

CONTAGEM DE FREQUÊNCIA DOS PADRÕES SILÁBICOS NO PORTUGUÊS

Resumo

A fim de subsidiar análises comparativas sobre o efeito da língua ambiente na aquisição dos padrões silábicos por crianças falantes de diversas línguas, e devido ao fato de inexistirem contagens estatisticamente controladas a respeito dos padrões silábicos encontrados no Português que sirvam para alimentar estudos comparativos desta (ou de outra) natureza, impôs-se fazer o levantamento sistemático da ocorrência dos tipos silábicos na língua. Para tanto, utilizou-se a base de dados escrita empregada por Albano et alli (1995) para o estudo de frequência dos sons no Português Brasileiro. A amostra analisada é constituída pelo conjunto completo de verbetes (27,074) do Minidicionário Aurélio, (Ferreira, 1977). Achados anteriores foram corroborados (Teixeira, 1997), confirmando ser CV o padrão mais recorrente na língua, seguido por (C)VC, CCV e V, e (C)VCC.

contagem de frequência; padrões silábicos; fonologia do português.

Abstract

In order to allow for comparative analyses about ambient language effects on the acquisition of syllable patterns by children from different language environments, and due to the inexistence of statistically controlled counts on Portuguese syllable patterns, a study of this type was felt necessary. For this purpose, the written database compiled by Albano et alli. (1995), for the study of segment frequency in Brazilian Portuguese was used. The corpus analysed consisted of the whole set of entries (27.074) from the Minidicionário Aurélio (Ferreira, 1997) The syllable patterns computed were, basically: V, CV, (C)VC, CCV e (C)VCC. Previous findings could be corroborated (Teixeira, 1997), confirming that CV is the most recurrent pattern in the language, followed by (C)VC, CCV, V e (C)VCC.

Considerações iniciais

A descoberta de generalizações relativas ao conteúdo e a estrutura dos inventários fonológicos das línguas tem sido um dos mais perseguidos objetivos dos estudos atuais. Estas generalizações têm sido levadas em conta, implícita ou explicitamente, na formulação de teorias fonológicas, na avaliação de reconstruções históricas, na construção de modelos de mudança lingüística e de aquisição da linguagem, além de ter estimulado grande parte da pesquisa lingüística voltada para a fonética.

Existem, basicamente, três tipos de fontes para observação de inventários fonológicos. O tipo de mais antiga tradição são os relatos impressionistas baseados na experiência de lingüistas com conhecimento sobre diversas línguas. Exemplos deste tipo de relatos são os estudos de Trubetsky (1939) Jakobson & Halle (1956), Ladefoged (1971) bem como os comentários incidentais encontráveis em artigos de numerosos autores. Embora estes relatos sejam calçados na familiaridade dos autores com as línguas estudadas, a validade das conclusões fica prejudicada principalmente pela falta de quantificação associada aos mesmos.

O segundo tipo consiste em amostras de línguas explicitamente compiladas para servirem a estudos específicos, como é o caso dos estudos de Ferguson (1963), Greenberg (1970) e Hyman (1977) sobre nasais, consoantes glotais e acento, respectivamente. Nesses casos, pode-se avaliar a qualidade das amostras e a significação das conclusões no que diz respeito a cada assunto específico.

O terceiro tipo de fonte de dados é um levantamento normatizado com um amplo espectro de objetivos. O mais conhecido trabalho deste tipo é o Arquivo Fonológico de Stanford, compilado na Universidade de Stanford sob a coordenação de Greenberg & Ferguson. O outro grande trabalho nesta linha é a Base de dados de Inventários Fonológicos Segmentais de UCLA (UPSID), coordenada por Maddison (1984).

A superioridade deste tipo de base de dados se prende à própria natureza da área de estudos. Este tipo de base de dados serve, primeiramente, não só para gerar observações relativas à frequência de diferentes tipos de segmentos e de diferentes atributos destes segmentos, como também para sua co-ocorrência nos inventários fonológicos. A partir daí, então, colocam-se hipóteses com postulações sobre a relação contingente entre a ocorrência de segmentos e conjuntos de segmentos e as limitações relativas a sua distribuição dentro dos inventários.

Embora o conhecimento estatístico sobre as línguas naturais, em geral, seja ainda bastante limitado, já existem muitos estudos sobre a frequência de ocorrência de segmentos e padrões sonoros em algumas línguas, tanto no que diz respeito à língua adulta como à língua infantil. Entre estes, destacam-se:

Em relação à língua adulta:

- BUEKELMAN, D.R., YORKSTON, K.M., POBLETE, M. & NARANJO, C. (1984) Frequency of word Occurrence in Communication Samples produced by Adult Communication Aid Users. *Journal of Speech and Hearing Disorders*, 49: 367.
- CARTERETTE, E. & JONES, M. (1974) *Informal Speech: Alphabetic and phonemic Texts with Statistical Analyses and Tables* Berkeley: University of California Press.
- KENT, R. D. (1994) *Reference Manual for Communicative Sciences and Disorders: Speech and Language*. Austin, Tx.: Pro-ed.
- MINES, M. HANSON, B. & SHOUP, J. (1978) Frequency of occurrence of phonemes in conversational English. *Language and Speech* 21, 221-241.
- SRIBERG, L.D. & KWIATKOWSKY, J. (1982) Descriptive Statistics for Continuous Speech Samples. *Journal of Speech and Hearing Disorders* 47, 268
- WEPMAN, J. M. & LOZAR, B. (1973) The most frequently used words in spoken English. *Journal of Psycholinguistic Research*, 2, 134-135.

Em relação à língua infantil, encontram-se estudos normativos e comparativos entre línguas, como, por exemplo:

- ANTHONY, A., BOGLE, D., INGRAM, T.T.S. & MCISAAC, M.V. (1971) *Edinburgh Articulation Test – Textbook*. Edinburgh: E.S. Livingstone.
- CAMAIONI, L., M.C. CASELLI, E. LONGOBARDI e V. VOLTERRA. A Parent report instrument for early language assessment. *First Language*, 11, 345-359, 1991.
- CASELLI, M.C. e P. CASADIO. A parent report study of lexical and grammatical development in Italian. Contribution to a symposium *Parent Report on Communicative Development in Five*

Languages, International Congress for the Study of Child language, Trieste, Italy, 1993.

- DALE, P. S. The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Sciences*, 34, 565-571, 1991.
- DALE, P. S., E. BATES, S. REZNICK e C. MORISSET. The validity of a parent report instrument of child language at 20 months. *Journal of Child Language*, 16, 239-250, 1989.
- FENSON, L., DALE, P. et. alli MacArthur Communicative Development Inventories: User's Guide and Technical Manual. San Diego, California: Singular Publishing Group, 1993.
- JACKSON-MALDONADO, D. et. alli. Early lexical development of Spanish-speaking infants and toddlers. *Journal of Child Language* (no prelo)
- MADER, J. (1954) The Relative frequency of occurrence of English Consonant Sounds in Words in the Speech of Children in Grades One, Two and Three, *Speech Monographs* 21, 294-300.
- MOE, J. HOPKINGNS, C.J. & RUSH, R.T. (1982) Frequency of word usage by First-grade children. In *The Vocabulary of First-Grade Children*, 94, Springfield, Ill: Charles C. Thomas.
- OGURA, T. et. alli. Some preliminary findings from the Japanese Early Communicative Development Inventories. Contribution to a symposium *Parent Report on Communicative Development in Five Languages*, International Congress for the Study of Child language, Trieste, Italy, 1993.
- REILLY, J.S., K. PROVINE e U. BELLUGI. The MacArthur Communicative Development Inventories. Contribution to a symposium *Parent Report on Communicative Development in Five Languages*, International Congress for the Study of Child language, Trieste, Italy, 1993.

Em Português, o primeiro estudo de contagem de frequência utilizando bases de dados da língua adulta foi o de ALBANO, E., MOREIRA, A., AQUINO, P., SILVA, H., & KAKINOHANA, R. (1995). Segment Frequency and Word Structure in Brazilian Portuguese.

Vol. 3, 346-349. Stockholm. Neste estudo, duas bases de dados foram utilizadas para a contagem de frequência da ocorrência de consoantes e vogais no Português Brasileiro:

- **a base de dados escrita:** constando de 27.074 verbetes ortográficos do Minidicionário Aurélio, envolvendo um total de 212.225 segmentos analisados,
- **e a base de dados falada:** constando de 57 fitas ortograficamente transcritas do projeto NURC, envolvendo um total de 1.681.133 segmentos analisados.

Ambos os corpora foram convertidos da ortografia oficial para um código fonológico abstrato com a ajuda de um programa de computador que determinava a silabificação e o acento lexical, ao mesmo tempo que resolvia ambigüidades ortográficas.

Mais recentemente, foi realizada uma contagem de frequência com alvos lexicais adultos encontrados no repertório infantil inicial. Para tanto, foi realizada uma adaptação para o Português da Parte I (PALAVRAS USADAS PELAS CRIANÇAS) do protocolo de PALAVRAS E SENTENÇAS dos **Inventários MacArthur de Desenvolvimento Comunicativo**. Os *Inventários* foram utilizados a fim de estabelecer uma contagem de frequência dos sons do sistema adulto-alvo com maior recorrência na fala infantil inicial, com o objetivo maior de verificar as influências da língua ambiente nas primeiras produções infantis no Português Brasileiro. Este trabalho - *'The MacArthur Inventory (CDI) Adapted to Brazilian Portuguese: Analysis of the Early Sound Patterns of Children's Targets'* - foi apresentado em forma de conferência promovida pelo Grupo de Pesquisa do Speech Production Lab, da University of Texas em Austin, em 1997. Mais recentemente, um outro trabalho sobre a adaptação desses inventários para o Português foi apresentado no II Congresso Nacional da ABRALIN, em fevereiro de 1999.

No estudo de 1997, a frequência dos seguintes aspectos foi estudada e comparada a dados de produção infantil:

- segmentos consonantais e vocálicos
- Ditongos
- Encontros Consonantais
- Extensão Lexical
- Padrões Silábicos
- Recorrência das Categorias Gramaticais no Léxico inicial

O trabalho de contagem de frequência dos padrões silábicos no português brasileiro

A fim de subsidiar análises comparativas sobre o efeito da língua ambiente na aquisição dos padrões silábicos por crianças falantes de diversas línguas (Teixeira & Davis 1999) (Davis, MacNeilage, Gildersleeve-Newmann & Teixeira 1999), e devido ao fato de inexistirem contagens estatisticamente controladas a respeito dos padrões silábicos encontrados no Português que sirvam para alimentar estudos comparativos desta (ou de outra) natureza, impôs-se fazer o levantamento sistemático da ocorrência dos tipos silábicos na língua.

Para tanto, utilizou-se a **base de dados escrita** empregada por Albano, Moreira, Aquino, Silva and Kakinohana (1995), gentilmente cedida ao PROAEP e foi criado um programa de computador, intitulado **PROAEP**, especialmente para o PROAEP, para se proceder à contagem dos padrões silábicos encontrados no corpus delimitado.

A amostra analisada é constituída pelo conjunto de 25.582 verbetes dos (27,074) constantes no *Minidicionário Aurélio*, (Ferreira, 1977). A redução deveu-se ao fato de o programa elaborado ter ignora-

do enunciados de formato fonológico idêntico, mesmo quando referindo-se a itens lexicais distintos, como por exemplo *creme* – substantivo masculino e *creme* – adjetivo; *melhor* – adjetivo, *melhor* – substantivo masculino e *melhor* – advérbio; *banda* – (s.f.) parte lateral e *banda* – (s.f.) grupo, facção. Foram ignorados ao todo 1492 itens.

Os seguintes padrões foram estabelecidos como parâmetros para a contagem: **V, VC, VS, CV, SV, CVC, CVS, CCV, CSV, CVCC, CVSC, CCVC, CCVS, CCSV, CSVC, CSVS, CCVCC, VSC, CCVSC**, onde **C** representa uma consoante, **V** uma vogal e **S** uma semivogal.

Foram computadas como **C precedendo V** todos os 19 elementos distintivos que ocorrem na cabeça da sílaba na língua: / **p b t d k g f v s z ʒ ʃ m n ɲ r l ʎ** /.

Foram computadas como **C seguindo V** os elementos consonantais que ocorrem na parte terminal da sílaba na língua: / **N S R L** /.

Foram computadas como **S**, as semivogais / **j w** /, que ocorrem como transições de *glide* tanto antes como depois de **C**.

Assim sendo, as sílabas de estrutura silábica (**C**)**VCC** indicam, quase sempre, os casos em que: o **primeiro C** da coda é o travamento nasal, e o **segundo C** é o elemento fricativo final, como, por exemplo MÃOS [mẽws] /mawɲʃ/, mas podem também incluir os casos, menos frequentes, de /**R**/ seguido de /**S**/, como em PERS – PECTIVA [pɛxs] /pɛRS/.

Além disso, alguns critérios, de base estritamente fonética, foram adotados para a análise dos padrões, e conseqüentemente, determinaram a forma utilizada para a separação das sílabas do corpus constituído. É importante ressaltar que este critérios deveram-se à necessidade de comparar os padrões reais de pronúncia da fala adulta às produções infantis.

1. Nas palavras com encontros consonantais “impróprios”, foi inserida uma vogal epentética, para representar a pronúncia mais recorrente no Português Brasileiro e indicar que, na verdade, foneticamente, não existe, nestes casos, encontro.

Ex. **ABNEGAR** [abine'gax] V.CV.CV.CVC
PSICOLOGIA [pɪsikolo'ʒjɛ] CV.CV.CV.CV.CSV

2. Palavras como oxigênio, fixo, foi inserida uma sílaba para representar a pronúncia recorrente no Português Brasileiro.

Ex. **OXIGÊNIO** [ɔkisi'geniw] V.CV.CV.CV.CVS
FIXO ['fikisu] CV.CV.CV

3. Em outras como tóxico, toxicologia, não foi inserido nenhum segmento devido à variação na pronúncia.

Ex. **TÓXICO** ['tɔxiku] CV.CV.CV ou
['tɔkɪsiku] CV.CV.CV.CV.
TOXICOLOGIA [toxi'kɔlɔ'gja] CV. CV. CV.
CV.CSV ou
[tɔkɪsɪkɔlɔ'gja] CV. CV. CV.
CV.CV.CSV.

Discussão dos resultados

Dos 25.582 verbetes, 23.360 já foram processados, equivalendo a 91,31% do total, faltando computar apenas as letras L e S.

PADRÕES	%
V	6,84%
VC	5,68%
VS	0,3%
CV	55,9%
SV	0,006%
CVC	19,1%
CVS	2,3%
CCV	4,5%
CSV	1,98%
CVCC	0,1%
CVSC	1,95%
CCVC	0,9%
CCVS	0,1%
CCSV	0,05%
CSVS	0,09%
CSVC	0,1%
CCVCC	0,09%
VSC	0,1%
CCVSC	0,03%

Os resultados deste estudo corroboram os achados anteriores (Teixeira, 1997) confirmando ser CV o

padrão mais recorrente na língua, seguido por (C)VC, CCV e V, e (C)VCC.

Referências Bibliográficas

DAVIS, B., MACNEILAGE, P., GILDERSLEEVE-NEWMANN & TEIXEIRA, E.R (1999) Cross-language studies of consonant-vowel co-occurrence constraints in infants and adults: Ambinet language effects in first words. In Resumos do *Twentieth Annual Child Phonology Conference*. University of Wales, Bangor, País de Gales, 18.

MADDIESON, I. (1984) *Patterns of Sounds*. Cambridge: Cambridge Univ. Press.

TEIXEIRA, E.R.(1999) A Adaptação dos Inventários MacArthur de Desenvolvimento Comunicativo (CDI's) para o Português Brasileiro. In *Cadernos de Resumos do II Congresso Nacional da ABRALIN*. Universidade federal de Santa Catarina, 192.

_____ & DAVIS, B. (1999) Phonetic patterns and Ambient Language influences in the speech acquisition of two Brazilian Portuguese Speakers. In Resumos do *VIII International Congress for the Study of Child Language*. Universidade do País Basco, San Sebastian, Espanha, 215.