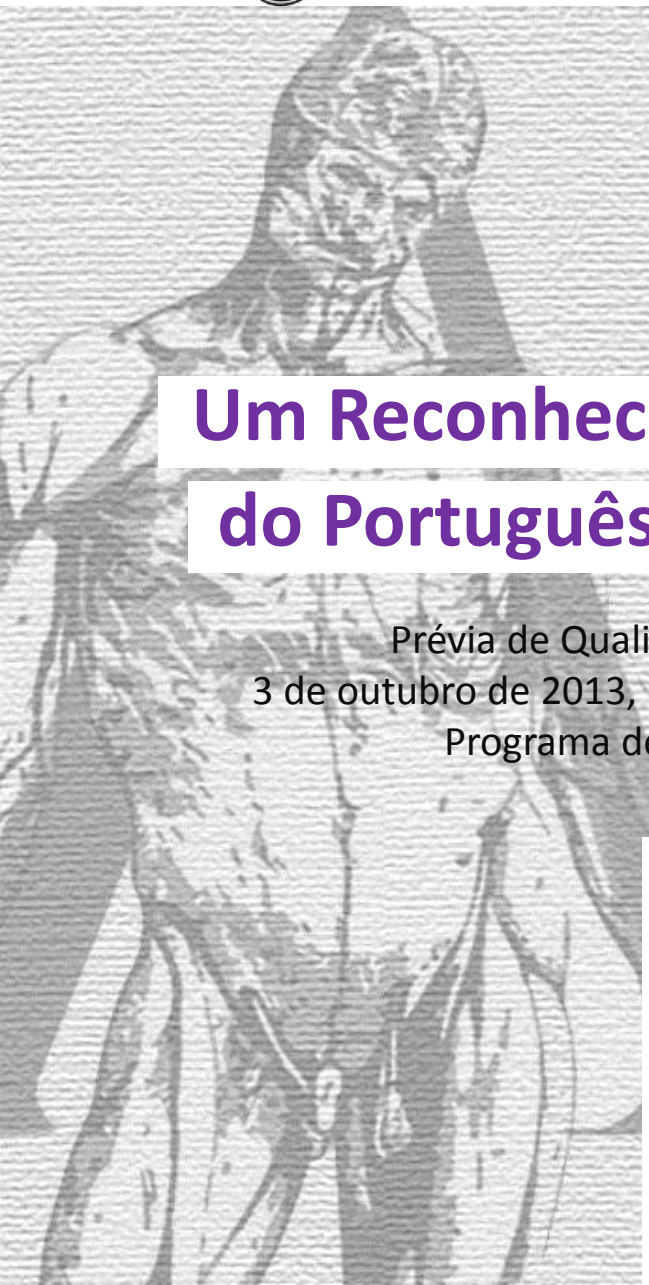


Listener:

Um Reconhecedor de Pronúncia para Falantes do Português Brasileiro Aprendizes de Inglês

Prévia de Qualificação (Introdução e Revisão Bibliográfica) apresentada em 3 de outubro de 2013, como trabalho da disciplina Metodologia em IA 2º/2013, no Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional do ICMC/USP.

Gustavo Augusto de Mendonça Almeida (USP)

gustavoauma@gmail.com

Orientadora: Profa. Dra. Sandra Maria Aluisio (USP)

sandra@icmc.usp.br

Co-orientador: Prof. Dr. Aldebaro Klautau Jr. (UFPA)

aldebaro.klautau@gmail.com

0. *Overview*

Seção 1: Introdução

Motivação, Objetivo, *Gap* e Hipótese de Pesquisa, Medidas de Avaliação

Seção 2: Fundamentação Teórica

2.1: Aquisição de Segunda Língua (ASL)

2.2: Reconhecimento de Pronúncia

Seção 3: Trabalhos Relacionados

3.1: Adaptações no Modelo Acústico (MA)

3.2: Adaptações no Modelo de Pronúncia (MP)

3.3: Adaptações no Modelo de Língua (ML)

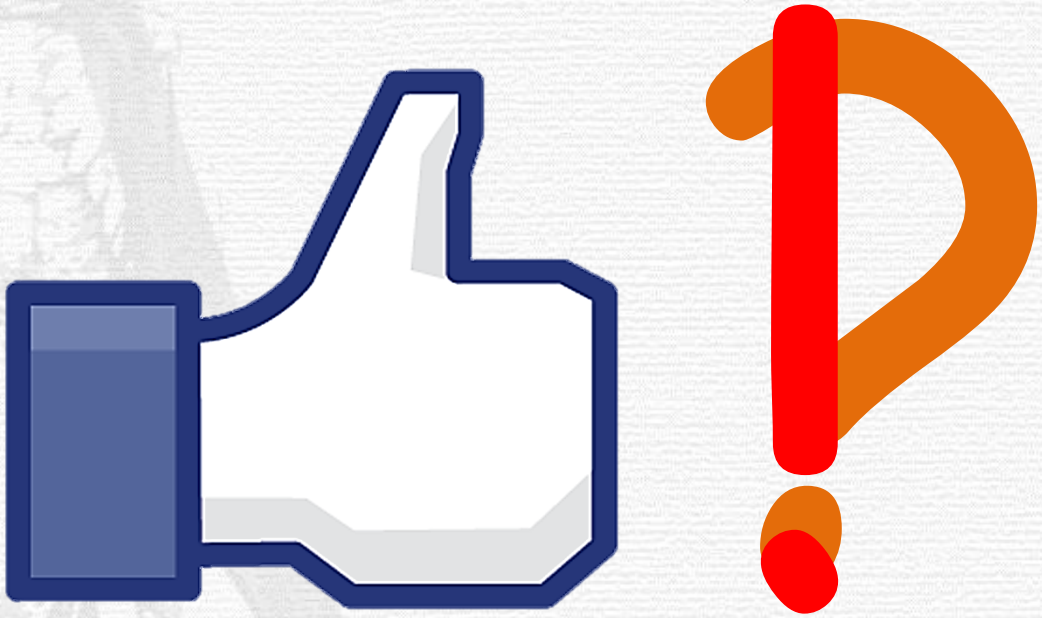
Seção 4: Considerações Finais

Seção 5: Referências Bibliográficas

Seção 1: Introdução

1. Introdução

QUAL O NÍVEL DE CONHECIMENTO DE INGLÊS DOS BRASILEIROS?



1. Introdução

Em uma pesquisa realizada pela *Global English* (2013), envolvendo 137.000 informantes sobre o conhecimento de inglês em empresas, o Brasil ocupou a 71ª posição em um ranking de 77 países.

(GLOBAL ENGLISH, 2013)

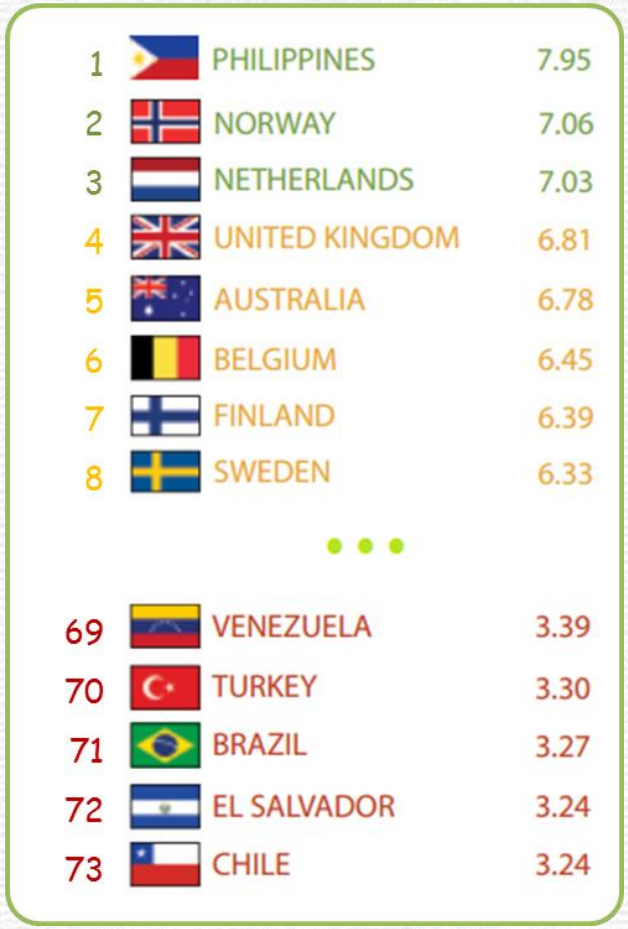


Figura 1. Ranking parcial da *Global English* (2013).

1. Introdução

O desempenho dos brasileiros correspondeu ao nível *beginner*, que constitui a pior das faixas consideradas pela pesquisa.

Essa faixa delimita indivíduos com conhecimento de inglês iniciante, com capacidades comunicativas bastante limitadas.

(GLOBAL ENGLISH, 2013)

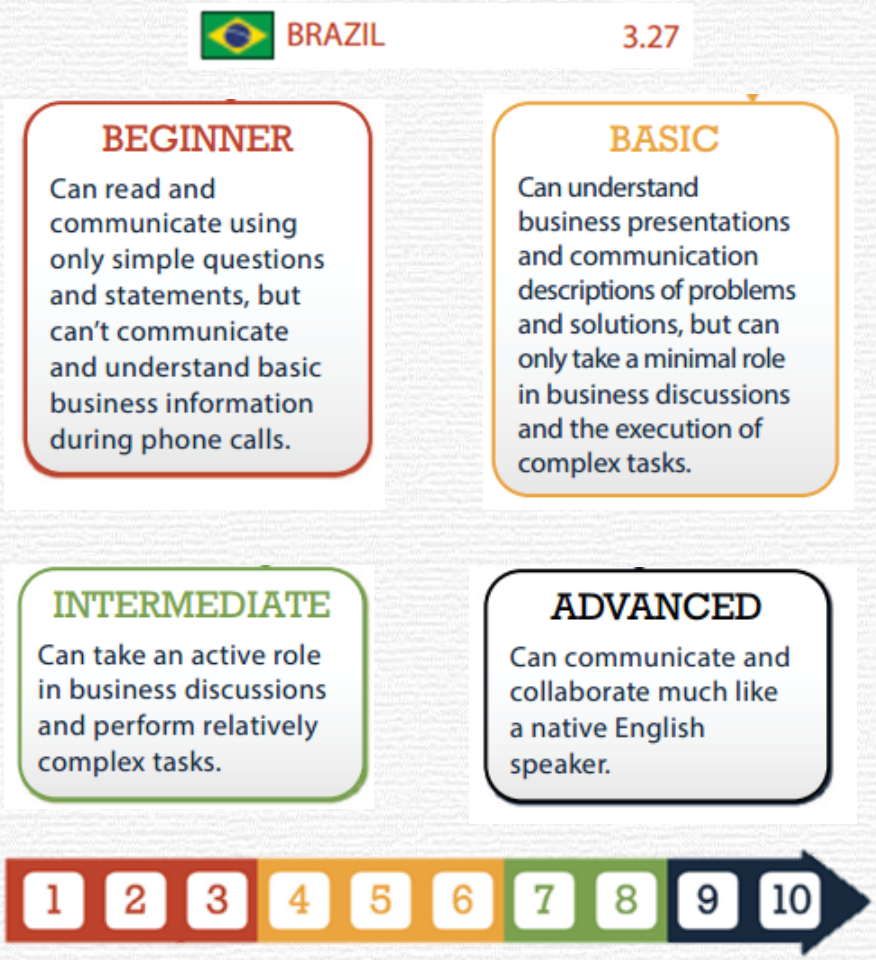


Figura 2. Faixas de desempenho consideradas pela *Global English* (2013).

1. Introdução

No Índice de Proficiência em Inglês, estabelecido pela agência *Education First* (EF), o Brasil, em 2012, foi classificado na 46ª posição de 54 países, sendo agrupado na faixa de países com proficiência muito baixa em inglês.

(EDUCATION FIRST, 2012)

Proficiência muito baixa

Classificação	País	EF EPI
39	Chile	48.41
40	Venezuela	47.50
41	El Salvador	47.31
42	Síria	47.22
43	Equador	47.19
44	Argélia	47.13
45	Kuwait	47.01
46	Brasil	46.86
47	Guatemala	46.66
48	Egito	45.92
49	Emirados Árabes Unidos	45.53
50	Colômbia	45.07
51	Panamá	44.68
52	Arábia Saudita	44.60
53	Tailândia	44.36
54	Líbia	42.53

Figura 3. Ranking de países com proficiência muito baixa.

1. Introdução

ÍNDICE DE PROFICIÊNCIA EM INGLÊS – EDUCATION FIRST (2012)

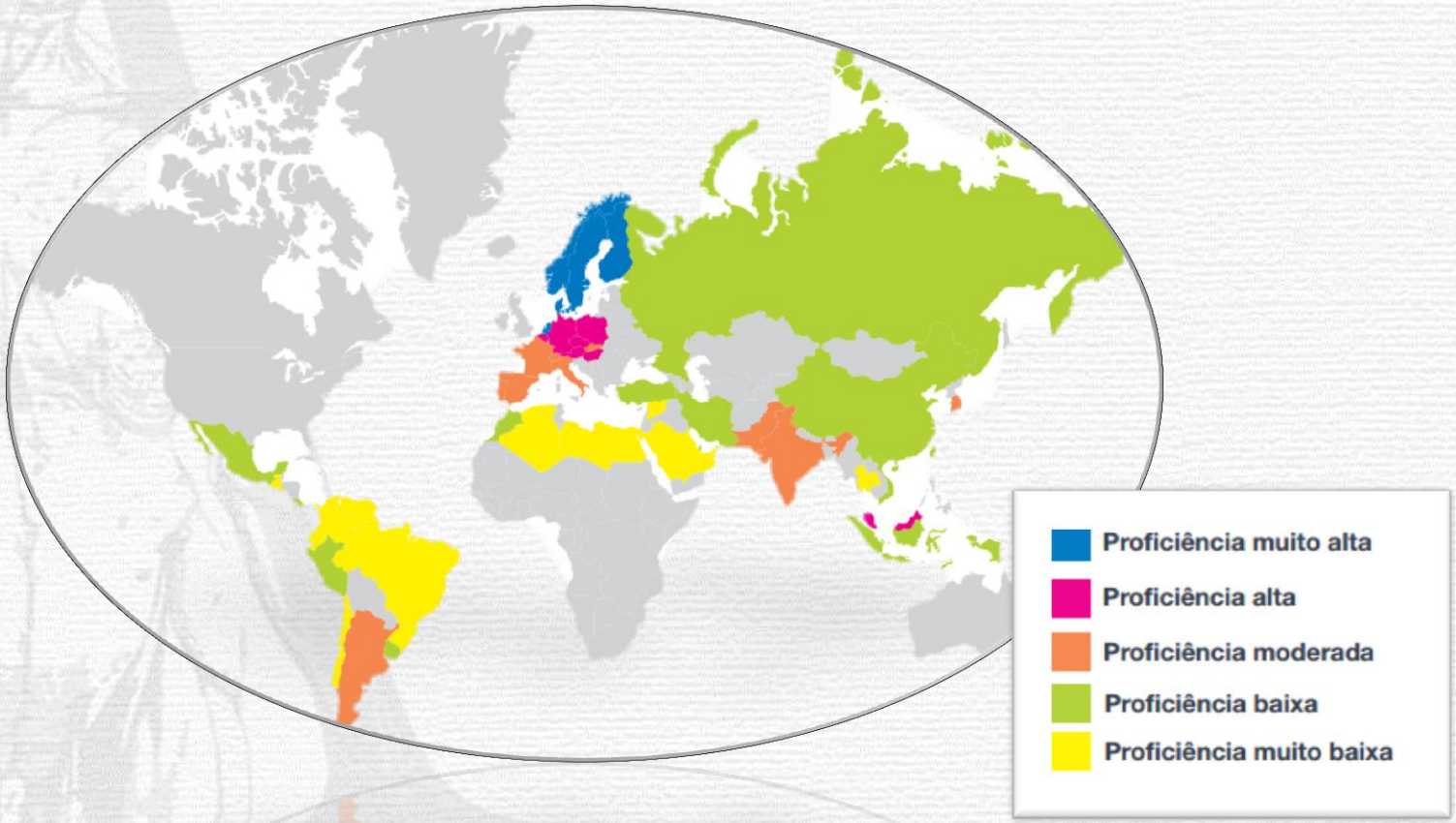


Figura 4. Mapa de Proficiência em Inglês.

(EDUCATION FIRST, 2012)

1. Introdução

Este projeto busca trazer contribuições para a melhoria desses índices. O objetivo é desenvolver um reconhecedor de pronúncia para falantes do português brasileiro (PB) aprendizes de inglês, chamado *Listener*, que seja capaz de fornecer ao usuário *feedback*, em tempo real, sobre a qualidade de sua pronúncia.

Recursos semelhantes já foram desenvolvidos para outras línguas, como o japonês (TSUBOTA et al., 2004), o espanhol (REIS & HAZAN, 2011), o holandês (STRIK et al., 2008; NERI et al., 2003) e o francês (GENEVALOGIC, 2006).

No entanto, para o PB, há ainda uma lacuna a ser explorada.

(TSUBOTA et al., 2004; REIS & HAZAN, 2011; STRIK et al., 2008; NERI et al., 2003; GENEVALOGIC, 2006)

1. Introdução

A hipótese de pesquisa é que é possível construir um reconhecedor de fala eficiente para analisar a pronúncia de inglês de falantes nativos do PB, através de:

(i) uma classificação de erros de pronúncia que leve em conta a transferência de padrões de L1 para L2;

(ii) um modelo acústico que agregue dados de fala do inglês tanto de nativos, quanto de aprendizes;

(iii) um dicionário de pronúncia que contenha a transcrição das pronúncias desviantes do aprendiz;

(iv) um modelo de língua que condiga com a sintaxe do aprendiz.

A eficiência do *Listener* será verificada a partir de medidas tradicionais para avaliação intrínseca de reconhecedores de fala: *Word Error Rate* (WER), *Character Error Rate* (CER) e Matrizes de Confusão para Fones e Palavras.

1. Introdução

A eficiência do reconhecedor de pronúncia será mensurada de modo intrínseco/*in vitro*, através das medidas:

- *Word Error Rate* (WER)

$$WER = \frac{S + D + I}{N}$$

- *Character Error Rate* (WER)

$$CER = \frac{C}{N}$$

- Matrizes de confusão de fones e palavras

	ϕ_1	ϕ_2	ϕ_3	...	ϕ_n
ϕ_1	#rec(ϕ_1, ϕ_1)	#rec(ϕ_1, ϕ_2)	#rec(ϕ_1, ϕ_3)		#rec(ϕ_1, ϕ_n)
ϕ_2	#rec(ϕ_2, ϕ_1)	#rec(ϕ_2, ϕ_2)	#rec(ϕ_2, ϕ_3)		#rec(ϕ_2, ϕ_n)
ϕ_3	#rec(ϕ_3, ϕ_1)	#rec(ϕ_3, ϕ_2)	#rec(ϕ_3, ϕ_3)		#rec(ϕ_3, ϕ_n)
...					
ϕ_n	#rec(ϕ_n, ϕ_1)	#rec(ϕ_n, ϕ_2)	#rec(ϕ_n, ϕ_3)		#rec(ϕ_n, ϕ_n)

Tais medidas serão analisadas por meio de *ten-fold cross validation*.

(CHEN et al., 1998; GORONZY, 2002; MOSTEFA et al., 2006)

Seção 2: Fundamentação Teórica

2. Fundamentação Teórica

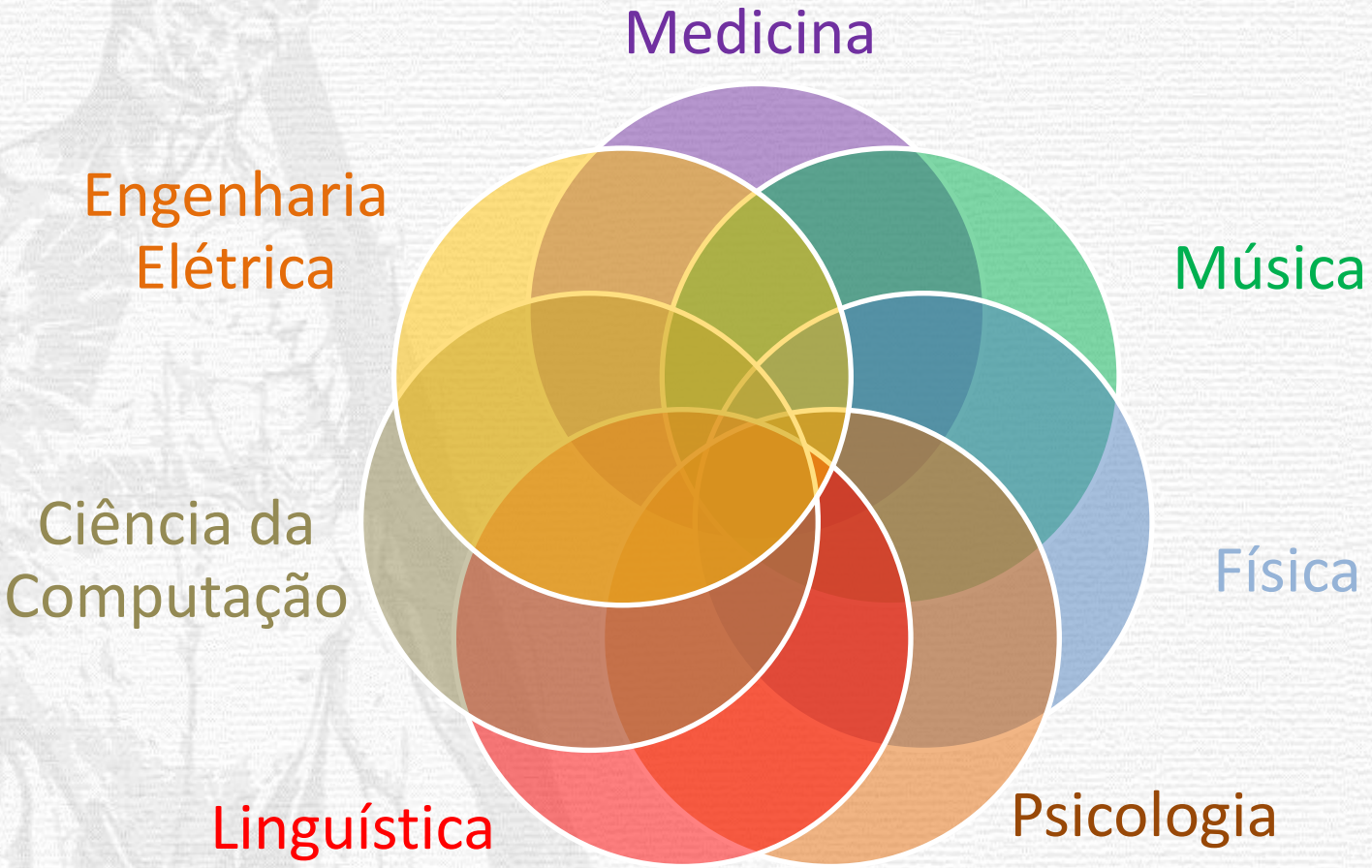


Figura 5. Áreas relacionadas ao Reconhecimento Automático de Fala.

2.1. Aquisição de Segunda Língua (ASL)

> “Quando nos deparamos com uma língua estrangeira, a tendência natural é que interpretemos seus sons a partir dos sons de nossa própria língua. Analogamente, quando falamos uma língua estrangeira, tendemos a utilizar os sons e os padrões sonoros de nossa língua nativa na realização. (WELLS, 2000)

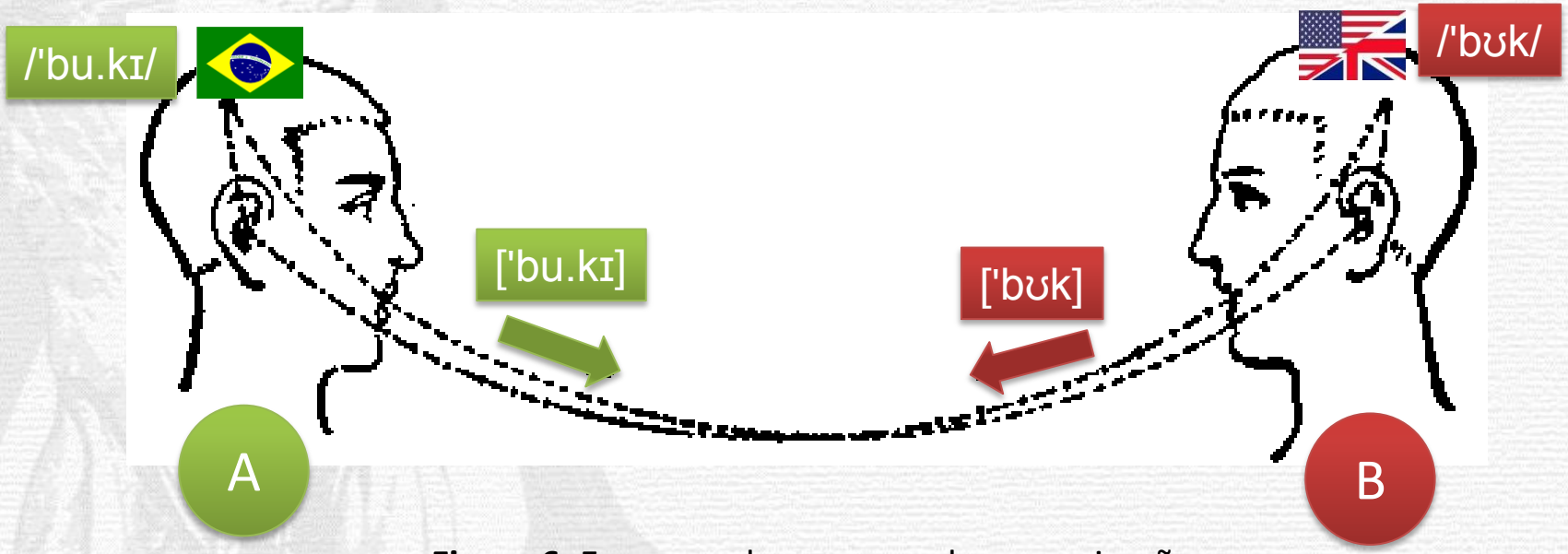


Figura 6. Esquema do processo de comunicação.

(SAUSSURE, 1916; WELLS, 2000)

2.1. Aquisição de Segunda Língua (ASL)

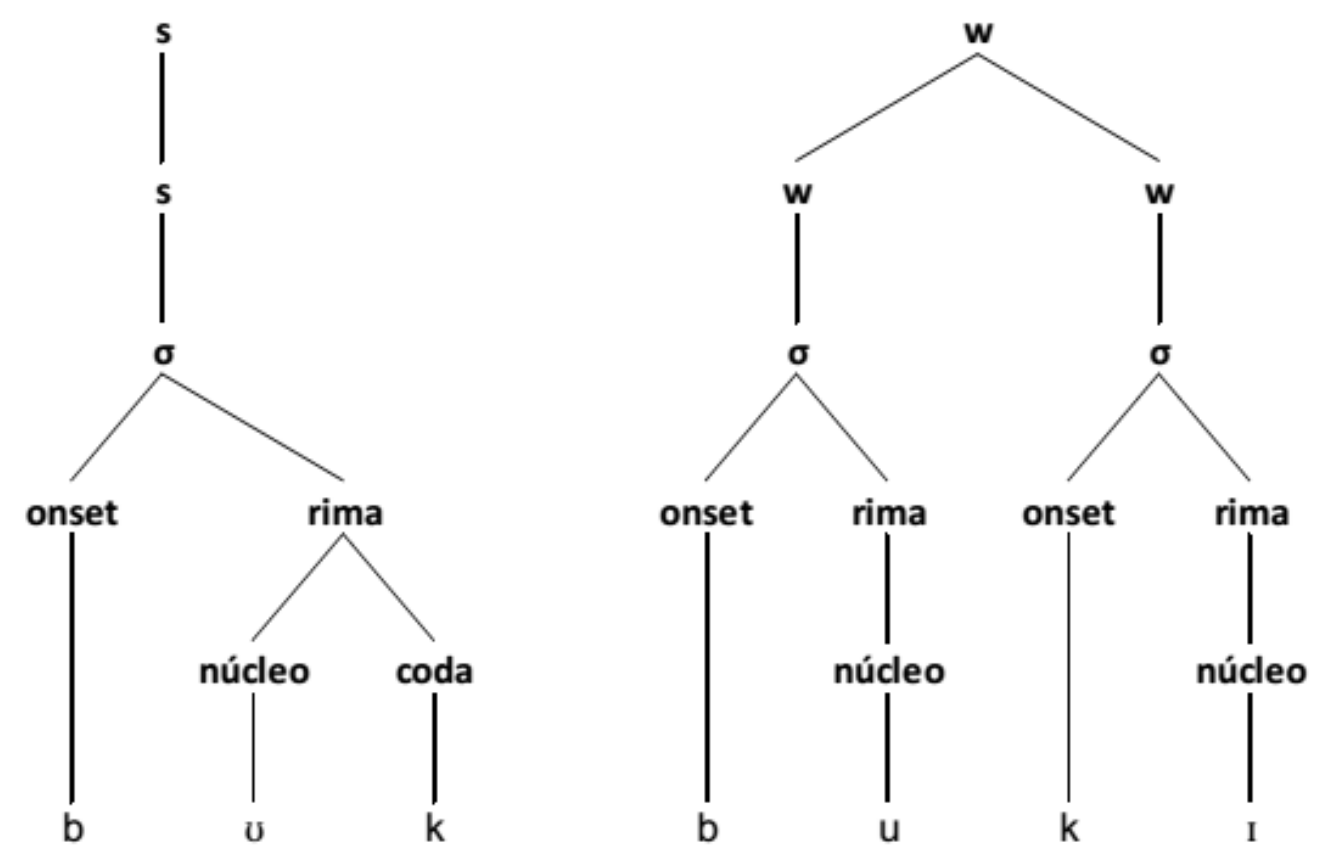


Figura 7. Realização da palavra 'book' na pronúncia padrão do inglês (esq.) e com transferência do PB para o inglês (dir.) – Representação autossegmental segundo Selkirk (1982).

(SELKIRK, 1982)

2.1. Aquisição de Segunda Língua (ASL)

No que diz respeito à pronúncia de uma língua estrangeira, há, principalmente, a transferência de padrões do sistema fonológico da L1 para a L2 e, também, a transferência de padrões de correspondência entre letra e som da L1 para a L2.

Quadro 1. Exemplo de influência de padrões de escrita na fala do aprendiz.


Forma Ortográfica	Forma Fonética *AmE*	Forma Fonética *PB>>AmE*	
time	['tʰaɪm]	['taɪ.mɪ]	
him	['hɪm]	['hĩ]	*['hĩ.mɪ]
nice	['naɪs]	['naɪ.sɪ]	
mass	['mæs]	['mɛs]	*['mɛ.sɪ]

(ZIMMER & ALVES, 2006)

2.1. Aquisição de Segunda Língua (ASL)

Quadro 2. Articulação da consoante nasal velar [ŋ].

FORMA ORTOGRÁFICA	FORMA FONÉTICA *AmE*	FORMA FONÉTICA *PB>>AmE*
king	['kɪŋ]	['kĩ.gɪ]
reading	['riː.dɪŋ]	['ri.dĩ]
writing	['raɪ.t̬ɪŋ]	['raɪ.ʧ̃ĩ]
singer	['sɪŋ.ɚ]	['sĩ.gɚ]
finger	['fɪŋ.gɚ]	['fĩ.gɚ]

Quadro 3. Articulação das consoantes fricativas dentais [θ] e [ð].

FORMA ORTOGRÁFICA	FORMA FONÉTICA *AmE*	FORMA FONÉTICA *PB>>AmE*
thank	['θæŋk]	['fẽ.kɪ]
thought	['θɑːt]	['tɔ.tʃɪ]
fought	['fɑːt]	['fɔ.tʃɪ]
then	['ðen]	['dẽ]
this	['ðɪs]	['dis]

(GODOY et al., 2006; ZIMMER et al., 2009; CRISTÓFARO-SILVA, 2012)

2.1. Aquisição de Segunda Língua (ASL)

Quadro 4. Inventário fonético consonantal do PB e do AmE.

	FONES CONSONANTAIS DO PORTUGUÊS BRASILEIRO							
	Bilabial	**Labiodental**	**Dental**	**Alveolar**	**Alveopalatal**	**Palatal**	**Velar**	**Glotal**
Oclusiva	p b			t d			k g	
Africada				tʃ dʒ				
Nasal	m			n		ɲ		
Vibrante								
Tepe				ɾ				
Fricativa		f v		s z	ʃ ʒ		x ɣ	h ɦ
Aproximante						j	w	
Lateral				l		ʎ		

	FONES CONSONANTAIS DO INGLÊS							
	Bilabial	**Labiodental**	**Dental**	**Alveolar**	**Alveopalatal**	**Palatal**	**Velar**	**Glotal**
Oclusiva	p b			t d			k g	
Africada				tʃ dʒ				
Nasal	m			n			ŋ	
Vibrante								
Tepe								
Fricativa		f v	θ ð	s z	ʃ ʒ			h
Aproximante				r		j	w	
Lateral				l				

(CRISTÓFARO-SILVA, 2005; OGDEN, 2012)

2.1. Aquisição de Segunda Língua (ASL)

Quadro 5. Inventário fonético vocálico do PB e o AmE.

	VOGAIS DO PORTUGUÊS BRASILEIRO					
	Anterior		Central		Posterior	
	Não-arr.	Arr.	Não-arr.	Arr.	Não-arr.	Arr.
Alta	i ɪ ĩ					u ʊ ũ
Média-alta	e ẽ					o õ
Média-baixa	ɛ					ɔ
Baixa			a ə ã			

	VOGAIS DO INGLÊS AMERICANO					
	Anterior		Central		Posterior	
	Não-arr.	Arr.	Não-arr.	Arr.	Não-arr.	Arr.
Alta	i: ɪ					u: ʊ
Média-alta						
Média			ə			
Média-baixa	ɛ		ɜ:		ʌ	ɔ:
Baixa	æ		a:			

(CRISTÓFARO-SILVA, 2005; OGDEN, 2012)

2.1. Aquisição de Segunda Língua (ASL)

A Linguística de Corpus é um método de investigação linguística, de base empirista, que propõe o estudo da língua a partir de exemplos reais de uso.

> “Na linguística, um *corpus* é uma coleção de textos (um “corpo” da língua) armazenado em um banco de dados eletrônico. Comumente, *corpora* são grandes coleções de textos legíveis, em formato legível por computadores, os quais contêm milhares ou milhões de palavras. (BAKER et al. 2006)

Tipos de *corpora* (KENNEDY, 1998):

- gerais/de referência vs. especializados;
- históricos vs. da língua atual;
- regionais vs. multidialetais;
- de aprendizes vs. de nativos;
- multilíngues vs. monolíngues;
- falado vs. escrito vs. transcrito.

(BAKER et al., 2006; KENNEDY, 1998)

2.2. Reconhecimento de Pronúncia

Um reconhecedor de pronúncia nada mais é do que um reconhecedor de fala voltado a uma tarefa específica, qual seja: compreender e analisar a pronúncia de um aprendiz.

> “ O propósito de um reconhecedor automático de fala (RAF) é transformar, de forma eficiente e precisa, o sinal acústico da fala em sua contraparte textual. (RABINER & SCHAFER, 2007)

$$\frac{\text{RAF}}{\text{Audição}} \propto \frac{\text{Aviação}}{\text{Vôo dos Pássaros}}$$

(RABINER & SCHAFER, 2007)

2.2. Reconhecimento de Pronúncia

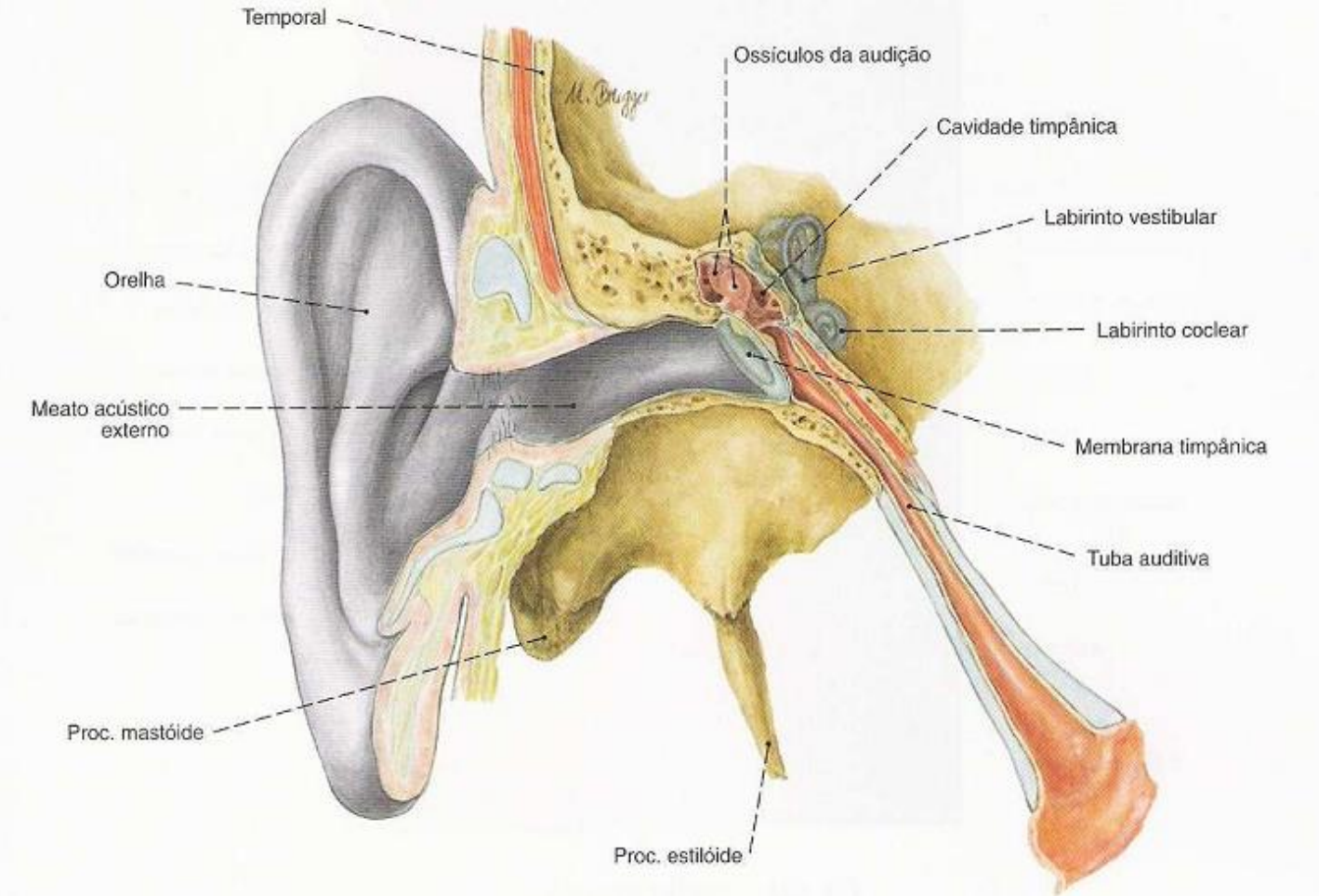


Figura 8. Visão geral do sistema auditivo humano.

(PUTZ & PABST, 2000)

2.2. Reconhecimento de Pronúncia

O paradigma majoritário em sistemas de RAF é estocástico, destacando-se, especialmente, a utilização de Modelos Ocultos de Markov, ou *Hidden Markov Models* (HMM).

Em tais modelos, a tarefa de reconhecimento é considerada a partir da metáfora do canal ruidoso, ou *noisy-channel*, em que se busca estimar, considerando-se uma língua $\mathcal{L}$, para uma sequência de palavras W, qual a sequência $\hat{W}$ mais provável, dado conjunto de estados acústicos observáveis O:

$$\hat{W} = \underset{W \in \mathcal{L}}{argmax}\, P(W|O)$$

Aplicando-se Bayes e eliminando-se o fator de normalização, obtém-se:

$$\hat{W} = \underset{W \in \mathcal{L}}{argmax}\, P(O|W)\, P(W)$$

(HUANG et al., 2001; JURAFSKY & MARTIN, 2009)

2.2. Reconhecimento de Pronúncia

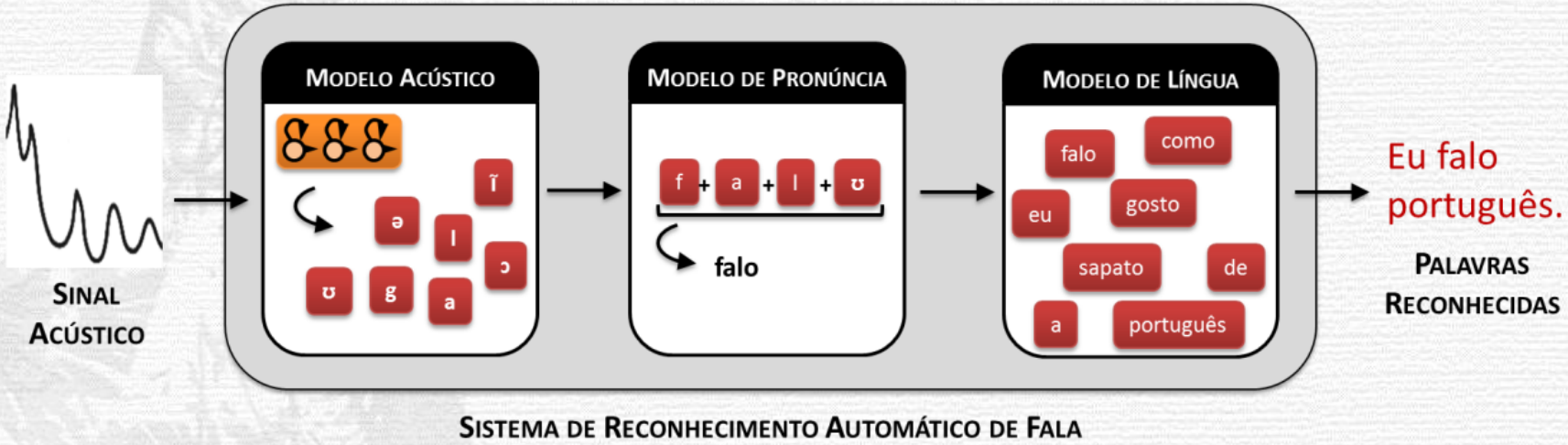


Figura 9. Arquitetura básica de um reconhecedor.

Estimada pelo Modelo Acústico (MA)

$$\hat{W} = \underset{W \in \mathcal{L}}{argmax}\, P(O|W)\, P(W)$$

Estimada pelo Modelo de Língua(ML)

(JURAFSKY & MARTIN, 2009)

2.2. Reconhecimento de Pronúncia

O PROBLEMA: Se o estado da arte em reconhecimento de fala já apresenta nível razoável de confusão para dados de fala de nativos, como reconhecer (e processar) a fala de não-nativos?

2.2. Reconhecimento de Pronúncia

Há diversas formas de se possibilitar um eficiente reconhecimento automático de fala de não-nativos, por exemplo, através do tratamento da variação nos vários do reconhecedor: no modelo acústico, no modelo de língua ou no modelo de pronúncia.

Modelo Acústico (MA)

- adaptação ao falante;
- modelos de interlíngua, ou combinados;
- modelos bilíngues.

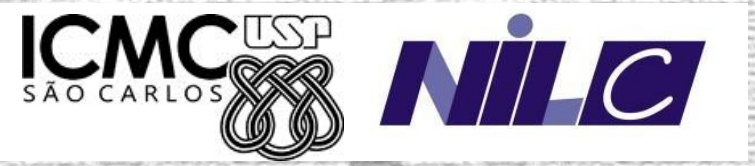
Modelo de Língua (ML)

- interpolação de modelos;
- especificação de restrições;
- utilização da informação de tópico;
- conhecimento semântico;
- modelos híbridos.

Modelo de Pronúncia (MP)

- dicionários multipronúncia.

(STRIK & CUCCHIARINI, 1999; WANG et al., 2003; KIM et al., 2008; BELLEGARDA, 2001)



Seção 3: Trabalhos Relacionados

3.1. Adaptações no Modelo Acústico (MA)

| ORIGEM DOS DADOS ACÚSTICOS DE TREINO:

| ABORDAGEM:

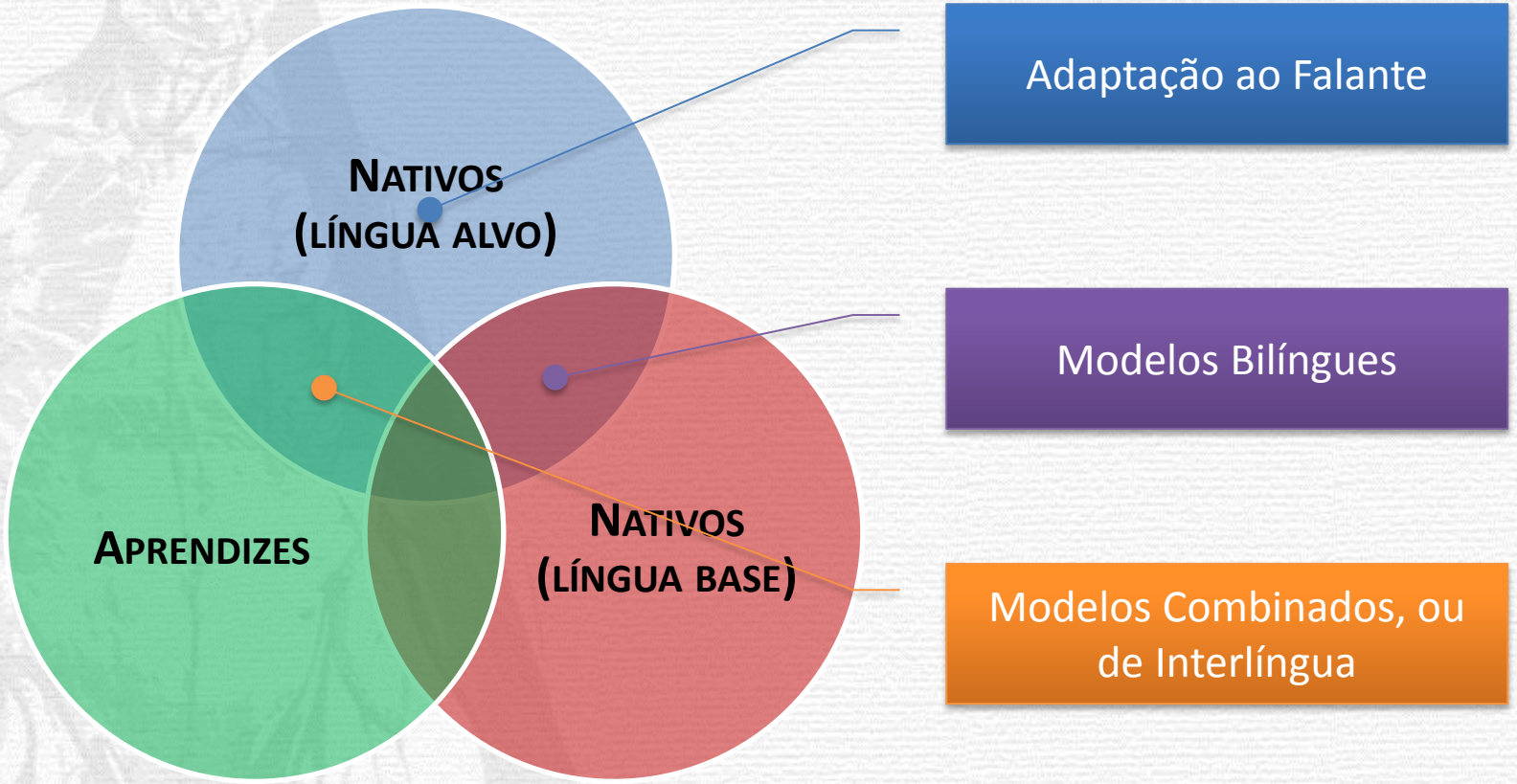


Figura 10. Abordagens para se adaptar o Modelo Acústico (MA) do reconhecedor a dados de não-nativos.

(WANG et al., 2003)

3.2. Adaptações no Modelo de Pronúncia (ML)

Quadro 6. Exemplo de entradas no dicionário de pronúncia do *VoxForge Speech Corpus,* com adição de pronúncias dos aprendizes.

ID DA PALAVRA	TRANSCRIÇÃO	FONÉTICA (ARPAbet)
DISANO	[dʒiza'pih]	d ix s aa n ow
DISANTI	[dʒiza'piɹ]	d ix s ae n t iy
DISANTIS	[diza'piɹ]	d ix s aa n t ix s
DISANTO	[DISANTO]	d ix s ae n t ow
DISAPPEAR	[dʒisa'piɹ]	d ih s ax p ih r
DISAPPEAR(2)	[DISAPPEAR]	d ih s ax p iy r
DISAPPEARANCE	[dizə'piɹ]	d ih s ax p ih r ax n s
DISAPPEARANCE(2)	[DISAPPEARANCE]	d ih s ax p iy r ax n s
DISAPPEARANCES	[disə'piɹ]	d ih s ax p ih r ax n s ix z
DISAPPEARANCES(2)	[dɪsə'pɪɹ]	d ih s ax p iy r ax n s ix z

3.2. Adaptações no Modelo de Pronúncia (ML)

Abordagem *Data-driven*

- Utilização da saída do Modelo Acústico (MA) do reconhecedor;
- Utilização de um vocabulário canônico para gerar variantes.

- Barata;
- Independente de língua;
- Facilmente replicável.

- Dependente da anotação do *corpus*;
- Pode aumentar muito a confusão do reconhecedor.

(KIM et al., 2008; WESTER, 2003)

Abordagem *Knowledge-based*

- Consulta a especialistas do domínio (linguistas);
- Consulta a base de dados já compiladas (dicionários ou tratados de pronúncia);

- Fiável;
- Específica para o propósito.

- Custosa;
- Demorada;
- Dependente de língua.

3.3. Adaptações no Modelo de Língua (ML)

> “ Modelos de língua podem ser vistos como um conjunto de restrições que é imposto às sequências de palavras de uma dada língua. (BELLEGARDA, 2003)

Modelos de língua estatísticos, baseados em *n-grama*, tendem a ser extremamente dependentes do domínio a partir do qual foram gerados.

↓

Um modelo de língua para conversas via telefone é mais eficiente se gerado a partir de **2 mi** de palavras desse gênero, do que a partir de **140 mi** de palavras do gênero apresentações de jornal via TV ou rádio.

(BELLEGARDA, 2003; ROSENFELD, 2000)

3.3. Adaptações no Modelo de Língua (ML)

Técnicas de Adaptação de um Modelo de Língua (ML)

- interpolação de modelos;
- especificação de restrições;
- utilização da informação de tópico;
- conhecimento semântico;
- modelos híbridos.

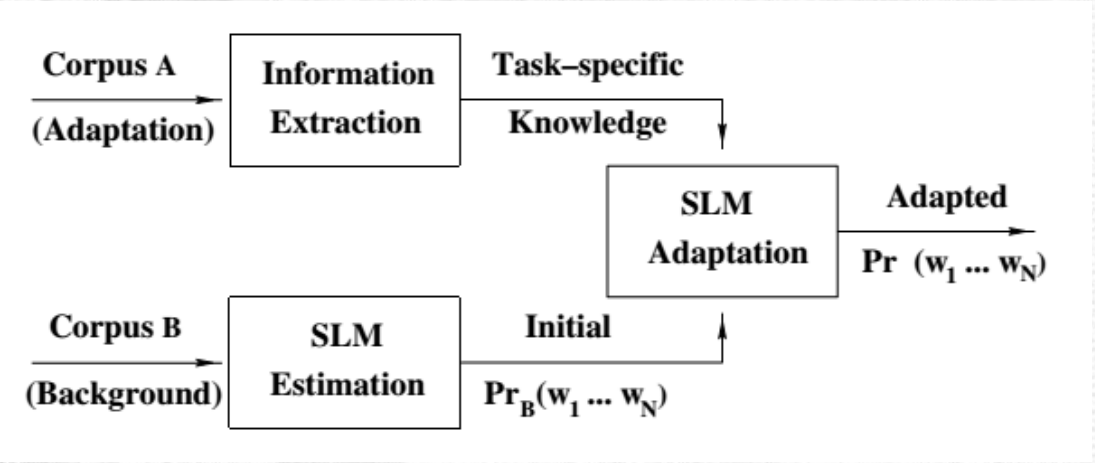


Figura 10. Esquema geral de adaptação de um Modelo de Língua Estatístico (SLM).

(BELLEGARDA, 2003)

Seção 4: Considerações Finais

4. Considerações Finais

Os problemas verificados com a revisão bibliográfica...

4. Considerações Finais

A DIFICULDADE DE AVALIAR OS MÉTODOS:

- As taxas de WER e CER reportadas não são, imediatamente, comparáveis, dado a diferença de *corpora* e língua empregados;
- A precisão de um reconhecedor de fala está atrelada à função para o qual ele foi concebido, de modo que reconhecedores destinados a propósitos distintos não podem ser avaliados com base um mesmo critério;
- As métricas existentes talvez não sejam tão boas;
- Não há um *benchmark*.

A DIFICULDADE DE SE OBTER DADOS DE FALA:

- Há poucos *corpora* de fala disponíveis;
- Os *corpora* de fala são caros e, em muitas vezes, sua qualidade e robustez não são ótimas;
- Compilá-los demanda MUITO trabalho.

(WANG et al., 2003)

4. Considerações Finais

A DIFICULDADE DE LIDAR COM A VARIAÇÃO LINGUÍSTICA:

- As línguas naturais são dinâmicas e modificam sua estrutura a todo o tempo, lidar com toda essa variação no reconhecimento de fala é problemático;
- Levantamentos linguísticos, raramente, são feitos de forma computacionalmente aplicável, sendo necessárias adaptações.

(WANG et al., 2003)

Obrigado!

Gustavo Augusto de Mendonça Almeida (USP)
gustavoauma@gmail.com

Orientadora: Profa. Dra. Sandra Maria Aluisio (USP)
sandra@icmc.usp.br

Co-orientador: Prof. Dr. Aldebaro Klautau Jr. (UFPA)
aldebaro.klautau@gmail.com

cenas de um próximo capítulo...

listener

Gustavo Augusto de Mendonça Almeida (USP)
gustavoauma@gmail.com

Orientadora: Profa. Dra. Sandra Maria Aluisio (USP)
sandra@icmc.usp.br

Co-orientador: Prof. Dr. Aldebaro Klautau Jr. (UFPA)
aldebaro.klautau@gmail.com