

Universidade do Algarve  
Universitat Autònoma de Barcelona

International Masters in  
NATURAL LANGUAGE PROCESSING  
&  
HUMAN LANGUAGE TECHNOLOGY

**DESCRIÇÃO E ANÁLISE FONÉTICA DO DIALETO DE BELÉM DO  
PARÁ ORIENTADAS À INCORPORAÇÃO DE CONHECIMENTO  
FONÉTICO À CONVERSÃO DE TEXTO EM FALA**

CARLA PATRICIA VALENZUELA CORREA



Faro, junho 2010

Universidade do Algarve  
Universitat Autònoma de Barcelona

International Masters in  
NATURAL LANGUAGE PROCESSING  
&  
HUMAN LANGUAGE TECHNOLOGY

**DESCRIÇÃO E ANÁLISE FONÉTICA DO DIALETO DE BELÉM DO  
PARÁ ORIENTADAS À INCORPORAÇÃO DE CONHECIMENTO  
FONÉTICO À CONVERSÃO DE TEXTO EM FALA**

CARLA PATRICIA VALENZUELA CORREA

Dissertação apresentada ao Curso de International Masters in Natural Language Processing and Human Language Technology da Universidade do Algarve e Universitat Autònoma de Barcelona como requisito parcial para a obtenção do grau de Mestre.

Orientado por: Dr. Joaquim Llisterra Boix

Co-orientado por: Dr. Jorge Baptista

Faro, junho 2010

This Project was supported by the European Commission, Education & Training, Erasmus Mundus: EMMC 2008-0083, Erasmus Mundus Masters in NLP & HLT programme.

*À Angela, meu porto seguro e maior  
fonte de inspiração.*

## AGRADECIMENTOS

Ao professor Dr. Joaquim Llisterri pela oportunidade de realizar este trabalho sob a sua direção, pela sua enorme paciência e gentileza e pelo entusiasmo com o que se prontificou desde os primeiros dias à orientação deste projeto.

Ao professor Dr. Jorge Baptista pelo esforço que fez sempre para que este trabalho fosse o melhor que podia ser.

A todos os professores da *Universitat Autònoma de Barcelona* e da *Universidade do Algarve* por ter me recebido e proporcionado o melhor ambiente de estudo possível.

A todos os queridos professores do LIP na Universidade de Brasília (UnB), por todo o apoio brindado durante a realização desta pesquisa. À minha querida amiga Theresa pelo apoio incondicional e à Carmen pela sua enorme ajuda com os informantes e com as gravações.

À querida Adriana por compartilhar comigo estes quase dois anos e ter me brindando a sua amizade e simpatia.

A todas as pessoas que me incentivaram e contribuíram de uma ou outra forma na realização deste trabalho, especialmente aos meus pais, esposa e amigos de Brasília pela presença constante na minha vida apesar da distância.

Finalmente agradeço ao consórcio organizador do programa IM in NLP&HLT e à Comissão Europeia pela bolsa de estudos Erasmus Mundus concedida para a realização deste projeto.

## **RESUMO**

Este trabalho descreve as características fonéticas do português falado na cidade de Belém do Pará, norte do Brasil, e pretende ser uma contribuição linguística para o desenvolvimento de tecnologias de conversão de texto em fala para esta variedade do português. Primeiramente, se apresenta uma revisão dos conceitos e métodos básicos envolvidos no processo de síntese da fala, posteriormente são descritas as características fonéticas que distinguem o dialeto de Belém dos outros dialetos do Português do Brasil e finalmente se apresentam os dados formalizados de maneira em que possam ser integrados ao módulo de conversão ortográfico-fonética de um sistema de conversão de texto em fala.

## **RESUMEN**

Este trabajo se centra en la descripción fonética de la variante del portugués hablado en la ciudad de Belém do Pará, en el norte de Brasil, y se orienta hacia el desarrollo de tecnologías de conversión de texto en habla para esta variedad del portugués. En primer lugar, se presenta una revisión de los conceptos y de los métodos básicos relacionados con el proceso de síntesis del habla, posteriormente se describen las características fonéticas que distinguen el dialecto de Belén de otros dialectos del portugués de Brasil y finalmente se presentan los datos formalizados de manera que puedan ser integrados al módulo de conversión ortográfico-fonética de un sistema de conversión de texto en habla.

# SUMÁRIO

AGRADECIMENTOS .....	ii
<b>Resumo</b> .....	iii
<b>Resumen</b> .....	iii
1 INTRODUÇÃO .....	1
1.1 Objetivos.....	1
1.2 Pressupostos.....	2
1.3 Motivação .....	3
1.4 A Pesquisa Fonética e a Síntese da Fala no Brasil .....	5
1.4.1 A Fala de Belém .....	7
1.5 Estrutura do Trabalho.....	8
2 CONTEXTUALIZAÇÃO TEÓRICA .....	10
2.1 A Síntese da Fala.....	10
2.1.1 Funcionamento de um Sistema de Síntese .....	13
2.1.2 A Síntese de Alto Nível.....	16
2.1.3 Transcrição Fonética Automática para o Português Brasileiro .....	18
2.1.4 A Síntese de Baixo Nível.....	20
2.1.4.1 Síntese a Partir de Corpus (ou Síntese por Seleção de Unidades).....	21
2.1.4.2 Síntese por Concatenação de Unidades.....	21
2.1.4.3 Síntese por Regras (Paramétrica).....	25
2.1.4.4 Síntese Articulatória .....	26
2.1.5 Técnicas de Geração do Sinal Acústico.....	27
2.1.5.1 Síntesis por Formantes .....	28
2.1.5.2 A Técnica PSOLA.....	29
2.2 Descrição Fonética do Português Brasileiro .....	29
2.2.1 Descrições do Dialeto de Belém do Pará.....	30
3 METODOLOGIA .....	33
3.1 Localização, período da pesquisa e informantes.....	33
3.2 Métodos, Instrumentos e Técnicas de Coleta de Dados .....	35
3.3 O Corpus .....	35
3.4 Técnicas de análise.....	37
3.4.1 Análise da Transcrição .....	37
3.4.2 A Análise Acústica .....	38
3.4.3 As Regras.....	40
4 RESULTADOS DA ANÁLISE .....	42
4.1 As Vogais .....	42
4.1.4 O grafema <a> .....	46

4.1.5	O Grafema <e>.....	49
4.1.6	O grafema <o> .....	51
4.2	As Consoantes .....	52
4.2.1	<d> e <t> antes de [i].....	53
4.2.2	<l> seguido de [i] .....	54
4.2.3	<n> seguido de [i].....	55
4.2.4	<s> em Coda .....	56
5	FORMALIZAÇÃO DOS RESULTADOS .....	58
5.1	Conclusões Desprendidas da Análise.....	58
5.2	A Incorporação do conhecimento fonético a um sistema de CTF .....	60
5.3	Avaliação .....	63
6	CONCLUSÃO .....	65
7	REFERÊNCIAS.....	66
8	ANEXO I – ALFABETO FONÉTICO INTERNACIONAL.....	72

## LISTA DE FIGURAS

Figura 1 - Diagrama simplificado de um sistema de conversão de texto em fala. Adaptado de Dutoit, 1997, p.14 .....	14
Figura 3- Segmentação em Praat da produção da frase “cultura popular”. (Informante 3, Parte 1) .....	39
Figura 4 - Amostra de Palavras Preparadas para a Aplicação das Regras de Transcrição .....	41
Figura 5 – Dispersão do [a] Tônico .....	48
Figura 6 – Dispersão do [ẽ] Tônico e Pretônico.....	48

## LISTA DE TABELAS

Tabela 1 - Correspondência entre Grafemas e Fonemas .....	4
Tabela 2 - Correspondência entre Fonemas e Grafemas .....	4
Tabela 3- Grafema sem um Correspondente Sonoro.....	4
Tabela 4 – Valores obtidos para as Vogais Tônicas no dialeto de Belém.....	42
Tabela 5 - Desvio Padrão e Coeficiente de Variação dos Valores de F1 e F2 das Vogais .....	43
Tabela 6 – Valores obtidos para as vogais átonas no dialeto de Belém .....	44
Tabela 7 - As vogais do português. (Câmara Jr., 1994, p. 41) .....	45
Tabela 8 - Realizações de <a> .....	46
Tabela 9 – Realizações de <e>.....	49
Tabela 10 - Realizações de <o> .....	51
Tabela 11 - Consoantes do Português de Belém .....	53
Tabela 12 - Realizações de <t> .....	53
Tabela 13 - Realizações de <d> .....	54
Tabela 14 - Realizações de <l>.....	54

Tabela 15 - Realizações de <n> .....	56
Tabela 16 - Realizações de <s> em coda.....	57
Tabela 17 - Convenções de Anotação.....	61
Tabela 18 – Correspondência dos fones do dialeto de Belém em IPA e X-SAMPA .....	62
Tabela 19 - Regras para Algoritmo de Transcrição Ortográfico-Fonética .....	63
Tabela 20 - Resultados Obtidos na Avaliação .....	63

# 1 INTRODUÇÃO

## 1.1 OBJETIVOS

Este trabalho visa a descrever o falar da comunidade de Belém, capital do estado do Pará, situada no norte do Brasil; com base nos preceitos da fonética acústica, apresento os principais aspectos que diferenciam o dialeto da cidade de outros falares do Brasil, com ênfase nas características que são relevantes para as tecnologias da fala, isto é, que possam permitir o desenvolvimento de sistemas melhor adaptados à fala da comunidade paraense.

O objetivo principal da investigação é, então, descrever os fenômenos fonéticos que são específicos à variante da língua portuguesa falada na cidade de Belém do Pará de forma que esta descrição possa ser utilizada no desenvolvimento de futuras aplicações de CTF (conversão de texto em fala, também conhecida pela sigla inglesa TTS, *Text-to-Speech*) dirigidas à população paraense que envolvam conhecimento fonético prévio do falar desta região.

São objetivos específicos da pesquisa, então:

- descrever os traços fonéticos acústicos que diferenciam a variedade de PB falada na cidade de Belém.
- propor a forma em que o conhecimento fonético da variante de Português brasileiro falada em Belém pode se aplicar a um sistema de CTF.

Para atingir tais objetivos elaborou-se um corpus; realizou-se a gravação e colheita das amostras de fala mediante a entrevista de três informantes; transcreveram-se foneticamente e analisaram-se acusticamente as gravações.

Posteriormente, uma seleção dos dados obtidos a partir da análise do corpus gravado foi formalizada de maneira que estes dados possam ser utilizados na CTF tanto no inventário de alofones quanto no módulo de regras de transcrição fonética automática.

## 1.2 PRESSUPOSTOS

A presente pesquisa partiu do pressuposto de que o falar de Belém se diferencia substancialmente do falar do resto do território brasileiro, não apenas no léxico, mas também nos aspectos fonéticos, tais como a realização de determinados fonemas e a prosódia. Pressupõe-se também, que estas características diferenciais do dialeto de Belém são suficientemente significativas para justificar um tratamento diferente ao momento de desenvolver um conversor de texto em fala dirigido a esta comunidade a fim de obter um resultado melhor aceito pelo público-alvo. Pesquisas como a de Dahlbäck et al. (2007) evidenciam o efeito da utilização de diferentes dialetos de um mesmo idioma na confiança e aceitação demonstrada por usuários de serviços de informação com base no paradigma de similaridade-atração. Outros trabalhos, como o de Pucher et al. (2009), corroboram estes resultados não só para o uso do dialeto do público-alvo de um sistema de CTF, senão também para socioletos.

Partiu-se também do pressuposto de que mediante a incorporação de conhecimento fonético ao desenvolvimento de um conversor de texto em fala é possível melhorar consideravelmente a qualidade da fala sintética (Llisterri, 2007a; Llisterri, 2007b; Llisterri, Carbó, Machuca, Mota, Riera, & Ríos, 2003).

### 1.3 MOTIVAÇÃO

A fala é o nosso principal meio de comunicação, a forma mais natural de interagir com outros. As línguas naturais são primeiramente orais, porém, pela necessidade de transmitir informações sem a presença física do indivíduo receptor da informação, se faz necessária uma forma de representação gráfica delas.

O mecanismo da fala é o de combinar de unidades sonoras distintivas (capazes de criar distinção entre uma palavra e outra) para a formação de palavras. Estes sons que criam contraste entre palavras são chamados de fonemas e o seu número varia de uma língua para outra, mas é sempre um conjunto finito.

No caso do Português, assim como em todas as outras línguas românicas, a escrita representa a combinação de fonemas da língua por meio da combinação das letras do alfabeto. Em Português, como se sabe, os grafemas deste alfabeto (letras) e os fonemas não apresentam uma correspondência biunívoca, isto é, não existe uma correspondência absoluta entre cada fonema da língua e a letra do alfabeto que o representa. Um mesmo grafema pode representar mais de um som, como vemos na TABELA 1

Tabela 1 - Correspondência entre Grafemas e Fonemas

Grafema	Fonema	Exemplo
<c>	/k/	cansa, acalma
	/s/	cinza, acima
<s>	/s/	mensagem, aspas
	/z/	paisagem, asas

Da mesma forma, um fonema pode ser representado por mais de um grafema ou combinação deles, como se mostra na TABELA 2:

Tabela 2 - Correspondência entre Fonemas e Grafemas

Fonema	Grafema	Exemplo
/s/	<s>	sopa, sapato
	<c>	acima, acento
	<ss>	essa, assim
	<ç>	alça, açúcar
/k/	<c>	calo, acorda
	<qu>	quiabo, aquele

Há ainda o caso do grafema <h> em posição inicial de palavra que não têm um correspondente sonoro como se vê nos exemplos abaixo:

TABELA 3- Grafema sem um Correspondente Sonoro

Fonema	Grafema	Exemplo
∅	<h>	homem, hábil

Cada língua possui seus próprios fonemas e tem formas diferentes de representá-los, o que faz com que um mesmo grafema possa representar diferentes fonemas dependendo da língua.

Assim sendo, antes de dar início ao desenvolvimento de qualquer sistema de síntese da fala que aceite como entrada qualquer tipo de texto, é necessário um estudo cuidadoso da fonética da língua em questão, pois para produzir fala sintética é preciso entender de que forma a sequência de grafemas que conformam a escrita se relacionada com os determinados fonemas que formaram finalmente as palavras.

#### 1.4 A PESQUISA FONÉTICA E A SÍNTESE DA FALA NO BRASIL

No Brasil, os estudos linguísticos orientados às tecnologias da fala iniciaram-se na década dos 90' quando tanto investigadores da área de tecnologia quanto linguistas perceberam a necessidade existente da interação de especialistas de ambas as áreas para conseguir criar sistemas próprios para o Português brasileiro (doravante PB). Um dos primeiros centros criados no país para a investigação linguística experimental direcionada às aplicações tecnológicas e, até hoje o mais importante, foi o LAFAPE (Laboratório de Fonética e Psicolinguística) da Universidade de Campinas, que nasceu, segundo informações obtidas no *website*<sup>1</sup>, pela cooperação da Faculdade de Engenharia Elétrica e do Instituto de Estudos da Linguagem. O trabalho desenvolvido neste laboratório impulsionou a pesquisa linguística computacional no Brasil durante a última década do século passado.

O crescente interesse dos linguistas brasileiros pelos estudos direcionados às tecnologias da fala não se tem refletido em resultados significativos no referente a produtos comerciais. Há uma evidente falta de produtos adequados à realidade linguística brasileira, pois apesar da demanda

---

<sup>1</sup> <http://www.lafape.iel.unicamp.br/> [05/05/2009]

que estes produtos têm, ainda há poucos sistemas disponíveis, principalmente, sistemas comerciais de qualidade. Na subsecção 2.1 pode-se encontrar uma lista de alguns dos produtos disponíveis em PB.

As investigações feitas nos núcleos e centros criados para o desenvolvimento das tecnologias linguísticas têm, no geral, obtido bons resultados como, por exemplo, o projeto inicial do LAFAPE, que tinha o objetivo de capacitar profissionais brasileiros na área da linguística computacional e que culminou na criação do conversor de texto em fala Aiuruetê (Barbosa, et al., 1999) ou as experiências feitas no NILC (Núcleo Interinstitucional de Linguística Computacional<sup>2</sup>) que possui atualmente vários projetos em andamento. Na subsecção 2.1 se encontram outros exemplos de projetos desenvolvidos posteriormente para o PB, porém, a pesquisa na área ainda é muito escassa e quase sempre centrada em dialetos específicos da região centro-oeste o que representa uma restrição às possíveis aplicações destes estudos a produtos destinados ao resto do Brasil.

A diversidade linguística no Brasil é um fato inegável que vem sendo destacado por gramáticos e linguistas há muitos anos. Nascentes (1922), por exemplo, dividiu o Português brasileiro em seis diferentes subfalares: amazônico, nordestino, baiano, fluminense, mineiro e sulista. Hoje em dia sabemos que os dialetos do Português falados no Brasil são muitos e que a distinção entre eles se estende muito além das diferenças no léxico (um exemplo clássico é o da *macaxeira* que dependendo da região pode também ser chamada de *mandioca doce* ou *aipim*), sempre tão destacadas pelos

---

<sup>2</sup> <http://www.nilc.icmc.usp.br/nilc/index.html> [05/05/2009]

manuais de gramática e livros escolares (*Português de Maia* ou *Novas palavras* de Amaral et al., por citar alguns).

Há no Brasil numerosos projetos e pesquisas que visam a descrição dos dialetos brasileiros de várias regiões do país nos níveis fonético, fonológico, morfológico, sintático além do lexical (Brandão & Moraes, 1995; Aragão, 2005; Cristiani & Encarnação, 2006). Muitas destas pesquisas já foram concluídas e outras tantas estão em andamento e graças a elas já se conhecem muitas características da língua portuguesa falada em diversas áreas do país, porém a grande maioria destes estudos tem fins sociolinguísticos enquadrando-se nas teorias que procuram compreender a variação dialetal do Português de forma qualitativa e social, normalmente com objetivos educacionais, sem nenhum interesse no desenvolvimento de tecnologias.

#### 1.4.1 A FALA DE BELÉM

A cidade de Belém, localizada no estado do Pará no Norte do Brasil, com quase um milhão e meio de habitantes, é a cidade com a maior concentração populacional da região norte e uma das maiores do Brasil<sup>3</sup>. Já o estado do Pará possui a maior concentração populacional da região norte com sete milhões e meio de habitantes.

Devido às suas origens históricas, a fala da cidade de Belém se diferencia em vários aspectos do PB de outras regiões do país. O exemplo mais evidente desta diferença se encontra no léxico, já que no seu dia-a-dia o belenense

---

<sup>3</sup> Informação obtida no *website* do IBGE (Instituto Brasileiro de Geografia e Estatística), <http://www.ibge.gov.br> [17/02/2010]

utiliza ao redor de 7200 palavras de origem tupi (Dias & Dias, 2007). Porém, esta não é a única forma em que o dialeto de Belém se distingue de outros. Esta variante do Português, além de apresentar algumas estruturas sintáticas que não são utilizadas no resto do território brasileiro, possui características fonéticas muito particulares, por exemplo:

- A manutenção dos ditongos *ou* e *ei* que no resto dos dialetos do Brasil sofrem monotongação (Paiva, 1996; Silva F. , 1997).
- A ditongação em sílabas terminadas em *-es* sendo /e/ tônico, como em *três* ou *inglês*, pronunciados respectivamente [tr'ej] e [ɪ'glej], são fenômenos que merecem um estudo mais aprofundado.

O estudo destes fenômenos pode permitir a identificação de padrões ou regras que possam ser formalizadas de maneira que sejam úteis para sua inclusão, por exemplo, no módulo ortográfico-fonético de um sistema de síntese da fala (cf. subsecção 2.1.2).

## 1.5 ESTRUTURA DO TRABALHO

O capítulo 2 é dedicado à revisão bibliográfica e subdivide-se em duas partes:

- Uma descrição dos principais métodos utilizados atualmente na síntese da fala, mais especificamente na CTF, com uma descrição dos procedimentos envolvidos nesta forma de síntese assim como das principais técnicas de conversão ortográfica-fonética.

- A revisão dos principais trabalhos realizados sobre a fonética do Português do Brasil, principalmente da fala de Belém.

No capítulo 3, apresenta-se a metodologia adotada na investigação junto a uma descrição do *corpus* e esclarecimentos sobre os critérios utilizados para a escolha dos informantes e a obtenção das amostras de fala.

O capítulo 4 relata os resultados obtidos na análise dos dados coletados e descreve as características fonéticas distintivas da fala de Belém.

No capítulo 5 se apresentam as propostas extraídas a partir da análise para a incorporação dos dados obtidos em um sistema de CTF.

Finalmente o trabalho é encerrado com um capítulo de considerações finais que apresenta uma avaliação final do trabalho desenvolvido durante esta pesquisa.

## 2 CONTEXTUALIZAÇÃO TEÓRICA

### 2.1 A SÍNTESE DA FALA

De acordo com Barbosa (1999), as primeiras aproximações à síntese da fala foram feitas ainda no século XVIII quando Kratzestein e, posteriormente, Von Kempelen apresentaram ao mundo os seus “instrumentos falantes”, mas foi só na segunda metade do século XX, na década dos 80, com o advento da informática e com o avanço da tecnologia em geral, que o homem foi realmente capaz de criar métodos para gerar fala sintética apoiando-se nos sistemas computacionais existentes e criar som a partir de programas de computador ou utilizando circuitos digitais especiais.

Do surgimento dos primeiros sintetizadores acadêmicos e comerciais (por exemplo, *Kurzweil Reading Machine for the Blind*, 1976; *SoftVoice TTS*, 1979) até hoje, tem havido um grande avanço tanto na teoria que fundamenta esta área quanto nos produtos finais. (Klatt D. H., 1987).

Nos últimos 20 anos as tecnologias da fala, mais especificamente a síntese, têm alcançado resultados de altíssima qualidade em línguas como o inglês, por exemplo. Produtos como *Natural Voices*, da AT&T possuem uma alta inteligibilidade e naturalidade nesta língua sendo aplicados principalmente no campo das telecomunicações e multimídia. Entretanto, línguas como o PB ainda não têm produtos suficientes para fazer realmente parte da vida diária do cidadão brasileiro comum.

Os estudos direcionados à síntese da fala no Brasil iniciaram-se na Escola Politécnica da Universidade de São Paulo, com a pesquisa

desenvolvida por Campos (1980) que visava criar um sistema de síntese para Português que aceitasse como entrada transcrições fonéticas. Alguns anos depois Esquivel (1985) apresentou um sistema de síntese a partir de texto que possuía um módulo que adicionava algumas informações fonéticas para a correta pronúncia das palavras.

Mais tarde, a Universidade de Campinas (UNICAMP) começou os estudos na área quando Egashira (1992) apresentou um trabalho sobre síntese da fala sem restrições para o texto de entrada.

Posteriormente, em 1993, a Universidade Federal do Rio de Janeiro desenvolveu um produto chamado Dosvox (Borges J. A., 1996).

Podemos dividir os produtos criados para a língua portuguesa do Brasil entre aqueles desenvolvidos no Brasil e os desenvolvidos fora do país. Entre os primeiros podem ser citados:

- Dosvox (1993), segundo informações obtidas no *website*<sup>4</sup> do projeto, é um sistema operacional para microcomputadores que possui um conjunto de programas especialmente criados para atender as necessidades dos deficientes visuais. O pacote Dosvox inclui editor de texto, calculadora, navegador de internet e é capaz de sintetizar fala a partir de qualquer texto.
- DeltaTalk<sup>5</sup> (1998) é um *software* para a CTF desenvolvido pela MicroPower, uma empresa com sede em São Paulo . É um dos poucos

---

<sup>4</sup> <http://intervox.nce.ufrj.br/dosvox/> [05/05/2009]

<sup>5</sup> <http://www.micropower.com.br/v3/pt/tradutores/sintetizadores/delta/index.asp> [05/05/2009]

sistemas comerciais em PB disponíveis para o usuário final e de fácil aquisição por meio da página web da empresa.

- Aiuruetê (Barbosa, et al., 1999) é um sistema de síntese criado especificamente para o PB, desenvolvido pelo LAFAPE da Universidade Federal de Campinas apenas com finalidades acadêmicas, não estando comercialmente disponível. Este sistema alia a concatenação de unidades à técnica PSOLA e à híbrida. Pode ser utilizado apenas no ambiente Windows.
- Orator<sup>6</sup> (2005) é um sistema desenvolvido pelo grupo LINSE da Universidade de Santa Catarina que se baseia na seleção de unidades de tamanho variável para realizar a síntese. As unidades que este sistema seleciona podem ser fonemas, sílabas, palavras, sintagmas e até frases curtas
- Liane TTS<sup>7</sup> (2010) é um conversor para o PB que tem como objetivo auxiliar os deficientes visuais nas sua interação. O sintetizador foi desenvolvido pelo SERPRO (Serviço Federal de Processamento de Dados, empresa pública, vinculada ao Ministério da Fazenda) em parceria com a UFRJ (Universidade Federal do Rio de Janeiro) e está atualmente em fase final de teste, foi apresentado ao público em setembro de 2009 e será lançado ainda em 2010.

Entre aqueles que foram desenvolvidos para o PB fora do Brasil se encontram:

---

<sup>6</sup> <https://www.linse.ufsc.br/> [07/04/2010]

<sup>7</sup> <http://www.serpro.gov.br/noticiasSERPRO/2009/setembro/treinamento-apresenta-sintetizador-de-voz-construido-pelo-serpro-e-pela-ufrj> [17/02/2010]

- Aculab Prosody TTS<sup>8</sup> é um sistema de CTF desenvolvido pela Aculab<sup>9</sup> que está disponível em vários idiomas, incluindo o PB, e se baseia na tecnologia concatenativa. Pode ser utilizado nos sistemas Windows 2000, Linux e Solaris.
- Real Speak<sup>10</sup> é um conversor de texto em fala para vários idiomas, incluindo o PB, desenvolvido pela Nuance Communications. Pode ser utilizado nos sistemas Windows, Linux e Solaris.
- LoquendoTTS<sup>11</sup> é um sistema comercial de conversão de em fala multilíngue e multilocutor baseado na concatenação de segmentos acústicos contextuais não uniformes diretamente extraídos da fala natural. (Quazza, et al., 2001, *apud* Llisterri, et al., 2004).
- Acapela<sup>12</sup> é um sistema comercial de CTF com vozes em vários idiomas, incluindo PB, que e se baseia na tecnologia concatenativa. Pode ser utilizado nos sistemas Windows.

### 2.1.1 FUNCIONAMENTO DE UM SISTEMA DE SÍNTESE

Um conversor de texto em fala é um sistema que produz fala sintética a partir de um texto escrito na forma ortográfica usual. Logicamente, o objetivo

---

<sup>8</sup> [http://www.mcct.com/aculab\\_text.html](http://www.mcct.com/aculab_text.html) [05/05/2009]

<sup>9</sup> <http://www.aculab.com/> [05/05/2009]

<sup>10</sup> <http://www.nuance.com/realspeak/> [05/05/2009]

<sup>11</sup> <http://www.loquendo.com> [05/05/2009]

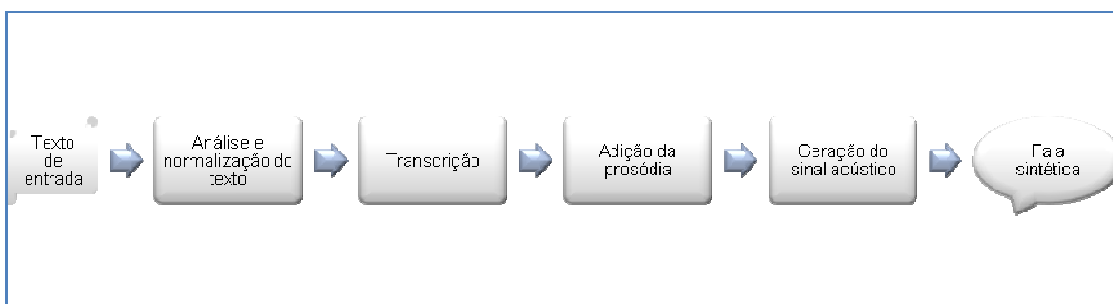
<sup>12</sup> <http://www.acapela-group.com/portuguese-brazil-46-text-to-voice.html> [25/05/2009]

final de um sistema deste tipo é ser capaz de produzir fala a partir de qualquer tipo de texto.

Intuitively, the ultimate goal of a text-to-speech (TTS) synthesizer is to read *any* text, whether it was directly introduced in the computer by an operator or scanned and submitted to an optical character recognition (OCR) system. Reading should be intelligible and natural. (Dutoit, An Introduction to Text-to-Speech Synthesis, 1997, p. 13)

Os sistemas de conversão de texto em fala já foram descritos em diferentes níveis de profundidade e detalhe por diversos autores (Dutoit, An Introduction to Text-to-Speech Synthesis, 1997; Sagisaka, 1997; Lemmetty, 1999; Llisterri, 2001; Bonafonte, Escudero, & Riera, 2006; Taylor, 2009). Apresento a seguir uma descrição básica e simplificada do seu funcionamento.

A FIGURA 1 representa de forma simplificada a estrutura básica de um sistema de CTF.



**FIGURA 1 - Diagrama simplificado de um sistema de conversão de texto em fala. Adaptado de Dutoit, 1997, p.14**

A síntese da fala é realizada num sistema em duas fases (Dutoit, An Introduction to Text-to-Speech Synthesis, 1997; Lemmetty, 1999):

A primeira, muitas vezes chamada de síntese de alto nível, é a que depois da análise do texto gera a transcrição fonética, isto é, de acordo com as regras

ortográficas da língua em questão, se transformam as letras, números e símbolos que constituem o texto numa série de símbolos que representam fones e alofones.

Depois disso, é preciso adicionar as informações de duração, melodia e pausas. O que se faz no módulo chamado de processador prosódico que irá distribuir essas informações entre os fones de cada trecho.

Este passo é o que irá conferir à fala sintética a naturalidade necessária para que a fala seja tanto interpretável como agradável aos ouvidos dos usuários. A prosódia dá informação sobre a estrutura do enunciado, ajuda a dividi-lo em blocos menores para a melhor compreensão, além de ajudar a resolver ambiguidades e identificar informação nova ou mais relevante num texto (Braga, Freitas, & Ferreira, *Processamento Linguístico Aplicado à Síntese da Fala*, 2003). Um sistema de conversão de texto em fala que pretenda reproduzir todo tipo de texto, não pode prescindir de adicionar informação prosódica à síntese.

Na segunda fase, conhecida como síntese de baixo nível, se gera o sinal acústico (a fala sintética) a partir da representação fonética anterior. Esta segunda parte envolve a utilização de modelos matemáticos, algoritmos de processamento digital de sinais (Dutoit, *An Introduction to Text-to-Speech Synthesis*, 1997, p. 14).

É na primeira fase descrita, no bloco de pré-processamento que o conhecimento fonético da língua pode ser incorporado e melhor aproveitado, principalmente no módulo de transcrição e no módulo prosódico, no qual o

conhecimento do funcionamento da língua pode ajudar a resolver os casos nos quais os métodos estatísticos ou de extração automática de padrões não tem sucesso.

### 2.1.2 A SÍNTESE DE ALTO NÍVEL

Antes de transformar um texto na sua representação fonética é necessário normalizá-lo, isto é, transformar todos os números, símbolos, abreviações, etc. na sua forma extensa para que o sistema seja capaz de lê-las posteriormente. Este pré-processamento pode parecer simples, porém, para um sistema que não é capaz de compreender o distinguir entre números, abreviações ou acrônimos, esta tarefa fica bem mais complicada. (Braga, Freitas, & Ferreira, *Processamento Linguístico Aplicado à Síntese da Fala*, 2003)

Em português, o processo de transformar números em palavras envolve algum grau de análise gramatical do texto. Um exemplo disto é quando se está diante de números que devam concordar em gênero com o nome que acompanham, como na frase *2 (duas) pessoas vieram ontem e 2 (dois) clientes vieram ontem*, a mesma diferença existe entre *1/2 xícara de açúcar* que deve ser transformada em *meia xícara de açúcar* enquanto a frase *1/2 litro de leite* deve ser transformada em *meio litro de leite*.

O bloco de pré-processamento para o Português deve incluir também algum grau de análise morfológica (mesmo que muito rudimentar) para não gerar problemas de concordância de gênero neste primeiro passo.

É preciso também resolver a questão da forma em que os números são expressos por extenso dependendo do tipo de número que se trata: por exemplo, quando falamos em R\$290.719,98 lemos essa cifra como *duzentos e noventa mil setecentos e dezenove reais e noventa e oito centavos*, já quando nos referimos a 29/07/1998 lemos *vinte nove de julho de mil novecentos e noventa e oito*, mas quando lemos 2907-1998 provavelmente diremos *dois, nove, zero, sete, um, nove, nove, oito* ou *vinte nove, zero sete, dezenove, noventa e oito*. Para escolher a forma certa de ler um número o pré-processador precisa ter informação sobre como interpretar símbolos como “R\$” “/” “-“ “,” (vírgula) entre números.

Isto feito, vem a etapa na qual se converte o texto à sequência de fones que o representam. Isto é feito pelo módulo chamado de conversor ortográfico-fonético. Este processo é essencial para que a fala sintética tenha uma boa qualidade, pois qualquer erro na representação fonética se traduzirá numa pronúncia incorreta.

Existem vários métodos para implementar o módulo de transcrição num sistema de CTF, algumas das formas já propostas para o inglês são: árvores de decisão (Lucassen, 1983), abordagens por redes neuronais (Sejnowski & Rosenberg, 1987), árvores de decisão treinadas automaticamente (Black, Lenzo, & Pagel, 1988), abordagens baseadas em dicionários (Coker, Church, & Liberman, 1990), modelos de Tabela look-up (Bosch & Daelemans, 1993), abordagens baseadas em regras linguísticas (Kaplan & Kay, 1994), modelos híbridos (Meng, Seneff, & Zue, 1994), por máquinas de estados finitos (Roche & Schabes, 1995), modelos estatísticos (Chotimongkol & Black, 2000), por cadeias escondidas de Markov (Taylor P. , 2005).

Estas abordagens, como já foi dito, foram propostas para o inglês, porém podem ser adaptadas para outras línguas de acordo às necessidades e características de cada uma.

### 2.1.3 TRANSCRIÇÃO FONÉTICA AUTOMÁTICA PARA O PORTUGUÊS BRASILEIRO

Como mencionado na subsecção 1.3, a relação que existe em PB entre grafemas e fones não é biunívoca, isto é, não há uma correspondência entre as letras que se escrevem e a forma em se pronunciam. Mesmo assim é possível estabelecer algumas regras de pronúncia para a maioria dos grafemas, por exemplo:

<c> → [s] /\_e/i (<c> se pronuncia [s] diante de <e> ou <i>)

<c> → [k] /\_a/o/u (<c> se pronuncia [k] diante de <a>, <o> ou <u>)

Em consequência disto, a conversão ortográfica fonética em Português não apresenta tanta dificuldade como em outras línguas (francês, inglês, etc.). (Braga & Resende, 2006)

Os problemas se apresentam em grafemas que podem representar vários fonemas e cuja regra não é clara, como é o caso do <x> e o ainda mais complicado caso de <e> e <o>, que podem representa tanto as vogais médias abertas /ɛ/ e /ɔ/ quanto médias fechadas /e/ e /o/.

Este tipo de grafema é considerado problemático para a síntese da fala, pois ao momento de realizar a transcrição fonética a partir de um texto, o módulo de transcrição fonético-ortográfico deve escolher o fonema correto para

ser reproduzido posteriormente, se neste processo se escolhe um fonema errado, a palavra não será produzida corretamente pelo sintetizador.

Palavras como *esc/ɔ/la*, *sac/ɔ/la*, *esm/ɔ/la*, *ceb/o/la*, *amp/o/la* representam um desafio para um sistema de síntese. Pela dificuldade que existe em saber, a partir de um texto escrito, se um determinado grafema representa uma vogal aberta ou fechada, muitas vezes se opta por adicionar este tipo de palavras a um dicionário.

Esta técnica por dicionário é bastante aplicada às línguas que não apresentam uma correspondência grafema-fonema biunívoca, como é o caso do Inglês ou do Francês e poderia ser uma forma de solucionar o problema que representa reconhecer quando uma vogal média é pronunciada aberta ou fechada, porém método não é o melhor já que em primeiro lugar, o dicionário teria que ser bastante extenso para conseguir dar conta de todas as palavras da língua. Além de que registrar a transcrição de todas as palavras mais usadas em Português, seria preciso também uma grande quantidade de memória para armazenar estes dados.

Em segundo lugar, não pode se esquecer que a língua não é um sistema fechado nem é estática, ela está em constante mudança e novas palavras são adicionadas ao léxico de uma língua constantemente. Por tanto, mesmo que alguém consiga registrar, por exemplo, a lista de todas as palavras escritas com <x> que se pronunciam /ks/ esta informação só seria válida para hoje, no futuro próximo pode ser adicionada uma nova palavra, já seja por empréstimo de outra língua ou por criação espontânea dos falantes, que tenha a mesma pronúncia e nesse caso a palavra não estaria contemplada no dicionário.

Uma possibilidade um pouco mais econômica em termos de esforço e espaço para a armazenagem dos dados é a de criar um sistema misto que receba um dicionário com transcrições e seja capaz de extrair dele regras linguísticas ou, por meio de uma análise estatística, extrair padrões que lhe permitam prever a transcrição das palavras que não estejam no dicionário.

Optou-se neste trabalho pela abordagem por regras, por várias razões:

- Apesar de não existir correspondência biunívoca entre grafemas e fonemas em Português, é possível, como já foi dito, identificar regras de pronuncia para a maioria das combinações de grafemas possíveis.
- Uma vez criadas estas regras, não é mais preciso se preocupar com novos vocábulos que possam ser criados na língua, pois não importa qual seja a palavra, a sua realização fonética obedecerá às regras da língua, por tanto o sistema será capaz de produzi-la.
- Um conjunto de regras de pronuncia é muito mais econômico, em termos de memória, que um dicionário de termos.

Justificativas similares apresentam Braga & Freixeiro (2007) para o idioma Galego, língua românica de sintaxe, morfologia e fonética muito similar ao Português.

#### 2.1.4 A SÍNTESE DE BAIXO NÍVEL

Depois de realizada a transcrição fonética, incluindo a informação prosódica, é preciso gerar o sinal da fala. Para realizar esta reconstrução da voz humana a partir de um texto há vários métodos que podem ser aplicados, alguns utilizam trechos de voz gravados como base, outros emulam o trato

vocal ou diretamente o seu produto, a fala. Cada um destes métodos possui vantagens e desvantagens que serão resumidas a seguir.

#### 2.1.4.1 SÍNTESE A PARTIR DE CORPUS (OU SÍNTESE POR SELEÇÃO DE UNIDADES)

Este método de síntese utiliza um *corpus*, gravado previamente por um locutor profissional, segmentado e etiquetado do qual são extraídas as unidades que melhor se ajustem ao texto que será sintetizado (Dutoit, 2008). Isto significa que as unidades não são fixas como na síntese por concatenação (cf. subsecção 2.1.4.2), senão que, de todas as possibilidades que o *corpus* oferece, o sistema seleciona aquele segmento que mais se adéqua à necessidade de cada transcrição sem importar o tamanho da unidade. Este método, à semelhança do de concatenação descrito na secção 2.1.4.2, precisa de um módulo de transcrição fonética adequado e, após a seleção das unidades, de um tratamento específico da melodia (Listerri, 2009).

Ao unir as unidades de som para formar palavras e frases, criam-se pontos de transição entre uma e outra unidade. Estes pontos resultam em falhas e efeitos perceptíveis desagradáveis que resultam das discontinuidades espectrais produzidas pela junção das unidades.

#### 2.1.4.2 SÍNTESE POR CONCATENAÇÃO DE UNIDADES

Uma estratégia para sintetizar a fala é a de utilizar como base pequenos trechos de voz natural pré-gravados que serão conectados ou concatenados para reproduzir qualquer texto (Simões, 1999, p. 78).

Neste método, é preciso segmentar trechos de fala que serão codificados e farão parte de um dicionário de unidades acústicas. Este dicionário deve conter suficientes unidades para gerar qualquer enunciado no idioma selecionado para a síntese.

O tipo de unidade que irá se adotar para a concatenação é fundamental para a qualidade do resultado final. Em teoria, quanto maior for o segmento adotado menos pontos de concatenação haverá, por tanto a qualidade será maior. Contudo, é preciso considerar que ao escolher trechos de fala maiores como unidade, são necessários mais deles, o que implica num incremento no tamanho da base de dados que irá armazenar estas unidades.

Outro aspecto essencial para se obter um bom resultado final com este método é considerar o efeito da coarticulação entre segmentos adjacentes. Por este motivo não é possível utilizar como unidades sons isolados (fonos, por exemplo),

Conseguir habla sintetizada de calidad concatenando sonidos aislados e intentando imitar el resultado acústico de los movimientos del tracto vocal en las transiciones entre sonidos es una operación sumamente difícil. Por tal motivo, las unidades a partir de las cuales se construye un sistema de síntesis no suelen ser sonidos aislados, sino combinaciones de sonidos. (Llisterri, 2001, p. 9)

Grande parte da naturalidade do sinal vai depender dos fenômenos coarticulatórios que desempenham também um papel importante para a eficácia da comunicação (Fowler, 1981; Whalen, 1990 *apud* Barbosa, 1999, p.8). Um mesmo fonema pode ser realizado de diversas formas dependendo do contexto no qual está inserido, este é mais um motivo pelo qual não podem ser utilizadas unidades isoladas, pois seria preciso considerar todos os contextos possíveis para capturar todos alofones (diferentes realizações de um

mesmo fonema) existentes o que resultaria numa quantidade imensa de unidades.

A unidade adotada normalmente é o dífone ou trífone que consiste na segmentação desde o ponto estável de um fone (normalmente o seu meio) até o ponto estável do fone seguinte, ou seja, a metade de cada fone. De acordo com Llisterra (2001, p. 9), desta forma ao concatenar se consegue que a união das unidades se produza pela parte em que há uma menor variação acústica e não pela parte onde há transição de um som a outro. Assim se preserva o ponto onde se manifestam os efeitos da coarticulação.

Para extrair as unidades de voz natural se utilizam gravações feitas normalmente com locutores profissionais que devem ser capazes de manter a voz o mais constante possível para não ocasionar mudanças bruscas no espectro.

Os fones, dífonos e trífonos não existem de forma isolada na língua, por este motivo é preciso escolher de que contexto serão extraídas estas unidades. As possibilidades são:

- a) De palavras isoladas.
- b) De diferentes frases que contenham ao menos uma unidade de interesse.
- c) De frases iguais repetidas várias vezes com a mudança de uma só palavra, que é exatamente a palavra que contém a unidade de interesse.

- d) De logátomos, isto é, palavras que possuem a unidade de interesse, mas que não necessariamente têm significado ou existem na língua, porém são fonologicamente possíveis.

Depois de gravar todo o material é preciso identificar ou etiquetar, na gravação, os fones e as unidades maiores (dífonos, trífonos, polifones) utilizadas na síntese e, posteriormente, escolher o ponto onde será realizado o corte.

Em um sistema de síntese por concatenação um gerador segmental se ocupa da concatenação efetiva dos dífonos, trífonos ou polifones, mas mesmo com a utilização de unidades maiores como são os polifones não é possível armazenar previamente unidades suficientes para cobrir todos os fenômenos coarticulatórios mais extensos.

É preciso considerar também que, se a fala sintética se faz apenas concatenando as unidades pré-cortadas, resultaria numa fala com inúmeros “pulos” e altos e baixos tanto de melodia quanto de intensidade cada vez que se passasse de uma unidade à seguinte. Por este motivo é preciso suavizar os pontos onde as unidades se juntam e dar à fala sintética uma uniformidade.

O ritmo da fala é dado pela modulação de duração ao longo da frase pelo que se faz necessária também a implementação de um modelo de geração automática da duração segmental.

Dois dos principais modelos de geração da duração são:

- a) O modelo de Dennis Klatt (1987) desenvolvido originalmente para a língua inglesa tem servido de base para muitos outros modelos (Van Santen, 1994). O princípio básico deste modelo consiste em associar uma duração específica a

cada segmento, posteriormente um conjunto de regras que modificará a duração será aplicado, mas sem que os segmentos possam ser comprimidos além de uma duração mínima estabelecida.

b) O modelo de Nick Campbell (1992), desenvolvido para o inglês britânico separa o controle do tempo ao nível da sílaba do cálculo da duração do segmento. Este modelo apresenta duas etapas: uma primeira em que a duração silábica que se obtém pelo aprendizado automático e uma segunda em que esta duração é distribuída entre os segmentos que formam a sílaba pelo uso de um modelo estatístico chamado de modelo de repartição (Barbosa, 1994 *apud* Barbosa, 1999, p. 7).

#### 2.1.4.3 SÍNTESE POR REGRAS (PARAMÉTRICA)

Este método simula de forma eletrônica um conjunto de parâmetros acústicos do sinal da fala. As regras são determinadas por parâmetros necessários para sintetizar uma expressão vocal desejada (Allen, 1987). O conversor recebe, então, estes parâmetros e produz fala sintética a partir de um texto apenas utilizando os parâmetros recebidos.

O controle dos parâmetros neste método podem ser obtidos por análise de segmentos de voz natural ou podem ser calculados de acordo com um conjunto de regras que dependem do contexto.

Pols (1990, *apud* Dutoit, 1997) destaca que os conversores de texto em fala baseados em regras constituem uma aproximação cognitiva, generativa do mecanismo de fonação e de articulação. Em outras palavras, existe ao criar um

sistema deste tipo, um esforço por compreender as características da fala para depois reproduzi-las.

#### 2.1.4.4 SÍNTESE ARTICULATÓRIA

Este método recria e modela a configuração que o trato vocal do falante vai assumindo enquanto fala. São recriados os movimentos dos órgãos articulatórios de forma que o sistema não se centra na fala em si, mas na forma como ela é produzida, isto evita ter que estudar acusticamente a fala.

Há diferentes tipos de síntese articulatória; a mais básica recebe os fonemas que representam o texto e os transforma em comandos musculares neuromotores, algo parecido ao que faz uma pessoa quando pensa numa mensagem que deseja transmitir e o seu cérebro envia os comandos aos músculos que controlam os articuladores que ao mudar de forma produzem a fala. Este método precisa de um modelo que relate os comandos musculares que serão utilizados no movimento do articulador e um algoritmo que vai converter as posições do articulador nas formas do trato vocal ou funções de área. (Teixeira et al., 2004).

Também é possível mapear diretamente os fonemas para as posições articulatórias e movimentos, pois é mais fácil obter dados experimentais para o movimento articulatório do que para os comandos musculares. Um sistema típico assume de 7 a 11 parâmetros que são suficientes para descrever o comportamento articulatório. Para obter o modelo de movimento dos articuladores, normalmente se utilizam sequências de raio-X do trato vocal, e depois se aproxima a forma do trato vocal a estes parâmetros. Os diferentes

fonemas são gerados a partir dos dados armazenados sobre os diferentes valores destes parâmetros para cada fonema, que são armazenados numa base de dados.

A FIGURA 2 mostra um modelo articulatório; as linhas pontilhadas indicam as relações entre os parâmetros e flechas mostram o grau de liberdade.

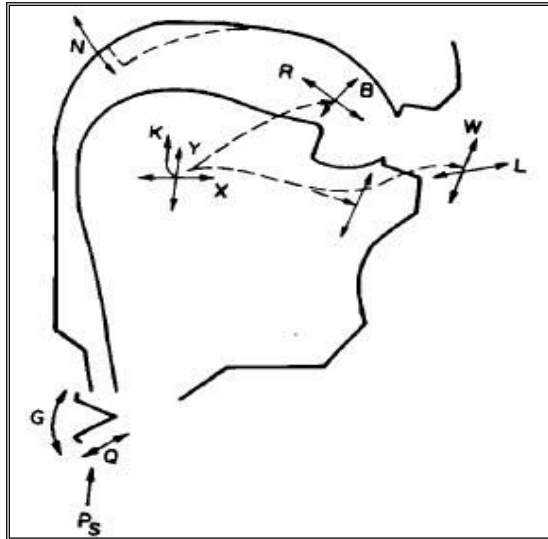


Figura 2 - Modelo Articulatório Simplificado (KLATT D. , 1987, P. 757)

Este método permite gerar uma síntese de muito alta qualidade, porém os parâmetros não são fáceis de obter nem controlar além dos modelos articulatórios apresentarem uma grande complexidade computacional que dificulta ainda mais a utilização deste sistema na síntese com fins comerciais.

#### 2.1.5 TÉCNICAS DE GERAÇÃO DO SINAL ACÚSTICO

Como explica Barbosa (1999, p. 8), além do método de síntese é preciso utilizar também uma técnica para a geração do sinal acústico. As mais utilizadas, segundo o autor, são a PSOLA (Pitch-Synchronous Overlap and

Add), a LPC (*Linear Predictive Coding*) e a representada por um sintetizador por formantes.

#### 2.1.5.1 SÍNTESES POR FORMANTES

Nestes métodos de síntese, a fala é produzida pela modelagem dos órgãos vocais mediante a simulação das frequências e amplitudes formantes (Styger & Keller, 1994). A base para este tipo de síntese é a teoria acústica da produção da fala (Fant, 1960), ou seja, a teoria que estima que a fala é o resultado da excitação de um filtro linear por uma ou mais fontes de som.

A síntese por formantes consiste então de fontes que simulam o que acontece no trato vocal e provocam a excitação de um conjunto de filtros que modelam o trato vocal. A excitação pode consistir de impulsos periódicos, para produzir os fonemas vozeados (sonoros); ruído, para produzir os fonemas não vozeados (surdos) ou uma combinação de ambos que irá produzir sons fricativos vozeados e oclusivos sonoros.

Uma forma de evitar ter que desenvolver todo o conjunto de regras de coarticulação que seria necessário é calcular automaticamente as propriedades espectrais da fala mediante coeficientes LPC (Coeficientes de predição linear). Nesta técnica o sinal de fala é produzido por uma fonte sonora aplicada a um filtro. O filtro é composto por um conjunto de parâmetros fixos que permitem a obtenção da amostra atual do sinal de fala a partir de amostras anteriores. (Barbosa P. A., 1999, p. 8). Com este método, a função de transferência do filtro que simula o trato vocal é controlada por parâmetros definidos por um código preditivo linear.

#### 2.1.5.2 A TÉCNICA PSOLA

O nome da técnica em inglês faz referência à sobreposição e soma e é exatamente isso o que a técnica faz: sobrepor e adicionar pequenas unidades de forma de onda utilizando dois períodos de frequência fundamental adjacentes. Esta técnica atua diretamente sobre o sinal da fala modificando a curva melódica (F0), duração e amplitude do sinal concatenado via multiplicação, redução, compressão ou expansão de períodos glotais (Barbosa P. A., 1999, p. 8).

## 2.2 DESCRIÇÃO FONÉTICA DO PORTUGUÊS BRASILEIRO

Um dos primeiros trabalhos sobre a fonética e fonologia do PB é a tese de doutorado de Câmara Jr. (1953) quem, seguindo as teorias do Círculo de Praga, realizou uma detalhada descrição do sistema fonológico do PB, que foi aperfeiçoada nas suas publicações posteriores.

Praticamente todas as análises fonológicas do PB posteriores (Rameh, 1961; Cyr, 1965; Lemle, 1966; Pontes, 1972; Cagliari, 1977), até mesmo as análises experimentais, se apoiam nas publicações de Câmara.

Depois de Câmara, outros linguistas já discutiram detalhadamente a classificação dos sons da língua portuguesa, entre eles Cunha e Cintra (1985), Callou e Leite (1990) e Casaes (1990). Há também outros trabalhos mais recentes sobre a fonética do Português, como o de Silva (1999) e o de Barbosa (2006), que trata especificamente da prosódia da língua portuguesa falada no Brasil.

No que diz respeito à fala regional há vários trabalhos dialectológicos como o de Marroquim (1934), no Nordeste; Amaral (1920), em São Paulo; e Nascentes (1922), no Rio de Janeiro. Além de outros mais atuais desenvolvidos à luz da sociolinguística, como o projeto NURC<sup>13</sup> (Callou, 1991; Callou & Lopes, 1992; Callou & Lopes, 1994; Mota & Rollemberg, 1994; Castilho & Preti, 1986; Castilho & Preti, 1987) e o PEUL<sup>14</sup> (Naro, 2001), na Universidade Federal do Rio de Janeiro; o VARSUL<sup>15</sup> (Vandersen, P. & Coelho, I., 2008), no Rio Grande do Sul, Santa Catarina e Paraná; e o VALP (Hora, 2005), na Universidade Federal da Paraíba.

#### 2.2.1 DESCRIÇÕES DO DIALETO DE BELÉM DO PARÁ

Sobre o dialeto de Belém do Pará existem alguns trabalhos que se inserem nas pesquisas geolinguísticas ou sociolinguísticas e dialetais. A seguir se apresenta uma breve descrição destes trabalhos:

a) Atlas Linguístico Sonoro do Pará (ALISPA)<sup>16</sup>

O projeto do qual surgiu o ALISPA teve início em 1996 com o objetivo de criar o *Atlas Geo-Sociolingüístico do Pará*. Os registros sonoros informatizados do projeto foram publicados em CD-ROM em 2004 com o nome de ALISPA (Razky, 2004).

---

<sup>13</sup> <http://www.letas.ufrj.br/nurc-rj/> [17/02/2010] <http://www.fflch.usp.br/dlc/nurc/index.html>

[17/02/2010] <http://www.ufpe.br/pgletras/programa-nucleos-nurc.htm> [17/02/2010]

<sup>14</sup> <http://www.letas.ufrj.br/peul/> [17/02/2010]

<sup>15</sup> <http://www.pucrs.br/fale/pos/varsul/index.php> [17/02/2010]

<sup>16</sup> <http://www.ufpa.br/alipa/> [17/02/2010]

O Atlas Sonoro foi elaborado segundo critérios estabelecidos em pesquisas dialetais e controla as variáveis sociolinguísticas como gênero e faixa etária (18 a 30 e 40 a 70 anos). Dez cidades do Estado do Pará foram selecionadas para a coleta de dados, que foram obtidos por meio de entrevistas, nas quais se utilizou um questionário composto por 157 questões, conhecido como Questionário Fonético-fonológico (QFF) do *Projeto do Atlas Lingüístico do Brasil — ALiB*, e que foi aplicado a 4 informantes de cada localidade.

O CD-ROM consta com um *software* que permite que o usuário visualize mapas, com os 157 vocábulos do QFF transcritos foneticamente e que ouça a voz do falante, a partir de um *click* sobre qualquer uma das transcrições fonéticas.

O objetivo deste atlas sonoro é constituir uma fonoteca dos falares paraenses armazenando os dados sonoros de um variado grupo de falantes.

b) Atlas Geo-Sociolingüístico do Pará<sup>17</sup>

Este projeto pretende registrar a fala de uma rede de 57 pontos do Estado do Pará. Para esta finalidade se utilizou um questionário que foi aplicado a informantes de ambos os sexos tanto das zonas rurais como urbanas.

c) O Atlas Lingüístico do Brasil (Projeto ALiB)<sup>18</sup>.

---

<sup>17</sup> <http://www.ufpa.br/alipa/> [17/02/2010]

<sup>18</sup> <http://www.alib.ufba.br/index.asp> [17/02/2010]

Além dos projetos mencionados acima que compreendem unicamente a fala da região do Pará, existe um projeto nacional em andamento desde 1996, para a elaboração do *Atlas Lingüístico do Brasil (ALiB)*.

A ideia do atlas linguístico a nível nacional existe oficialmente desde o 20 de março de 1952 quando o Decreto 30.643, Art. 3º, indicou como finalidade prioritária da Comissão de Filologia da Casa de Rui Barbosa a *elaboração do Atlas Lingüístico do Brasil*.

Para a coleta de dados deste projeto são escolhidos quatro informantes por cada localidade que devem ser falantes do dialeto do local. Controlam-se as variáveis de sexo e idade dos informantes (um homem e uma mulher de 18 a 30 anos e outro casal de 50 a 65 anos) além da variável de nível de escolaridade mínimo e máximo.

O questionário aplicado é organizado em três seções: fonético-fonológico, com 159 perguntas; semântico-lexical, com 207 questões e morfossintático, com 80 questões.

A criação destas bases de dados tem impulsionado o desenvolvimento de pesquisas linguísticas, principalmente nas subáreas da sociolinguística e dialetologia, pela facilidade que representa ter um *corpus* oral com variáveis controladas.

### 3 METODOLOGIA

Parte do objetivo principal de esta investigação é descrever foneticamente a língua portuguesa falada na cidade de Belém do Pará. Está implícito neste objetivo que se partiu do pressuposto de que o Português de Belém do Pará se diferencia de outros dialetos do Português falados no território brasileiro. Por este motivo o primeiro passo foi revisar tanto a bibliografia relacionada à descrição fonética do PB quanto os trabalhos de descrição do Português de Belém já existentes.

Após esta aproximação inicial, foi constatado que as poucas descrições existentes do dialeto de Belém têm fins sociolinguísticos enquadrando-se principalmente nos estudos dialetológicos no marco da teoria variacionista laboviana (cf. Labov, 1972, 1994, 2000) ou então se resumem apenas a descrições da diversidade lexical do falar da região norte (cf. subsecção 2.2.1). Estes estudos não apresentam por tanto nenhuma contribuição para o desenvolvimento de tecnologias da fala, que é a segunda parte do objetivo principal deste trabalho.

Para descrever foneticamente o dialeto de Belém, fez-se necessário então obter amostras da fala de belenenses com o intuito de analisá-las, inclusive, acusticamente.

#### 3.1 LOCALIZAÇÃO, PERÍODO DA PESQUISA E INFORMANTES

A coleta de dados foi realizada em Setembro de 2009 na própria cidade de Belém.

A seleção dos informantes obedeceu tanto a critérios tradicionais no âmbito da pesquisa fonética tais como:

- a) ser nativo da localidade;
- b) não apresentar problemas de articulação ou de fonação;
- c) não se ter afastado da localidade por mais de um terço de sua vida;
- d) não ser analfabeto.

Além desses critérios básicos, foi decidido que os informantes deveriam ser do sexo masculino e de idade entre os 20 e 35 anos, pois os estudos já existentes sobre as características acústicas do PB são mais abundantes para sujeitos com estas características, além de representarem, segundo dados obtidos no *website* do IBGE<sup>19</sup>, o grupo etário mais representativo do estado do Pará.

Foram selecionados três voluntários que concordaram em ter a sua fala gravada e servir como informantes para esta pesquisa. Os informantes serão identificados de aqui em diante da seguinte forma:

**a) Informante 1**

**Idade:** 21  
**Escolaridade:** Terceiro grau incompleto.  
**Outros onde residiu:** Castanhal, Pará (dos 6 aos 11 anos de idade).  
**Naturalidade dos pais:** Ambos de Belém.  
**Profissão:** Estudante de graduação em Matemáticas.  
**Nível socioeconômico:** Médio.

**b) Informante 2**

**Idade:** 23  
**Escolaridade:** Terceiro grau incompleto.  
**Outros onde residiu:** Nenhum.  
**Naturalidade dos pais:** Mãe de Belém, pai desconhecida.

---

<sup>19</sup> [www.ibge.gov.br/](http://www.ibge.gov.br/) [17/02/2010]

**Profissão:** Estudante de graduação em História.  
**Nível socioeconômico:** Médio-alto.

**c) Informante 3**

**Idade:** 25  
**Escolaridade:** Segundo grau completo.  
**Outros onde residiu:** Santarém, Pará (dos 14 aos 16 anos).  
**Naturalidade dos pais:** Ambos de Belém.  
**Profissão:** Auxiliar de construção civil.  
**Nível socioeconômico:** Meio.

### 3.2 MÉTODOS, INSTRUMENTOS E TÉCNICAS DE COLETA DE DADOS

As gravações se realizaram com o software *Wave Studio*<sup>20</sup> a 44.1 kHz (mono, 16-bit) no formato WAV. Foram conduzidas numa sala acusticamente isolada utilizando um microfone unidirecional *LeSonSM58* diretamente conectado à placa de som X-Fi (compatível com Sound Blaster) de um computador portátil *Dell Inspiron 1525*.

Foi feito primeiramente um teste de som para verificar a qualidade da fala gravada. Uma vez que o teste entregou resultados aceitáveis, se procedeu a instruir os informantes (um de cada vez) quanto à distância que deveriam manter o microfone e o volume ideal da voz, que deveriam tentar manter constante.

### 3.3 O CORPUS

O corpus que será analisando neste trabalho consiste de duas partes:

---

<sup>20</sup> <http://www.creative.com> [16/03/2010]

- a) Uma primeira que corresponde à leitura, por parte de cada um dos três informantes, de dois breves textos.

O primeiro texto se encaixa no domínio de descrição e história da cidade e trata, principalmente, das origens da cidade de Belém e alguns dos problemas urbanos que a cidade apresenta atualmente em decorrência do processo migratório gerado pela sua localização privilegiada.

O segundo texto se encaixa no domínio da identidade cultural e trata especificamente da língua falada como elemento de expressão da identidade cultural.

- b) Na segunda parte, os informantes relatam o que conteúdo de um dos textos lidos anteriormente. Desta forma, após a leitura dos textos, se obtiveram frases produzidas semi-espontaneamente com as características prosódicas típicas da fala e não da leitura, que costuma apresentar um padrão mais regular que a fala espontânea (Esser, 1988). Também desta forma se obtém várias realizações da mesma palavra.

Cada um dos informantes falou durante uma média de 2 minutos sobre um dos textos que havia lido anteriormente.

Para a elaboração deste *corpus* foram considerados os seguintes critérios:

- A existência de todos os 8 fones orais e os 5 fones nasais em todos os contextos.
- A existência de pelo menos 5 exemplos de cada tipo de ditongo.

- A aparição de cada um dos 21 fones consonantais em posição inicial, medial e final de palavra (antes de espaço) em contextos de ataque e coda sempre que possível. Os grafemas <t>, <d>, <s>, <z>, <l>, <n> e <r> receberam atenção especial, cuidando que existissem no *corpus* exemplos deles precedendo todas as vogais.

O *corpus* gravado contém 2309 palavras, sendo 1.527 provenientes da leitura dos textos e 782 correspondentes à produção semi-espontânea dos informantes.

### 3.4 TÉCNICAS DE ANÁLISE

#### 3.4.1 ANÁLISE DA TRANSCRIÇÃO

As gravações foram transcritas em ortografia convencional e posteriormente se realizou uma transcrição fonética estreita (desde que se utilizaram alofones) utilizando o IPA (*International Phonetic Alphabet*) (ver o Anexo I), como mostra o seguinte exemplo:

A língua é o instrumento que o povo utiliza para expressar-se.

[aːlĩgwa    ɛ    u    ʔĩftruːmẽtu    kiuːpɔvutʃiːliza  
paraːɛʃpreːsaxsi]

[Informante 3. Parte 2]

Depois, se procedeu a uma análise da transcrição fonética e se separaram os alofones que apresentam uma realização diferente da

considerada convencional no dialeto de PB geralmente utilizado para a CTF (a variedade paulista).

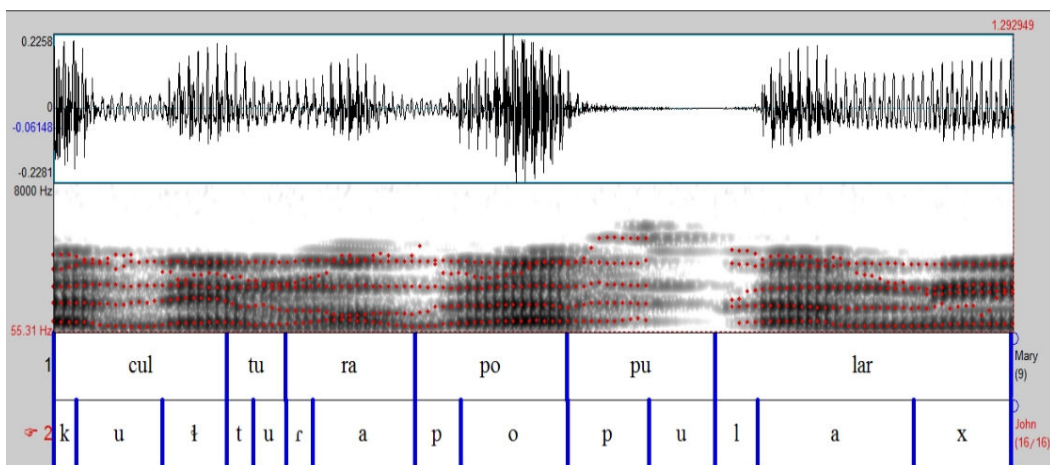
Uma vez separados estes alofones, se analisou o contexto no qual cada um aparece a fim de verificar se a substituição de alofones se dá sempre no mesmo contexto. Para a análise do contexto se consideraram os seguintes critérios:

- Que tipo de fone antecede o alofone de interesse.
- Que tipo de fone há imediatamente depois o alofone de interesse.
- Se a sílaba de aparição é tônica ou átona. Se átona, de que tipo.
- Se o fone está na posição inicial, medial ou antes de pausa.
- A posição do fone dentro da sílaba.

Após analisar o contexto de aparição de cada fonema de interesse se procedeu a verificar se é possível extrair algum padrão ou regra de correspondência entre o grafema e o fone de acordo ao contexto de aparição para depois formalizá-las.

#### 3.4.2 A ANÁLISE ACÚSTICA

FIGURA 3- Segmentação em Praat da produção da frase “cultura popular”. (Informante 3, Parte 1)



Para realizar a análise acústica, se utilizou o programa Praat v5.1.20 no qual se abriram os arquivos de som para obter tanto o espectrograma da sequência de sons quanto o oscilograma e para, desta forma, proceder à segmentação das sílabas e fones e à sua transcrição e anotação, como mostra a Figura 3.

Na análise acústica se consideraram principalmente os valores de f1, f2 e a duração dos segmentos.

Os valores médios da primeira e segunda frequências formantes, F1 e F2 respectivamente, permitem caracterizar os sons segundo o posicionamento da língua e da cavidade oral no momento da sua produção. O F1 está relacionado com a altura e deslocamento da língua no plano vertical da cavidade oral e com a abertura da mesma, já o F2 se relaciona com o deslocamento da língua no plano horizontal da cavidade, isto é, o grau de variação no sentido anteroposterior.

Estes dados permitem distinguir entre dois alofones diferentes nos casos em que há dúvida sobre a produção ou nos casos em que se pretende

confirmar uma avaliação perceptiva anterior que indicava o uso de um alofone não utilizado no dialeto de PB considerado padrão.

### 3.4.3 AS REGRAS

Depois de definidas a partir da análise as características do dialeto consideradas interessantes para a transcrição ortográfico-fonética, estas características foram analisadas de forma a extrair padrões que permitiram estabelecer correspondências ortográfico-fonéticas considerando o contexto de aparição dos alofones. Estas correspondências foram transformadas numa lista de 16 regras codificadas e formalizadas para a sua implementação ao módulo ortográfico fonético e a sua aplicação a um texto já normalizado ao qual hajam já sido aplicadas as regras para o PB.

Esta série de regras foram posteriormente testadas em 206 palavras que foram manualmente transcritas em X-SAMPA (Wells, 1994), receberam marcação de fronteiras (início e fim de palavra) e de sílabas de forma a emular o estado em que se encontrariam se houvessem passado já por uma parte do pré-processamento estando, então, prontas para a aplicação das regras específicas para o dialeto de Belém, como se vê na Figura 4

FIGURA 4 - Amostra de Palavras Preparadas para a Aplicação das Regras de Transcrição

Palavra em ortografia convencional	Palavra antes da aplicação das regras
acontecimento	#a.k~o.te.si."m~e.tu#
acontecimentos	#a.k~o.te.si."m~e.tus#
alterações	#aw.te.4a."s~ois#
amazonas	#a.ma."zo.nas#
amenizar	#a.me.ni."zax#
apareceu	#a.pa.4e."sew#
apesar	#a.pe."zax#
apresenta	#.ap4e."z~e.ta#
apresentada	#a.p4e.z~e.ta.da"
apresentado	#a.p4e.z~e."ta.du#
áreas	#"a.4jas#

As 206 palavras que serviram para testar as regras foram retiradas do corpus e selecionadas de acordo a dois critérios:

1. Palavras cuja transcrição demonstrasse cumprir uma ou mais das condições estabelecidas nas regras da subsecção 5.2.
2. Palavras que contivessem algum dos grafemas afetados pelas regras da subsecção 5.2, porém, cujo contexto fosse diferente do estipulado pelas regras.

De acordo com as regras da subsecção 5.2, algumas das palavras selecionadas deveriam, então, ter a sua transcrição alterada, por cumprir as condições estipuladas nas regras, e outras delas deveriam manter a sua transcrição intata. O segundo grupo de palavras foi escolhido para testar se a aplicação das regras não interfere com a transcrição de palavras às quais os fenômenos descritos no capítulo 4 não se aplicam.

As regras foram aplicadas às 206 palavras na mesma ordem em que são apresentadas as subsecção 5.2, substituindo de forma sistemática os alofones da transcrição em PB geral pelos alofones (ou combinação deles)

estabelecidos pela regra. Após este processo, a transcrição obtida para cada palavra foi comparada com a transcrição esperada, que foi obtida da produção dos informantes.

## 4 RESULTADOS DA ANÁLISE

### 4.1 AS VOGAIS

Mediante análise acústica realizada, se verificou que, na posição tônica, as vogais têm as seguintes características:

TABELA 4 – Valores obtidos para as Vogais Tônicas no dialeto de Belém

Vogal	Duração Média (ms)	F1 Médio(Hz)	F2 Médio(Hz)	Nº Casos Analisados
[a]	108,9	828,8	1580,1	208
[ã]	149,8	589,8	1625,2	41

[ɛ]	90,1	566,6	1847,0	48
[e]	112,0	469,8	1966,2	24
[ê]	163,8	437	2098,2	17
[i]	69,4	308,4	2180,0	30
[ĩ]	162,5	289,9	2595,6	28
[ɔ]	119,1	571,1	1035,6	50
[o]	94,3	484,3	998,2	50
[ô]	165,3	469,6	905,3	39
[u]	74,0	330,9	814,4	30
[ũ]	169,9	279,4	803,0	26

**TABELA 5 - Desvio Padrão e Coeficiente de Variação dos Valores de F1 e F2 das Vogais**

Vogal	DP F1 (Hz)	CV F1	DP F2 (Hz)	CV F2
[a]	23,6	0,028 (2,8%)	54,2	0,034 (3,4%)
[ã]	52	0,088 (8,8%)	45,1	0,027 (2,7%)
[ɛ]	32,7	0,057 (5,7%)	86,8	0,046 (4,6%)
[e]	27,7	0,058 (5,8%)	79,8	0,025 (2,5%)
[ê]	23,9	0,052 (5,2%)	86,3	0,041 (4,1%)
[i]	19,5	0,063 (6,3%)	54,9	0,025 (2,5%)
[ĩ]	31	0,1 (10%)	87,7	0,033 (3,3%)
[ɔ]	43,4	0,075 (7,5%)	68,3	0,068 (6,8%)
[o]	17,8	0,036 (3,6%)	97,8	0,097 (9,7%)
[ô]	36,2	0,077 (7,7%)	89,8	0,099 (9,9%)
[u]	22,1	0,066 (6,6%)	50,5	0,062 (6,2%)
[ũ]	34,5	0,12 (12%)	63,2	0,078 (7,8%)

Estas características acústicas coincidem, de forma geral, com as características relatadas algumas vogais de outros dialetos de PB em trabalhos anteriores:

TABELA 9 – Valores médios das Vogais do PB

Vogal	F1 (Hz)	F2 (Hz)	Fonte
[a]	620	1320	Cagliari (1977)
	738		Souza (1994)
	829,9	1583,9	Nassif et. Al (2009)
[ã]	500	1425	Cagliari (1977)
	630		Souza (1994)
[ɛ]	566,6	1847	Nassif et. Al (2009)
[i]	270	2190	Cagliari (1977)
	291		Souza (1994)
	308,4	2180	Nassif et. Al (2009)
[ɪ]	270	2600	Cagliari (1977)
	279		Souza (1994)
[ɔ]	571,1	1035,6	Nassif et. Al (2009)
[u]	280	870	Cagliari (1977)
	299		Souza (1994)
[ũ]	250	800	Cagliari (1977)
	310		Souza (1994)
	330,9	814,4	Nassif et. Al (2009)

TABELA 6 – Valores obtidos para as vogais átonas no dialeto de Belém

Vogal	Duração Média (ms)	F1 Médio(Hz)	F2 Médio(Hz)	Nº Casos Analisados
-------	--------------------	--------------	--------------	---------------------



Média	/e/	/o/	(2° grau)
Média	/ɛ/	/ɔ/	(1° grau)
Baixa	/a/		
<b>Vogais em posição pretônica</b>			
Alta	/i/	/u/	
Média	/e/	/o/	
Baixa	/a/		
<b>Vogais em posição postônica</b>			
Alta	/i/	/u/	
Média	/e/		
Baixa	/a/		
<b>Vogais em posição postônica final</b>			
Alta	/i/	/u/	
Baixa	/a/		

#### 4.1.4 O GRAFEMA <a>

TABELA 8 - Realizações de <a>

	[a]		[ɐ]		[ẽ]		Total
	Tônica	Átona	Tônica	Átona	Tônica	Átona	
Acentuado com ~	-	-	-	-	31 (100%)	-	31
Antes de consoante nasal em coda	-	-	-	-	11 (52,38%)	10 (47,62%)	21
Sem ~ nem consoante nasal em coda	208 (52%)	94 (13,5%)	-	*98 (24,5%)	-	-	400

Total	208	94	98	41	10	452
-------	-----	----	----	----	----	-----

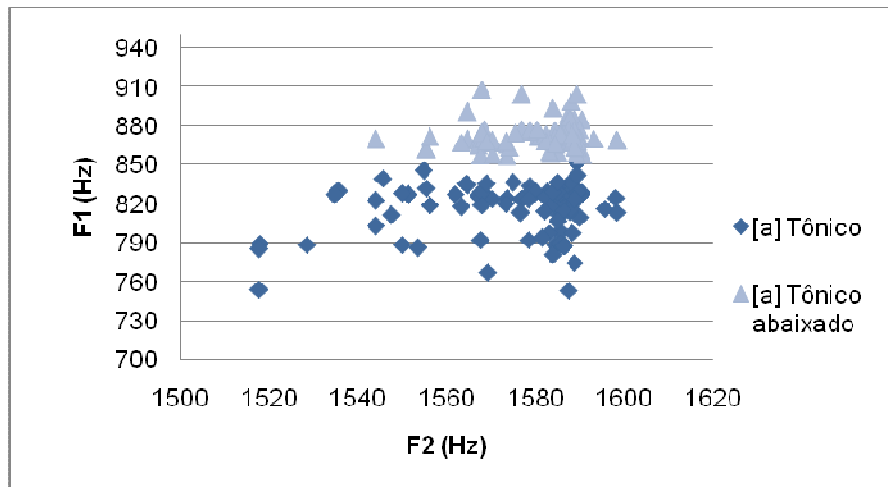
\*Este dato corresponde apenas à aparição de <a> em sílaba átona em posição de final de palavra (antes de espaço), em todos os outros contextos, <a> em sílaba pretônica ou postônica não final é sempre realizada como [a].

O grafema <a> representa a vogal central baixa que, como se vê na Tabela 8, no dialeto de Belém é realizada como [a] em todas as posições quando aparece numa sílaba tônica. Quando em posição de final de palavra (antes de espaço) é sempre realizado como [ɐ], se aparece numa sílaba átona, como em *malha* ['maʎɐ] ou *situada* [si'twadə]. Precedendo uma consoante nasal em coda ou graficamente acentuado com *til* (~) é sempre realizado como [ẽ].

Foi constatado nos dados analisados que esta vogal é frequentemente abaixada e ligeiramente alongada na posição tônica em nomes próprios, nomes comuns e, em menor quantidade, em verbos, fenômeno que não se encontra registrado na bibliografia revisada referente aos dialetos do resto do país.

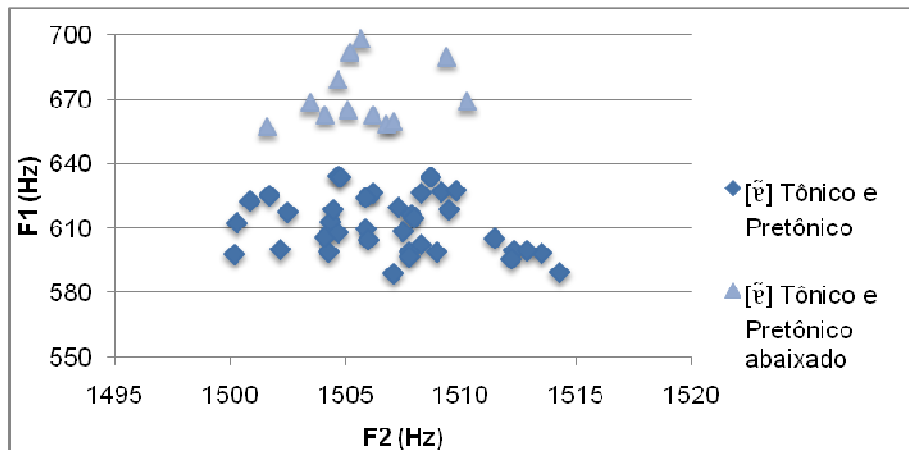
Das 208 vogais [a] em posição tônica analisadas, 97 (36,63%) apresentaram a tendência ao abaixamento em relação aos valores médios do dialeto para [a] em sílaba tônica, como se vê na Figura 5.

FIGURA 5 – Dispersão do [a] Tônico



Este mesmo fenômeno se verifica na vogal nasal [ẽ] para as mesmas categorias de palavras tanto em posição tônica quanto pretônica. A Figura 6 mostra a dispersão encontrada em 52 sons [ẽ] analisados. Destes 52 sons, 17 (32,69%) apresentaram um abaixamento e discreto alongamento em relação aos valores médios do dialeto para [ẽ].

FIGURA 6 – Dispersão do [ẽ] Tônico e Pretônico



Embora não seja do interesse desta pesquisa tentar descobrir as razões para este fenômeno, cogitou-se a possibilidade se tratar de um recurso

utilizado pelos falantes para enfatizar ou intensificar uma palavra considerada chave dentro do enunciado, tendo em vista que, no corpus analisado, o abaixamento se manifesta apenas em palavras pertencentes às categorias nominal e verbal. Contudo, chama a atenção o fato de não se tratar de um fenômeno do estilo particular de um falante, senão de uma constante verificada na fala dos três informantes, o que levanta a suspeita de se estar diante de um fenômeno sistemático do falar da região.

#### 4.1.5 O GRAFEMA <e>

TABELA 9 – Realizações de <e>

	[e]	[ɛ]	[i]	[ê]	[ej]	[ɛĩ]	Total
Em sílaba tônica com acento agudo (´)	-	25 (100%)	-	-	-	-	25
Em sílaba tônica com acento circunflexo (^)	24 (100%)	-	-	-	-	-	24
Antes de consoante nasal	-	-	-	25 (100%)	-	-	25
Em sílaba pretônica antes de tônica fechada	42 (85,7%)	3 (6,1%)	4 (8,2%)	-	-	-	49
Em sílaba pretônica antes de tônica aberta [a] [ɛ] [ɨ]	2 (8%)	23 (92%)	-	-	-	-	25
Em sílaba pretônica antes de pretônica medial alta [i] [u]	13 (59%)	3 (13,6%)	6 (27,2%)	-	-	-	22
Em sílaba pretônica antes de tônica alta [i] [u]	-	9 (75%)	3 (25%)	-	-	-	12
Em sílaba pretônica inicial terminada em [ʃ]	-	-	12 (100%)	-	-	-	12
Em sílaba postônica	25 (100%)	-	-	-	-	-	25
Em sílaba postônica em final de palavra (antes de espaço)	-	-	25 (100%)	-	-	-	25
Em sílaba tônica terminada em [ʃ]	-	-	-	-	15 (100%)	-	15
Com acento agudo em sílaba tônica terminada consoante nasal	-	-	-	-	-	9 (100%)	9
Em sílaba tônica terminada em consoante nasal	-	-	-	-	12 (100%)	-	12
Total	106	63	50	25	27	9	280

Este grafema representa as vogais médias e, como se aprecia na Tabela 9, no dialeto de Belém, tem seis realizações possíveis. Em sílaba tônica <e> é realizada da mesma forma que em outros dialetos de PB; salvo em posição tônica de sílabas terminadas em [ʃ] ou consoante nasal, nas quais tanto a vogal média fechada quanto a aberta são sempre realizadas como um ditongo, como se observa na Tabela 9.

Na mesma tabela, é possível observar que quando <e> aparece imediatamente antes de uma vogal tônica fechada, há uma forte tendência à realização fechada [e] (85,7% dos casos). Uma exceção constitui a palavra *terreno* que apresentou em 100% dos casos a realização aberta [ɛ], e constitui o único exemplo desta realização no contexto mencionado no corpus analisado.

Antes de vogal tônica aberta, <e> pretônico é realizado, na grande maioria dos casos (92%), como [ɛ]. É importante destacar que esta abertura da vogal pretônica antes de tônica aberta não é um fenômeno generalizado no PB, senão uma característica exclusiva de alguns dialetos da região norte e nordeste (como o dialeto pernambucano, por exemplo).

Quando a vogal da sílaba seguinte é alta em posição pretônica medial há concorrência da vogal pretônica inicial média fechada [e], da aberta [ɛ] e da alta [i]. Já quando a vogal média aparece precedendo a vogal alta em posição tônica, concorrem apenas a vogal média fechada [e] e a alta [i].

Outro fator que influencia o alçamento da vogal média anterior provocando que <e> seja realizada como [i] é ela aparecer em posição inicial

imediatamente antes da consoante [ʃ] em coda, por exemplo, em palavras como *está* [iʃ'ta] e *espanhola* [iʃpa'ŋɔle].

Das 280 vogais médias anteriores analisadas, 12 se encontram no referido contexto tendo sido confirmado após análise perceptivo e acústico que 100% delas foram realizadas como [i]. É importante destacar que na grande maioria dos dialetos de PB a consoante [ʃ] não costuma aparecer em coda.

Na posição postônica, a vogal média, anterior, oral passa a ser realizada exclusivamente como [e], da mesma forma que acontece no resto dos dialetos de PB. Segundo Câmara (1970; 1994), em posição postônica final, o quadro vocálico se reduz a ainda mais, sendo possível encontrar apenas três fonemas nesta posição em PB.

#### 4.1.6 O GRAFEMA <O>

TABELA 10 - Realizações de <o>

	[o]	[ɔ]	[õ]	[u]	Total
Em sílaba tônica com acento agudo (´)	-	25 (100%)	-	-	25
Em sílaba tônica com acento circunflexo (^)	25 (100%)	-	-	-	25
Acentuada com til (~)	-	-	25 (100%)	-	25
Antes de consoante nasal em coda	-	-	25 (100%)	-	25
Em sílaba pretônica	27 (40%)	37 (56,9%)	-	2 (3%)	65
Pretônica depois de consoante velar	-	15 (100%)	-	-	15
Em sílaba postônica	25 (100%)	-	-	-	25
Em sílaba postônica em final de palavra (antes de espaço)	-	-	-	25 (100%)	25
Total	78	77	50	25	230

O grafema <o> representa as vogais médias, posteriores, orais cuja realização no dialeto de Belém, na posição tônica, não difere da de outros dialetos do PB.

Na posição pretônica, de 80 fones analisados, o 65% foram realizados como vogal média, posterior, oral, aberta [ɔ], demonstrando um claro predomínio deste fone em dita posição.

Analisando apenas os 52 casos em que a vogal média aberta foi produzida em posição pretônica foi possível verificar que um dos fatores que influencia a abertura da vogal é ela aparecer depois de uma consoante velar.

É interessante destacar que em alguns dialetos mais difundidos de PB a vogal pretônica, posterior, oral costuma sofrer um alçamento sendo realizada como [u] (Callou, Moraes, & Leite, 2002; Celia, 2004; Borges G. L., 2008). Tal fenômeno no *corpus* analisado se verificou apenas de forma pontual, ocorrendo somente na palavra *bonita* [bu'ɲita] na qual a vogal pretônica foi realizada como vogal posterior, alta em 100% dos casos.

## 4.2 AS CONSOANTES

TABELA 11 - Consoantes do Português de Belém

Ponto de articulação / Modo articulação	Bilabial		Labio-dental		Dental		Alveolar		Alveolopalatal		Palatal		Velar		
	surd	son	Surd	son	surd	son	surd	son	surd	son	surd	son	surd	son	
Orais	Oclusiva	<b>p</b>	<b>b</b>			<b>t</b>	<b>d</b>							<b>k</b>	<b>g</b>
	Africadas									<b>tʃ</b>	<b>dʒ</b>				
	Fricativa			<b>f</b>	<b>v</b>			<b>s</b>	<b>z</b>	<b>ʃ</b>	<b>ʒ</b>			<b>x</b>	
	Lateral aproximante								<b>l</b>				<b>ʎ</b>		
	Vibrante simples								<b>r</b>						
Nasais		<b>m</b>						<b>n</b>				<b>ɲ</b>			

Após a análise acústica e perceptiva do dialeto foi possível identificar 21 alofones consonantais para o dialeto, como se vê na Tabela 11.

No referente às consoantes, a característica mais marcante do dialeto de Belém é a realização palatalizada de alguns fonemas. Como se mostra a seguir, este é um fenômeno linguisticamente significativo e que acaba caracterizando bastante o falar da região.

O quadro de consoantes do dialeto de Belém se diferencia do do PB mais geral não pela quantidade ou tipo dos alofones, mas pelos contextos em que eles aparecem. (Cf. Barbosa & Albano, 2004).

#### 4.2.1 <d> E <t> ANTES DE [i]

TABELA 12 - Realizações de <t>

	[t]	[tʃ]	Total
Em posição de ataque	25 (100%)	-	25

Em posição de ataque antes de [i]	-	25 (100%)	25
Em coda	-	5 (100%)	5
Total	25	30	55

TABELA 13 - Realizações de <d>

	[d]	[dʒ]	Total
Em posição de ataque	25 (100%)	-	25
Em posição de ataque antes de [i]	-	25 (100%)	25
Total	25	25	50

Os segmentos que conformam a fala, isto é, os fones, não se produzem de forma isolada, pelo contrário, a fala é um contínuo no qual estes segmentos influenciam os segmentos adjacentes, gerando assim grande quantidade de alofones para cada fonema. Este fenômeno é chamado de coarticulação. (Hardcastle & Hewlett, 1999).

No dialeto de Belém, assim como em muitos outros dialetos de PB, a coarticulação dos fonemas dentais com o segmento vocálico [i] dá origem aos alofones alveolopalatais [tʃ] e [dʒ], como mostram as Tabela 12 e Tabela 13. Como se aprecia nos dados analisados, a palatalização de [t] e [d] diante de [i] no dialeto de Belém é sistemática, dos 105 segmentos analisados, 100% tiveram e uma realização palatalizada. Contudo, é importante mencionar que este fenômeno, se bem é comum a muitos dialetos de PB, não se produz em todos os dialetos do Brasil.

#### 4.2.2 <l> SEGUIDO DE [i]

TABELA 14 - Realizações de <l>

	[l]	[ʎ]	Total
Em posição de ataque	45 (100%)		49

Em posição de ataque antes de [i]	39 (100%)	39
Total	45	84

Uma característica interessante do dialeto encontrada no corpus e que não aparece registrada na bibliografia consultada sobre o dialeto de Belém é a realização de <l> antes de vogal, alta, anterior.

O fone palatal, lateral, aproximante, sonoro existe em todos os dialetos de português, porém, nos outros dialetos de PB, ele está sempre associado ao fonema /ʎ/ e representa o dígrafo <lh>. O fonema /ʎ/ não aparece em PB na posição inicial (exceto em vocábulos de origem estrangeiro, como *lhama* /ʎama/), por tanto o fone [ʎ] não aparece em posição inicial nestes dialetos. No português de Belém, por influência da vogal anterior alta, o fonema /l/ pode ser realizado como [ʎ], isto viabiliza a aparição de [ʎ] em posição inicial.

Como se observa na Tabela 14, a sequência /l/ seguido de [i] é sempre realizada como [ʎi] na fala de Belém.

Durante a análise se realizou uma comparação das características acústicas do fone [ʎ] nas palavras *realidade*, *linguagem*, *malha* e *melhor* com o intuito de verificar a semelhança de ambos fones.

Esta análise não mostrou diferenças acústicas relevantes entre [ʎ] como alofone palatal de /l/ ou como realização de /ʎ/.

#### 4.2.3 <N> SEGUIDO DE [i]

Outro caso muito similar ao descrito na subsecção 4.2.2 o constitui o <n> precedendo [i].

TABELA 15 - Realizações de <n>

	[n]	[ɲ]	Total
Em posição de ataque	50 (100%)		50
Em posição de ataque antes de [i]		33 (100%)	33
Total	50	33	83

O fone palatal, nasal, sonoro [ɲ] existe em PB associado ao fonema /ɲ/ que representa dígrafo <nh>. Este dígrafo não aparece em posição inicial, portanto o fone [ɲ] não é produzido nesta posição em outros dialetos de PB.

No português de Belém, homologamente ao que acontece ao fonema //, o fonema /n/ sofre influência da vogal anterior alta quando a precede, sendo realizado como [ɲ], inclusive em posição inicial. Como se observa na Tabela 15, este fenômeno acontece em todos os 33 fones analisados, ou seja, <ni> foi realizado como [ɲi] em 100% dos casos.

Foi realizada uma análise acústica comparativa dos fones produzidos nas palavras *amazônia* e *ninguém* e os produzidos em *espanhola* e *nenhuma* sem que se encontrassem diferenças significativas.

#### 4.2.4 <S> EM CODA

TABELA 16 - Realizações de <s> em coda

	[s]	[ʃ]	[z]	[ʒ]	[ø]	Total
Antes de fone sonoro	-	-	-	18 (100%)	-	18
Antes de fone surdo	-	49 (100%)	-	-	-	49
Em contexto prepausal	-	19 (86,3)	-	-	3 (13,6%)	22
Total	-	68	-	18	3	89

Na posição de coda existem em PB pelo menos quatro possíveis realizações para a consoante <s>. Perceptivamente, o dialeto de Belém se enquadra entre os dialetos que palatalizam o fonema /s/ neste contexto.

Na análise dos dados coletados foi possível comprovar que o dialeto de Belém apresenta uma forte tendência à palatalização de /s/ como se pode observar nos exemplos da Tabela 16.. Este fenômeno só não acontece nos dados analisados nos casos em que o grafema <s> é de origem flexional, pois este contexto morfológico parece facilitar o apagamento do fonema /s/. Esta variação não é relevante para o presente estudo, pois não se configura como um fenômeno típico e expressivo do dialeto em si, senão que se trata de um fenômeno que atinge um determinado socioleto do PB em geral. (Guy, 1981).

É interessante apontar que, em PB, <s> na posição de final de palavra (antes de espaço), quando não em contexto prepausal, costuma sofrer influência do fone inicial da palavra seguinte. Se a palavra seguinte tiver como primeiro som uma consoante surda, o fone alveolar, fricativo produzido será também surdo [s]; se a palavra seguinte tiver como primeiro som uma consoante sonora ou uma vogal, o fone alveolar, fricativo produzido será sonoro [z].

No caso da fala de Belém, a mesma regra se aplica, porém a fricativa sofre palatalização, por tanto o som produzido não será mais o alveolar, senão o alveopalatal, fricativo surdo [ʃ] ou sonoro [ʒ], segundo seja o caso. Um exemplo disto se vê no segundo caso mostrado na expressão *das Américas* na qual o fonema fricativo final de *das* foi produzido 100% das vezes como [ʒ].

## 5 FORMALIZAÇÃO DOS RESULTADOS

### 5.1 CONCLUSÕES DESPRENDIDAS DA ANÁLISE

Depois de analisar os dados obtidos sobre as características segmentais da fala de Belém, é possível desprender as seguintes conclusões sobre o dialeto de Belém:

1. A vogal [a] tônica é alongada e abaixada (possui valores de F1 5% mais altos que a média) com o intuito de enfatizar algum vocábulo no enunciado.
2. A vogal [ɛ] tônica ou pretônica é alongada e abaixada (possui valores de F1 5% mais altos que a média) com o intuito de enfatizar algum vocábulo no enunciado.
3. A aparição da vogal [ɛ] se restringe apenas ao contexto de final de palavra quando a vogal baixa é átona.
4. Diferentemente de outros dialetos do Brasil (os do centro-oeste e sul, por exemplo), o dialeto de Belém apresenta vogais médias abertas em posição pretônica.
5. A vogal [e] sofre ditongação em posição tônica de palavras agudas terminadas em [ɲ] e seguida de consoante nasal em sílabas tônicas.
6. A realização das vogais médias pretônicas -especialmente a média anterior- sofre forte influência da vogal tônica contigua apresentando uma tendência a se harmonizar com ela.
7. Quando a vogal /e/ precede uma vogal tônica aberta, há uma forte tendência à realização [ɛ].
8. A vogal média anterior aberta [ɛ] não aparece em posição pretônica quando a vogal tônica é alta.

9. A vogal média anterior aberta [ɛ] não aparece em posição postônica.
10. Há uma forte tendência a realizar a vogal /e/ como media-fechada [e] quando ela precede uma vogal tônica fechada.  
A consoantes oclusivas, dentais, surda /t/ e a sonora /d/, se realizam como alveopalatais, africadas, [tʃ] e [dʒ] respectivamente, antes de [i].
11. A consoante /l/ se realiza como [ʌ] antes de [i].
12. A consoante /n/ se realiza como [ɲ] antes de [i].
13. Na posição de coda, o fonema /s/ é realizado como [ʃ].
14. Na posição de coda, o fonema /z/ é realizado como [ʒ].

## 5.2 A INCORPORAÇÃO DO CONHECIMENTO FONÉTICO A UM SISTEMA DE CTF

A incorporação do conhecimento linguístico a um sistema de CTF se faz preferencialmente no módulo de conversão ortográfico-fonético que, como já foi mencionado, se encontra dentro do bloco de pré-processamento do texto, no componente que realiza a transcrição fonética do texto ortográfico para símbolos que representam fones.

Esta incorporação se faz normalmente mediante uma série de regras que são combinadas em algoritmos que recebem como entrada o texto em ortografia normal e mediante a aplicação linha a linha das regras neles contidos transformam o texto em símbolos ou sequências deles que representam fones.

Cada algoritmo criado utiliza uma anotação própria que é determinada pelo criador do mesmo, a seguir se apresentam as convenções de anotação utilizadas neste trabalho:

TABELA 17 - Convenções de Anotação

Símbolo	Significado
/.../	qualquer fonema
[...]	qualquer fone
<...>	qualquer grafema
C	consoante
V	vogal
CN	consoante nasal
VN	vogal nasal
x(y)	y é uma características de x
x(atq)	x está na posição de ataque
x(cod)	x está na posição de coda
x(#_)	x é inicio de palavra (é precedido por um espaço)
x(_#)	x é final de palavra (é seguido de um espaço)
"xy	xy é uma sílabatônica
x("_")	x é uma vogal postônica
("_)x	x é uma vogal pretônica
[x(son)]	x é um fone sonoro
x(sur)	x é um fone surdo
{ }_	sílaba anterior
_ { }	sílaba seguinte
_ {/i(")/}	A sílaba seguinte contém o fonema /i/ em posição tônica
,	separador de opções
+	seguido de
-	não é seguido de
→	então

Na descrição das regras se utilizou o alfabeto X-SAMPA para representar os 36 fones do Português de Belém. A seguir se apresenta a correspondência dos símbolos X-SAMPA e IPA:

TABELA 18 – Correspondência dos fones do dialeto de Belém em IPA e X-SAMPA

IPA	X-SAMPA	Descrição
p	p	Oclusiva, bilabial, surda
b	b	Oclusiva, bilabial, sonora
m	m	Nasal, bilabial
f	f	Fricativa, labiodental, surda
v	v	Fricativa, labiodental, sonora
t	t	Oclusiva, dental, surda
d	d	Oclusiva, dental, sonora
s	s	Fricativa, alveolar, surda
z	z	Fricativa, alveolar, sonora
l	l	Lateral, alveolar
r	4	Vibrante, alveolar
n	n	Nasal, alveolar
tʃ	tS	Africada, alveopalatal, surda
dʒ	dZ	Africada, alveopalatal, sonora
ʃ	S	Fricativa, alveopalatal, surda
ʒ	Z	Fricativa, alveopalatal, sonora
ʎ	L	Lateral, palatal
ɲ	J	Nasal, palatal
k	k	Oclusiva, velar, surda
g	g	Oclusiva, velar, sonora
x	x	Fricativa, velar, surda
a	a	Baixa, central
ɐ	6	Chuá, aberta
e	e	Média-fechada, anterior
ɛ	E	Média-aberta, anterior
i	i	Alta, anterior
o	o	Média-fechada, posterior
ɔ	O	Média-aberta, posterior
u	u	Alta, posterior
ã	~6	Baixa, central, nasal
ẽ	~e	Média-aberta, anterior, nasal
ĩ	~i	Alta, anterior, nasal
õ	~o	Média, posterior, nasal
ũ	~u	Alta, posterior, nasal

As regras de pronúncia específicas para o português de Belém extraíveis da análise podem ser formalizadas da seguinte maneira (Tabela 19):

TABELA 19 - Regras para Algoritmo de Transcrição Ortográfico-Fonética

Linha	Regra	Comentário
1.	Se <a(")(_#)> → [6]	Para <a>
2.	Se <e("> + <s(_#)> → [ej]	Para <e>
3.	Se <e("> + CN(_#) → [ej]	
4.	Se <e("> + _{[i,u]} → [e]	
5.	Se <e> + _{[i,u(")]} → [e]	
6.	Se <e("_)> → [e]	
7.	Se <e("_)> + _{[a,6,E,O(")]} → [E]	
8.	Se <é, ê> + CN → [E~i]	
9.	Se <o> + _{[i,u(")]} → [o]	Para <o>
10.	Se <t> + [i] → [tS]	Para <t> e <d>
11.	Se <d> + [i] → [dZ]	seguidos de [i]
12.	Se <l> + [i] → [L]	Para <l> seguido de [i]
13.	Se <n> + [i] → [J]	Para <n> seguido de [i]
14.	Se <s(cod)> → [S]	Para <s> e <z> em coda
15.	Se <z(cod)> → [Z]	
16.	Se [S(cod)] + [...(son)] → [Z]	

### 5.3 AVALIAÇÃO

Os resultados obtidos após a aplicação das regras a um total de 206 palavras foram os seguintes:

TABELA 20 - RESULTADOS OBTIDOS NA AVALIAÇÃO

Transcrições corretamente alteradas pela aplicação das regras	118 (57,28%)	96,11%
Transcrições corretamente não alteradas pela aplicação das regras	80 (38,83%)	
Transcrições incorretamente alteradas pela aplicação das regras	0	3,88%
Transcrições incorretamente não alteradas pela	8 (4,37%)	

aplicação das regras		
Total	206	

Como é possível observar na Tabela 20, aplicando as regras propostas a uma seleção de 206 palavras do corpus, foi possível obter um 96,11% de acerto na transcrição final em relação à transcrição esperada.

Na mesma tabela é possível observar também que as regras não alteraram nenhuma transcrição que já estivesse correta, porém, em 3,88% as regras não alteraram a transcrição de palavras por não haver sido considerada na sua formulação a variação alofônica que algum fone do contexto poderia sofrer, como se vê a continuação:

- A regra número 12 estipula que: Se <n> + [i] → [J]. Porém esta regra não contempla a semi-vogal [j] que aparece em palavras como *amazônia* [ama'zoŋje] e *colônia* [ko'loŋje], que não foram alteradas ao se aplicar a regra número 12.

Este é um problema de fácil solução, bastando apenas alterar a regra para resolver 6 dos 8 erros detectados, diminuindo assim a percentagem de erro para apenas 0,97%.

Os outros 2 casos de erro correspondem às palavras bonita [bu'ŋjtɛ], terreno [tɛ'xenu] que constituem exceções às regras propostas e a sua transcrição deveria ser adicionada a uma lista de exceções.

## 6 CONCLUSÃO

Neste trabalho, se descreveram as características fonéticas do português falado na cidade de Belém do Pará que diferenciam este dialeto de outros falados no Brasil.

No capítulo 2 se descreveu o funcionamento de um sistema de CTF, a sua arquitetura básica com a sua divisão em blocos e módulos e se explicaram as diferentes abordagens, métodos e técnicas existentes tanto para a CTF.

No capítulo 4 se descreveram os resultados obtidos na análise dos dados coletados durante a pesquisa e se verificou a existência de regras que regem a pronúncia diferenciada desta região, a partir desta análise, no capítulo 5 se propuseram regras para a incorporação destas características ao algoritmo de transcrição ortográfico-fonética do módulo de transcrição de um sistema de CTF construído desde uma abordagem linguística e se demonstrou que mediante a aplicação destas regras é possível adaptar um sistema de CTF feito para o PB à variedade dialetal de Belém com uma percentagem de acerto próxima ao 99,3%.

Como trabalho futuro, seria possível expandir a investigação ao módulo de geração de prosódia onde o conhecimento das características supra-segmentais do dialeto poderia ser melhor aproveitado.

## 7 REFERÊNCIAS

- Albano, E. (1999). O Português Brasileiro e as Controvérsias da Fonética Atual: Pelo Aperfeiçoamento da Fonologia Articulatória. *DELTA*, vol.15.
- Allen, J. (1987). Synthesis of Speech from Unrestricted Text. *Proceedings of the IEEE*, v. 64, n. 4, (pp. 433-442).
- Amaral, A. (1920). *O dialecto caipira*. São Paulo: Casa editora O livro.
- Aragão, M. S. (2005). *Os Estudos Geolinguísticos no Brasil: Dos Atlas Regionais ao AliB*. Acesso em 25 de Maio de 2009, disponível em [http://www.sbpcnet.org.br/livro/57ra/programas/CONF\\_SIMP/textos/msocorroaragao1.htm](http://www.sbpcnet.org.br/livro/57ra/programas/CONF_SIMP/textos/msocorroaragao1.htm)
- Barbosa, P. A. (2006). *Incursões em torno do Ritmo da Fala*. Campinas: Pontes Editores - FAPESP.
- Barbosa, P. A. (1999). Revelar a Estrutura Rítmica de uma língua construindo máquinas falantes: pela integração entre ciência e tecnologia de fala. In: E. Scarpa, *Estudos de Prosódia* (pp. 21-52). Campinas: Editora da Unicamp.
- Barbosa, P. A. (2000). 'Syllable timing in Brazilian Portuguese': uma crítica a Roy Major. *D.E.L.T.A.* , 369-402.
- Barbosa, P. A., Violaro, F., Albano, E. C., Simões, F. O., Aquino, P. A., Madureira, S., et al. (1999). Aiuruetê: a High-Quality Concatenative Text-to-Speech System for Brazilian Portuguese with Demisyllabic Analysis-Based Units and a Hierarchical Model of Rhythm Production. *Proceedings of the Eurospeech'99*, (pp. 2059-2062). Budapeste.
- Barbosa, P. & Albano, E. (2004). Illustrations of the IPA: Brazilian portuguese. *Journal of the International Phonetic Association* , 227-232.
- Black, A., Lenzo, K., & Pagel, V. (1988). Issues in Building General Letter to Sound Rules. *3rd ESCA Workshop on Speech Synthesis*, (pp. 77-80). Jenolan Caves, Australia.
- Bonafonte, A., Escudero, D., & Riera, M. (2006). La conversión de texto en habla. In: J. Llisterri, & M. J. Machuca, *Los sistemas de diálogo* (pp. 177-206). Bellaterra - Soria: Universitat Autònoma de Barcelona, Servei de Publicacions - Fundación Duques de Soria.
- Borges, G. L. (2008). Uma Análise Sobre as Vogais Pretônicas do Mumincípio de Uberaba/MG. *A MARGem - Estudos* , 79-93.
- Borges, J. A. (1996). Dosvox - Um Novo Acesso dos Cegos à Cultura e ao Trabalho. . *Revista Benjamin Constant - IBCENTRO/MEC* , 24-29.
- Bosch, A. & Daelemans, W. (1993). Data-oriented methods for grapheme-to-phoneme conversion. *Sixth Conference of the European Chapter of the Association for Computational Linguistics (EACL'93)*, (pp. 45-53). Utrecht, Holanda.

- Braga, D. & Freixeiro, X. (2007). Algoritmos de Conversão Grafema-Fonema em Galego para Sistemas de Conversão Texto-Fala. In: B. Baltrusch, & G. Pérez, *Estudos galegos de Tradución & Poartradución no século XXI*. Vigo: Xerais (No Prelo).
- Braga, D., & Resende, F. (2006). Módulos de processamento de texto baseados em regras para sistemas de conversão Texto-Fala em PE. *Actas do XXII Encontro da Associação Portuguesa de Linguística* (pp. 141-155). Coimbra: APL.
- Braga, D., Freitas, D., & Ferreira, H. (2003). Processamento Linguístico Aplicado à Síntese da Fala. *Proceedings of III Congresso Luso-Moçambicano de Engenharia*, (pp. 1349-1360). Maputo.
- Brandão, S. F., & Moraes, J. A. (1995). A geolingüística no Brasil: resultados e perspectivas. *Terceira Margem III*, 218-222.
- Cagliari, L. C. (1977). *An Experimental Study of Nasality with particular reference to Brazilian Portuguese*. Thesis (Ph.D). University of Edinburgh.
- Cagliari, L. C. (2003). Da importância da prosódia na descrição de fatos gramaticais". In: I. Rodolfo, *Gramática do português falado. Vol. II: Níveis de análise lingüística*. 2ª ed. Campinas: Editora da UNICAMP.
- Cagliari, L. C. (1981). *Elementos de fonética do português brasileiro*. Tese (Livre Docência em Linguística). UNICAMP.
- Cagliari, L. C., & Abaurre, M. B. (1986). Elementos para uma investigação instrumental das relações entre padrões rítmicos e processos fonológicos no português brasileiro. *Cadernos de Estudos Lingüísticos*, 39-57.
- Callou. (1991). *A linguagem falada culta na cidade do Rio de Janeiro. Materiais para seu estudo. vol I. Diálogo entre dois informantes*. Rio de Janeiro: UFRJ/FUJB.
- Callou, D., & Leite, Y. (1990). *Iniciação à Fonética e à Fonologia*. São Paulo: Jorge Zahar Editor Ltda.
- Callou, D., & Lopes, C. (1994). *A linguagem falada culta na cidade do Rio de Janeiro. Materiais para seu estudo. vol III. Diálogo entre dois informantes*. UFRJ/CAPES/CNPq.
- Callou, D., Moraes, J., & Leite, Y. (2002). A elevação das vogais pretônicas no português do Brasil: processo(s) de variação estável. *Letras de Hoje*, v. 37, n. 127, 9-24.
- Câmara Jr., J. M. (1970). *Estrutura da língua portuguesa*. Petrópolis: Vozes.
- Câmara Jr., J. M. (1953). *Para o estudo da fonêmica portuguesa*. Rio de Janeiro: Organizações Simões.
- Câmara Jr., J. M. (1994). *Problemas de lingüística descritiva*. Petrópolis: Vozes.
- Câmara, J. M. (1953). *Para o estudo da fonêmica portuguesa*. Rio de Janeiro: Organizações Simões.
- Campbell, N. (1992). Syllable-based Segmental Duration. In: G. Bailly, C. Benoit, & T. Sawallis, *Talking Machines: Teories, Models and Designs* (pp. 211-224). Oxford: Elsevier.
- Campos, G. L. (1980). *Síntese de Voz para o Idioma Português*. Tese (Ph.D). Escola Politécnica, Universidade de São Paulo.

- Casaes, E. J. (1990). *Descrição Acústico-Articulatória dos Sons da Voz para um Modelo dos Sons do Português do Brasil*. Thesis (Ph.D). Faculdade de Filosofia, Universidade de São Paulo.
- Castilho, A. T., & Preti, D. (1986). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo Vol. I - Elocuções formais*. São Paulo: T. A. Queiroz.
- Castilho, A. T., & Preti, D. (1987). *A linguagem falada culta na cidade de São Paulo: materiais para seu estudo. Vol. II - Diálogos entre dois informantes*. São Paulo: T. A. Queiroz/FAPESP, 1987. 264 p. São Paulo: T. A. Queiroz/FAPESP.
- Celia, G. F. (2004). *As vogais médias pretônicas na fala culta de Nova Venécia - ES*. Dissertation (MA). Campinas, SP: Universidade Estadual de Campinas, Instituto de Estudos da Linguagem.
- Cereja, W. R., & Magalhães, T. C. *Português: Linguagens. 5ª edição*. Editora Saraiva.
- Chotimongkol, A., & Black, A. (2000). Statistically trained orthographic to sound models for Thai. *Proceedings of ICSLP2000*. (pp 11-18). Beijing, China.
- Coker, C. H., Church, K. W., & Liberman, M. Y. (1990). Morphology and rhyming: Two powerful alternatives to letter-to-sound rules for speech synthesis. *ESCA Workshop on Speech Synthesis*, (pp. 83-86). Aufrans, France.
- Cristiani, A. C., & Encarnação, M. R. (2006). *De Antenor Nascentes ao Projeto Atlas Lingüístico do Brasil - ALiB: conquistas da Geolingüística no Brasil*. Acesso em 1 de Maio de 2009, disponível em <http://www.letramagna.com/geolingüistica.pdf>
- Cunha, C., & Cintra, L. (1985). *Nova Gramática do Português Contemporâneo*. Rio de Janeiro: Ed. Nova Fronteira.
- Cyr, J. J. (1965). *Some Intonational Features of Brazilian Portuguese*. Dissertation (MA). Texas: University of Texas at Austin.
- Dahlbäck, N., Wang, Q., Nass, C., & Alwin, J. (2007). Similarity is more important than expertise: Accent effects in speech interfaces. *Proc. SIGCHI conference on human factors in computing systems*, 1553–1556.
- Dias, C. S. Belém do Pará: História, Urbanismo e identidade. *Planejamento Urbano e Regional: ensaios acadêmicos co CAUFAG*. Cascavel: Smolarek Arquitetura.
- Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. Dordrecht: Kluwer Academic Publishers.
- Dutoit, T. (2008). Corpus-Based speech synthesis. In: J. E. Benesty, M. M. Sondhi, & Y. Huang, *Springer Handbook on Speech Processing* (pp. 437-520). Berlin - Heidelberg: Springer.
- Egashira, F. (1992). *Síntese de Voz a Partir de Texto*. Dissertation (MA). Universidade Estadual de Campinas.
- Esquivel, A. S. (1985). Um Sistema de Síntese de Voz. *Congresso Nacional de Informática* (pp. 776-782). São Paulo: Sucesu.
- Esser, J. (1988). *Comparing Reading and Speaking Intonation*. Amsterdam: Editions Rodopi B.V.
- Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

- Fowler, C. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *J.S.H.R.* , 127-139.
- Guy, G. R. (1981). *Linguistic variation in Brazilian Portuguese: aspects of the phonology, syntax, and language history*. Philadelphia: University of Pennsylvania.
- Hall Jr., R. A. (1943). The units phonemes of Brazilian Portuguese. *Studies in Linguistics*, 15 , 1-16.
- Hardcastle, W. J., & Hewlett, N. (1999). *Coarticulation*. Cambridge.
- Kaplan, R. M., & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics* 20 , 331-378.
- Klatt, D. H. (1987). Review of text-to-speech conversion for English. *The Journal of the Acoustical Society of America*, 82(3) , 737-793.
- Klatt, D. (1987). Review of text-to-speech conversion in English. *Journal of the Acoustical Society of America* 82 , 737-793.
- Leite, Y. (1974). *Portuguese Stress and Related Rules*. Thesis (PhD). Austin: Language Department, University of Texas.
- Lemle, M. (1966). *Phonemic System of the Portuguese of Rio de Janeiro*. Dissertation (MA). University of Pensilvania.
- Lemmetty, S. (1999). *Review of speech synthesis technology*. (Master's Thesis). Department of Electrical and Communications Engineering. Helsinki University of Technology.
- Llisterri, J. (2001). La conversión de texto en habla. *Quark, Ciencia Medicina, Comunicación y Cultura* , 79-89.
- Llisterri, J., Carbó, C., Machuca, M. d., Mota, C., Riera, M., & Ríos, A. (2003). El papel de la lingüística en el desarrollo de las tecnologías del habla. In: M. Casas, & C. Varo, *VII Jornadas de Lingüística* (pp. 137-191). Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- Llisterri, J., Machuca, M., Madrigal, N., Mancini, F., Massimino, P., Mota, C., et al. (2004). Aspectos lingüísticos en el diseño de un conversor de texto en habla en castellano y en catalán: El sistema Loquendo TTS®. *6º Congreso de Lingüística General*. (pp. 521-532). Santiago de Compostela: Área de Lingüística Xeral, Universidade de Santiago de Compostela.
- Llisterri, J. (2007a). El papel de la fonética en las tecnologías del habla. In: M. G. González, E. Fernández, & B. González, *Actas do III Congreso Internacional de Fonética Experimental* (pp. 23-37). Santiago de Compostela: Xunta de Galicia.
- Llisterri, J. (2007b). Màquines que parlen i que escolten: el paper de la fonètica en el desenvolupament des les tecnologies de la parla. In: J. Carrera, & C. Pons, *Aplicacions de la fonètica. Catorzè Col·loqui Lingüístic de la parla* (pp. 139-169). Barcelona: Promociones y Publicaciones Universitarias.
- Llisterri, J. (2009). Las tecnologías del habla en las lenguas románicas ibéricas. *Studies in Hispanic and Lusophone Linguistics* 2(1), 133-180.
- Lucassen, J. M. (1983). *Discovering Phonemic Base Forms Automatically: An Information Theoretic Approach*. Yorktown Heights, NY: IBM Thomas J. Watson Research Center.
- Major, R. C. (1985). Stress and Rhythm in Brazilian Portuguese. *Language* , 259-282.

- Major, R. C. (1981). Stress-timing in Brazilian Portuguese. *Journal of Phonetics* , 343-352.
- Marroquim, M. (1934). *A língua do Nordeste: Alagoas e Pernambuco*. Curitiba: HD Livros.
- Massini-Cagliari, G. (1992). *Acento e Ritmo*. São Paulo: Contexto.
- Mateus, M. H. (1982). *Aspectos da Fonologia Portuguesa*. 2a. ed. Lisboa: Instituto Nacional de Investigação Científica/ Centro de Lingüística da Universidade de Lisboa.
- Meng, H. M., Seneff, S., & Zue, V. (1994). Phonological parsing for bi-directional letter-to-sound/sound-to-letter generation. *ARPA Human Language Technology Workshop*. Princeton.
- Moraes Barbosa, J. (1962). Les voyelles nasales portugaises: interprétation phonologique. *Proceedings of the Fourth International Congress of Phonetic Sciences* (pp. 691-708). The Hague: Mouton.
- Moraes, J. A., & Leite, Y. (1992). Ritmo e velocidade da fala na estratégia do discurso: uma proposta de trabalho. In R. Ilari, *Gramática do Português Falado. Volume II: Níveis de Análise Lingüística* (pp. 65-77). Campinas: UNICAMP.
- Mota, J., & Rollemberg, V. (1994). *A linguagem falada culta na cidade de Salvador: materiais para seu estudo. Vol. I: Diálogos entre informante e documentador*. Salvador: Instituto de Letras, UFBA.
- Naro, J. A. (2001). *Implementação e Encaixamento da Mudança Lingüística. Relatório Final do Projeto PEUL/UFRJ*. CNPq.
- Nascentes, A. (1922). *O linguajar Carioca*. Rio de Janeiro: Simões.
- Nassif, M., Gama, A., Ferreira, C., & Dolabella, J. (2009). Análise acústica das vogais: estudo comparativo de F1 e F2 em indivíduos glossectomizados parciais e no grupo controle. *Arquivos de ciências da saúde. Vol 16.* , 26-30.
- Paiva, M. A. (1996). Supressão das semivogais nos ditongos decrescentes & Atuação das variáveis sociais na supressão da semivogal anterior nos ditongos decrescentes. In: G. M. Oliveira e Silva, & M. (. Scherre, *Padrões Sociolingüísticos: análise de fenômenos variáveis do português falado na cidade do Rio de Janeiro* (pp. 217-236). Rio de Janeiro: Tempo Brasileiro.
- Pontes, E. (1972). *Estrutura do verbo no português coloquial*. Rio de Janeiro: Vozes.
- Pucher, M., Schuchmann, G., & Fröhlich, P. (2009). Regionalized text-to-speech systems: Persona design and application scenarios. In: A. Esposito, A. Hussain, M. Marinaro, & R. Martone, *Multimodal signals: Cognitive and algorithmic issues* (pp. 216-222). Berlin - Heidelberg: Springer. .
- Quazza, S., Donetti, L., Moisa, L., & Salza, P. (2001). Actor@: a multilingual unitselection speech synthesis system. *Proceedings of 4th ISCA Tutorial and Research Workshop on Speech Synthesis 2001*. Edimburgh.
- Rameh. (1961). *Contrastive Analysis of Portuguese Intonation*. Dissertation (MA). Universidade de Georgetown.
- Reed, W. D., & Leite, Y. (1956). The segmental phonemes of Brazilian Portuguese: Standard Paulista Dialect. In K. Pike, *Phonemics: a technique for reducing language to writing* (pp. 194-201.). Ann Arbor: The University of Michigan Press, 5ª edição.

- Roche, E., & Schabes, Y. (1995). *Exact Generalization of Finite-State Transductions: Application to Grapheme-to-Phoneme Transcription*. Technical Report TR-95-08. Cambridge, USA: Mitsubishi Electric Research Laboratories.
- Sagisaka, Y. (1997). Overview - Spoken Output Technologies. In: R. A. Cole, J. Mariani, H. Uszkoreit, A. Zaenen, & V. Zue, *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press.
- Sejnowski, T., & Rosenberg, C. R. (1987). Parallel networks that learn to pronounce English Text. *Complex Systems*, v. 1 , 145-168.
- Silva, F. (1997). *O processo de monotongação em João Pessoa*. (MA). João Pessoa: UFPB.
- Silva, T. C. (1999). *Fonética e Fonologia do português: roteiro de estudos e guia de exercícios*. 4ª. Ed. São Paulo: Contexto.
- Simões, F. O. (1999). *Implementação de um sistema de conversão texto-fala para o português do Brasil*. Dissertation (MA). Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas.
- Souza, E. G. (1994). *Para a caracterização fonéticoacústica da nasalidade no português do Brasil*. Dissertation (MA). Campinas: IEL/UNICAMP.
- Sproat, R. (1997). *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*. Dordrecht: Kluwer Academic Publishers.
- Styger, T., & Keller, E. (1994). *Fundamentals of speech synthesis and speech recognition: Basic concepts, state of the art, and future challenges*. Chichester: John Wiley & Sons.
- Taylor, P. (2005). Grapheme-to-phoneme conversion using Hidden Markov models. *Proceedings of Interspeech 2005*. (pp. 1973–1976). Lisboa, Portugal.
- Taylor. (2009). *Text-To-Speech synthesis*. Cambridge: Cambridge University Press.
- Teixeira, A., Martinez, R., Silva, L., M.T., L., & Vaz, F. (2004). Articulatory synthesis of Portuguese. *The International Workshop Dedicated to the Memory of Farkas Kempelen (Wolfgang von Kempelen)*. Budapest.
- Van Santen, J. (1994). Using statistics in text-to-speech system construction. *Proceedings of the ESCA/IEEE Workshop on Speech Synthesis*, (pp. 240-243). Mohonk NY.
- Wells, J. C. (1994). Computer-coding the IPA: a proposed extension of SAMPA. *Speech, Hearing and Language, Work in Progress*. (University College London, Department of Phonetics and Linguistics). V. 8 , 271-289.
- Whalen, D. (1990). Coarticulation is largely planned. *Journal of Phonetics* 18 , 3-35.





Education and Culture

**Erasmus Mundus**