintanilhaa@gmail.com

Resumo

Este artigo tem por objetivo, à luz do algoritmo de aprendizagem vinculado ao modelo formal da Teoria da Otimidade Estocástica – OT (BOERSMA & HAYES, 2001), descrever e analisar a produção da epêntese vocálica após plosivas em codas mediais de palavras, por parte de aprendizes de Português Brasileiro, que são falantes nativos de Espanhol (colombianos). Para a constituição do *corpus*, foi proposto um instrumento composto de palavras cognatas do Português e do Espanhol para a leitura em frases veículo, gravadas no Programa AUDACITY 1.3.5 (ANSI). Foram escolhidos, para este trabalho, quatro estudantes colombianos que permaneceram no Brasil por um período médio de seis meses, em mobilidade acadêmica no Sul do Brasil, com o cuidado de serem homogeneizados fatores extralinguísticos, como sexo, idade e nível de escolaridade. Foi confirmada a

hipótese de que, mesmo em palavras cognatas nas duas línguas analisadas, há diferente tratamento às plosivas em coda, dependendo do sistema linguístico que está sendo utilizado pelo falante de Espanhol aprendiz de PB. Tal fato foi captado, à luz da OT Estocástica, via hierarquia de restrições universais.

Palavras-chave

Epêntese Vocálica; Restrições Universais; Teoria da Otimidade Estocástica

Introdução

Neste trabalho, procuramos estabelecer qual seria a tendência de um falante de Espanhol Colombiano adquirindo Português, diante de palavras cognatas nas duas línguas, em um contexto de plosivas em codas mediais de palavras. A hipótese é a de que os sujeitos aplicarão o processo de epêntese, assim como nós, brasileiros, por mostrar-se processo prevalente na gramática em aquisição, ainda que possa ser variável na língua-alvo, em tal contexto fonológico. Trataremos a epêntese como um processo fonológico caracterizado pela inserção de um som vocálico no *output* das palavras, som esse não presente no *input*. Tal fenômeno vai servir para tornar menos marcada a estrutura silábica de uma dada língua, como podemos perceber em palavras como *et[i]nia* e *ac[i]ne*, no Português Brasileiro.

Para verificar a hipótese suscitada de que os estudantes recorrerão à epêntese na produção das palavras selecionadas, foram gravadas produções de quatro sujeitos colombianos que estiveram na cidade de Pelotas (RS/Brasil) realizando intercâmbio por um período médio de seis meses. Eles foram submetidos a uma tarefa de leitura de sequências recorrentes que continham o contexto fonológico para o aparecimento do fenômeno da epêntese no Português Brasileiro. Após a verificação acústica, os dados gravados foram dispostos para análise através da Teoria da Otimidade Estocástica (BOERSMA & HAYES, 2001), visando a verificar se tal modelo teórico se mostra capaz de formalizar o fenômeno levantado.

A fim de atender ao nosso objetivo geral de, à luz da OT Estocástica, descrever e analisar a produção da epêntese vocálica num contexto de plosivas em codas mediais de palavras, por parte de aprendizes de Português Brasileiro, que são falantes nativos de Espanhol (colombianos), apresentamos as Questões Norteadoras que guiaram a análise:

- 1 – A epêntese vocálica que ocorre no PB, entre as sequências de segmentos /pt, pn, kt, kn, tn/, em codas medias de palavras, será produzida por aprendizes colombianos de PB, cuja L1 admite segmentos plosivos em coda?
- 2 – No que diz respeito à análise à luz da Teoria da Otimidade Estocástica, quais restrições serão relevantes para o processo de aquisição dos padrões silábicos do PB como L2, por parte de falantes do Espanhol Colombiano?
- 3 – O algoritmo consegue convergir em gramáticas que reflitam os dados de aquisição?

A seguir, abordamos os pressupostos teóricos que fornecem subsídios e suporte ao desenvolvimento do trabalho, bem como realizamos a análise desse fenômeno com base no modelo teórico da OT Estocástica.

1. Fundamentação Teórica

1.1 Pressupostos da OT Estocástica e do Algoritmo de Aprendizagem Gradual

A Teoria da Otimidade Estocástica (BOERSMA & HAYES, 2001) apresenta restrições distribuídas em uma escala numérica hierarquizada. Assim, na OT Estocástica, a hierarquia é estabelecida a partir desses índices numéricos que são atribuídos às restrições.

As restrições podem ser entendidas como exigências para que se alterem ou preservem estruturas que podem ou não ser violadas. De acordo com McCarthy (2002, p.13), as restrições são de dois tipos: (a) Marcação – que devem impor exigências na formação estrutural de *output*, ou seja, levam a *outputs* menos marcados. Exs.: As sílabas têm de ser abertas (No-Coda), oposição à consoante plosiva em coda (*Stop-Coda) e (b) Fidelidade – que devem garantir que as formas ótimas de *output* sejam as mais semelhantes possível ao *input* recebido, conspirando contra qualquer tipo de mudança. Ex.: oposição à adição de segmento no *output* (DEP), oposição ao apagamento de um segmento no *output* (MAX).

O *tableau* a seguir, que apresenta restrições hierarquizadas por relações de dominância, e os *outputs* possíveis a partir da forma subjacente, representam esse modelo:

Tableau 1 – Exemplo de *tableau* no Modelo da OT Estocástica

Output/ Restrições	(40) Restrição A	(30) Restrição B	(15) Restrição C
[output 1]	*!		*
☞ [output 2]		***	

No exemplo 1, o candidato “1” foi eliminado, pois violou a restrição que, em função dos valores numéricos a ela atribuídos por meio do algoritmo de

aprendizagem associado ao modelo (OT-GLA), se encontra mais altamente ranqueada, deixando para o candidato “2” o *status* de candidato ótimo, já que viola uma restrição mais baixa no *ranking* e obedece à restrição mais altamente ranqueada, mesmo que tenha em seu histórico três violações à Restrição B. Isso quer dizer, também, que a Restrição A domina as Restrições B e C (RA>>RB>>RC).

Na análise via OT Estocástica, podemos, ainda, com o mesmo sistema de análise, ou seja, com as mesmas restrições, demonstrar a trajetória percorrida na aquisição da língua materna ou estrangeira. A relação entre os estágios desenvolvimentais é, nesse modelo, representada pelo deslocamento contínuo das restrições. Tal movimentação é tarefa do algoritmo de aprendizagem vinculado ao modelo, que tem a responsabilidade de guiar o sistema linguístico do aprendiz, demovendo ou promovendo as restrições. O referido algoritmo, a ser utilizado na presente análise, serve para dar conta da variação apresentada pelos aprendizes, bem como da gradualidade apresentada no processo de aquisição da língua.

O algoritmo OT-GLA é sensível aos erros do aprendiz, sendo, portanto, do tipo *error-driven*. Dessa forma, vai alterar o valor numérico das restrições quando o *output* da sua gramática se mostrar diferente do padrão encontrado na linguagem ambiente (erro). Esse índice de alteração se dá em função de um valor de plasticidade¹ que é adotado pelo algoritmo. Assim, para a atribuição dos valores numéricos das restrições, o algoritmo vai diminuir (demover) o valor numérico das restrições que são violadas pelas formas encontradas de produção e vai aumentar (promover) o valor das restrições violadas pelo *output* “errado”, na hierarquia de restrições vigente.

Sob tal algoritmo, cada restrição apresentará, na verdade, dois valores numéricos: o valor que corresponde ao ponto central da faixa de valores - valor central ou valor de ranqueamento (*ranking value*) — e, o ponto de seleção, que é um valor assumido dentro da faixa de valores, que corresponde ao valor exibido, no momento de avaliação dos candidatos. É, ainda, através dos valores centrais das Restrições, que este algoritmo vai representar *outputs* variáveis ou categóricos. Valores centrais bem afastados, ou seja, valores com distância superior a 10 pontos entre os valores da Restrição DEP e as demais Restrições dadas, representam a ocorrência de um *output* categórico, pois diferentes momentos de produção linguística, não causarão cruzamento na faixa de valores destas Restrições. Esse algoritmo é disponibilizado no *software* PRAAT, e serve para expressar o processo de aquisição e as possíveis variações do *output*, conforme podemos verificar no exemplo a seguir.

$$\text{Ex.: } 45 \overset{49}{\underset{\boxed{\text{R1}}}{50}} 55 \qquad 50 \overset{52}{\underset{\boxed{\text{R2}}}{55}} 60$$

¹ Plasticidade é um valor numérico através do qual o algoritmo vai ajustar o *ranking* das restrições. Esse valor é fornecido durante a simulação computacional, através da função *Learn* do PRAAT.

Através do exemplo acima, evidenciamos duas Restrições com valores centrais diferentes, porém próximos. O valor de ruído produzido pelo algoritmo deve dar conta de tal fato, garantindo que, numa faixa estabelecida de 10 pontos (exemplos acima Restrição 1 = 45 a 55 e Restrição 2 = 50 a 60), o ponto de seleção varie, permitindo que ora R1 esteja acima, ora R2 esteja acima na hierarquia, o que vai garantir que tenhamos variação.

No que diz respeito ao processo de epêntese, que constitui o fenômeno fonológico com que estamos trabalhando, será necessário compreendermos, sobretudo, a estrutura da sílaba dos idiomas em foco (Português Brasileiro e Espanhol). No Português Brasileiro, os segmentos plosivos não são permitidos em coda de sílaba, diferentemente do que ocorre no Espanhol, que permite² essa sequência na estrutura de uma sílaba. Dessa forma, retomando os módulos da OT Estocástica, vamos ver que o que vai distinguir o sistema linguístico do Português Brasileiro para o Espanhol são as hierarquias das restrições, que vão caracterizar os sistemas, dando a cada um a sua especificidade.

2. A Análise dos Dados

2.1. Descrição dos Dados e Formalização das Restrições

Antes de iniciarmos a reflexão sobre as restrições, adiantamos que, para os grupos de sequências heterossilábicas de segmentos [p.n - apneia, k.n - acne, t.n - etnia, p.t - rapto, k.t - cacto] a serem analisados, trabalhamos com a hipótese de que o contato silábico está influenciando na epêntese pelos aprendizes de Português Brasileiro. Logo, a tabela dos dados abaixo, bem como a discussão das restrições que segue, refletem essa ideia.

Tabela 1 – Produção das sequências [p.n, k.n, t.n, p.t, k.t] no EC (L1)

Consoante em coda	Epêntese	Apagamento
Consoante em coda [k.n, p.n, t.n] 100%	Epêntese [ki.n, pi.n, ti.n] 0	Apagamento total [s.n] 0
Consoante em coda [k.t, p.t] 100%	Epêntese [ki.t, pi.t] 0	Apagamento total [s.t] 0

² Em função do Fenômeno de Afrouxamento de Condição de Coda (BISOL, 1999), codas mediais com plosivas podem vir a ser produzidas variavelmente no português brasileiro. Entretanto, conforme será visto na seção seguinte, para fins desta análise, uma vez que não contamos com um corpus em Português Brasileiro da região em que foi realizado o estudo, assumiremos, para fins teóricos, que o alvo da aquisição corresponde a 100% da epêntese.

Tabela 2 – Produção das sequências [p.n, k.n, t.n, p.t, k.t] no PB (L2)

Consoante em coda	Epêntese	Apagamento
Consoante em coda [k.n, p.n, t.n] 86,52%	Epêntese [ki.n, pi.n, ti.n] 13,48%	Apagamento total [s.n] 0
Consoante em coda [k.t, p.t] 98,96%	Epêntese [ki.t, pi.t] 1,04%	Apagamento total [s.t] 0

A Tabela 1, referente às produções dos nossos informantes de palavras na sua língua materna, serve para confirmar que, no Espanhol Colombiano, para uma situação de plosiva em posição de coda medial, não teremos epêntese. Já na Tabela 2, que se refere às produções por parte dos falantes nativos de Espanhol Colombiano em palavras no Português Brasileiro, os dados se apresentam diferentes, uma vez que há a produção, em caráter variável, de vogais epentéticas.

Conforme já explicitado, nossa análise nos levou a crer que contato silábico estaria influenciando nas produções com epêntese dos informantes, nas produções em Português. As sequências neste sentido, consideradas mais marcadas, ([k.n, p.n, t.n] — segmento plosivo em coda seguido de nasal) apresentaram, nas produções em Português, maior número de epêntese (13,48%), com relação às sequências consideradas menos marcadas ([k.t, p.t] – *plateau* de sonoridade, ou o encontro de duas consoantes plosivas – 1,04%). Assim, para a formalização das Restrições que comporão nossa análise, consideramos este fenômeno.

No que diz respeito às Restrições de Fidelidade, devemos estabelecer restrições que busquem a preservação do *input*. Para o caso ora posto, de plosiva em posição de coda medial de palavra, argumentamos que as restrições de fidelidade DEP e MAX (McCARTHY e PRINCE, 1995) conseguem dar conta dos padrões encontrados. No que se refere às restrições de Marcação, a primeira decisão tomada foi a da utilização do Mecanismo de Alinhamento Harmônico (HA) de Prince & Smolensky (1993), que associa elementos da escala de sonoridade a posições silábicas, favorecendo, portanto, elementos de mais alta sonoridade em coda, pois esse mecanismo propiciará um conjunto de restrições que pode mostrar quais segmentos se apresentam mais aptos a ocorrer em coda silábica a partir de uma escala linguística.

Quanto às restrições de Contato Silábico, que vão servir para dar conta da influência da sonoridade causada pelo *onset* da próxima sílaba na inserção de epêntese nos nossos dados, baseamo-nos no Mecanismo de Alinhamento Relacional de Gouskova (2004), que vai combinar a escala de sonoridade de coda com a de *onset*, determinando quais distâncias são mais marcadas entre os segmentos heterossilábicos que propomos para este estudo. Gouskova (2004) vai transformar em índices numéricos as diferenças de sonoridade entre coda e *onset* da sílaba seguinte, uma vez que, em termos de contato silábico,

é melhor termos uma redução do valor do segmento da coda para o *onset* da próxima sílaba. Basicamente, é um esquema de Distância de Sonoridade entre duas consoantes (*DISTANCE), onde *Dist (0) >> *Dist (-1) >> *Dist (-2), já que, nesse caso, quanto maior a queda de sonoridade, mais harmônica é a distância. Em outras palavras, é melhor termos um encontro [k.t] do que um [k.n], ou um encontro [p.t] a um [p.n]. Gouskova (2004) não sugere relações de estringência, mas um *ranking* fixo entre tais restrições. Em nossa análise, ao seguirmos Alves (2008), pensamos, para esta e as demais restrições utilizadas, numa organização que trabalhe esta relação em estringência, para que possamos demonstrar as relações de marcação existentes entre os membros de uma escala sem a estipulação extrínseca de um *ranking* fixo ao algoritmo, formalizando a ordem de aquisição de cada um desses membros.

Acrescemos ainda aos nossos candidatos a *output* as sequências [s.n] e [s.t], que, no PB, não são passíveis de epêntese, e procuramos diferenciá-las das sequências heterossilábicas [p.t, k.t] e [p.n, k.n, t.n], através da ação de restrições conjuntas e de restrições de proibição a plosivas em coda, que construímos para se oporem apenas às sequências com tais segmentos em final de sílaba. No que segue, demonstramos as restrições de Fidelidade e Marcação formalizadas para a análise.

RESTRIÇÕES DE FIDELIDADE

MAX e DEP

RESTRIÇÕES DE MARCAÇÃO

*Dist {+3}, *Dist {+3, +2}, *Dist {+3, +2, +1}, *Dist {+3, +2, +1, 0}, *Dist {+3, +2, +1, 0, -1}, *Dist {+3, +2, +1, 0, -1, -2}, *Dist {+3, +2, +1, 0, -1, -2, -3}, *{stop}_{coda}, *{stop, fric}_{coda}, *{stop, fric, nas}_{coda}, *{stop, fric, nas, liq}_{coda}, *{stop, fric, nas, liq, glide}_{coda}, *Dist {1, 0} & *{stop}_{coda} e *Dist {1} & *{stop}_{coda}.

2.2 A simulação computacional dos dados de aquisição

Ao considerarmos que o primeiro estágio de aquisição da L2 é o sistema da L1, as simulações precisam expressar, primeiramente, a aquisição da língua materna. Assim, foi necessário verificar se o algoritmo seria capaz de convergir na L1 dos aprendizes, que caracteriza um caso de produção categórica, de 100% de plosiva em coda (Espanhol Colombiano), bem como na língua estrangeira, que apresenta 100% de produção da epêntese vocálica (Português Brasileiro). A partir daí, pudemos verificar o estágio desenvolvimental dos aprendizes colombianos em direção ao alvo (aquisição do PB), com as produções da vogal epentética ocorrendo variavelmente. Por essa razão é que trabalhamos sempre com três diferentes simulações para cada algoritmo:

- Simulação 1 – Aquisição do sistema do Espanhol Colombiano (L1);
- Simulação 2 – Aquisição plena do sistema do Português Brasileiro (L2),

p.n/k.n/t.n				*	*	*	*	*	*	*	*	*	*	*	*
pi.n/ki.n/ti.n			*!												
n			*!												
n.t										*	*	*	*	*	*
ni.t			*!												
t			*!												
l.t												*	*	*	*
li.t			*!												
t			*!												
w.t														*	*
wi.t			*!												
t			*!												
s.t									*	*	*	*	*	*	*
si.t			*!												
t			*!												
s.n					*			*	*	*	*	*	*	*	*
si.n			*!												
n			*!												

Nesta primeira simulação (conjunto de *Tableaux* 1), o algoritmo aprendeu o Espanhol Colombiano. Sendo assim, é necessário que DEP seja promovida, assumindo uma posição na hierarquia que supere as restrições vinculadas às consoantes em coda, de modo a fazer com que o candidato com epêntese receba uma violação fatal, por violar uma restrição mais altamente ranqueada.

Juntamente com a promoção da restrição de fidelidade DEP, devemos ter as restrições MAX e *Dist {+3} e *Dist{+3, +2}, que se referem, respectivamente, a apagamento e a contatos silábicos inexistentes no EC (consoante nasal em posição de coda seguida de glide, plosiva em coda seguida de glide e plosiva em coda seguida de uma consoante líquida), acima das demais restrições de marcação, independentemente dos pesos que assumam, para que, neste modelo de dominância estrita, recebam posições mais altas no ordenamento das restrições e excluam candidatos com apagamento e epêntese da disputa pelo *output* ótimo.

Quanto às restrições de marcação, temos valores de ranqueamento que respeitam um ordenamento de marcação, que atribuem índices mais baixos às restrições do conjunto de estringência que fazem oposição a candidatos menos marcados e mais gerais. Conseguimos demonstrar, contudo, uma aquisição que obedece à ordem de marcação, onde [t.n] (Dist {+1}) é mais marcada do que [t.t] (Dist {0}), que, por sua vez, é mais marcada do que [n.t] (Dist {-1}) e [l.t] (Dist {-2}) e [w.t] (Dist {-3}), por estringência. As Restrições que fazem oposição aos candidatos menos marcados, dessa forma, devem ser mais violadas e, portanto, mais demovidas, assumindo posições mais baixas no *ranking*.

Assim, as restrições *Dist {+3,...-3} e *{stop, ...glide}_{coda}, que abrangem todas as proibições de distância entre os segmentos em coda e *onset* e todas

as proibições de modo de articulação em coda, são as restrições mais baixas na hierarquia, demovidas por estringência. Logo, o algoritmo consegue convergir para uma gramática que se apresenta categórica na simulação 1 da OT Estocástica, a partir da qual os segmentos consonantais em codas mediais são produzidos sem epêntese (estágio final da aquisição), seguindo uma lógica de marcação que é determinada por estringência.

2.2.2 Simulação 2 – Aquisição do Português Brasileiro (L2)

A aquisição plena do PB vai ter no estágio final de aquisição, conforme por nós definido, a produção de 100% de epêntese após as consoantes plosivas em posição de coda. Ao alimentarmos o sistema com os valores numéricos das restrições obtidos na simulação anterior (estágio final do Espanhol Colombiano e inicial para o Português Brasileiro), temos que o estágio inicial da L2 vai contar com as restrições de fidelidade (MAX e DEP) dominando as restrições de marcação, na hierarquia das restrições propostas para lidar com a aquisição de sequências com consoantes em coda medial de palavra. O estágio final é dado no conjunto de *Tableaux 2*.

Conjunto de *Tableaux 2* – Simulação 2 OT/Aquisição do Português Brasileiro (100% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
*Dist {+3, +2}	100.000	99.949	1.000000
*Dist {+3}	100.000	98.415	1.000000
Max	64.120	61.648	1.000000
*{stop}coda	56.644	55.124	1.000000
*Dist {1} & {stop}coda	55.997	54.909	1.000000
*Dist {1,0} & {stop}coda	56.644	54.779	1.000000
Dep	44.951	49.044	1.000000
*Dist {+3, ...+1}	32.053	35.545	1.000000
*Dist {+3, ... 0}	15.993	15.302	1.000000
*{stop, fric}coda	15.993	13.117	1.000000
*{stop, ... nas}coda	3.203	5.804	1.000000
*Dist {+3, ...-1}	3.203	2.758	1.000000
*{stop, ... liq}coda	-2.091	2.306	1.000000
*Dist {+3, ...-2}	-2.091	-2.040	1.000000
*Dist {+3, ...-3}	-9.071	-8.477	1.000000
*{stop, ... glide}coda	-9.071	-11.179	1.000000

(44.951), nesta simulação, só precisou encontrar um valor central que fosse inferior e mantivesse um distanciamento (distância de 10 pontos) das restrições específicas à plosiva em coda ($*\text{Dist}\{1\} \& *\{\text{stop}\}_{\text{coda}} = 55.997$, $\{\text{stop}\}_{\text{coda}} = 56.644$ e $*\text{Dist}\{1,0\} \& *\{\text{stop}\}_{\text{coda}} = 56.644$), ao mesmo tempo em que fosse superior e mantivesse um valor de distanciamento das demais restrições de marcação.

Assim, em se tratando da aquisição plena da língua estrangeira – Português Brasileiro - pelos aprendizes colombianos, o algoritmo conseguiu convergir, sob o modelo da OT Estocástica, em uma gramática com *outputs* categóricas exibindo epêntese.

2.2.3 Simulação 3 – Interlíngua (outputs variáveis)

Nesta simulação, devemos observar e analisar o tratamento do algoritmo de aprendizagem frente à gramática em desenvolvimento dos aprendizes, com *outputs* variáveis. Diferentemente das simulações 1 e 2, os valores centrais das restrições DEP e das restrições específicas às sequências com plosiva em coda, neste momento, devem ser muito próximos, para explicar a variação no *output*. As alterações nos valores dos pontos de seleção, diante de avaliações sucessivas, não representarão apenas uma mera variação numérica dos pesos das restrições, mas motivação para que tenhamos mudança no *ranking* das restrições, conforme podemos ver no conjunto de *tableaux* 3.

Conjunto de *Tableaux* 3 – Simulação 3 OT / Interlíngua (*outputs* variáveis)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
*Dist {+3}	100.000	100.659	1.000000
*Dist {+3, +2}	100.000	98.071	1.000000
Max	65.078	65.989	1.000000
Dep	50.650	49.310	1.000000
*{stop}coda	43.366	44.198	1.000000
*Dist {1} & {stop}coda	47.595	44.164	1.000000
*Dist {1,0} & {stop}coda	43.366	42.893	1.000000
*Dist {+3, ...+1}	30.272	32.248	1.000000
*{stop, fric}coda	9.336	10.670	1.000000
*Dist {+3, ... 0}	9.336	5.783	1.000000
*Dist {+3, ...-1}	-3.454	-1.435	1.000000
*{stop, ... nas}coda	-3.454	-3.391	1.000000
*Dist {+3, ...-2}	-8.748	-8.044	1.000000
*{stop, ... liq}coda	-8.748	-10.300	1.000000
*{stop, ... glide}coda	-15.728	-14.812	1.000000
*Dist {+3, ...-3}	-15.728	-17.790	1.000000

Conjunto de *Tableaux* 4 – Simulação 3 OT / Interlíngua – Segunda Rodada

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
*Dist {+3}	100.000	101.327	1.000000
*Dist {+3, +2}	100.000	101.133	1.000000
Max	65.078	63.423	1.000000
*Dist {1} & {stop}coda	47.595	51.028	1.000000
Dep	50.650	49.625	1.000000
*Dist {1,0} & {stop}coda	43.366	46.476	1.000000
*{stop}coda	43.366	42.035	1.000000
*Dist {+3, ...+1}	30.272	29.754	1.000000
*Dist {+3, ... 0}	9.336	10.949	1.000000
*{stop, fric}coda	9.336	7.922	1.000000
*{stop, ... nas}coda	-3.454	0.353	1.000000
*Dist {+3, ...-1}	-3.454	-4.720	1.000000
*{stop, ... liq}coda	-8.748	-8.050	1.000000
*Dist {+3, ...-2}	-8.748	-9.427	1.000000
*{stop, ... glide}coda	-15.728	-15.449	1.000000
*Dist {+3, ...-3}	-15.728	-17.979	1.000000

	*Dist{+3,...-3}	*{stop,...glide}coda	*Dist{+3,...-2}	*{stop,...lq}coda	*Dist{+3,...-1}	*{stop,...nas}coda	*{stop, fric}coda	*Dist{+3,...0}	*Dist{+3,...+1}	*{stop}coda	*Dist{1,0} & {stop}coda	DEP	*Dist{1} & {stop}coda	MAX	*Dist{+3, +2}	*Dist{+3}	
p.t/k.t	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
pi.t/ki.t											*!						
t														*!			
p.n/k.n/t.n	*	*	*	*	*	*	*	*	*	*	*	*!	*	*	*	*	*
pi.n/ki.n/ti.n											*						
n														*!			
n.t	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
ni.t												*!					
t														*!			
l.t	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
li.t												*!					
t														*!			
w.t	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
wi.t												*!					
t														*!			
s.t	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
si.t												*!					
t														*!			
s.n	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
si.n												*!					
n														*!			

No momento em que a restrição $*\text{Dist}\{1\} \& *\{\text{stop}\}_{\text{coda}'}$, específica das sequências ([p.n, k.n, t.n]), é promovida na hierarquia das restrições, superando DEP no ordenamento, poderá haver a emergência do candidato com epêntese. Conforme já comentamos, um dos aspectos diferenciais deste algoritmo de aprendizagem diante do modelo *standard* da OT é sua capacidade de trabalhar com *outputs* variáveis, dados pela alteração dos pontos de seleção (*disharmony*). Sendo assim, estando DEP e $*\text{Dist}\{1\} \& *\{\text{stop}\}_{\text{coda}'}$ com valores de ranqueamento próximos, essas podem alterar sua posição no *ranking*, em qualquer momento de avaliação, respeitando as probabilidades de ocorrência de cada restrição.

Assim como DEP foi capaz de alterar sua posição com a restrição específica das sequências formadas de consoante plosiva seguida de nasal, ex.: etnia), vemos, no *ranking*, que a restrição de fidelidade DEP também está suscetível diante da sequência de consoante plosiva seguida de outra plosiva [p.t, k.t]). Uma vez que é necessária uma diferença de mais de 10 pontos entre as restrições para que haja variação, [p.t, k.t] também é capaz de emergir, graças à também proximidade de $*\text{Dist}\{1,0\} \& *\{\text{stop}\}_{\text{coda}'}$ (43.366) com DEP (50.650), de modo a respeitar a probabilidade de variação de 98,96% para o candidato [p.t, k.t] ser o vencedor e 1,04% para [ki.t, pi.t] emergir, conforme dados de produção dos estudantes colombianos (tabela 2).

Dado que as sequências [s.t], [s.n], [w.t], [l.t], [n.t] são produzidas no Português Brasileiro e no Espanhol Colombiano, de modo a emergirem sem a inserção de epêntese, é necessário que DEP, frente às restrições que representam estas sequências ($*\{\text{stop}, \dots \text{fric}\}_{\text{coda}'}$, $*\{\text{stop}, \dots \text{nas}\}_{\text{coda}'}$, $*\{\text{stop}, \dots \text{liq}\}_{\text{coda}'}$, $*\{\text{stop}, \dots \text{glide}\}_{\text{coda}'}$, $*\text{Dist}\{+3, \dots +1\}$, $*\text{Dist}\{+3, \dots 0\}$, $*\text{Dist}\{+3, \dots -1\}$, $*\text{Dist}\{+3, \dots -2\}$, $*\text{Dist}\{+3, \dots -3\}$), apresente ranqueamento superior e com um distanciamento de mais de 10 pontos. Assim, em se tratando de uma gramática em desenvolvimento, com a presença de *outputs* variáveis, o algoritmo de aprendizagem será capaz de convergir em uma gramática que reflita os padrões encontrados nos dados dos aprendizes.

3. Considerações finais

Nosso objetivo de descrever e analisar a produção da epêntese vocálica após plosivas em codas mediais de palavras, por parte de aprendizes colombianos de Português Brasileiro, resultou na formulação de três Questões Norteadoras, que nos fizeram refletir desde a descrição dos dados empíricos até as restrições relevantes para tal processo de aquisição.

A primeira Questão Norteadora foi extremamente relevante para que se chegasse à conclusão de que o contato silábico estava se mostrando significativo na inserção da vogal. Assim, ao demonstrarmos que podem ser produzidas vogais de reparo, mesmo que de forma tímida, tivemos condições de trabalhar com uma teoria que lidasse com a nossa suposição de que o Contato Silábico estaria influenciando na inserção da vogal após a plosiva em coda, o que nos

possibilitou responder ao segundo questionamento proposto para esta investigação, questionamento esse que versava sobre as restrições que seriam relevantes para a aquisição dos padrões silábicos do PB como L2.

Para tanto, trabalhamos com teorias que envolviam cada uma das estruturas que abrangem cada língua, principalmente para que pudéssemos perceber que o estágio de interlíngua dos nossos aprendizes de PB estava mais próximo à língua materna dos estudantes. Consideramos cabíveis as restrições pertencentes à família Dist, proposta por Gouskova (2004), como base para nossas restrições de marcação à análise final, com as sequências [k.t, p.t, k.n, p.n, t.n], como forma de oposição a dadas produções de estruturas de *outputs*.

Abordamos, na sequência, a terceira Questão Norteadora, que, questionava se o algoritmo consegue convergir em gramáticas que reflitam os dados de aquisição. As hierarquias de restrições utilizadas nas simulações conseguiram dar conta tanto das gramáticas categóricas da língua materna, com 0% de epêntese, quanto das gramáticas de língua estrangeira, com 100% de epêntese, até a gramática de *outputs* variáveis, da interlíngua.

Verificamos que, na L1 dos aprendizes, as restrições de fidelidade DEP e MAX assumem um valor central que as mantém afastadas das restrições de marcação, de tal forma que, mesmo se fizéssemos avaliações sucessivas, não veríamos variação no que diz respeito ao status hierárquico estabelecido entre as restrições, ou seja, diferentes momentos de produções linguísticas não modificariam as relações hierárquicas estabelecidas entre as restrições. A demissão das restrições de marcação e a promoção de fidelidade no modelo teórico da OT Estocástica demonstrou que, no Espanhol Colombiano, as restrições de fidelidade assumem as posições mais altamente ranqueadas, o que caracteriza o *ranking* no Espanhol.

A hierarquia correspondente à aquisição plena da língua estrangeira — Português Brasileiro — já teve suas alterações marcantes para que se pudesse assumir outro candidato ótimo e, também, categórico. Na aquisição plena do PB, assumimos que o candidato ótimo deveria ser aquele com 100% de epêntese e sem variação; logo, DEP teve de ser demovido para que a epêntese pudesse emergir como a estratégia de reparo capaz de adaptar, ao padrão de PB, os *inputs* que resultariam em plosivas em codas mediais das palavras dadas.

A simulação 3, que pretendia demonstrar os *outputs* variáveis das produções dos aprendizes no PB, teve de restabelecer os pesos das restrições, que se reorganizaram em ranqueamento, demonstrando semelhança à hierarquia de restrições de língua materna. Esse fato sugeriu que a interlíngua dos aprendizes se encontra ainda próxima ao sistema da língua materna, no que se refere à aquisição das codas mediais de consoantes plosivas no PB. Destaquemos, no que diz respeito a esta simulação, a diferença fundamental de que tal simulação resultou em valores próximos para o candidato dado como ótimo e o candidato com epêntese, proximidade essa que indicava a probabilidade de variação entre tais candidatos, permitindo a produção variável do segmento

epentético. O algoritmo consegue, assim, retratar os estágios desenvolvimentais por que passaram os aprendizes.

Feitas essas considerações, concluímos que o estudo de aquisição da língua estrangeira à luz do modelo da Teoria da Otimidade Estocástica permitiu-nos uma explicação formal para os padrões de produção encontrados em nossos dados. ☒

AZEVEDO, R. Q.; MATZENAUER, C. L.; ALVES, U. K. EPENTHESIS IN THE PRODUCTION OF WORD-MID STOP CODAS BY COLOMBIAN LEARNERS OF BRAZILIAN PORTUGUESE: AN OPTIMALITY-BASED APPROACH

Abstract

This study aims to describe and analyze, according to the framework of Stochastic Optimality Theory – OT (BOERSMA & HAYES, 2001), the production, by Colombian Learners of Brazilian Portuguese (BP), of stop segments in word-mid codas, liable for the production of epenthetical segments. In order to build the corpus, an instrument composed of Portuguese and Spanish cognate words was proposed, so that vehicle phrases could be recorded in AUDACITY 1.3.5 (ANSI) software. Four Colombian students, who remained in Southern Brazil for a period of six months, were chosen to take part in the investigation. In the selection of the participants, extralinguistic factors such as sex, age and scholarity were homogenized. We confirmed the hypothesis that even in cognate words, there is a different treatment to plosives in word-mid codas in both languages, depending on the language system which is being used by the speaker. This fact was accounted for, according to the Stochastic Optimality Theory framework, through hierarchies of universal constraints.

Keywords

Stochastic Optimality Theory; Universal Constraints; Vowel Epenthesis

Referências bibliográficas

ALVES, U. K. *A Aquisição das Sequências Finais de Obstruintes do Inglês (L2) por Falantes do Sul do Brasil: Análise via Teoria da Otimidade*. Tese de doutorado inédita. Porto Alegre: PUCRS, 2008.

_____. *A Epêntese Vocálica na Aquisição das Plosivas Finais do Inglês (L2): Tratamento pela OT Estocástica e Pela Gramática Harmônica*. Trabalho apresentado no II SIS Vogais. Belo Horizonte, 2009.

BISOL, L. A Sílabas e seus Constituintes. In: NEVES, M. H. M. (Org.). *Gramática do Português Falado*, v. 7. Campinas: Ed. da UNICAMP, 1999.

_____. (org). *Introdução a Estudos de Fonologia do Português Brasileiro*. 4. ed. Porto

Alegre: EDIPUCRS, 2005.

_____; Schwindt, L.C (org.). *Teoria da Otimidade: Fonologia*. Campinas, SP: Pontes Editores, 2010.

BOERSMA, P. HAYES, B. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86, 2001.

BOERSMA, P.; PATER, J. Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. Amsterdam: University of Amsterdam, UMass Amherst, 2008. (Manuscript).

CLEMENTS, G. N. The Role of the Sonority Cycle in Core Syllabification. In: KINGSTON, J. & BECKMAN, M. (orgs.). *Papers in Laboratory Phonology I*. Cambridge: CUP, p. 283-333, 1990.

_____.; HUME, E. V. The Internal Organization of Speech Sounds. In: GOLDSMITH, J. *The Handbook of Phonological Theory*. London: Blackwell, 1995.

COLLISCHONN, G. *Análise Prosódica da Sílabo em Português*. Tese de doutorado inédita. Porto Alegre: PUCRS, 1997.

_____. A Sílabo em Português. In: BISOL, Leda (org.). *Introdução a Estudos de Fonologia do Português Brasileiro*. 4. ed. Porto Alegre: EDIPUCRS, p. 101-133, 2005a.

GOUSKOVA, M. Relational Hierarchies in Optimality Theory: The case of Syllable Contact. *Phonology* 21:2, p. 201-250, 2004.

HARRIS, J. W. *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge, Mass: MIT Press, 1983.

JESNEY, K.; TESSIER, A. Re-evaluating Learning Biases in Harmonic Grammar. In: BECKER, M. (org.). University of Massachusetts Occasional Papers in Linguistics, 36: *Papers in Theoretical and Computational Phonology*. Amherst: GLSA, 2007. p.1-42.

LUCENA, R. M.; ALVES, U. K. Implicações Dialetais (Dialeto Gaúcho vs. Paraibano) na Aquisição de Obstruintes em Coda por Aprendizes de Inglês: Uma análise Variacionista. *Letras de Hoje*, v. 45, p. 35-42, 2010.

_____; SORACE, A.; SMOLENSKY, P. The Optimality-Theory – Harmonic Grammar Connection. In: SMOLENSKY, P.; LEGENDRE, G. *The Harmonic Mind*. The MIT Press, 2006.

MCCARTHY. *A Thematic Guide to Optimality Theory*. Cambridge University Press, 2002.

_____; PRINCE, A. Faithfulness and Reduplicative Identity. In: BECKMAN, J. et al (eds.). *University of Massachusetts Occasional Papers in Linguistics* 18. Amherst, MA: GLSA Publications, 1995. [Disponível: ROA 60].

MATZENAUER, C.; MIRANDA, A. R. Traços Distintivos. In: BISOL, Leda; SCHWINDT, Luiz Carlos (org.). *Teoria da Otimidade: Fonologia*. Campinas: Pontes Editores, 2010, p. 11-55.

MURRAY, R.; VENNEMANN, T. Sound Change and Syllable Structure [: problems] in Germanic Phonology. *Language* 59, 1983.

PRINCE, A; SMOLENSKY, P. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004. [Disponível: ROA-573].

QUINTANILHA, R; ALVES, U. A Epêntese no Português Brasileiro (L2) em Segmentos Obstruintes em Códigos Mediais de Palavras, por Falantes Nativos do Espanhol Colombiano (L1): Uma Análise via Gramática Harmônica. Palhoça-S.C: Unisul, 2010, p. 248.

SELKIRK, E. The syllable. HULST, H. V. D., SMITH. *The structure of phonological representations (part II)*. Foris, Dordrecht, p. 337-383, 1982.

SMOLENSKY, P.; LEGENDRE, G. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge: MIT, 2006.