

# ASPA – UMA FERRAMENTA DE APOIO AOS ESTUDOS DA SONORIDADE SEGMENTAL

**Autores: Thaís Cristófaró Silva (UFMG, CNPq) – thaicristofaro@gmail.com**  
**Gustavo Mendonça (UFMG) – gustavoama@gmail.com**  
**Leonardo Almeida (UFMG, CNPq) – sion701@gmail.com**  
**Eduardo Gonçalves (UFMG) – goncalves.ef@gmail.com**

## OBJETIVO

Este trabalho apresenta o desenvolvimento de uma ferramenta de busca para a investigação de padrões sonoros do português brasileiro: o Buscador do ASPA. Apresentam-se, no artigo, informações sobre a primeira formulação do Buscador do ASPA, bem como modificações que foram feitas no intuito de aprimorar tal ferramenta. Indicam-se as categorias que foram utilizadas no banco de dados e as buscas que são possíveis de serem feitas a partir dessas categorias.

## A PRIMEIRA FORMULAÇÃO DO BANCO DE DADOS DO ASPA

- Dados oriundos de uma lista de palavra fornecida, em 2004, pelo Projeto DIRECT-PUC-SP, contendo: 199.864 palavras (ou tipos) e 228.766.402 ocorrências (ou tokens).
- Transcrição sonora obtida por meio do conversor ortográfico-fonético LETRASOM [Almeida 2005].
- Somente palavras com frequência maior ou igual a 6 foram consideradas.
- Siglas (FHC, EMBRAPA), palavras com pronúncia instável ou não-inferível (Bunchen, Husa), erros ortográficos (mamobra), palavras estrangeiras (future, salati) foram desconsideradas e armazenadas em tabela à parte no Banco de Dados.
- Dados finais utilizados pelo Buscador do ASPA que foram transcritas pelo LETRASOM totalizam: 124.248 palavras e 209.836.004 ocorrências.
- As buscas eram realizadas via MySQL e Matlab, mediante solicitação feita por e-mail aos organizadores do Projeto ASPA.
- Codificação dos segmentos sonoros feita por meio de um código de 4 letras, exemplos:

Segmento	Cód. 4 letras	Categoria por extenso
p	COBD	Consoante Oclusiva Bilabial Desvozeada
s	CFAD	Consoante Fricativa Alveolar Desvozeada
m	CNBV	Consoante Nasal Bilabial Vozeada
e	VMBA	Vogal Média-Baixa Anterior
ê	VMEN	Vogal Média E Nasal
E	VMNA	Vogal Média Neutralizada Anterior

## AS MODIFICAÇÕES FEITAS NO BANCO DE DADOS

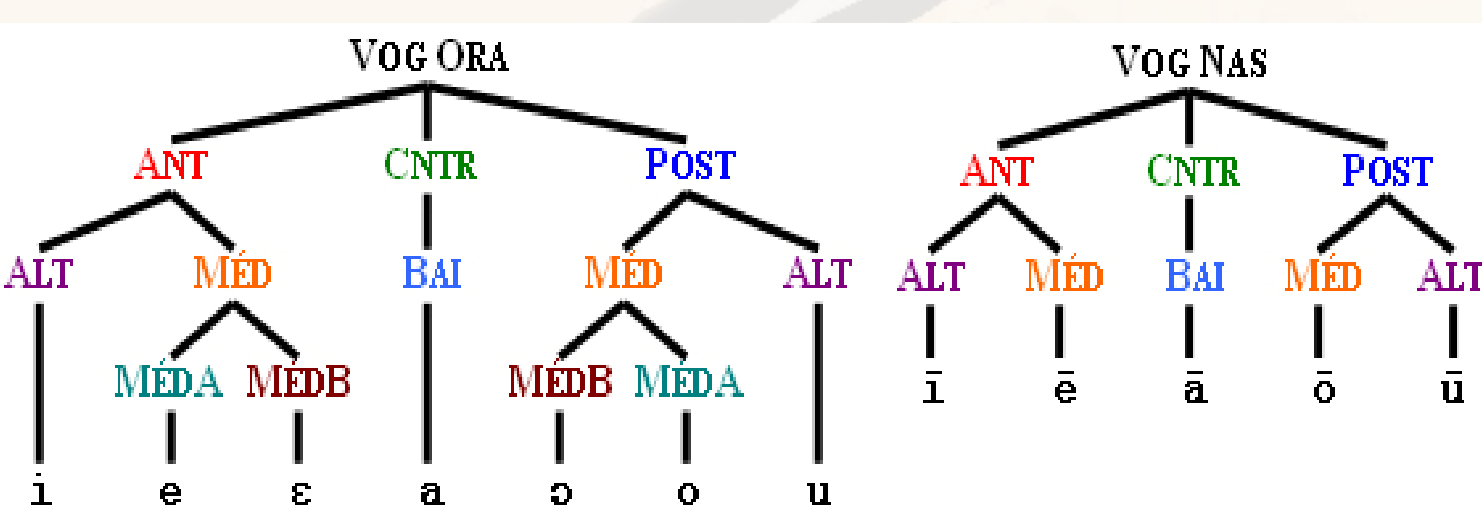
Mudança de um código de 4 letras para um código de 6 letras para classificar os segmentos sonoros, exemplos:

Classe principal	Cavidade ressonante	Tipo de vogal	Posição da língua		
			Dimensão horizontal	Dimensão vertical	
Vogal	Oral Nasal	Plena Glide Epentética Neutralizada	Anterior	Alta	Média-alta
			Central	Média	Média-baixa
			Posterior	Baixa	

A utilização de um código de 6 letras aumenta a possibilidade de classificação dos segmentos sonoros. É possível realizar buscas a partir de um maior número de categorias as quais podem ser visualizadas a partir de diagramas arbóreos, similares aos elaborados por Clements [1995, 2001]:

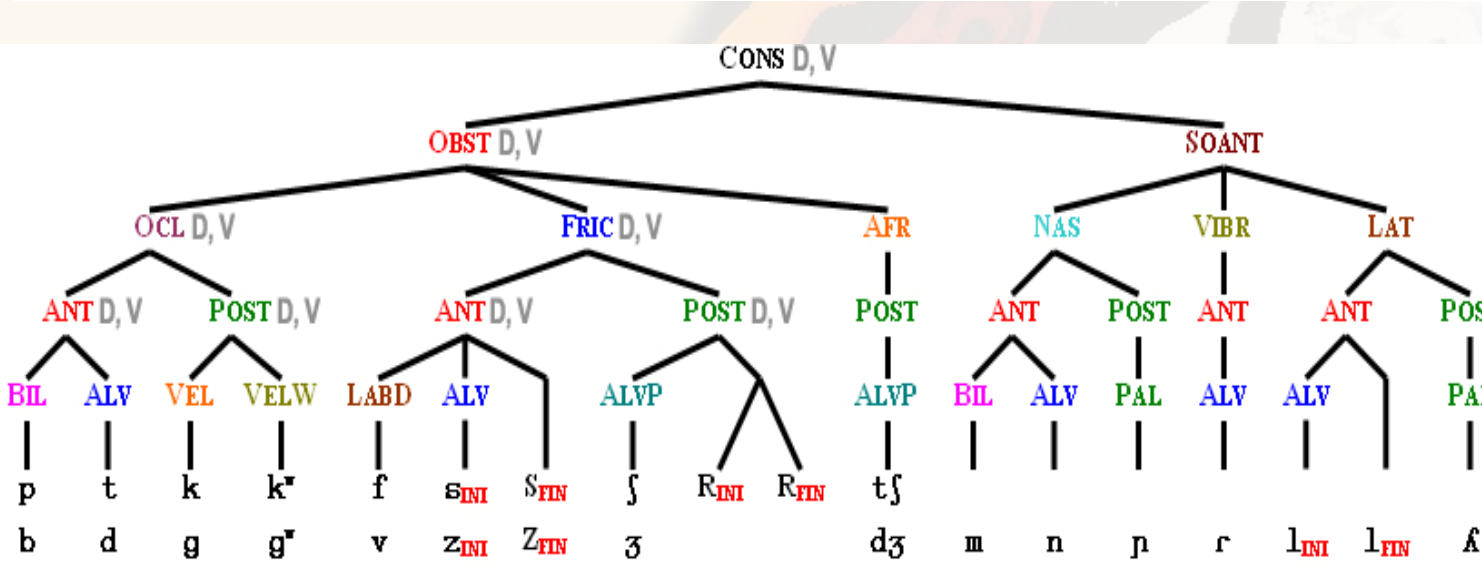
Vogais:

Classe principal	Cavidade ressonante	Tipo de vogal	Posição da língua		
			Dimensão horizontal	Dimensão vertical	
Vogal	Oral Nasal	Plena Glide Epentética Neutralizada	Anterior	Alta	Média-alta
			Central	Média	Média-baixa
			Posterior	Baixa	



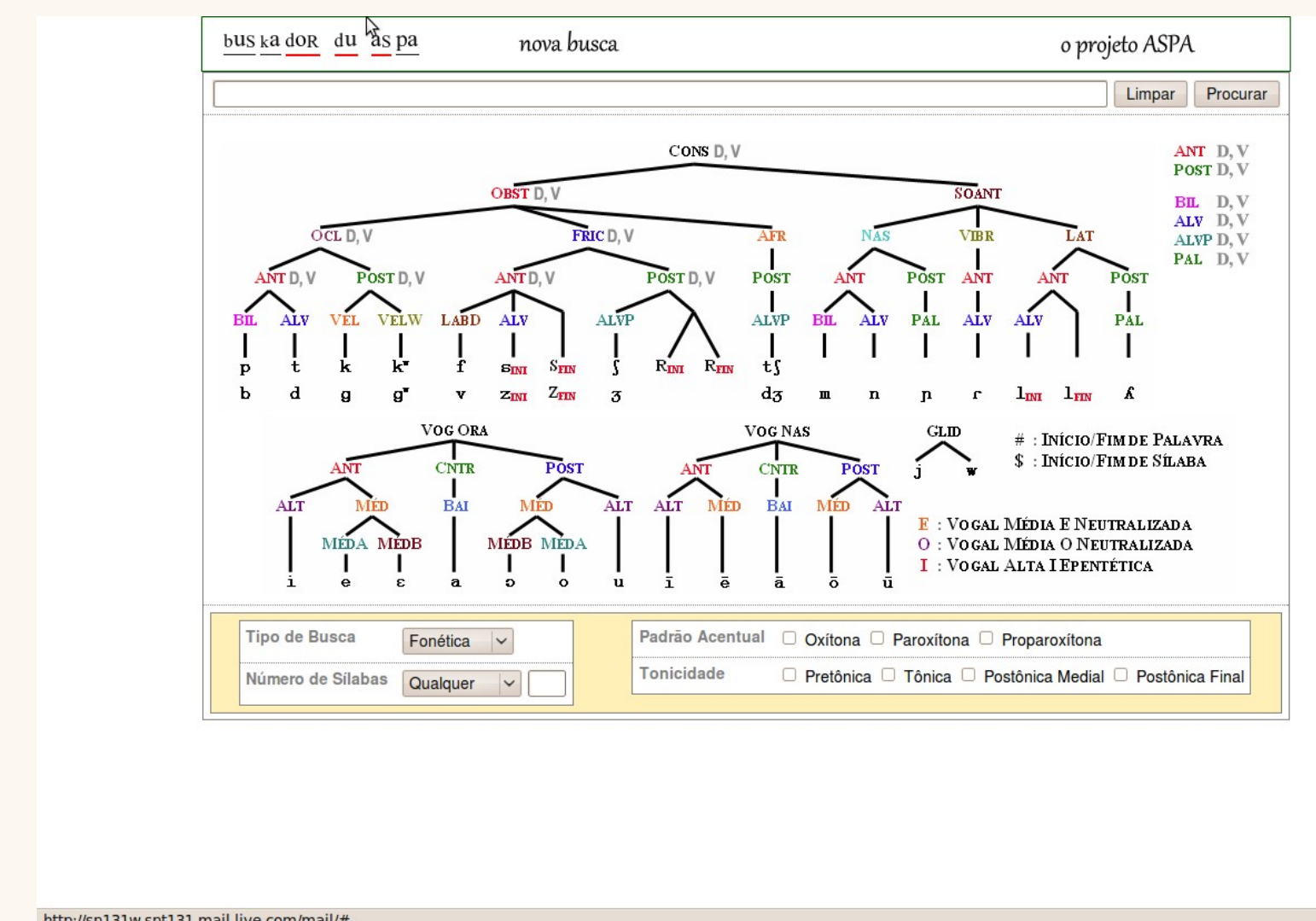
Consoantes:

Classe Principal	Propriedade de classe	Modo de articulação	Lugar de articulação	Grav de vozeamento
Consoante	Obstruente Soante	Oclusivo Fricativo Africado Nasal Vibrante Lateral	Bilabial Labio-dental Alveolar Alveopalatal Palatal Velar Velar-labial.	Vozeado Desvozeado



- **Adição de informação sobre o padrão acentual de cada palavra do banco de dados.** Um algoritmo desenvolvido em PHP determinou automaticamente, a partir do nº de sílabas pós-tônicas, o padrão acentual das palavras.
- **Adição de informação sobre o número de sílabas de cada palavra.** Um algoritmo em PHP percorreu as colunas de cada sílaba das palavras, realizando sua contagem.
- **Elaboração de interface gráfica para o Buscador do ASPA,** que facilita o acesso e a disponibilização do banco de dados. A interface gráfica é de fácil acesso e não demanda conhecimento de SQL. A seguir, há exemplos dessa interface:

Exemplo da página de busca:



Exemplo da página de resultado:

## SOBRE AS BUSCAS QUE PODEM SER REALIZADAS NO BUSCADOR DO ASPA

O estado atual do Buscador do ASPA permite a realização de buscas a partir dos seguintes filtros, os quais podem ser conjugados para restringir ainda mais os resultados:

- \* busca pela forma ortográfica;
- \* busca pela forma fonética;
- \* busca pelo número de sílabas;
- \* busca pelo padrão acentual;
- \* busca pela tonicidade da sílaba.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Bybee, J. (2001) *Phonology and Language Use*. Cambridge, CUP.
- Cristófaró-Silva, T.; Almeida, L.; Fraga, T. (2005) "ASPA: a formulação de um banco de dados de referência da estrutura sonora do português contemporâneo". In: *Anais do XXV Congresso da Sociedade Brasileira de Computação*. São Leopoldo, Sociedade Brasileira de Computação. v. 1. p. 2268-2277.
- Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge, CUP.
- Sardinha, T. B. (2004) *Linguística de Corpus*. Barueri, Manole.