

REAVALIAÇÃO DE RESULTADOS EXPERIMENTAIS SOBRE A FONOTAXE DO PORTUGUÊS BRASILEIRO: TRANSIÇÕES ENTRE FONES E GRAU DE ACEITABILIDADE EM LOGATOMAS*

Aglael GAMA ROSSI (Pontifícia Universidade Católica de São Paulo)

gamarossi@uol.com.br

Soluá Simões de ALMEIDA (Pontifícia Universidade Católica de São Paulo)

soluaalmeida@hotmail.com

ABSTRACT: *The paper presents an analysis of how the transitions of sounds (or phonotactics) inside nonsense words can influence their acceptability in an experiment. Evidences in favor of the Obligatory Contour Principle are found.*

KEYWORDS: *Transitions between and among sounds; Nonsense Words; Acceptability Judgments; OCP*

0. Introdução

Este trabalho insere-se na área da Psicolinguística Experimental, lugar onde se buscam evidências empíricas para o processamento lingüístico-cognitivo. Seu interesse diz respeito ao levantamento de hipóteses sobre os fatores envolvidos nos julgamentos de gramaticalidade das combinações de sons/letras na língua (fonotaxe). A partir de seqüências de sons ou letras, aceitas ou preteridas, pelos falantes do português brasileiro (PB), são feitas suposições sobre que tipo de conhecimento fonológico está por detrás do processamento lexical. Na análise a ser aqui relatada, são levantadas as freqüências de ocorrência de transições ou seqüências de tamanho variável dentro de logatomas (palavras passíveis de serem pronunciadas ou lidas, mas sem significado). Como exemplo, no logatoma *daba*, encontram-se as seguintes transições diretas: *da, ab, ba, dab, aba*, além das indiretas: *d-b, a-a*. O logatoma como um todo, durante o julgamento de gramaticalidade, pode ainda ser contraposto a pares mínimos, palavras da língua que dele diferem por um som ou letra, como por exemplo: *baba, data* e *taba*. A questão deste trabalho diz respeito ao papel dessas transições e dos pares mínimos no grau de aceitação dos logatomas.

* Este trabalho liga-se aos projetos de pós-doutorado “Realidade psicológica da fonotaxe do português brasileiro e o estatuto das representações lexicais” (LIAACC/LAEL/PUC-SP/FAPESP, processo n°. 00/00906-0) e iniciação científica “O controle de variáveis no estudo experimental da fonotaxe do português brasileiro”, de Soluá Simões de Almeida (PIBIC-CNPq).

1. *Corpus* de logatomas e coleta dos julgamentos de gramaticalidade

Na análise das transições entre sons/letras, foram utilizados 307 logatomas que faziam parte um *corpus* original de 396, oferecidos, num primeiro experimento (Gama-Rossi e Madureira, 2002), a 38 sujeitos adultos, 9 homens e 29 mulheres, com no mínimo terceiro grau completo, numa tarefa de leitura silenciosa, na qual eles respondiam se o logatoma poderia ou não ser uma palavra do PB. Foi utilizado como meio para a realização do experimento o programa EX-PE6 (Pallier, Dupoux e Jeannin, 1997), que anotava a lista de logatomas apresentada, a resposta ao logatoma (sim ou não) e o tempo em milésimos de segundo entre o aparecimento dele no centro da tela do computador e o apertar do *joystick* pelo sujeito.

Quanto aos critérios que serviram de base à confecção dos logatomas, a feição destes é de “substantivos”, dissílabos paroxítonos, com estrutura silábica C1V1.C2V2, na qual V1 é uma das sete vogais orais do PB e V2 é sempre “a”. O levantamento da frequência de ocorrência das 18 consoantes que puderam ocupar a posição C1 (excetuando-se a vibrante simples “r” que ocorre somente em posição intervocálica no PB) e das 19 consoantes que ocuparam a posição C2 foi realizado dentro do subconjunto de dissílabos paroxítonos com estrutura silábica CV.CV, terminados em “a”, no *MiniDicionário Aurélio*, que contém como base 27.078 palavras, as quais podem ser buscadas a partir de diferentes critérios no Programa LISTAS, confeccionado no LAFAPE (IEL/UNICAMP), sob orientação da Profa. Dra. Eleonora Albano, que nos cedeu uma cópia do mesmo.

Os logatomas foram compostos, na primeira sílaba, pela combinação entre as consoantes que em C1 eram de baixa ocorrência no subconjunto pesquisado (correspondentes aos fones /v/, /ʒ/, /n/, /g/, /ʃ/, /d/, /z/, /k/, /ŋ/, grafados respectivamente com “v”, “j” antes de “a, o, u” e “g” antes de “e, i”, “n”, “g” antes de “a, o, u” e “gu” antes de “e, i”, “ch”, “d”, “z”, “lh” e “nh”), seguidas pelas 7 vogais orais do PB em V1, e na segunda sílaba, com todas as consoantes, seguidas sempre por “a” em V2. A forma ortográfica dos logatomas (por exemplo, a decisão de usar “j” ou “g” antes de “e,i” ou “ss” ou “ç” entre vogais) foi também pesquisada no LISTAS e utilizada a de maior frequência de ocorrência.

A hipótese de trabalho, no primeiro experimento, era que os logatomas, sempre formados por tônicas de baixa frequência, seriam mais aceitos quando seguidos por pós-tônicas de alta frequência. Porém, essa hipótese nem sempre se confirmou. Exemplos disso são os altos índices de aceitação recebidos por logatomas terminados em sílabas pós-tônicas de baixa frequência no subconjunto pesquisado (Gama-Rossi e Madureira, op. cit.). Os logatomas iniciados com /ʃ/, /k/ e /ŋ/ receberam um alto índice de rejeição, independentemente da frequência de ocorrência de C2. Por isso, eles foram retirados desta pesquisa das transições, a qual lidou com um conjunto de 307 logatomas.

Durante o experimento, os sujeitos referiram que, ao julgar o logatoma, na tarefa de leitura silenciosa, avaliavam sua *sonoridade* ou *facilidade de pronúncia*, sugerindo que as formas acústica, articulatória, e visual ou grafêmica das palavras encontram-se intimamente ligadas em sujeitos com alto grau de escolaridade.

2. Critérios para o levantamento das transições dos logatomas no LISTAS

O levantamento das transições foi realizado com base na transcrição fonêmica oferecida pelo LISTAS (Albano, Moreira, Silva, Aquino e Kakinohana, 1999). São, portanto, transições entre fones, algumas delas constituindo unidades de análise lingüística. Retomando o exemplo de *daba*, *da* corresponde à sílaba tônica; *ba*, à sílaba pós-tônica; *ab*, ao GIPC ou Grupo *Inter-Perceptual Center*, uma unidade de produção e percepção da fala, que tem como base o *onset* da vogal e todas as consoantes que a ela se seguem até a próxima vogal (Barbosa, 1996); *d-b*, coarticulação entre consoantes, e *a-a*, entre vogais, as quais podem ser influenciadas pelo Princípio do Contorno Obrigatório (*Obligatory Contour Principle – OCP*), segundo o qual dois sons idênticos não podem ser adjacentes a menos que estejam separados por uma fronteira de palavra (Goldsmith, 1990). Já no que concerne às seqüências *dab* e *aba* nem sempre é possível afirmar, quando se considera o conjunto de logatomas, que elas constituam unidades lingüísticas, tais como morfemas ou palavras. Quanto aos pares mínimos, os mesmos foram levados em conta para verificar até que ponto o número de pares mínimos que o logatoma possui com palavras da língua poderia intervir em sua aceitabilidade.

Cada tipo de transição, no conjunto de 307 logatomas, foi levantado no LISTAS (27.078 palavras), a partir de critérios criados especialmente para tais buscas. Como exemplos, são citadas as buscas do GIPC. Ainda tomando como base *daba*, foram realizadas buscas com e sem divisão silábica, a saber: *a.b* e *ab*. No primeiro caso, o GIPC é contado entre sílabas ("a" e "b" pertencem a sílabas separadas) e no segundo caso, dentro da mesma sílaba. Com a marca da divisão silábica, foram então contadas as transições *a.b* de palavras tais como: *abacate*, e sem a marca da divisão silábica foram contadas as transições *ab*, como em: *abdicar*. Outro exemplo de critério adotado no levantamento das transições refere-se àquele entre V1V2. Neste caso, as buscas envolveram três tipos de estrutura silábica: 'V1.C2V2, 'C1V1.C2V2, 'C1C1V1.C2V2, isto porque o critério adotado foi que não poderia haver mais de uma consoante (C2) entre as vogais (V1 e V2), não importando o número de consoantes que antecedessem à primeira vogal (V1). Quanto à transição entre consoantes, as buscas foram feitas de modo a excluir consoantes de coda na primeira sílaba, embora na segunda pudesse haver grupo consonantal (respectivamente, 'C1V1.C2 ou 'C1V1.C2C2V2).

3. Análise das transições contidas nos logatomas

As buscas mostram uma variação muito grande em relação aos valores de frequência de ocorrência das transições dentro de cada logatoma. Essa variação parece diminuir dos valores absolutos para os valores de log10 (uma transformação de base imposta aos absolutos).

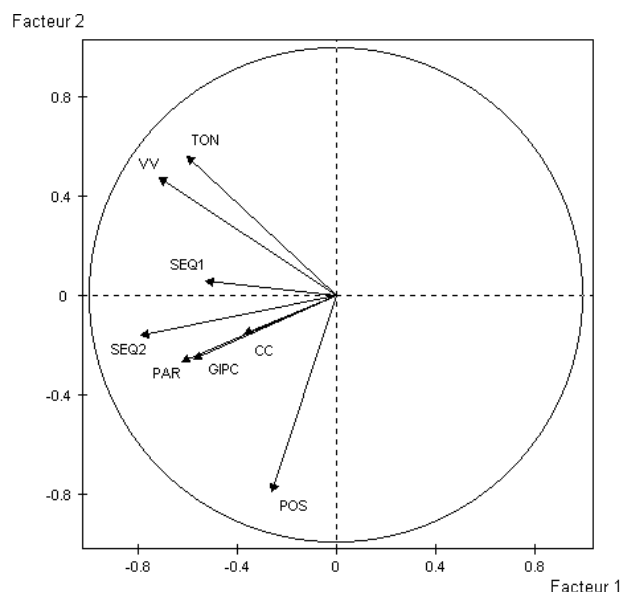


Gráfico 1: Influência das variações entre as transições para valores absolutos

Uma análise fatorial¹ apontou a variação que uma transição exercia sobre a outra, para os três valores (absolutos, percentuais e log10). Por exemplo, no gráfico 1, colocado acima, que tem como base valores absolutos, as variações das transições Par Mínimo, GIPC e C1C2 influenciam-se fortemente, conforme pode ser visto por meio de seus vetores inteiramente sobrepostos uns aos outros.

Segundo a análise, a variação da Tônica parece exercer influência sobre a variação de V1V2. Já a variação da Pós-tônica não influencia ou sofre influência da variação de nenhuma das demais transições. As variações das transições referentes à Sequência 1, Sequência 2, Par Mínimo, GIPC e C1C2 influenciam-se, apesar de fazerem-no em diferentes graus em função do valor de transformação de base adotado. Para valores percentuais, Par Mínimo, GIPC e C1C2 também sobrepõem-se, sendo que Sequência 2 parece ser mais influenciada pela variação dessas três transições que Sequência 1. Para valores de log10, GIPC e Sequência 1 estão sobrepostos e o vetor relativo a C1C2 aproxima-se daquele da Pós-Tônica. Um teste de correlação mostrou as seguintes correlações em torno ou acima de 50%:

Tabela 1: Correlações em torno e acima de 50% entre as transições levantadas

	Tônica			GIPC			Sequência 2		
	abs	per	log	abs	per	log	abs	per	log
Seq 2	-	-	-	0,52	0,53	0,81	-	-	-
V1V2	0,59	0,59	0,52	-	-	-	0,46	0,47	0,41
Par Mín	-	-	-	-	-	-	0,50	0,49	0,44

Os resultados na tabela 1 mostram, para valores de log10, uma alta correlação entre Seqüência 2 e GIPC, o que é totalmente esperado, já que o GIPC está contido na Seqüência 2 (por exemplo, *ab* em *aba* de *daba*). Há ainda correlação entre V1V2 e Sílabas Tônicas, o que parece ressaltar o papel da vogal tônica (V1). Mas seria esperado então encontrar uma forte correlação também entre Tônica e GIPC e V1V2 e GIPC, o que, como mostra a tabela 2, colocada abaixo, não ocorre. Finalmente, na tabela 1, são observadas correlações entre Par Mínimo e Seqüência 2, o que pode ser entendido pelo fato de a Seqüência 2 constituir pares mínimos com alguns logatomas, quando deles se extrai uma letra ou som, tais como: *epa* contida em *gepa*; *opa* contida em *gopa*; *upa* em *nupa*; *eba* contida em *neba*; *oba* em *zoba*.

Tabela 2: Correlações abaixo de 50% entre as transições levantadas

	GIPC			Seq1		
	ab	per	log	ab	per	log
Tônica	0,12	0,11	0,27	0,23	0,21	0,33
GIPC	-	-	-	0,19	0,20	0,32
Pós-tônica	0,21	0,21	0,26	0,05	0,04	0,08
Seqüência 2	-	-	-	0,28	0,28	0,26
V1V2	0,30	0,30	0,33	0,26	0,26	0,15
C1C2	0,06	0,06	0,15	0,19	0,20	0,28
	C1C2			Par Mínimo		
	ab	per	log	ab	per	log
Tônica	0,29	0,28	0,11	0,22	0,23	0,26
GIPC	-	-	-	0,17	0,17	0,31
Pós-tônica	0,25	0,26	0,30	0,26	0,25	0,26
Seqüência 2	0,12	0,13	0,00	-	-	-
V1V2	0,11	0,11	0,00	0,25	0,26	0,10
C1C2	-	-	-	0,15	0,13	0,17

De acordo com a tabela 2, os valores transformados em log10, a princípio menos variáveis, apontam para outras correlações entre as transições pesquisadas. Há uma melhora das correlações entre GIPC e Tônica, Seqüência 1 e Tônica, Seqüência 1 e GIPC, o que pode sugerir um papel importante para a vogal tônica (V1), elemento comum a todas essas transições. A piora nas correlações entre Seqüência 1 e V1V2, em oposição ao aumento na correlação entre Seqüência 1 e C1C2, parece mais plausível, uma vez que Seqüência 1 contém C1C2. Também parece esperada a falta de correlação entre C1C2 e V1V2. Houve ainda um aumento na correlação entre GIPC e Par Mínimo, nos valores de log10, para o qual não encontramos explicação. Portanto, a análise das correlações entre as frequências de ocorrência das transições dentro de cada logatoma ressaltou influências, muitas delas já esperadas, entre transições que são sabidas relacionadas, mas também apontou para a falta de importância de algumas unidades, como parece ser o caso da Pós-tônica, que não se correlaciona a nenhuma das outras transições.

3.1. Cruzamento entre transições e resultados experimentais

O objetivo das análises abaixo é contrapor o grau de aceitação recebido pelo logatoma, no experimento, com os valores de cada uma de suas transições, tentando verificar qual(is) dela(s) pode(m) ter contribuído para a maior aceitação do logatoma ou até se nenhuma delas participou da mesma. Vale notar que a média de aceitação no experimento foi 20, considerando os logatomas iniciados em “ch”, “nh” e “lh”, que tiveram maior índice de rejeição, e sem eles, a média foi 29.

Uma inspeção dos dados mostrou que havia claras evidências a favor da influência do Princípio do Contorno Obrigatório (*OCP*) na aceitabilidade dos logatomas, o que significa um papel importante das transições C1C2 e V1V2. Sendo a dissimilaridade um fator facilitador da articulação de sons numa seqüência e da saliência perceptual dessa, tomemos alguns exemplos de logatomas formados por consoantes que apresentam mesmo modo e ponto de articulação, o que significa um contorno consonantal ruim, ou uma violação de *OCP*, e também de logatomas formados por consoantes com maior grau de dissimilaridade, medido através da quantidade de categorias fonéticas (modo, ponto e vozeamento) que separam os fones consonantais ou vocálicos do logatoma.

Quadro 1: Logatomas com diferentes graus de dissimilaridade entre C1C2

logatoma	ton	pós	GIPC	seq1	seq2	V1V2	C1C2	par
guessa	8	194	2434	0	29	316	22	0
Zessa	29	194	2434	0	29	316	14	0
Zissa	112	194	1505	2	34	574	14	1
Gissa	127	194	1505	2	34	574	6	2

Entre os logatomas do quadro 1, *guessa* recebeu uma aceitação de 29 contra 21 de *zessa*. Uma vez que as vogais estão controladas entre ambos os logatomas, pode-se dizer que a maior aceitação de *guessa* está relacionada ao maior grau de dissimilaridade entre suas consoantes, oclusiva velar sonora e fricativa alveolar surda, ao passo que em *zessa* tem-se duas fricativas alveolares na seqüência, que diferem apenas pelo vozeamento. Note que a seqüência C1C2 (no quadro 1) é maior para *guessa*, embora a Tônica seja maior para *zessa*. Na comparação entre *zessa* e *zissa*, ambos receberam basicamente o mesmo grau de aceitação, respectivamente: 21 e 22. Ambos apresentam o mesmo valor para C1C2, mas a tônica “zi” tem uma ocorrência quase quatro vezes maior que a tônica “ze”, sendo V1V2 também maior para “zi”. Apenas o GIPC é maior para *zessa*, o que não parece suficiente para torná-lo mais aceito. No caso de *gissa* versus *zissa*, apesar de a freqüência de ocorrência de C1C2 ser maior em *zissa*, o *OCP* é melhor em *gissa* (fricativa palatal sonora e fricativa alveolar surda), o que pode explicar sua maior aceitação: 28 em oposição a 22, para *zissa*. Uma volta ao LISTAS mostrou que entre as 14 palavras com C1C2 correspondente a /z/-/s/, apenas quatro ocorrências contêm essa seqüência em seu radical: opositorista (contado duas vezes, como substantivo e adjetivo), exacerbar e exacerbação. Nas demais, a seqüência /z/-/s/ aparece em formas derivadas, devido aos prefixos

des- (desacertar, desassimilação, desassimilar, desassisado, desassociar, desassombrado, desassossegado, desassossegar) e trans- (transoceânico). Quanto às seis palavras com a seqüência /z/-/s/, em apenas duas, ela surge em formas derivadas, intrujice e rabugice, sendo as demais: gesso, jaça, jaçanã e joça.

Quadro 2: Logatomas com pouca dissimilaridade entre C1C2

logatoma	ton	pós	GIPC	seq1	seq2	V1V2	C1C2	par
zassa	119	194	1123	11	119	781	14	3
zessa	29	194	2434	0	29	316	14	0
zissa	112	194	1505	2	34	574	14	1
zussa	15	194	305	0	5	369	14	0

No caso dos logatomas pares mínimos, *zassa*, *zessa* e *zissa* (quadro 2), parece que nem os altos valores de GIPC nem os de transição V1V2 ajudaram em sua aceitação, respectivamente: 16, 21 e 22. *Zassa*, o menos aceito, corresponde também à seqüência, entre as três, de menor dissimilaridade, tanto entre vogais como entre consoantes, apesar de ele apresentar o maior valor para a transição V1V2. Também as diferenças entre Tônica, GIPC, e V1V2 de *zessa* e *zissa* não colaboraram no desempate dos mesmos. Esse parece mais um exemplo de que o grau de dissimilaridade entre os fones da seqüência funcionou como um fator importante na aceitabilidade dos logatomas, uma vez que neles a maior distinção reside entre as vogais /i/-/a/ e /e/-/a/. Entre *zassa* e *zussa*, que receberam respectivamente a aceitação de 16 e 24, *zussa* ultrapassa a média de aceitação do experimento, apesar do contorno consonantal ruim e dos baixos valores de suas transições como um todo, sugerindo que o contorno vocálico agiu sobre sua aceitação. A análise de *zussa* insinua ainda que a freqüência de ocorrência de suas transições não influenciou sua aceitação.

Quadro 3: Logatomas com diferentes graus de dissimilaridade entre C1C2

logatoma	ton	pós	GIPC	seq1	seq2	V1V2	C1C2	par
zacha	119	59	85	0	34	781	0	2
zaja	119	34	258	17	11	781	21	3
zalha	119	114	104	2	53	781	6	6

Outro exemplo do que parece ser a influência de um contorno consonantal dissimilatório no julgamento de aceitabilidade dos logatomas pelos sujeitos pode ser visto entre os logatomas do quadro 3, que receberam os seguintes índices de aceitação, respectivamente: 26, 22 e 27. Entre /z/-/k/ (*zalha*), há uma diferença de modo e ponto de articulação (fricativa alveolar sonora e lateral palatal); entre /z/-/ʃ/ (*zacha*), uma diferença entre ponto de articulação e vozeamento (fricativa alveolar sonora e fricativa palatal surda), e finalmente entre /z/-/ʒ/ (*zaja*), há uma diferença apenas de ponto de articulação (fricativa alveolar sonora e fricativa palatal sonora), apesar de aqui serem encontrados valores mais altos de GIPC e 'C1C2.

4. Considerações Finais

A reavaliação dos resultados levando em conta o cruzamento entre as frequências de ocorrência das transições (ou unidades lingüísticas) dentro de cada logatoma e seu julgamento de aceitabilidade pelos sujeitos apontou para o fator da facilidade articulatória e saliência perceptual, ou seja, quanto mais dissimilares os fones que constituem os logatomas (boa formação quanto a *OCP*), maior sua aceitação. É provável que isso ocorra pelo fato de a estrutura silábica dos logatomas ser muito simples, C1V1.C2V2. O papel do contorno vocálico torna-se importante quando o contorno consonantal é ruim. E a Pós-Tônica, que no primeiro experimento foi tomada como parâmetro para a aceitação dos logatomas, mostrou-se de pouca importância na correlação com as demais transições e, por conseguinte, na aceitação dos mesmos.

REFERÊNCIAS BIBLIOGRÁFICAS:

- ALBANO, Eleonora Cavalcante; MOREIRA, Agnaldo; SILVA, Adelaide; AQUINO, Patrícia; KAKINOHANA, Régis. Um conversor ortográfico-fônico e uma notação prosódica mínima para síntese de fala em língua portuguesa. In: E. M. SCARPA (org.) *Estudos de prosódia*. Campinas: Editora da UNICAMP, 1999.
- BARBOSA, Plínio Almeida. At least two macrorhythmic units are necessary for modeling Brazilian Portuguese duration: emphasis on segmental duration generation. *Cadernos de Estudos Lingüísticos*, 31:33-53, 1996.
- GAMA-ROSSI, Aglael; MADUREIRA, Sandra. Julgamentos de aceitabilidade da fonotaxe do português brasileiro: primeira análise. *Estudos Lingüísticos*, v.XXXI, Grupo de Estudos Lingüísticos do Estado de São Paulo (GEL), publicação em CD-ROM, <file:///F:\htm\comunica\CiIII26a.htm>, 2002.
- GOLDSMITH, John. *Autosegmental & Metrical Phonology*. Oxford: Blackwell Publishers, 1990.
- PALLIER, Christophe; DUPOUX, Emmanuel; JEANNIN, Xavier. EXPE: an Expandable Programming Language for On-line Psychological Experiments. *Behavior Research Methods, Instruments and Computers*, 29 (3): 322-327, 1997.

NOTAS

¹ Agradecemos a análise estatística à Profa. Dra. Yara Castro, Consultora em Estatística da PUC-SP.