



UFMG

Universidade Federal de Minas Gerais
Faculdade de Letras

Problemas metodológicos no cadastramento de dados do



Natália Fontes de Oliveira, (FALE/UFMG).
Jaqueline de Castro Oliveira, (FALE/UFMG).
Profª Drª Thaís Cristóvão Silva, (FALE/UFMG).

INTRODUÇÃO

Este trabalho discute alguns problemas metodológicos no cadastramento de palavras em um banco de dados denominado ASPA - Avaliação Sonora do Português Atual. Tal projeto, ainda em fase de desenvolvimento, é coordenado pela professora Thaís Cristóvão Silva (FALE-UFMG) e Leonardo Almeida (CEFALA-UFMG). O projeto ASPA reflete a sonoridade do português brasileiro contemporâneo e tem como objetivo maior desenvolver uma ferramenta de apoio às pesquisas de fonética e fonologia, oferecendo contribuições para a análise do mapeamento de tipos silábicos e segmentais do português brasileiro atual.

As palavras que alimentam o banco de dados do projeto ASPA foram originadas da lista de palavras do corpus de escrita do LAEL (PUC-SP). Um programa elaborado pelos coordenadores do projeto cadastrou as palavras oferecendo a conversão ortográfica → fonética e algumas opções a serem selecionadas pelos participantes do projeto. Durante o processo de cadastramento verificava-se se a transcrição fonética estava correta e quando pertinente faziam-se as correções ou a enviava para ser avaliada pelos coordenadores. Os estagiários adicionalmente deviam cadastrar uma determinada palavra em relação às seguintes categorias:

- Transcrição fonética (verificar) e tonicidade;
- Classe de palavra (substantivo, adjetivo, verbo, ..., numeral, etc.);
- Morfologia (flexão verbal, flexão plural, derivado e original);
- Origem (nenhuma, indígena, africana);

Nessa primeira fase foram observadas algumas restrições para se cadastrar um dado. Uma das limitações foi o fato de que o ASPA é alimentado por um banco de dados que contém uma lista de 200.000 palavras. A lista não oferece o contexto em que a palavra ocorre. Conseqüentemente, para algumas palavras houve dificuldade de análise por parte dos colaboradores do projeto. Alguns exemplos serão discutidos neste pôster.

OBJETIVO GERAL

Apresentar alguns problemas metodológicos observados durante o cadastramento de palavras em um banco de dados denominado ASPA - Avaliação Sonora do Português Atual

RESULTADOS E DISCUSSÃO

Dentre os fatores que apresentaram problemas no cadastramento temos :

❖ Conhecimento lingüístico do colaborador:

Algumas palavras a serem cadastradas poderiam ser cadastradas como mais de uma categoria gramatical. Por exemplo, "ilustrado", pode ser cadastrada como adjetivo ou verbo (particípio).

Observe no exemplo abaixo o acesso de cadastramento apresentado aos colaboradores:

imbuí	Dúvida Substantivo Verbo	Dúvida Flex Verbal Flex Plural	Nenhuma	i	5	b	u	4	i	1	
ilustrado	Dúvida Substantivo Verbo	Dúvida Flex Verbal Flex Plural Derivado	Nenhuma	i	5	l	u	4	t	r	a

No caso de "ilustrado" idealmente o/a colaborador/a deveria selecionar as duas opções: substantivo e verbo. Contudo, alguns colaboradores selecionavam apenas uma opção de cadastro. Essa diferença no cadastramento das palavras pode ter prejudicado o escopo do banco de dados como um todo. Isso porque, num processo de busca para palavras terminadas em "-ado" que sejam adjetivos, a palavra "ilustrado" não apareceria se tiver sido cadastrada somente como verbo. Na expectativa de suprir a lacuna gerada pelo conhecimento lingüístico do colaborador os coordenadores do projeto elaboraram um programa que tenta solucionar este tipo de problema oferecendo consistência ao material disponibilizado pelo ASPA.

❖ Regionalidade e pronúncia individual:

Diante da variação dialetal foram incorporados à análise os símbolos (E,O) que representam a neutralização das vogais médias [ɛ, e, ɔ, o] nas posições pretônicas e postônicas. Por exemplo c[O]mida e núm[E]ro ilustram o uso dos símbolos E,O. Observe os exemplos abaixo que apresentam E,O em sílaba tônica:

sede	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	a	e	1	d	i	3
sede	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	e	3	1	d	i	3
forma	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	f	o	r	1	m	a
forma	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	f	o	r	1	m	a

❖ Para palavras como "poça" que pode ser pronunciada com uma vogal aberta ou fechada, infelizmente o cadastro contou com a opção individual da forma de falar do colaborador.

▪ Epêntese:

opção	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	õ	5	p	i	4	sãw	1
tétraco	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	t	e	k	i	m	i	2

Encontros consonantais foram preservados somente quando duas consoantes formam encontros tautossilábicos tais como [pr, tr, kl, etc] e quando os arquifonemas [R, S, L] ocorrem em final de sílaba como em ca[R]ta, pa[S]ta, sa[L]to. Em outros encontros consonantais foi acrescentada uma vogal epentética [ɪ] como em t[ɪkɪ]nico.

▪ Alternância de acentuação:

zeroz	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	z	e	1	z	k	2	z
zeroz	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	z	e	2	z	k	1	z
recorde	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	h	e	1	k	2	d	3
recorde	Dúvida Substantivo Verbo	Dúvida Flex Plural Derivado Original	Nenhuma	h	e	4	k	2	d	3

A falta de contexto dentro do texto escrito original contribuiu para que uma forma fosse escolhida em detrimento da outra.

▪ Duas pronúncias potenciais:

a) Alternância de acento devido a categoria gramatical:

O acesso a uma palavra como "piloto" oriunda de uma lista de palavras descontextualizadas não oferecia ao colaborador a possibilidade de decidir se numa seqüência gráfica como "piloto" a vogal média seria aberta (verbo) ou fechada (substantivo). Nestes casos optou-se para cadastrar tais palavras com o símbolo O (que representa potencialmente tanto a vogal média fechada quanto a vogal média aberta).

b) Limitação em relação aos casos potenciais de ditongo-hiato:

[ma.i.ze.na] ou [maj.ze.na]?

Em casos de ditongo e hiato percebe-se que a pronúncia de uma palavra como "maisena" varia ora como ditongo[maj.ze.na] ora como hiato[ma.i.ze.na] dependendo do falante ou mesmo de estilos de fala diferentes. Assim, diante da possibilidade de variação, optou-se, nestes casos, pelo uso do hiato. O hiato foi também preferido e adotado como registro no banco de dados em palavras proparoxítonas com seqüência de vogais postônicas: sábia, farmácia, pátio, biênio, etc.

[sábja] ou [sábja]?

CONCLUSÃO

Dentre os aspectos metodológicos de cadastramento das palavras do projeto ASPA discutidos podemos destacar:

- ✓ A análise da sonoridade partiu de um corpus escrito
- ✓ As palavras que foram submetidas à análise estavam descontextualizadas
- ✓ Os processos de variação da fala acontecem mais rapidamente que na escrita