



UNIVERSIDADE
ESTADUAL DE LONDRINA

EDIO ROBERTO MANFIO

**AVALIAÇÃO DE DISPOSITIVOS ACIONADOS POR VOZ E
TEXTO PARA O PORTUGUÊS BRASILEIRO**

Londrina
2016

EDIO ROBERTO MANFIO

**AVALIAÇÃO DE DISPOSITIVOS ACIONADOS POR VOZ E
TEXTO PARA O PORTUGUÊS BRASILEIRO**

Tese apresentada à Banca de Qualificação do Programa de Pós-graduação em Estudos da Linguagem (Doutorado) da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Doutor em Estudos da Linguagem.

Orientadora: Profa. Dra. Vanderci de Andrade Aguilera

Londrina
2016

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Manfio, Edio Roberto .

Avaliação de dispositivos acionados por voz e texto para o Português Brasileiro / Edio Roberto Manfio. - Londrina, 2016.
121 f. : il.

Orientador: Vanderci de Andrade Aguilera.

Tese (Doutorado em Estudos da Linguagem) - Universidade Estadual de Londrina, Centro de Letras e Ciências Humanas, Programa de Pós-Graduação em Estudos da Linguagem, 2016.

Inclui bibliografia.

1. Comando de voz - Teses. 2. ALiB - Teses. 3. Geolinguística - Teses. 4. Processamento de Linguagem Natural - Teses. I. Aguilera, Vanderci de Andrade . II. Universidade Estadual de Londrina. Centro de Letras e Ciências Humanas. Programa de Pós-Graduação em Estudos da Linguagem. III. Título.

EDIO ROBERTO MANFIO

AVALIAÇÃO DE DISPOSITIVOS ACIONADOS POR VOZ E TEXTO
PARA O PORTUGUÊS BRASILEIRO

Tese apresentada à Banca de Qualificação do Programa de Pós-graduação em Estudos da Linguagem (Doutorado) da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Doutor em Estudos da Linguagem.

BANCA EXAMINADORA

Orientadora: Profa. Dra. Vanderci de Andrade
Aguilera
Universidade Estadual de Londrina- UEL

Profa. Dra. Dircel Aparecida Kailer
Universidade Estadual de Londrina- UEL

Profa. Dra. Cinthyan Renata Sachs Camerlengo de
Barbosa
Universidade Estadual de Londrina- UEL

Profa. Dra. Rosa Maria Viccari
Universidade Federal do Rio Grande do Sul -
UFRGS

Prof. Dr. Abdelhak Razky
Universidade Federal do Pará - UFPA

Londrina, 15 de março de 2016

MANFIO, Edio Roberto. **Avaliação de dispositivos acionados por voz e texto para o Português Brasileiro**. 2016. 121f. Tese (Doutorado em Estudos da Linguagem) – Universidade Estadual de Londrina, Londrina, 2016.

RESUMO

O objetivo deste trabalho foi avaliar como equipamentos acionados com *comandos por voz* reagem a diferentes realizações linguísticas de determinadas regiões do Brasil. Embora dispositivos comandados por voz não sejam novidade, a popularização dos computadores e o barateamento de algumas tecnologias permitiram que tais recursos se tornassem mais presentes no cotidiano das pessoas. A hipótese, no entanto, previa que apenas uma minoria desses equipamentos pode ser acionada de modo eficiente por qualquer falante do Português Brasileiro sem exigir treinamento prévio quanto à pronúncia ou adições constantes de registros em áudio para as bibliotecas dos softwares envolvidos. A realização deste estudo baseou-se em dados contidos no Projeto ALiB - Atlas Linguístico do Brasil pelo fato de que um equipamento desta natureza teoricamente deve funcionar bem com realizações de fala de qualquer região do Brasil. Paralelamente à pesquisa bibliográfica, três representantes da categoria equipamentos/dispositivo foram testados: o *IBM Via Voice*, o buscador por voz da Google e o *Voice Recognition Module V2*. Tais testes foram realizados à luz de teorias voltadas às áreas de Sociolinguística, Geolinguística Pluridimensional, Processamento de Linguagem Natural e outras interdisciplinares e diretamente relacionadas à área de Desenvolvimento de Sistemas. Por extensão, abriu-se a oportunidade de desenvolver um aplicativo denominado Tical com o qual todos os testes feitos na tríade de experimentos citados puderam ser recriados, dando ainda mais lastro àquilo que se buscava demonstrar: conseguiu-se um projeto de aplicação amplo e ao mesmo tempo um objeto de pesquisa.

Palavras-chave: Comandos por voz. ALiB. Geolinguística Pluridimensional. Processamento de Linguagem Natural.

MANFIO, Edio Roberto. **Device evaluation driven by voice and text to Brazilian Portuguese.** 2016. 121p. Thesis (Doctorate in Language Studies) – Universidade Estadual de Londrina, Londrina, 2016.

ABSTRACT

The aim of this work was to evaluate how equipment fired with voice commands react to different linguistic realizations of certain regions of Brazil. Although voice-controlled devices are not new, the popularization of computers and the cheapening of some technologies have allowed such resources to become more present in the daily lives of people. The hypothesis, however, predicted that only a minority of such equipment can be thrown effectively by any Brazilian Portuguese speaker without requiring prior training regarding pronunciation or constant additions of audio records for libraries of software involved. This study was based on data contained in the ALiB Project - Linguistic Atlas of Brazil - by the fact that such equipment theoretically should work fine with speech achievements for any region of Brazil. In addition to bibliographical research, three representatives of the equipment/device category were tested: the IBM Via Voice, Google voice search and Voice Recognition Module V2. Such tests were carried out in the light of theories pointing to the areas of Sociolinguistics, Pluridimensional Geolinguistics, Natural Language Processing and others interdisciplinary and directly related to the Systems Development area. By extension, opened up the opportunity to develop an application named Tical with which all tests done in the triad of experiments cited could be rebuilt, giving even more ballast to what if sought to demonstrate: a large application project and at the same time an object of research.

Keywords: Voice commands. ALiB. Pluridimensional Geolinguistics. Natural Language Processing.

LISTA DE FIGURAS

Figura 01 - Disposição de áreas e subáreas do PLN.....	29
Figura 02 - Tela de apresentação do <i>DosVox</i>	47
Figura 03 - Menu primário do <i>DosVox</i>	47
Figura 04 - Menu de arquivos do <i>DosVox</i>	48
Figura 05 - Utilização de sistema de áudio e captação por microfone externo.....	64
Figura 06 - Utilização de um cabo de áudio.....	64
Figura 07 - Fragmento de áudio com a palavra luz sendo copiado.....	65
Figura 08 - O ícone do ‘microfone’ em destaque.....	72
Figura 09 - <i>Voice Recognition Module V2</i>	83
Figura 10 - Gabinete cromático.....	85
Figura 11 - Impressão de tela inteira do Professor Tical.....	91
Figura 12 - Falhas de digitação, grafia e pontuação na pergunta.....	97
Figura 13 - Interface convencional de Tical com uma amostra de pergunta e resposta.....	101
Figura 14 - Exemplo de interface com consultas feitas a Tical pelo WhatsApp.....	102
Figura 15 - Exemplo de consulta pelo WhatsApp com mais de 15 palavras.....	103
Figura 16 - Tela exibindo uma lista disponível no banco de dados.....	107

LISTA DE TABELAS

Tabela 01 - Critérios para escolha dos equipamentos analisados.....	14
Tabela 02 - Gênese da Dialetoлогия no Brasil e relação de coexistência com a Geolinguística.....	25
Tabela 03 - Relação do texto e pausas por hifens e vírgulas.....	36
Tabela 04 - Relação entre a frase, ênfase e modalidade.....	37
Tabela 05 - Nomes dos arquivos e respectivas transcrições.....	50
Tabela 06 - Resultado do experimento.....	67
Tabela 07 - Critério para os indicadores.....	67
Tabela 08 - Smartphones.....	71
Tabela 09 - Grandezas elétricas e componentes eletrônicos.....	74
Tabela 10 - Automóveis e peças de reposição.....	75
Tabela 11 - Domínios da Linguística e variações.....	75
Tabela 12 - Variações para tangerina - cartas L05a a l05e	76
Tabela 13 - Cores e palavras com proximidade fônica.....	76
Tabela 14 - Conjunto de cinco possíveis comandos por voz.....	79
Tabela 15 - Comparativo entre comandos.....	81
Tabela 16 - Possíveis comandos por voz em uma sala de estar.....	82
Tabela 17 - Amostra de possíveis grupos com cinco peças (comandos) cada	84
Tabela 18 - Relação dos comandos e cores utilizados no experimento	86
Tabela 19 - Variantes linguísticas dos comandos utilizados no experimento	87
Tabela 20 - Comparativos entre as peças de áudio e execuções aproximadas ou similares	87
Tabela 21 - Paralelo demonstrativo de perguntas e respostas.....	94
Tabela 22 - Paralelo demonstrativo de entradas e respostas x possibilidades.....	96
Tabela 23 - Paralelo demonstrativo entre palavra e sinônimo.....	96
Tabela 24 - Bloco de perguntas equivalentes já cadastrado e sua resposta.....	96
Tabela 25 - Sinônimos para perguntas sobre o ALiB.....	97
Tabela 26 - Entradas e respostas sobre o ALiB.....	97
Tabela 27 - Equipamentos analisados e Tical.....	112

SUMÁRIO

INTRODUÇÃO	09
1 PROCEDIMENTOS TEÓRICO-METODOLÓGICOS	12
2 ESTUDOS SOBRE VARIAÇÃO DO PORTUGUÊS BRASILEIRO	16
2.1 DIALETOLOGIA	16
2.2 ESTUDOS DIALETOLÓGICOS E O PRESENTE TRABALHO	19
2.3 VARIAÇÃO E TECNOLOGIA	20
2.4 ATLAS LINGÜÍSTICOS	24
2.5 ATLAS LINGÜÍSTICO DO BRASIL – ALiB	26
3 RELAÇÃO ENTRE VARIAÇÃO LINGÜÍSTICA E PROCESSAMENTO DE SINAIS DA FALA	29
3.1 COMO FUNCIONAM ALGUNS FONEMAS NO APLICATIVO <i>BALABOLKA</i>	32
3.1.1 <i>O Balabolka</i>	33
3.1.2 <i>Balabolka</i> e a Prosódia	34
3.1.3 As Realizações de Fala e o Registro	37
3.1.4 Alguns dos Fonemas	40
3.2 ASPECTOS FONÉTICOS NO <i>DOSVOX</i> ENQUANTO APLICATIVO TIPO TEXTO-FALA	42
3.2.1 <i>O DosVox</i>	44
3.2.2 Estabelecendo Comparações	45
3.3 PROCESSAMENTO DE LINGUAGEM NATURAL, PROCESSAMENTO DE SINAIS DA FALA, GEOLINGÜÍSTICA	53
3.3.1 PLN, PSF e Geolinguística	54
4 DADOS DO ALiB EM FUNCIONAMENTO: PLN, ROBÔS DE CONVERSAÇÃO E INTELIGÊNCIA ARTIFICIAL	59
4.1 O ALiB NO <i>IBM VIA VOICE</i> : PESQUISA GEOLINGÜÍSTICA ASSOCIADA A APLICATIVO DE RECONHECIMENTO DE FALA	59
4.1.1 O ALiB e Algumas possibilidades em Âmbito Tecnológico	59
4.1.2 O <i>IBM Via Voice</i> e Outras possibilidades	62
4.2 TECNOLOGIAS ON-LINE COM COMANDOS POR VOZ	70

4.2.1	As Muitas Tecnologias Disponíveis.....	70
4.2.2	O Buscador por Voz da Google.....	72
4.2.3	Testes e Resultados.....	73
4.3	RELAÇÃO ENTRE RECONHECIMENTO E COMPREENSÃO DE VOZ: EXPERIMENTO PARA ANÁLISE LINGUÍSTICA	77
4.4	PROFESSOR TICAL EM TRÊS ESTÁGIOS.....	90
4.4.1	Robô de Conversação Sobre Dialetologia e Geossociolinguística.....	91
4.4.2	Tical no WhatsApp.....	99
4.4.3	Tical por Comandos e Síntese de Voz.....	104
5 CONSIDERAÇÕES FINAIS.....		111
REFERÊNCIAS		114
ANEXOS		121
ANEXO A – Endereços Eletrônicos dos Arquivos de Áudio Reproduzidos pelo <i>Balabolka</i>		121

INTRODUÇÃO

Uma boa teoria científica descreverá uma ampla gama de fenômenos com base em poucos postulados simples e fará previsões claras que possam ser testadas. Se as previsões concordarem com as observações, a teoria sobreviverá ao teste (...) se as observações discordarem das previsões, a teoria terá de ser descartada ou modificada (HAWKING, 2009, p. 31)

Uma infinidade de equipamentos acionados com *comandos por voz*¹ está presente no cotidiano de muitas pessoas de diferentes países do mundo, criados, sobretudo, nas últimas três décadas. Entre eles encontram-se desde celulares multifuncionais ou smartphones, funções básicas em automóveis, aplicativos de computadores, máquinas industriais ou robôs cirúrgicos até brinquedos robóticos voltados exclusivamente ao entretenimento. Alguns aparelhos de telefonia móvel, por exemplo, têm sistemas de *reconhecimento de voz* com uma notável flexibilidade em termos de variação linguística e prosódica² – *acento, ritmo, tons, entoação, pausa, tessitura, quantidade/duração* entre outros (CAGLIARI; CAGLIARI, 2001; LOPES, 1999). Dentre estes equipamentos acionados com comandos por voz, há aqueles que são notadamente de uso público - como sistemas de atendimento em telefonia e de busca, embarcados em portáteis - e que estão se mostrando essenciais no século XXI. Desses, principalmente, espera-se um funcionamento minimamente adequado quando em contato com os diferentes registros³ linguísticos e variedades⁴ da fala em qualquer língua e isso obviamente deve valer para o Português Brasileiro.

Contudo, tais equipamentos são desenvolvidos para operar primeiramente em grandes centros urbanos, locais onde se concentra maior renda *per capita* e, teoricamente, fala-se o Português tido como privilegiado linguisticamente, em detrimento de outras regiões ou cidades menores (FARACO, 2004; CASTILHO, 2004; BORBA, 1991).

Dessa forma, torna-se interessante verificar se as realizações linguísticas de outra região menos privilegiada economicamente acionam o dispositivo com a mesma eficácia, afinal, as variantes individual ou regional não resultam da aleatoriedade de uma utilização

¹ *Comando por voz* não é o mesmo que *reconhecimento de voz* e/ou *reconhecimento de fala*, apesar de depender diretamente desses dois últimos.

² Importante lembrar que, ainda que tenham um papel distintivo no que diz respeito ao sentido dos enunciados eles não são fonemas, mas realizações distintivas deles (LOPES, 1999).

³ Neste texto entendemos *registro* como cada uma das modalidades linguísticas adotadas pelos falantes nos diferentes contextos em que se manifestam, muitas vezes como uso individual e com claras distinções estilísticas (CAMARA, 2004).

⁴ Entendemos *variedade* neste contexto como dialeto classe-social em que diferenças gramaticais existentes entre os falantes dão pistas sobre seus contextos sociais (TRUDGILL, 1974).

arbitrária e/ou inconsciente dos falantes, mas de “um uso sistemático e regular de uma propriedade inerente aos sistemas lingüísticos, que é a possibilidade de variação” (CAMACHO, 2001, p. 50). É importante não esquecer, portanto, que a variação é inerente à língua que “em sentido lato se subdivide em dialetos regionais, dialetos sociais e registros” (CAMARA JR, 2004).

A relevância deste trabalho, portanto, está em avaliar o grau de abrangência em termos de variação linguística do sistema de *reconhecimento de voz* de alguns desses equipamentos, uma vez que a possibilidade ao acesso de tais recursos pode depender diretamente do modo de falar do indivíduo que dele faz o uso.

Norteia esta tese, a seguinte hipótese: apenas uma pequena parte de todos esses equipamentos pode de fato ser acionada de modo eficiente por qualquer falante do Português Brasileiro sem exigir que haja considerável treinamento prévio quanto à pronúncia ou adições constantes de novos registros em áudio para as bibliotecas⁵ de determinado software ou linguagem de programação.

Se isso for comprovado, paralelamente, é possível verificar que uma parcela dos falantes do Português Brasileiro não tem como acessar tais equipamentos de modo espontâneo⁶, natural ou livre do incômodo de um treinamento prévio⁷ – ainda que o conceito de espontaneidade ou naturalidade na fala seja muito subjetivo. Isso aumentaria ainda mais a necessidade e a importância de divulgar e aplicar em outras áreas de conhecimento os estudos já existentes sobre a Geolinguística do Brasil.

A partir dessa hipótese, algumas perguntas de pesquisa podem ser feitas: (I) os equipamentos que operam com comandos por voz funcionam bem para qualquer variação linguística no Brasil? (II) Quais são os critérios lingüísticos adotados pelos desenvolvedores desses equipamentos? (III) O que ocorre de fato é reconhecimento de voz ou outro tipo de processamento de sinais de áudio?

⁵ As bibliotecas são, *grosso modo*, um conjunto de especialidades absolutamente essenciais na maior parte das linguagens de programação. Têm esse nome justamente pelo funcionamento análogo às bibliotecas tradicionais cujo conteúdo, em potencial, pode ser consultado sempre que necessário.

⁶ Câmara Jr (2006) comenta sobre certa pronúncia mais ou menos artificial, facilmente encontrável e perceptível na dicção normativa teatral, de professoras primárias em aula, de locutores de rádio, de narradores de futebol entre outros. Curiosamente, isso também é perceptível em pessoas que utilizam ou testam alguns equipamentos acionados com comandos por voz.

⁷ Profissionais da área de informática distinguem basicamente os sistemas de comando e reconhecimento em dois tipos: **com** dependência de locutor e **sem** dependência de locutor. Isso não implica a ausência ou presença do locutor – aquele que fala ao equipamento – mas um quesito importante na funcionalidade: quanto mais amostras dispostas nos bancos de dados, mais eficiente se torna um **locutor genérico**, ou seja, menor a dependência do sistema em relação à locução.

A partir de questionamentos como estes, o presente estudo tem por objetivos específicos (i) procurar trazer informações que possam auxiliar profissionais e pesquisadores da área de Linguística a aplicar mais pontualmente teorias como Fonética e Fonologia às tecnologias de comando de voz vindouras, condição que tem o potencial de (ii) ajudar a aproximar pesquisadores das áreas de humanas e exatas tendo em vista que os profissionais das áreas mais técnicas como Desenvolvimento de Sistema e Eletrônica Aplicada podem estar concebendo projetos associados a tecnologias de *comando por voz* com conhecimentos insuficientes sobre Dialetologia, Fonética e Fonologia.

Nesse âmbito, tem por objetivo geral avaliar como alguns equipamentos acionados com comandos por *voz* reagem a diferentes realizações linguísticas do Brasil. Dispositivos comandados por voz humana não são novidade em termos de desenvolvimento tecnológico, porém, com a popularização dos computadores e o barateamento de algumas tecnologias eles estão se tornando mais eficientes e essenciais na busca pelo conforto, segurança e qualidade de vida, principalmente quando se trata de pessoas com necessidades especiais. Por esse motivo, cada vez mais editores eletrônicos de texto, automóveis, eletrodomésticos, robôs cirúrgicos entre outros são fabricados com o recurso. Resta saber se as variações - diatópicas, diastráticas, diafásicas, entre outras - estão sendo minimamente consideradas quando da concepção e construção dos aparatos tecnológicos com tal recurso.

1 PROCEDIMENTOS TEÓRICO-METODOLÓGICOS

Da mesma forma que o biólogo não se preocupa em responder perguntas complexas como o que é a vida, o lingüista não se ocupa necessariamente de responder o que é linguagem (LYONS, 1987)

Para esta Tese percorreu-se o seguinte caminho: (i) pesquisa bibliográfica de natureza interdisciplinar envolvendo uma literatura específica para cada área. Para a área de humanas, várias obras sobre Fonética e Fonologia, Sociolinguística, Dialetologia, Geolinguística Pluridimensional, Linguística Geral entre outras figuram entre as referências. Para a área de exatas, muitos conteúdos voltados ao Processamento de Linguagem Natural, Processamento de Sinais da Fala, Inteligência Artificial, Eletrônica, softwares e linguagens de programação foram consultados; (ii) estudo do banco de dados do Projeto ALiB - Atlas Linguístico do Brasil com a finalidade de subsidiar a avaliação quanto à abrangência/eficiência de equipamentos que funcionam com comandos por voz no Brasil, tendo em vista que um equipamento desta natureza teoricamente deve funcionar com realizações de fala de qualquer região; (iii) seleção de algumas amostras das gravações feitas pela equipe do ALiB em 17 localidades do Paraná e nas 25 capitais; (iv) escolha de três representantes da categoria equipamento/dispositivo que operam com *comandos por voz*, avaliados à luz de teorias voltadas à Sociolinguística e à Geolinguística Pluridimensional: o *IBM Via Voice* (software/aplicativo), o *Voice Recognition Module V2* (dispositivo) e o buscador por voz da Google (aplicativo).

A organização dos capítulos e seções desta Tese segue uma lógica em que se considera a qualidade interdisciplinar da pesquisa. Em outras palavras, como se trata de um trabalho que pode interessar a pesquisadores das distintas áreas de humanas e exatas, foi necessário elencar seus capítulos e seções de modo que não faltassem os elementos introdutórios mínimos a cada uma das áreas envolvidas. A ordem, portanto, foi disposta para que todos os conteúdos pudessem ficar devidamente concatenados e completarem-se uns aos outros.

Assim sendo, o trabalho, inicia-se com uma parte introdutória na qual figuram alguns dados sobre o estado da arte dos assuntos envolvidos, a hipótese e as respectivas perguntas por ela geradas, a relevância e a justificativa do estudo, um objetivo geral - avaliar como alguns equipamentos acionados com *comandos por voz* reagem a diferentes realizações linguísticas de determinadas regiões do Brasil – e dois objetivos específicos.

O capítulo 2 contém seções que tratam, respectivamente, de informações gerais sobre o conceito de Dialetoлогия, da variação linguística e sua relação com as tecnologias que permeiam todo o estudo, os estudos dialetológicos e sua relação com o presente trabalho, um panorama sucinto sobre o desenvolvimento dos atlas linguísticos mais conhecidos no Ocidente e, ao final, especificamente, uma seção sobre o Atlas Linguístico do Brasil - ALiB. Este capítulo, cujo conteúdo é bastante familiar aos linguistas de um modo geral, tem grande potencial esclarecedor aos profissionais da área de exatas, uma vez que apresenta um horizonte bastante amplo dos estudos linguísticos mais atuais e o quanto podem auxiliar no desenvolvimento de novas tecnologias.

O capítulo 3 é composto de uma tríade de seções. As duas primeiras tratam dos aspectos e do funcionamento de fonemas em aplicativos que convertem texto em fala, ou seja, colocam em foco, sob o ponto de vista essencialmente linguístico, o conceito de *síntese de voz*. Esse conceito, embora bastante familiar a alguns desenvolvedores de sistemas, é abordado também sob o crivo da Linguística, fato que colabora quanto à capacidade de apontar eventuais falhas metodológicas quando do desenvolvimento de novas tecnologias. Ao linguista, mostra procedimentos essenciais à compreensão geral de como as máquinas podem fragmentar e manipular sequências de fonemas para formar textos inteligíveis. A terceira seção fecha o capítulo 3 tratando dos conceitos de Processamento de Linguagem Natural, Processamento de Sinais da Fala e Geolinguística e fornecendo informações sobre como esses conceitos se inter-relacionam nas tecnologias envolvidas.

No capítulo 4 concentram-se todos os experimentos realizados para esta Tese. As três primeiras seções tratam, respectivamente, do trio de equipamentos que operam com *comandos por voz* citados há pouco: o *IBM Via Voice*, o *V2* e o buscador por voz da Google. A escolha desses equipamentos obedeceu a critérios específicos e alinhados à proposta da pesquisa: que ao menos (i) não seja projetado para outra língua; (ii) seja conhecido por uma quantidade significativa de pessoas; (iii) possa atender a algumas especificidades do uso; (iv) não tenha caído em plena obsolescência. Esses critérios estão organizados junto às especificidades de cada equipamento na Tabela 01.

Tabela 01- Critérios para escolha dos equipamentos analisados

	Critérios	IBM Via Voice	V2 Voice Recognition Module	Buscador por voz da Google
(i)	compatibilidade com o idioma	possui suporte para Português Brasileiro	suporte para voz humana em geral	'reage' bem ao Português Brasileiro
(ii)	grau de difusão entre usuários	é um clássico (ocidental) para edição de textos	bastante difundido na área técnica	mundialmente conhecido por usuários na internet
(iii)	banco de dados atualizável	possui léxico básico integrado manipulável	não possui banco de dados integrado mas é manipulável	possui amplo léxico integrado não manipulável
(iv)	grau de utilização atual	ainda utilizado por projetistas e estudantes para Domótica e robótica	amplamente utilizado por projetistas e hobistas para Domótica e robótica	abrange a maior quantidade e variedade de equipamentos disponíveis na ocasião da pesquisa

Fonte: Tabela organizada pelo autor

A última seção deste capítulo versa sobre o aplicativo Tical, desenvolvido para testar, colocar à prova e recriar vários itens dos experimentos discutidos nas três seções anteriores. Trata-se do protótipo de um robô de conversação que responde por escrito ou por *síntese de voz* a perguntas sobre Linguística por meio do teclado, Whatsapp e/ou por *comandos por voz*. O robô foi desenvolvido por este pesquisador em parceria com o professor Fábio Carlos Moreno, pesquisador da área de Desenvolvimento de Sistemas.

Para todos os experimentos presentes no capítulo 4, houve o cuidado de aplicar no processo ao menos as realizações de fala mais comuns em Português Brasileiro - entre elas as variantes de /R/ e /S/ em coda silábica. A seleção das palavras utilizadas nessa avaliação dependeu da natureza do equipamento em questão, ou seja, variou de acordo com as funções que cada um deles realiza, com o respectivo banco de dados (de palavras) disponível e/ou com a possibilidade de programá-los com novas palavras.

Os testes foram feitos de três modos diferentes: (i) com o uso de gravações já coletadas em entrevistas do ALiB, como foi o caso do experimento com o IBM Via Voice na seção 4.1; (ii) com a presença de falante nativo - o próprio pesquisador - como verificado nos testes com o buscador por voz discutido na seção 4.2; (iii) ou somando aos modos anteriores algumas simulações de realizações de fala, tal como pode ser verificada na seção 4.3.

Uma escolha produtiva foi aplicar primeiramente algumas das variantes paranaenses – seção 4.1 - e paulistas - seção 4.3 -, não só em função de haver à disposição um instrumental bastante satisfatório para coleta de dados em pesquisa de campo (AGUILERA, 2005; CARUSO, 1983; COMITÊ NACIONAL DO ALiB, 2001; CARDOSO et al. 2014a; 2014b),

mas também pela proximidade geográfica que mantém com a Universidade Estadual de Londrina.

Esperava-se, primeiramente, que este estudo apresentasse uma significativa avaliação de como alguns equipamentos acionados com comandos por *voz* reagem a diferentes registros linguísticos de determinadas regiões do Brasil. Ao final, essa avaliação permitiu verificar se as realizações linguísticas de outras regiões acionavam o dispositivo com a mesma eficácia, o que colocou em evidência o grau de abrangência em termos de variação linguística do sistema de reconhecimento de voz de alguns desses equipamentos⁸.

Tais dados e informações têm o propósito de auxiliar profissionais e pesquisadores da área de Linguística a aplicar de modo mais pontual conhecimentos de Geolinguística às tecnologias de comando e/ou reconhecimento de voz vindouras. Isso pode, por exemplo, ajudar ainda mais a aproximar linguistas e programadores tendo em vista que, aparentemente, os profissionais das áreas mais técnicas como Desenvolvimento de software e Eletrônica Aplicada - inevitavelmente associadas a tecnologias de acionamento de dispositivos com *comando por voz* – podem estar construindo seus projetos com conhecimentos muito básicos sobre Dialetoлогия, Fonética e Fonologia.

Além disso, considerando-se que o Português é a sexta língua mais falada no mundo⁹ e que a variante brasileira é a maior representante mundial do idioma - hegemonias política, econômica e demográfica - é notoriamente interessante que mais pesquisas sejam feitas para que não seja necessário, por exemplo, que um usuário brasileiro tenha de pronunciar uma palavra em outra língua para acionar um equipamento comandado por voz, tendo em vista que esse tipo de recurso é mais presente em países onde as pesquisas na área estão mais avançadas. Nesse sentido, será pouco adequado que, nessa próxima década, um morador de uma cidadezinha do interior tenha de pronunciar *four* para que o elevador vá ao quarto andar.

⁸ Como disposto no **objetivo** deste projeto, é possível que uma parcela dos falantes da língua portuguesa do Brasil - não necessariamente apenas aqueles distantes dos grandes centros irradiadores de cultura (BORBA, 1991) - tenha dificuldades em utilizar tais equipamentos com grau considerável de espontaneidade ou sem treinamento prévio, e aí reside mais um motivo importante para divulgar e aplicar em outras áreas de conhecimento os estudos já existentes sobre a Sociolinguística, a Dialetoлогия e a Geolinguística do Brasil.

⁹ Não é consensual a real posição do Português Brasileiro entre as diferentes instituições e órgãos que mensuram esses números pois isso depende de uma série de fatores, entre eles, da perspectiva em que se observa o fenômeno. A *Ethnologue: Languages of the World*, por exemplo, cujo critério não discrimina entre falantes nativos e não nativos, é um desses órgãos que traz o português em sexto lugar.

2 ESTUDOS SOBRE VARIAÇÃO DO PORTUGUÊS BRASILEIRO

(...) o sucesso da caminhada está relacionado diretamente à fidelidade aos princípios assumidos, à ação de reflexão crítica que tem e terá de ser exercida continuamente, à capacidade de avaliar e rever, com rigor, com clareza e com a simplicidade dos sábios, cada passo (AGUILERA, 2005, p. 11)

2.1 DIALETOLOGIA

Ciência que surgiu no final do século XIX, a Dialectologia, como o próprio nome sugere, trata do estudo de dialetos. É importante salientar que o conceito de dialeto varia entre os diferentes pesquisadores, sejam eles da mesma nacionalidade ou pertencentes a diferentes países. Borba (1991), por exemplo, nos alerta sobre o fato da noção de dialeto ser ‘fugidia’. Antes, porém, de apresentar alguns conceitos sobre o que seja dialeto, talvez seja interessante delimitar mais o domínio da Dialectologia, cuja metodologia considera sempre mais de um universo de pesquisa, mais de um local.

Dialectologia pode ser considerada como o “estudo do arrolamento, sistematização e interpretação dos traços linguísticos dos dialetos” e é responsável também pela técnica de “descrição dos falares por meio de monografias dedicadas a uma dada região”, com as quais se constroem gramáticas e glossários regionais (CAMARA JR, 1981, p. 95). Como toda disciplina, a Dialectologia evolui e incrementa outros domínios. Nesse sentido, a Dialectologia e a Geolinguística - ou Geografia Linguística - se complementam, embora tradicionalmente a Geolinguística seja considerada particularização da Dialectologia (RECTOR, 1975).

Em perspectivas similares, “a técnica mais moderna de pesquisa e exposição em Dialectologia é a chamada Geografia Linguística, que consiste em levantar mapas da distribuição geográfica de cada traço linguístico dialetal” (CAMARA JR, 1981, p. 94-95), ou “estudo sistemático de todas as formas de dialeto, em especial o dialeto regional” e, por esse motivo, também pode ser denominada ‘geografia linguística’ ou ‘geografia dialetal’ (CRYSTAL, 1988, p. 81). A subsequente associação de mapas acaba por constituir aquilo que é conhecido como o atlas linguístico do território estudado. Mesmo com metodologias diferenciadas, fácil identificar zonas de intersecção entre Dialectologia e Geolinguística.

Geolinguística, por sua vez, é um “método dialectológico e comparativo que (...) pressupõe o registro em mapas especiais de um número relativamente elevado de formas linguísticas (fônicas, lexicais ou gramaticais) comprovadas mediante pesquisa direta e unitária

numa rede de pontos de um determinado território”, ou que ao menos, considere a distribuição das formas em cada espaço geográfico que corresponda respectivamente à língua, às línguas, aos dialetos ou aos falares estudados (COSERIU, 1982, p. 79). Mais sucintamente, pode-se ainda entender a Geolinguística como “o estudo de todas as variações ligadas à implantação, ao mesmo tempo social e espacial, dos usuários da linguagem” (DUCROT e TODOROV, 2001, p. 65).

É notável, nestas definições de Dialetologia e Geolinguística, a recorrência do objeto de estudo dialeto, que pode ser entendido como a diferenciação regional de uma língua, uma diversificação ligada aos fatores geográficos, ou seja, diatópicos (BORBA, 1991), como formas características que uma língua assume regionalmente (CUNHA, 1972, 2001) ou como um falar regional, muitas vezes constituído de uma multidão de falares em uma nação onde predomina oficialmente outro falar (DUCROT E TODOROV, 2001).

Alguns linguistas - muitos deles brasileiros - preferem não utilizar o termo dialeto de modo amplo - para dialeto regional ou dialeto urbano - e chamam muitas vezes cada uma dessas variedades locais de *registro* linguístico, termo cunhado por Câmara Jr (BORBA, 1991). O dialeto, ainda, “mantém com a língua de que promana, dita padrão, semelhança de traços linguísticos essenciais e um sentimento de comunidade por parte de seus usuários” (JOTA, 1981, p. 104) e de acordo com esse mesmo autor, há dois dialetos do Português: **o brasileiro e o europeu** (grifo nosso). Esta última e dicotômica afirmação é bastante generalizante e discutível. Como dito anteriormente, é curioso o quanto o conceito de dialeto pode variar entre os diferentes teóricos e o quanto as nuances e subcategorias do mesmo conceito podem se mostrar diversas.

Paralelamente aos dialetos regionais - e mesmo nacionais - coexistem os dialetos urbanos, objeto de estudo da Sociolinguística, responsável por estudar “todos os aspectos da relação entre língua e sociedade”, incluindo-se identidade linguística de grupos sociais, atitudes sociais em relação à língua, formas padrão ou não padrão entre outras (CRYSTAL, 1988).

Diferentemente da Dialetologia, a metodologia de pesquisa da Sociolinguística contempla um único universo de pesquisa, um único local. Um dos trabalhos mais consistentes e considerados como referência nesse domínio de conhecimento é o do norte-americano William Labov, cuja dissertação de mestrado denominada *The social history of a sound change on the island of Martha's Vineyard* (LABOV, 2008), publicada em 1963, é considerada um marco inicial para a Sociolinguística (ALKMIM, 2001). Outra não menos importante publicação deste autor e que é referência para aqueles que procuram se aprofundar

nessa disciplina é *The Social Stratification of English in New York City*, publicada em 1966 (LABOV, 2008), resultado de sua tese de doutorado. Esses e vários outros de seus trabalhos de pesquisa sempre consideram o fato de que linguagem e sociedade são coisas indissociáveis.

De fato, a relação destes dois elementos jamais foi negada pela comunidade científica em geral, mesmo considerando que, para alguns fins metodológicos, elas sejam vistas como entidades que podem ser observadas separadamente. “Segundo o ponto de vista mais tradicional, é a sociedade que determina a linguagem” (DUCROT e TODOROV, 2001, p. 70). Partindo deste pressuposto, o estudo de variantes linguísticas possibilita visualizar mais claramente as variantes sociais ou culturais que deram origem a elas. Num sentido inverso, observar as relações sociais pode ajudar a explicar o surgimento ou manifestação de alguns fenômenos linguísticos em uma sociedade local (DUCROT e TODOROV, 2001).

Numa acepção mais moderna, a Sociolinguística seria um ramo da Linguística que leva em conta o aspecto social das línguas. Porém, como não é possível distinguir entre uma linguística geral para as línguas e uma Sociolinguística apenas para os aspectos sociais, “a Sociolinguística é a Linguística” (CALVET, 2002, p. 33). Quase na mesma direção, ela trata de “um modelo teórico-metodológico que assume o “caos” lingüístico como objeto de estudo. Como esse modelo, por princípio não admite a existência de uma ciência da linguagem que não seja social, o próprio título “Sociolingüística” fica redundante” (TARALLO, 2001, p. 06-07). Considerando as colocações de Calvet e de Tarallo, mais estreitas ainda se tornam as relações entre os domínios da Dialetologia, Geolinguística e Sociolinguística.

A denominação Geolinguística Pluridimensional, contudo, talvez seja a mais conveniente para categorizar obras como os mais modernos atlas linguísticos, em especial o ALiB, uma vez que não são desconsiderados quaisquer domínios e respectivos aspectos comentados até este ponto. Em outras palavras, a Geolinguística Pluridimensional, como o próprio nome sugere, pode contemplar as diversas dimensões e contextos de variação linguística identificáveis: diatópica, diafásica, diastrática, diassexual, diagenérica, diageracional, diarreferencial e diamésica (AGUILERA, ALTINO, 2012).

Entre estudiosos das áreas de Humanas e principalmente linguistas, essa conceituação e respectiva terminologia parecem óbvias. No entanto, em domínios notadamente da área de exatas, como Análise e Desenvolvimento de Sistemas, Ciência da Computação, Engenharia Elétrica ou Eletrônica e alguns de seus subdomínios como Inteligência Artificial, Redes de Computadores ou Segurança da Informação, há pouquíssimas menções sobre o conceito de dialeto e as distinções que aqui foram feitas, mesmo quando se trata de trabalhos voltados a

reconhecimento ou *comandos por voz*, eminente e inevitavelmente interdisciplinares. No mais das vezes, quando ocorrem, conceitos como ‘sotaque’, ‘dialeto’ e ‘fala’ são quase sempre inseridos não apenas no mesmo campo semântico, mas na mesma dimensão de equivalência de significado, ou seja, conotam quase o mesmo referente, a mesma ideia. Praticamente não há menção sobre Dialectologia, Sociolinguística ou Geolinguística; parece haver um desconhecimento generalizado sobre a existência dos atlas linguísticos - incluindo-se o ALiB – e os critérios para distinguir e aplicar as variedades linguísticas brasileiras que entram em discussão são questionáveis em termos linguísticos.

A Análise de Sistemas - e, claro, Engenharias - necessita(m) encontrar caminhos mais bem trilhados para que possa transitar mais facilmente pela Dialectologia. É historicamente latente que pesquisadores de diversos domínios concentrem-se em suas áreas e explorem, em alguns casos, seus objetos de estudo à exaustão sem sequer atentar a pesquisas que, embora pertencentes a outros domínios de conhecimento, podem ter muito a contribuir em caráter interdisciplinar. De modo semelhante, a Dialectologia, a Sociolinguística, a Geolinguística e todas as outras parceiras necessitam de profissionais que possam incrementá-las, sempre que possível, com as tecnologias emergentes mais variadas.

2.2 ESTUDOS DIALETOLÓGICOS E O PRESENTE TRABALHO

Embora tenham se iniciado notadamente muito mais tarde que seus predecessores na Europa, os estudos sobre variação linguística no Brasil evoluíram muito no século XX e continuam em franca expansão nesse início do século XXI. Em nosso país, os maiores obstáculos, como a grande extensão territorial, patrocínio ou incentivo para pesquisa insuficientes e dificuldade em adquirir ou reproduzir tecnologias e eletrônicos afins, não fizeram parar os pesquisadores que se propuseram, cada qual em seu campo, seguir com seus estudos e mostrar seus resultados.

Esses obstáculos ainda existem e continuarão a existir, uma vez que não é possível diminuir as distâncias territoriais, é amiúde penoso adequar conjunturas políticas de fomento ou financiamento às necessidades de pesquisa, e novas tecnologias são sempre onerosas e proporcionalmente complexas para serem fabricadas ou adquiridas de modo veloz. A obstinação, curiosidade científica e necessidade de estudar a própria língua, no entanto, sempre levaram o homem em todo o mundo a se dedicar com afinco em suas empreitadas e, no Brasil, não foi e não é diferente.

Neste âmbito, este trabalho é uma das muitas manifestações de desejo e vontade de estudar e elencar como se projetam variedades de nossa língua não só nos aspectos sociais e linguísticos, mas também no tocante à sua permeabilidade na Sociedade da Informação como dados que precisam ser processados pelos mais diversos recursos tecnológicos que se destinam a esse fim. No século XXI, os estudos sobre variedades do Português Brasileiro já compreendem, portanto, mais que a extensa e múltipla obra de Sociolinguistas, Geolinguistas, Geossociolinguistas, Foneticistas entre muitos, mas também toda contribuição Psicolinguistas, Analistas do Discurso, Analistas e Desenvolvedores de Sistemas, Engenheiros Eletricistas, Eletrônicos, Eletroacústicos e tantos outros profissionais cuja expertise possa adicionar elementos chave a um propósito relevante a toda comunidade de fala.

Com quase 203 milhões de falantes (IBGE, *online*, 2015), o Português Brasileiro fica tranquilamente entre os dez idiomas mais falados no mundo. Considerando o Português como um sistema relativamente regular - em termos de regras ortográficas - entre os países lusófonos que o têm como língua oficial, fica entre as seis línguas mais utilizadas no planeta.

Para nós, brasileiros, assim como para qualquer falante nativo em todo o mundo, é muito confortável e adequado que tenhamos conhecimento sobre as variedades existentes no país, que toda literatura sobre fauna, flora, indumentária, gastronomia, leis, regulamentações e equipamentos em geral possa atender a essa diversidade e que sejamos respeitados em toda comunidade mundial pelos feitos científicos que digam respeito à consideração e ao valor que atribuímos ao nosso próprio idioma. O Atlas Linguístico do Brasil (doravante ALiB) contribui com grande parte de tudo isso e, a partir dele, é possível expandir os dados já tabulados para outras Tecnologias da Informação. Entre muitos exemplos, e não necessariamente o mais relevante, talvez não seja necessário que existam nas ruas de cidades turísticas no Brasil indicações como “Sinal, Sinaleiro, Farol, Semáforo” na mesma placa, mas uma atualização personalizada em Português Brasileiro para aparelhos de GPS em que o usuário escolhe a opção com a variedade correspondente à sua cidade - e/ou região - de origem é, no mínimo, muito conveniente.

2.3 VARIAÇÃO E TECNOLOGIA

A linguagem é inerente ao homem. “A história da humanidade é a história de seres organizados em sociedades e detentores de um sistema de comunicação oral” (ALKMIM, 2001, p. 21). Desde há muito tempo, quando do surgimento das primeiras civilizações, toda

vez que uma região se torna mais rica e influente politicamente, a língua ali falada passa a ter certo privilégio em relação à de outra mais pobre ou menos influente politicamente. Em outras palavras “a organização política sediada numa cidade (assim tornada capital) pode valorizar a modalidade da língua aí falada e o poder centralizador da administração contribui para fixá-la como padrão” (BORBA, 1991, p. 52).

Em se tratando do Português Brasileiro a situação não é diferente. *Grosso modo*, como em qualquer outro idioma, o Português padrão escrito é aquele ditado pelas gramáticas normativas, dicionários e alguns manuais de redação - instrumentos linguísticos legitimados pelas instituições (NUNES, 2002 In NUNES & PETTER, 2002; ORLANDI, 2002; FARACO, 2004) - e o Português padrão falado, além de atender a muitos dos preceitos ditados pelas gramáticas, fundamenta-se a partir do padrão utilizado pelas classes cultas ou de maior privilégio sócio-político-cultural (BORBA, 1991; CAMACHO, 2001). “Falantes de qualquer língua prestigiam ou marginalizam certas variantes regionais (ou pelo menos não as discriminam) a partir da maneira pela qual as sequências sonoras são pronunciadas” e podem ser determinadas como *variantes de prestígio* ou *variantes estigmatizadas* (SILVA, 1999, p.12).

Guardadas as devidas proporções, basta verificar o padrão - falado e escrito - utilizado nos meios jurídicos, telejornais de grande abrangência, documentários, jornais e revistas de grande circulação – sem falar ainda daqueles tidos como padrão por excelência, como o do sistema escolar e a linguagem oficial do Estado. Portanto, quanto se trata estritamente de norma linguística, observa-se uma abordagem altamente seletiva – conceitos de correção, adequação e até elegância – que rejeita imperativamente tudo o que é considerado incorreto ou inadequado.

No cotidiano, porém, pessoas de diferentes idades, classes sociais, graus de instrução e/ou regiões comunicam-se verbalmente sem maiores problemas - mesmo aquelas consideradas como analfabetas funcionais. Isso ocorre, entre outras razões, porque de fato há um esforço mútuo entre as pessoas de se entender e se fazer entender, afinal, “falar uma determinada língua implica um conhecimento que certamente transcende o escopo puramente linguístico” (SILVA, 1999, p. 11).

Nesse sentido, torna-se também necessário saber como uma máquina pode ser ajustada para também fazê-lo. Esse conhecimento tem rendido muitas pesquisas e avanços significativos nas últimas décadas e alguns de seus aspectos estão descritos a seguir.

Equipamentos acionados com comandos por voz, na maior parte das vezes, são aparelhos usuais (do cotidiano) com um novo dispositivo de entrada de dados¹⁰ e/ou uma nova interface humano-computador¹¹. Eles, porém, não pensam¹² como seres humanos. Quase a totalidade deles sequer chega perto do conceito básico de pensar - que por si só já apresenta um grau de complexidade absurdo. A maioria compara o sinal sonoro que recebeu a um banco de dados que possui gravado em sua memória: não sabe o que é *frente* ou *trás*. Apenas identifica o sinal elétrico captado pelo microfone e correspondente a [frētʃi] e o associa a um comando elétrico especificado pelo software ou circuito eletrônico, quando é o caso. “Um sistema reconhecedor de fala é basicamente um comparador de padrões acústicos de voz associados a um conjunto de símbolos que codificam uma mensagem” (DAMASCENO, 2005, p. 38).

O banco de dados, também conhecido como biblioteca (ver nota 05), não só é bastante limitado em quantidade de palavras como também, na maior parte das vezes, fundamenta-se na norma linguística. Se a norma rejeita o que é considerado *incorreto*, algumas pessoas precisam treinar a pronúncia de cada um dos comandos para poder utilizá-lo. Isso não é, em princípio, um problema para o usuário, uma vez que a quantidade de comandos é menor que dez e, muitas vezes, foi o próprio usuário quem decidiu utilizar o equipamento - como é o caso de alguns dispositivos com *comando por voz* para automóveis.

Porém, quando a utilização não é assim tão opcional como é o caso de cadeiras de rodas para pessoas que não têm, também, os movimentos nas mãos - e, portanto, não podem acionar botões ou alavancas - a situação pode ser diferente. Em vez de decorar comandos, o usuário grava com sua própria voz os comandos que deseja e então o uso fica bem mais espontâneo, além de não necessitar seguir a norma linguística, tampouco a dimensão semântica convencional: ‘frente’ ou ‘para frente’ por, simplesmente ‘vai’. Note-se que para

¹⁰ Os dispositivos de entrada (de dados) mais básicos em se tratando de computadores pessoais são o mouse e o teclado. Outros, bastante comuns na última década são as câmeras (webcams) e os microfones (embutidos ou não). Para usos mais avançados, há os sensores biométricos que ‘leem’ digitais ou íris, entre outros.

¹¹ Interface Humano-Computador não é o mesmo que Interação Humano-Computador, ainda que suscitem áreas de pesquisa próximas e relacionadas. Por Interface Humano-Computador entende-se todo dispositivo que permita ao homem acompanhar (aproveitar, utilizar) as tarefas realizadas por um computador: o monitor de vídeo, a impressora, as caixas de som e, mais modernamente, o leitor braille (que substitui o monitor para pessoas que não enxergam), óculos e capacetes digitais, entre outros. Por extensão, aquilo que versa linguagens de máquina para linguagens humanas (PREECE, 2005, p. 40).

¹² Como exemplo, ainda que as redes neurais - concebidas em analogia ao sistema cerebral humano - com suas várias ‘entradas de dados’ e seus ‘pesos sinápticos’ tenham funcionamento bastante satisfatório, enquanto máquina, apenas simulam algumas situações de processamento de sinais acústicos da fala. A Psicolinguística, por exemplo, do ponto de vista metodológico, ainda enfrenta o problema de não ter acesso direto ao processamento mental desse tipo de sinal (BALIEIRO Jr, 2001).

esse caso o uso é exclusivo, ou seja, não só a cadeira está disponível a uma só pessoa como não há outro querendo utilizá-la quando ela está parada.

Para esses dois casos: um, com banco de dados limitado e sequências pré-determinadas que requerem treino na pronúncia; e outro, em que o usuário grava a sequência que quiser para cada comando, não há muitos problemas quanto à utilização.

Quando se trata, porém, de um equipamento que deve funcionar com sua própria biblioteca em todas as regiões do país, talvez seja necessária maior atenção.

Considerando as sequências [ˈãdɐ] e [ˈparɐ] que acionam respectivamente comandos relativos a fazer algo mover-se para frente e parar, elas atendem à norma - modos imperativos afirmativos *anda* e *para* (tu) - e o mais importante: teoricamente funcionariam, por exemplo, em quase todo o Brasil¹³ sem maiores problemas pois a realização fonética é muito similar – o ‘a’ nasal de *anda* e o tepe intervocálico de *para*.

Os paranaenses de União da Vitória (sul), Guarapuava (centro) ou Londrina (norte) provavelmente acionariam esse comando de modo significativamente espontâneo, sem necessidade de treino considerável. O mesmo ocorreria com paulistas de Assis (oeste), Itapetininga (sudeste) e São Paulo (capital).

O processo, contudo, torna-se bem mais complexo quando se adicionam sequências correspondentes a outros comandos vetoriais e/ou direcionais como *direita* e *esquerda*. Para a palavra *direita* as realizações fonéticas talvez não sejam muitas, considerando-se as regiões há pouco citadas. A palavra *esquerda*, porém, traria várias possibilidades de realização entre paranaenses como [esˈkerdɐ] mais ao sul, [isˈkerdɐ] ou [ˈskerdɐ] mais ao centro, [isˈke.lɐ] mais ao norte, entre outros (AGUILERA, 2005). Em território nacional a situação fica ainda mais complexa e, para o correto funcionamento desse equipamento acionado com comandos por voz, só para dar conta da palavra *esquerda*, sua biblioteca precisaria ser bem mais ampla, ou seja, para cada palavra, a amostragem para um treinamento adequado de sistema deve ser grande: dezenas de locutores, sendo metade de cada sexo (YNOGUTI, 1999; DAMASCENO, 2005; HUGO 1995; TEVAH, 2006). Esse tipo de amostragem pode atender minimamente às variações entre os sexos (diassexual) e idades (diageracional) e das diferentes camadas sociais (diatráticas), porém, pode deixar muito a desejar quanto à variação relativa ao espaço geográfico (diatópica). Isso talvez ocorra porque o material

¹³ Temos à disposição, por exemplo, as variações *ande* [ˈãdɪ] e *pare* [ˈpaɾɪ]

normalmente é colhido *in loco* – os próprios desenvolvedores dispõem-se como locutores na amostragem.

Tendo em vista tais fatos, há muitas dúvidas sobre a importância que os respectivos fabricantes dão a essas questões de variação linguística. Seria ingênuo não inferir que todo fabricante quer de fato que seu equipamento funcione em todo o território nacional. Porém, talvez se mostre bastante produtivo procurar saber se há consultas a literaturas específicas como textos sobre Linguística Geral, Fonética e Fonologia, atlas linguísticos entre outros por parte dos desenvolvedores.

2.4 ATLAS LINGUÍSTICOS

Os atlas linguísticos são “reuniões de cartas em que o material linguístico está distribuído topograficamente”, situação na qual “cada carta apresenta um instantâneo dialetal da área explorada” (SILVA NETO, 1958) ou conjuntos de mapas contendo traços fonéticos, lexicais e morfossintáticos típicos de uma língua em determinado espaço geográfico (BRANDÃO, 1991). Assim, como dito anteriormente, estudos realizados por geolinguistas normalmente geram como resultados os atlas. Embora os primeiros passos em Geolinguística para a criação de um Atlas da Alemanha tenham sido dados por George Wenker (1881), o primeiro atlas linguístico significativamente criterioso e bem construído foi o ALF - *Atlas Linguístico da França* em 1910 (DUBOIS, 2004). Jules Gilliéron, linguista suíço, adotando critérios bastante produtivos para a época, conseguiu não só mapear aspectos de variação linguística em território francês como também proporcionou aos pesquisadores vindouros uma base bastante sólida em que se apoiar.

Outros, entre vários, como o ALI - *Atlante Linguistico Italiano* de Matteo Bartoli (antes de 1914), o ALC – *Atlas Linguistic de Catalunya*, de P. Antoni Griera (1912-1922), o ALEIC - *Atlante Linguistico-etnografico Italiano della Corsica*, de Gino Bottiglioni (1933-1942) ou o ALR - *Atlas Linguístico Romeno*, de Sextil Puscariu (1930-1937) apresentaram seus mapas numa perspectiva bastante similar àquela de Gilliéron. Silva Neto (1958), interessadamente, considera o *Atlas Linguístico-Etnográfico da Itália e Suíça Meridional* de Karl Jaberg e Jakob Jud - dois discípulos de Gilliéron - como “o maior monumento da Geografia Linguística” por sua extensão e valor metodológico (DUBOIS, 2004).

No Brasil, os estudos dialetológicos iniciaram-se com *O Dialeto Caipira* de Amaral (1920). Em seguida, Nascentes e Silva Neto, cada qual com suas obras e respectiva contribuição para esses estudos, constituem-se até os dias de hoje bases sólidas de consulta e

pesquisa: Silva Neto com o *Guia para estudos dialetológicos* (1958) e Nascentes, propondo uma eficiente distribuição de falares para o Brasil com *O Linguajar Carioca* (1924), *Bases para elaboração do Atlas Linguístico do Brasil* (1958) e *O idioma nacional* (1960).

A Tabela 02 apresenta alguns dos primeiros estudos dialetológicos que tiveram expressiva repercussão entre os dialetólogos em nosso país e que, certamente, incentivaram a integração dos mapas em estudo e impulsionaram o projeto de um atlas linguístico para o Brasil. A disposição em ordem cronológica nesta Tabela torna fácil observar que os trabalhos distribuem-se quase regularmente ao longo das décadas. Junto a eles, destacadas em negrito, figuram obras que dizem respeito a alguns dos estudos geolinguísticos já realizados¹⁴, o que faz saltar aos olhos a relação de coexistência em termos cronológicos com os resultados de investigação em dialetologia tradicional.

Tabela 02 - Gênese da Dialetologia no Brasil e relação de coexistência com a Geolinguística

Obras	Autores
A língua do nordeste (1934)	Marroquim
O falar mineiro (1938)	Teixeira
Estudos de dialetologia portuguesa: linguagem de Goiás (1944)	
APFB – Atlas Prévio dos Falares Baianos (1963)	Rossi, Nelson
Estudos da Dialetologia do Rio Grande do Sul (1969)	Bunse
Traços arcaicos do português popular do Brasil (1971)	Penha
Aspectos da linguagem de São Domingos (1972)	
Dialeto caipira da região de Piracicaba (1974)	Rodrigues
EALMG – Esboço de um Atlas Linguístico de Minas Gerais (1977)	Zagari et al.
Comportamento linguístico do dialeto rural (1982)	Veado
Aspectos sintáticos do dialeto caipira na região de Morrinhos (1985)	Vilefort
ALPB – Atlas Linguístico da Paraíba (1984)	Aragão e Menezes
ALSE – Atlas Linguístico de Sergipe (1987)	Ferreira et al.
ALPR – Atlas Linguístico do Paraná (1994)	Aguilera
ALSE II – Atlas Linguístico de Sergipe II (2005)	Cardoso
ALERS – Atlas Linguístico-Etnográfico da Região Sul do Brasil (2002)	Koch, Klassmann, Altenhofen
ALAM – Atlas Linguístico do Amazonas (2004)	Cruz
ALISPA – Atlas Linguístico Sonoro do Pará (2004)	Razky
ALMS - Atlas Linguístico do Mato Grosso do Sul (2007)	Oliveira et al.

Fonte: Tabela elaborada pelo autor com base em Aguilera (1997), Altino (2007) e Romano (2013)

Os atlas linguísticos proliferaram em vários países e continentes e contam, a cada dia, com tecnologia mais avançada. Os métodos de coleta e transcrição meramente auditivos e manuais feitos por Edmont sob a orientação de Gilliéron, no final do século XIX, deram espaço a gravadores digitais, editores eletrônicos de texto e sistemas de processamento de

¹⁴ No momento em que este texto é redigido, há outros atlas em andamento, alguns dos quais quase finalizados, relativos aos demais Estados do Brasil.

dados – áudio, imagem, caracteres etc – que poupam muito tempo do pesquisador. Mais que isso, os dados de um informante podem ser coletados a milhares de quilômetros e enviados ao destino em apenas alguns segundos.

É evidente que isso não elimina a necessidade da escolha adequada de informantes e pontos geográficos ou que qualquer pessoa possa fazer a entrevista e com quaisquer critérios, mas é possível contar cada vez mais com ajuda de colaboradores de outras áreas de conhecimento. Especificamente nesse trabalho, o ALiB está quase sempre associado às novas tecnologias, seja na dimensão meramente teórica e probabilística – como a possibilidade de um analisador automático de fonemas-, seja num âmbito mais aplicável e utilitarista – como é o caso de Tical.

2.5 ATLAS LINGUÍSTICO DO BRASIL - ALIB

O ALiB é uma obra inédita e única na história da Geolinguística brasileira, não só pela abrangência em termos territoriais como também pela acurácia dos dados coletados. Ele é resultado do *Projeto Atlas Linguístico do Brasil* - ou, simplesmente Projeto ALiB – que tem como marco inicial o *Seminário Caminhos e Perspectivas para a Geolinguística no Brasil* realizado na cidade de Salvador, em novembro de 1996.

Lançado oficialmente em outubro de 2014, na cidade de Londrina, Paraná, o ALiB (CARDOSO et al., 2014a; 2014b) tem como finalidade apresentar um panorama dos hábitos linguísticos - fônicos, lexicais, gramaticais – de falantes das diferentes regiões do Brasil. A descrição dessa realidade linguística tem um potencial de aplicabilidade imensurável e interessa a todos, mas principalmente a todos os estudiosos de língua portuguesa (CARDOSO, 1998).

Como tudo no ALiB tem como ponto de partida uma pesquisa de campo, na qual é coletada por meio de gravadores de áudio a fala de pessoas de diferentes regiões a partir de questionários padronizados, é perfeitamente possível comparar de imediato as pronúncias da mesma palavra em diferentes disposições diatópicas, além de possibilitar observação da diversidade lexical para o mesmo conceito (lexia).

Tais questionários padronizados foram elaborados pelo Comitê Nacional do Projeto ALiB especificamente para esse fim (COMITÊ, 2001) e são ao todo três¹⁵: **Questionário**

¹⁵ Válido lembrar que a elaboração desses questionários não surgiu do nada. Além das obras de cunho dialetológico, os atlas estaduais e o Atlas Linguístico e Etnográfico da Região Sul (KOCH et al, 2002; ALTENHOFEN et al, 2011), o modelo de entrevista adotado por Labov para coletar dados em Martha's

Fonético-Fonológico, Questionário Semântico-Lexical e Questionário Morfossintático. Além desses três questionários há também uma parte com **Questões de Pragmática, Temas** para discursos semidirigidos, **Perguntas Metalinguísticas** e **Texto** para leitura.

O **Questionário Fonético-Fonológico** contém 159 perguntas e tem por função obter dos informantes as pronúncias de cada uma das palavras nele elencadas. Como exemplo, a primeira delas é “Qual é o tipo de moradia mais comum aqui da região?”, com a qual se espera obter como resposta “casa”. Dessa palavra, especificamente, deseja-se avaliar o fonema /z/ intervocálico quando de sua transcrição ou o caso de reificação [’kaha], como ocorre em algumas localidades do Nordeste (ARAGÃO, 2013).

Quanto ao **Questionário Semântico-Lexical**, seu objetivo é levar o informante a apresentar uma correspondência lexical a determinado conceito disponibilizado pelo inquiridor. O conceito de “...um rio pequeno, de uns dois metros de largura?”, que é a primeira pergunta do Questionário, pode ser associado às palavras *córrego*, *igarapé*, *sanga* ou *riacho*, dependendo do uso que se faz na região do informante. O potencial de enriquecimento de dados desse questionário, no entanto, é bastante significativo, pois, quando se questiona como se chamam “as frutas menores que a laranja, que se descascam com a mão e, normalmente, deixam um cheiro na mão?”, pergunta de número 39, as respostas não são apenas *tangerina* ou *mexerica* no país¹⁶.

Finalmente, o **Questionário Morfossintático** permite coletar os diferentes usos gramaticais dos informantes. A pergunta de número 3, por exemplo, dá ao pesquisador a possibilidade de verificar como varia o gênero da palavra *alface* em diferentes regiões. “Como é que se chama aquela folha verde que se come geralmente na salada?” pode resultar como resposta *o alface* ou *a alface*.

As **Questões de Pragmática** referem-se à aplicabilidade de determinados tipos de tratamento no dia a dia quando se trata de pessoas desconhecidas. Uma proposição e uma pergunta como “Um objeto (...) caiu do bolso de um rapaz jovem e ele não viu. Como um outro rapaz jovem chama a atenção desse idoso?” pode resultar termos como *moço*, *senhor*, *tio* ou *cara*.

Em relação aos **Temas** para discursos semidirigidos, trata-se de um procedimento adotado pelo Comitê (COMITÊ, *online*, 2001) que proporciona ao informante a possibilidade de ficar mais à vontade, descontraído, situação em que a fala se torna mais espontânea,

Vineyard, que previa um questionário lexical, perguntas acerca de juízo de valor e um texto para leitura especial (LABOV, 2008) também serviu de referência para o Comitê.

¹⁶ Para as variantes de *tangerina* especificamente, há uma aplicação na seção 4.2.3.

aumentando a acurácia dos dados coletados. Um desses **Temas**, por exemplo, é o **Relato Pessoal**, em que o informante é convidado a expor um acontecimento marcante de sua vida.

As **Perguntas Metalinguísticas** procuram voltar a atenção do informante para questões relativas à sua própria linguagem e local em que vive, convidando-o a comparar seu modo de falar com outros de outras regiões que conheça ou tenha ouvido falar. Isso pode ser obtido perguntando “Como se chama a língua que você / o(a) senhor(a) fala?”

O **Texto** para leitura é outra maneira de coletar pronúncias dos informantes. Para esse caso, o grau de monitoramento é bastante alto em relação aos **Questionários**, aos **Temas** ou às **Perguntas Metalinguísticas** uma vez que a “Parábola dos sete vimes”, disposta no volume, leva o leitor a prender-se às formas ali presentes e isso faz que, intuitivamente, tente reproduzi-las do modo mais formal possível.

Portanto, a precisão dos dados coletados com os **Questionários**, as **Questões de Pragmática**, os **Temas**, as **Perguntas Metalinguísticas** e o **Texto** para leitura é bastante grande em função do cuidadoso procedimento de escolha dos informantes, que devem atender sempre a quesitos como faixa etária, escolaridade e sexo - bastante relacionados com os estudos sociolinguísticos¹⁷ de Labov - além de outras características extralinguísticas como terem nascido na localidade, possuírem profissão definida que não exija grande mobilidade e que esteja inserida no contexto local e serem preferencialmente filhos de pais da região.

Os dados do ALiB, como já posto na Introdução deste trabalho, são essenciais, afinal, pode se tornar bastante difícil verificar adequadamente a abrangência e a eficiência de equipamentos que funcionam com comandos por *voz* no Brasil sem as informações de um bom atlas, pois, equipamentos desta natureza, teoricamente, devem funcionar a partir do comando de qualquer falante, independentemente da região à qual pertença.

Ainda em termos de tecnologia, ter as entrevistas digitalizadas em arquivos de áudio não só permite consultar quantas vezes for necessário ao longo dos anos e por pesquisadores de gerações diferentes, como também realizar diversificados experimentos com o uso de processadores de áudio, tal como apresentado na seção 4.1. Gilliéron talvez sequer tenha imaginado essas possibilidades no final do século XIX, mas, certamente, ficaria extasiado e teria aproveitado muito bem os recursos caso os possuísse.

¹⁷ Para saber mais sobre esse conceito, seu surgimento e a influência de Labov, interessante consultar Alkmim (2001), Camacho (2001) e Labov (2008).

3 RELAÇÃO ENTRE VARIAÇÃO LINGÜÍSTICA E PROCESSAMENTO DE SINAIS DA FALA

O linguista que entra no mundo só pode concluir que o ser humano é o herdeiro legítimo da estrutura incrivelmente complexa que nós agora estamos tentando analisar e compreender (LABOV, 2008, p. 18)

O Processamento de Sinais da Fala (doravante apenas PSF) é uma área de conhecimento bastante ampla que engloba outras muito conhecidas, como *síntese de voz*, *reconhecimento de voz*, *comandos por voz* e *autenticação de voz*. O PSF, no entanto, mesmo sendo uma grande área, pertence a outra ainda maior: o Processamento de Linguagem Natural (doravante apenas PLN). A Figura 01 ilustra como as áreas correlacionadas se organizam hierarquicamente.

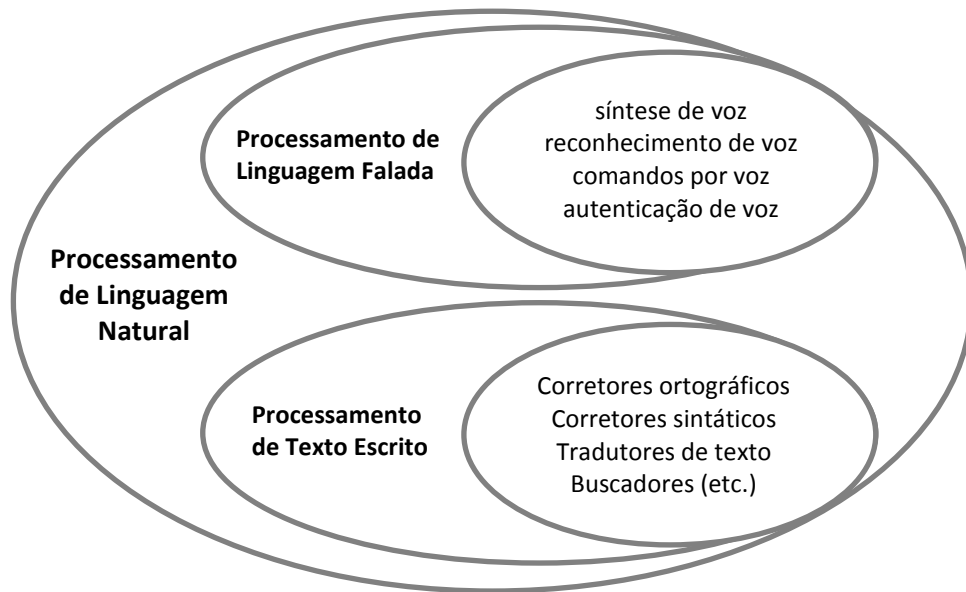


Figura 01 – Disposição de áreas e subáreas do PLN
Fonte: Baseada em Barbosa, 2004; Rich, 1993; Schildt, 1989.

Em termos gerais, o PLN busca desenvolver aplicações de informática capazes de processar os fenômenos da linguagem natural (HAGÈGE; DUARTE, 1995). A aplicação mais comum permite ao usuário consultar dados no computador em seu próprio idioma – como o Português – digitando as perguntas, por exemplo. Outros tipos de aplicação permitem ‘conversar’ com o computador por meio de ordens (FILGUEIRAS, 1995), ou seja, de outras maneiras que não apenas pelo teclado alfabético, como é o caso do controle por voz.

Não se trata, portanto, quando se envolve PLN, de um simples buscador de palavras ou expressões. Para processar os fenômenos da linguagem, é necessário construir gramáticas formais, instrumentos que permitem “descrever ou gerar todas as frases possíveis da linguagem e eventualmente dar um significado destas frases” (HAGÈGE; DUARTE, 1995, p. 72).

O PLN está intimamente ligado à Inteligência Artificial que, por sua vez, consiste em estudar maneiras de “como fazer os computadores realizarem coisas que, no momento, as pessoas fazem melhor” (RICH, p. 03) ou, de acordo com John McCarthy, “fazer a máquina comportar-se de tal forma que seja chamada inteligente caso fosse esse o comportamento de um ser humano” (apud BABINI & MARRANGHELLO, 2007, p. 18). Para alguns outros pesquisadores, a Inteligência Artificial ainda englobaria o PLN e outras tantas disciplinas, incluindo a Linguística - privilegiando a Gramática Gerativo Transformacional/competência - e Psicologia Cognitiva - funcionamento do cérebro humano/performance (ARARIBÓIA, 1988). Defini-la é uma tarefa árdua, pois, em vários casos, uma aplicação em PLN, seja qual for, pode ser classificada como dotada de algum tipo de Inteligência Artificial. No entanto, se um sistema, para ser considerado inteligente, deve ser flexível e capaz de se adaptar ao ambiente aprendendo e atualizando periodicamente seu banco de dados de modo minimamente automático (VICCARI, 1989), nem todas as aplicações em PLN terão um grau significativo de Inteligência Artificial – e talvez seja o caso de Tical, tal como se discute em toda a seção 4.4.

Aplicações do PLN são tão difundidas no cotidiano de seus usuários que passam despercebidas pela maior parte deles. Buscadores comandados por voz ou por texto escrito, dicionários eletrônicos, jogos, caixas eletrônicas, reconhecedores de caracteres, recursos automáticos em editores de texto, entre muitos, são apenas alguns dos exemplos mais comuns (BARBOSA, 2004; RICH, 1993; SCHILDT, 1989). Note-se, portanto, que o presente trabalho está também incluído na área do PLN, embora a mais comentada seja uma de suas duas subdivisões: o Processamento de Linguagem Falada, que origina o PSF¹⁸.

Como a matéria prima e referência absoluta do PSF é a voz humana, muito difícil desvinculá-lo da Linguística. No entanto, tradicionalmente, o PSF é mais voltado à Informática e ao Processamento de Dados que às outras com as quais ele se relaciona, como Fonética, Fonologia, Acústica, Eletroacústica, Fonoaudiologia, Gramática Gerativa

¹⁸ Para muitos pesquisadores, Processamento de Linguagem Falada e Processamento de Sinais da Fala são quase a mesma coisa, pois ambos trabalham com a voz humana.

Transformacional, entre outras. Isso pode ser dito porque, embora desde o início do PSF alguns linguistas e simpatizantes tenham participado da criação ou aprimoramento de alguns projetos, no geral, ao longo de várias décadas, sua contribuição foi bastante tímida e é questionável se, de fato, os conhecimentos linguísticos foram aplicados de modo significativamente criterioso em vários casos. Nas obras que versam sobre PSF consultadas, há em seu referencial bibliográfico mormente obras sobre ‘introdução à Fonética e Fonologia’ cujo montante, em média, não ultrapassa 5 % de toda bibliografia consultada.

Um dos indícios de que a participação de linguistas tem sido pequena é o fato de que muitos desses projetos - entre eles os três dispositivos analisados nesta Tese - normalmente não resistem em última análise ao crivo de linguistas que se propõem a avaliar dispositivos e equipamentos disponibilizados para uso do público em geral: eles ainda não operam de modo amplamente eficaz. O bom entendimento de como são estabelecidos alguns dos critérios na avaliação de equipamentos que operam com comandos por voz, no entanto, depende de algumas explicações básicas de como a máquina pode ser programada para processar a Linguagem Natural.

Importante salientar que, historicamente, a Linguagem Natural, quando tomada como objeto de estudo científico, “tem recebido enfoques variados no decorrer do tempo”. Em outras palavras, “não há conclusões definitivas sobre as características, o funcionamento e o uso deste objeto” denominado Linguagem Natural (ARARIBÓIA, 1988, p. 149). Aqui ela será entendida como toda linguagem que pode ser reproduzida basicamente em palavras faladas ou escritas, nem sempre pelos humanos. Então, por extensão, PLN é, basicamente, dar ao computador a capacidade de processar línguas humanas (ARARIBÓIA, 1988).

Assim sendo, as duas seções subsequentes tratam cada qual de um aplicativo que, embora não opere com *comandos por voz*, apresenta o recurso de *síntese de voz*, que fornece a base da compreensão de como uma máquina pode montar e desmontar a fala humana durante o processamento. Importante lembrar que, das quatro - *síntese*, *reconhecimento*, *comandos* e *autenticação de voz* - a *síntese de voz* desenvolveu-se primeiro, na ocasião do surgimento da era da eletrônica, momento a partir do qual a voz humana não apenas pode ser gravada e estudada de modo mais científico, mas também editada – fragmentada, misturada, sobreposta etc. Os primeiros sintetizadores de voz – controlados por máquinas dedicadas ou softwares específicos – só foram possíveis depois desses experimentos mais primitivos. Nesse âmbito, verifica-se que a relação entre variação linguística e processamento de sinais da fala é tão íntima que os desenvolvedores quase sempre querem ouvir o computador ‘falando’ como a si

próprios, sem muitas vezes levar em conta que dominam apenas uma das variantes linguísticas existentes.

3.1 FUNCIONAMENTO DE FONEMAS NO APLICATIVO *BALABOLKA*

A tecnologia na área da computação certamente deu um salto gigantesco a partir da segunda metade do século XX. Depois da invenção do transistor e de sua progressiva miniaturização, foi possível criar os microprocessadores que estão cada vez mais rápidos e menores. Em função disso, alguns sonhos da ficção científica tornaram-se realidade ainda no final do século passado: o videofone a baixo custo e em tempo real e o computador que fala, tal como podem ser apreciados no clássico de Stanley Kubrick *2001 Uma Odisséia no Espaço* (2001, Uma Odisséia, 1968), parecem já estar por aí há algum tempo.

Os computadores atuais com recurso de fala, no entanto, merecem ressalvas. O Hal 9000, do filme de Kubrick, capaz de interagir verbalmente sobre muitos assuntos com os tripulantes da nave, é tão sofisticado que, além dessa tarefa cotidianamente humana, também era responsável pelo funcionamento de todos os dispositivos da nave, inclusive o controle das vidas de outros tripulantes em hibernação. Ocorre, porém, que ele adquire vontade própria e acaba por colocar toda a missão em xeque. Os nossos computadores pessoais da atualidade ainda não fazem tudo isso.

Em termos de softwares comerciais, sejam eles proprietários ou livres, os que estão disponíveis no mercado são alguns poucos e cada qual com recursos bastante limitados se comparados à boa e velha ficção científica. Porém, desconsiderando-se a farta imaginação e a criatividade dos livros e do cinema de ficção, esses poucos softwares (algumas dezenas) costumam atender às funções para as quais foram criados – não fosse isso seriam simplesmente descontinuados.

A evolução, no entanto, não para e mesmo os softwares que operam apenas com a *síntese de voz* mudam constantemente. Assim, o propósito desta seção é justamente discutir, à luz de algumas teorias envolvidas com a Sociolinguística, Geografia Linguística e Dialetoлогия, o funcionamento de ao menos um desses vários softwares disponíveis: o *Balabolka*.

Para tanto, foram avaliados alguns sons e prosódias de palavras, frases e fragmentos retirados desta mesma seção, todos reproduzidos pela voz sintética do *Balabolka*. As palavras avulsas utilizadas para o mesmo fim são provenientes de exercícios e exemplos presentes em livros sobre Fonética e Fonologia – mormente Silva (1999) – e as frases utilizadas como

exemplos têm como ponto de partida algumas dessas mesmas palavras e exercícios de outras obras, como Lopes (1999) e Cagliari e Cagliari (2001, In MUSSALIM e BENTES, 2001).

3.1.1 O *Balabolka*

Antes de começar a falar sobre o *Balabolka*, é importante delimitar ao menos um termo que será utilizado para fazer referência à natureza desses tipos de softwares. Portanto, para o bem da melhor utilização terminológica – em um movimento contrário ao da simplificação – o *Balabolka* – e outros de natureza similar que porventura sejam mencionados ou avaliados – não será chamado de software a partir de agora. A denominação *aplicativo* é mais adequada, primeiramente porque software é um termo bastante geral que abarca muitas modalidades e categorias; depois, porque *programa*, ainda que bastante usual no cotidiano, pode conotar diretamente as linguagens de programação que são acessíveis e manipuláveis apenas a iniciados no ramo de programação.

Assim sendo, o *Balabolka* é um aplicativo comercial, ou seja, não é de uso restrito (militar ou institucional). Também pode ser baixado diretamente do sítio oficial que oferece suporte e atualizações (BALABOLKA, *online*, 2012). Outra característica é que ele se enquadra como aplicativo de *síntese de voz*, ou seja, utiliza-se de partes gravadas da voz humana¹⁹ para formar as palavras que pronuncia.

Tal tipo de aplicativo é conhecido como *text-to-speech*, ou simplesmente TTS. Aqui será denominado *texto-fala* justamente porque é isso mesmo o que ele faz: transforma o texto escrito em fala. Outros aplicativos com características comuns a ele, como o *Dosvox* ou o *IBM Via Voice*, também trazem o recurso TTS, como o *Balabolka*, porém possuem outras aplicações e especificidades. Estes aplicativos são discutidos nas seções 3.2 e 4.1, respectivamente.

Não fica difícil, portanto, imaginar para que serve o *Balabolka*. Basta inserir um texto qualquer, escolher uma das vozes disponíveis para o idioma que se deseja e ele reproduz as palavras com boa performance fonética, se comparado, é claro, a outros aplicativos com voz nitidamente mais robótica como o *Dos Vox*.

É válido lembrar, no entanto, que o aplicativo não faz tradução. Para que ele reproduza em Português é necessário que o texto esteja em Português e que a voz escolhida – homem ou

¹⁹ Os diferentes softwares – cada qual baseado em uma linguagem de programação diferente – usam maneiras diversas de combinar sons. Os difones e trifones são conceitos bastante comuns entre programadores e consistem em partes da fala humana recortadas com critérios diferentes: ora final de vogal com início de consoante, ora final de consoante com início de vogal, ora partes maiores (TEVAH, 2007, YNOGUTI, 2000).

mulher – não só esteja disponível em sua biblioteca, mas também que seja aquela desenvolvida para o Português.

Estando, porém, o aplicativo configurado, as atualizações requeridas todas instaladas e o sistema de áudio do computador funcionando adequadamente, ele de fato desempenha bem sua função. Se os dispositivos que funcionam com comando por voz permitem ao usuário usufruir do *hands free*, ou seja, fica com as mãos livres, o *Balabolka* certamente contribui com algo como o *eyes free*, não é mais necessário ler o texto, basta ouvir. É válido lembrar que sistemas *hands free* já existem há um bom tempo: várias montadoras equipam os modelos mais sofisticados de automóveis com esse recurso, que pode ser desde um simples comando para ligar ou desligar algo, até um sistema completo de acesso à internet com leitura automática de correios eletrônicos, entre outras funções.

Para o ouvinte mais exigente e, eventualmente, apreciador de ficção científica, talvez o *Balabolka* possa deixar a desejar. O Hal 9000 do filme há pouco citado - seu nome é inspirado na expressão em inglês *Heuristically programmed ALgorithmic computer* ou Computador Algorítmico Heuristicamente Programado - certamente parece-se muito mais com um ser humano falando. Considerando-se, no entanto, que no filme é de fato um ser humano - evidentemente - que empresta sua voz ao HAL, tanto na versão original em inglês quanto na dublada em Português, vale a pena ao menos dar uma olhada – ou melhor, ouvir, no *Felipe*, que é uma das vozes masculinas em Português Brasileiro disponíveis para instalação.

3.1.2 O *Balabolka* e a Prosódia

Um usuário comum, interessado nas vantagens de ter alguém - ou alguma máquina - lendo o texto para si certamente verá muitas vantagens no *Balabolka*. Ele pode - bastando para isso alguns pequenos, fáceis e rápidos ajustes - falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto para amanhã - como fazem a mãe, o pai ou os avós que leem histórias para as crianças.

Grosso modo, se por um lado tem o potencial de diminuir o incentivo à leitura - lembrando que, sabidamente, ler contribui para assimilação da ortografia, do vocabulário e da grafemática de modo geral - por outro, pode auxiliar em muito pessoas que sofrem de pouca visão, labirintite, postura ou outro problema de saúde que as impeçam de ficar horas diante de um monitor de computador. Salvo os problemas de saúde ou a pura preguiça de ler, outro fator que favorece o uso talvez seja o de ficar com os olhos livres, como foi dito anteriormente, o que é bastante útil para profissionais que trabalham com aplicativos de animação, desenho,

arquitetura, entre outros. A versão mais aprimorada de Tical, que se tornou operacional até o término deste trabalho, pode fazer isso.

Analisando o funcionamento do *Balabolka* com olhos e ouvidos de linguista, porém, a abordagem é diferente - e os comentários também.

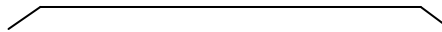
A prosódia é incrivelmente próxima da situação real da fala – como já foi dito, se comparado a outros aplicativos de mesma natureza. As frases afirmativas, os começos e finais de frase, as frases apositivas, entre outras, são muito bem gerenciadas pelo aplicativo.

Importante lembrar que, mesmo sendo o Português Brasileiro um idioma não tonal - como o Mandarim, por exemplo -, os aspectos prosódicos são importantíssimos na distinção entre um modo e outro de pronunciar a mesma frase. Nas palavras de Lopes (1991, p. 121), “a entonação é um meio importante para diferenciar sentidos a partir da variabilidade da altura da voz durante as emissões sonoras, e a partir das variações musicais que se apreendem de modo relativo – umas em relação às outras.”

Uma frase como ‘Nós estamos utilizando o *Balabolka* hoje’, por exemplo, é reproduzida de modo a atender os procedimentos prosódicos básicos. Em termos de tonalidade (altura)²⁰, há certa ênfase ao início e um final mais baixo. As nuances de volume também são controladas, pois, não só as sílabas tônicas são obedecidas como também algumas ênfases em certas palavras podem ser percebidas: os pronomes, os artigos, preposições e conjunções são quase sempre pronunciados de modo mais sutil – tonalidade mais baixa.

Representando graficamente de modo simplificado temos no exemplo (a)²¹:

‘Nós estamos utilizando o *Balabolka* hoje’ (a)



²⁰ No cotidiano, em se tratando de som, as pessoas confundem **altura** com **volume**. Independentemente das nomenclaturas utilizadas pelos diferentes teóricos, é válido lembrar que **altura** está relacionada ao tom, ou seja, a frequência das vibrações das cordas vocais perceptível quase que somente quando da pronúncia de vogais. O **volume** diz respeito à pressão sonora: fala normal, sussurrada, aos gritos, vociferada, entre outras. Borba (1991) prefere o termo **intensidade** a **volume** e Lopes (1999) utiliza-se do conceito de **tonalidade** a **altura**.

²¹ Para cada um dos exemplos (a), (b), (c), Tabela 01 e Tabela 02 há um arquivo de áudio salvo pelos recursos do próprio *Balabolka* disponível para consulta. Os endereços eletrônicos estão no Anexo A deste trabalho. Para ouvi-los a partir do **Dropbox**, basta clicar no canto superior direito em ‘download’, depois ‘direct download’ e repetir a operação se for necessário.

Em se tratando de palavras, expressões ou frases apositivas o aplicativo novamente se sai bem. Com uma frase relativamente curta, dotada de uma oração com valor apositivo, pode-se notar as pausas bem delimitadas no balisamento feito pelas vírgulas:

‘O *Balabolka*, que é um programa de fácil utilização, mostra-se bastante útil a diversas pessoas.’ (b)

Em se tratando de enumeração, ou seja, elementos separados por vírgula seguidos pela conjunção ‘e’ antes do último item, as pausas, volume e tonalidade também são satisfatórios:

‘O *Balabolka* reproduz as pausas, o tom, o volume e as sílabas tônicas.’ (c)

Um fragmento desta mesma seção, também dotado de frases com características apositivas, foi reproduzido de modo bastante satisfatório, tanto usando vírgulas quanto hifens. Note-se que, neste fragmento, disposto na Tabela 03, há também itens enumerados como em (c):

Tabela 03 – Relação do texto e pausas por hifens e vírgulas

Texto	condição
Um usuário comum, interessado nas vantagens de ter alguém - ou alguma máquina - lendo o texto para si certamente verá muitas vantagens no <i>Balabolka</i> . Ele pode - bastando para isso alguns pequenos, fáceis e rápidos ajustes - falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto pra amanhã - como fazem a mãe, o pai ou os avós que leem estórias para as crianças.	hifens por vírgulas
Um usuário comum, interessado nas vantagens de se ter alguém, ou alguma coisa, lendo o texto para si certamente verá muitas vantagens no <i>Balabolka</i> . Ele pode, bastando para isso alguns pequenos, fáceis e rápidos ajustes, falar mais lentamente, em um tom mais alto, com mais ou menos volume e não vai reclamar que está cansado ou que vai deixar o resto pra amanhã, como fazem a mãe, o pai ou os avós que leem estórias para as crianças.	vírgulas por hifens

Fonte: Texto organizado pelo autor

As perguntas, ou seja, os modos interrogativos, no entanto, deixam bastante a desejar. Por exigirem muito mais nuances da prosódia, estas não são reproduzidas da forma como são utilizadas no cotidiano ou como imaginadas na fala de um personagem literário, por exemplo – situação típica em que a prosódia não pode ser representada pela grafemática e requer inferência do leitor. Isso pode ser visualizado analisando a Tabela 04, a seguir, que se baseia

em exemplos encontráveis em Lopes (1999) e Cagliari e Cagliari (2001, in MUSSALIM e BENTES, 2001).

Tabela 04 – Relação entre a frase, ênfase e modalidade

Frase	ênfase	entoação
<u>Ele</u> virá aqui hoje.	quem	assertiva
Ele virá <u>aqui</u> hoje.	onde	
Ele virá aqui <u>hoje</u> .	quando	
<u>Ele</u> virá aqui hoje?	quem	interrogativa
Ele virá <u>aqui</u> hoje?	onde	
Ele virá aqui <u>hoje</u> ?	quando	

Fonte: Lopes (1999) e Cagliari e Cagliari (2001, in MUSSALIM & BENTES, 2001)

A ênfase individual em cada uma das palavras sublinhadas pode mudar completamente a afirmação ou a pergunta e isso o *Balabolka* não consegue fazer.

É importante ressaltar, no entanto, que a complexidade envolvida no processo de falar cada uma das frases da Tabela 04 é bastante grande e nos faz pensar no quanto tempo cada um de nós demora para ter destreza linguística suficiente a ponto de dominá-la. Veja-se, por exemplo, que crianças até determinadas idades podem não distinguir as nuances como estas apresentadas.

Evidentemente, para reproduzir a maior parte das possibilidades prosódicas utilizadas no cotidiano por falantes do Português Brasileiro, ele teria de possuir um banco de dados muito maior e o computador, onde ele estivesse instalado, grande poder de processamento.

Isso se dá, resumidamente porque, como foi dito anteriormente, o *Balabolka* se enquadra como aplicativo de *síntese de voz* e, para tanto, utiliza partes gravadas da voz humana para formar as palavras que pronuncia: quanto mais nuances prosódicas, mais partes devem ser gravadas e gerenciadas pelo aplicativo.

3.1.3 Realizações de Fala e o Registro

O processo de montagem de palavras do *Balabolka* é artificial, ainda que procure reproduzir com a maior presteza possível cada um dos fonemas da língua portuguesa. Após ouvir vários textos a partir dele, tornam-se notórios diferentes graus de artificialidade na cadeia da fala. Fosse uma criança, bastaria dar alguns exemplos e fazê-la repetir algumas vezes. Quanto ao *Balabolka*, a única maneira de corrigi-lo é aumentando seu banco de dados – uma tarefa que também envolve programadores.

As realizações de fala do *Balabolka* - se é que, para o caso de um aplicativo seja conveniente falar em realizações de fala - podem remeter a um registro linguístico específico.

O *Felipe* - uma das vozes masculinas em Português Brasileiro disponíveis – não parece ser do nordeste ou extremo sul do Brasil. Também não faz lembrar alguém nativo de capitais como Cuiabá, Belo Horizonte ou Curitiba. Lembra sim, em uma primeira audição, um apresentador de jornal televisivo de abrangência nacional.

O cuidado em não denominar categoricamente os sons produzidos pelo aplicativo como *realização de fala* pode não se justificar ao levar em consideração que a montagem feita por ele segue um algoritmo construído por falantes nativos - os programadores – e, portanto, tende a agradar os ouvidos dos outros ouvintes que, igualmente, se identificam com a mesma língua. Assim sendo, os sons, em última análise, são perceptivelmente artificiais.

Os sons do aplicativo, no entanto, podem mesmo nos remeter a um registro linguístico específico. Muito do que se encontra em termos de realizações da fala em programas televisivos de grande abrangência também dão essa impressão. Basta, no cotidiano, alguém ter a voz um pouco mais empostada, pronúncia mais apurada e certa correção ao colocar pronomes e conjugar verbos, para virar chacota de outros colegas por parecer-se com apresentador de telejornal, ator de teatro (novela, cinema) ou narrador futebolístico. Isso ocorre nesses meios porque - entre outras razões - a fala é sempre mais tensa (BORBA, 1991).

No cotidiano, porém, em que ocorre a fala menos tensa ou frouxa (BORBA, 1991) isso pode soar um tanto artificial (CAMARA JR, 2006; 2004), lembrando mesmo as pronúncias ensaiadas (TV, teatro, cinema), estilísticas (narrações) e didáticas (docentes de ensino fundamental), e é o que ocorre com o *Felipe*.

Resta responder porque então o *Felipe* e os apresentadores de Telejornal - entre outros, como colocado - não só soam artificiais comparados ao uso linguístico cotidiano como também notoriamente não parecem ter ‘origem’ sulista ou nordestina.

Historicamente, é fato que todas as regiões de maior prestígio sócio-político-cultural exercem grande influência sobre as outras menos privilegiadas, em qualquer continente ou país. Em uma importante capital de determinado país, não só a língua ali falada como o registro linguístico em que se manifesta passa a ter certo privilégio em relação à de outras cidades menos influentes econômica ou politicamente. Para o caso do Português Brasileiro, o padrão falado, além de atender a muitos dos preceitos ditados pela gramática normativa, fundamenta-se a partir do padrão utilizado pelas classes cultas que se concentram nessas capitais (BORBA, 1991; CAMACHO, 2001; LOPES, 1999; SILVA, 1999).

O registro encontrado no eixo Rio-São Paulo parece há muito tempo influenciar o modo como as maiores mídias operam com uma espécie de padronização dialetal, provavelmente, por serem tidas, desde há muito, como grandes centros irradiadores de cultura

(BORBA, 1991). Basta verificar o quanto é raro – ou quase impossível, numa abordagem mais franca e empírica - ouvir tradicionais apresentadores de telejornais de abrangência nacional veiculados pelos maiores canais de TV aberta utilizando um ‘erre’ retroflexo quando do exercício de sua função. Não fica, portanto, difícil verificar o motivo pelo qual o *Felipe* pode fazer lembrar locutores ou apresentadores de telejornal em geral.

O processo de melhoramento de um aplicativo como este, portanto, não se restringe apenas a fazê-lo funcionar nas diferentes versões dos sistemas operacionais mais conhecidos (Windows, Linux ou OS). Além das diferentes plataformas (sistemas operacionais) devem também operar nas configurações dos computadores mais modernos e ainda atender aos quesitos de *usabilidade e ambiência amigável*.

A propósito, os conceitos de *usabilidade e ambiência amigável* utilizados em todo este trabalho estão relacionados a um processo denominado *design* de interação (PREECE, 2005). Assim sendo, a *usabilidade* prevê metas em que os sistemas devem ser eficazes, eficientes, seguros, úteis, apreensíveis/assimiláveis, memorizáveis (PREECE, 2005). A *ambiência amigável*, por sua vez, relaciona-se às *metas decorrentes da experiência do usuário* e prevê que os sistemas sejam “satisfatórios, agradáveis, divertidos, interessantes, úteis, motivadores, esteticamente apreciáveis, incentivadores de criatividade, compensadores, emocionalmente adequados” (PREECE, 2005, p. 40).

Note-se que, diferentemente da *usabilidade*, pautada em metas mais objetivas, a *ambiência amigável* vincula-se à experiência que os sistemas interativos proporcionam aos usuários, ou seja, como eles se sentem ao interagir com um sistema e “isso envolve explicar a natureza da experiência do usuário em termos subjetivos” (PREECE, 2005, p. 40). As vozes sintéticas, entre muitos outros recursos digitais, operam a partir de aplicativos e, assim, são analisáveis também a partir desses parâmetros. Atendidas, portanto, todas essas necessidades, resta tornar a síntese de voz mais convincente aos ouvidos.

O *Felipe* não é a única voz disponível. Há, por exemplo, a *Juliana* e o *Alexandre*, porém, muito inferiores em qualidade. Assim, comparativamente, o *Felipe* já é uma evolução: a voz é mais empostada, a montagem é mais próxima à da cadeia da fala, a qualidade do som é superior. Para melhorar, lembrando que outros aplicativos proprietários²² já são superiores, a nova voz teria de ser menos susceptível a críticas, ou seja, as análises prosódicas e fonético/fonológicas precisam apontar menos discrepâncias em relação a um locutor humano,

²² Em Informática, o conceito de *proprietário* se opõe ao de *livre*. O software proprietário é aquele atrelado a “um modelo econômico baseado na exploração de licenças de uso” (SILVEIRA, 2004), ou seja, não é propriedade do usuário: não pode ser alterado, compartilhado ou vendido.

independentemente da região à qual ele pertence. Para tanto, a seguir há uma breve análise, desta vez, voltada a alguns sons do Português Brasileiro.

3.1.4 Alguns dos Fonemas

Há muitos fonemas que poderiam ser analisados no *Balabolka*, porém, o espaço destinado a esta seção não comporta. Assim, é conveniente iniciar com os sons do rótico, que se realizam de diversas maneiras na mesma palavra e de modo diverso nas regiões do Brasil.

Um dos motivos para a escolha dos róticos é que, no Brasil, ainda que tenha fonologicamente falando poucos valores distintivos se comparados às várias possibilidades de realização em termos fonéticos, constitui muitas vezes motivo para preconceito linguístico, principalmente quando se trata do contraste capital/interior.

Para as transcrições realizadas nesta seção e em todo o trabalho, foi utilizada a fonte SILDoulos IPA, disponível gratuitamente na página do *Summer Institute of Linguistics* - SIL International (2012) e em vários sites relacionados às disciplinas de Fonética, Fonologia, Sociolinguística, Geografia Linguística, Dialetoлогия, entre outras.

As três posições do /r/ mais analisadas e encontráveis em palavras do Português Brasileiro são em ataque ou início de sílaba ou palavra; em coda, no interior e no final de sílaba ou de palavra (SILVA, 1999; CAGLIARI; CAGLIARI, 2001; AGUILERA, 1990; CALLOU; LEITE, 2001).

Em posição de ataque (início de sílaba) como *arara*, *barato*, *carango*, *parada* ou *marajá*, *caro*, *caroteno*, *maracujá* e nos encontros consonantais como *prata*, *graxa*, *brava*, *cara*, *barata* - exemplos extraídos de Silva (1999, p. 48) - o *Felipe* parece utilizar o tepe²³, como em quase todo o Brasil. Nem sempre o tepe tem boa qualidade e isso é verificável quando mantém algumas vizinhanças. A análise sistemática do problema será feita oportunamente em outro estudo pelo fato de não se tratar exatamente de ponto de articulação mas do modo como os sons foram concatenados.

Em palavras como *rata*, *rapaz*, *rama*, *rala* (SILVA, 1999, p. 49) e outras como *rato*, *remo*, *rede*, *roda*, *rima*, *runa* utilizadas como exemplo o som é nitidamente velar [x] ou glotal [h], independentemente da vogal que o segue – ainda que, evidentemente, esteja aberto a discussões.

²³ Para todos esses exemplos, há, também, um arquivo de áudio salvo pelos recursos do próprio *Balabolka* disponível para consulta. Os endereços eletrônicos estão no Anexo deste trabalho.

Em coda silábica externa - final de palavra - algumas considerações extras precisam ser feitas. Ainda que o ponto de articulação seja claramente alveolar, como no tepe, há uma pequena fricção, muito utilizado pelos apresentadores de Telejornal. O tepe realiza-se e é seguido por uma fricção criada pela passagem do ar entre a língua e a região alveolar. Está mais próxima de uma fricativa lateral, tal como pode ser verificado na tabela do Alfabeto Fonético Internacional (SILVA, 1999, p. 41) ou em comentário feito por Borba (1991, p. 121). Essa realização, no entanto, só ocorre quando a palavra é pronunciada separadamente. Quando em contexto, momento em que o ‘erre’ mantém vizinhança com outros fonemas, a realização é diferente.

A palavra ‘mar’, por exemplo, quando pronunciada isoladamente realiza-se [maʁ], tal como descrito anteriormente como uma fricativa vibrante palatalizada. Essa mesma palavra, quando pronunciada em contexto como na frase ‘o mar está calmo’, manifesta-se com um tepe propriamente dito formando sílaba como o segmento fônico seguinte (V+S), ou seja, [’maris]. Esse detalhe não nos dá ainda margem para afirmar que o locutor humano, utilizado na coleta do banco de sons para o *Felipe*, seja paulistano, por exemplo. Primeiro porque, ainda que os paulistanos natos façam, em sua maioria, o uso do tepe em final de palavras como [mar] (CARDOSO et al., 2014b), não realizam na fala cotidiana [maʁ], com fricção. Além disso, mesmo no oeste paulista, onde ‘mar’ realiza-se [maʁ], sua realização em contexto como ‘o mar está calmo’ torna-se também [maris] ou, em transcrição completa, [u, ’maris’ta, ’kawmu].

De acordo com Silva (1999), os sons tendem a ser modificados pelo ambiente em que se encontram. Trata-se, portanto, da primeira premissa que deve ser levada em consideração quando se trata de transcrição fonética. Palavras como *desdém* ou *vesga* são também bons exemplos para isso: [dez’dẽj] e não [des’dẽj].

A propósito, o *Felipe* faz uma boa distinção também desses tipos de vizinhança fônica, ou seja, no interior da mesma palavra. Nas frases ‘Ele não comia **desde** ontem’ e ‘Ele não comia **deste** ontem’, as palavras *desde* e *deste* realizam-se de modos distintos, tal como ocorre na fala de nativos do Português Brasileiro. O aplicativo gerencia adequadamente o fato de, no primeiro caso, o fonema fricativo /s/ avizinhar-se da oclusiva **vozeada** [d] e ter de ser realizada também com vozeamento em [z]. Da mesma maneira faz a adequação quando o mesmo fricativo /s/ avizinha-se da oclusiva **desvozeada** [t], mantendo a igualmente desvozeada [s].

Note-se, por essas diversificações de pronúncia dentro ou fora do contexto, que o *Balabolka* de fato surpreende quanto à funcionalidade. Em programas como *Dosvox* ou *IBM Via Voice*, isso funciona de modo um pouco diverso.

Muitos outros sons do *Felipe*, no *Balabolka*, podem ser comentados e discutidos. Quanto aos ditongos, por exemplo, o *Balabolka* deixa bastante a desejar. Normalmente, os glides presentes em palavras como *noite*, *caixa*, *baixo*, *árdua*, *mágoa*, *tênue*, *côngrue*, no Português Brasileiro são transcritos como [j] e [w]. Para o caso do Felipe, o que seriam glides nessas palavras são fonemas vocálicos pronunciados separadamente: soam muito artificiais - a qualquer ouvido - e resultam em transcrições como [ˈnoitʃɪ] ou [ˈkaiʃɐ] e não [ˈnojtʃɪ] ou [ˈkajʃɐ], tal como ocorre nas transcrições do ALiB.

Em suma, a *síntese de voz*, ainda que tenha evoluído muito, não passa despercebida aos linguistas de modo geral. Há vários relatos curiosos do dia-a-dia em que, por exemplo, algumas pessoas chegam ao ponto de dizer obrigado ao ‘robô’ que fala do outro lado da linha. Isso ocorre porque a URA, ou Unidade de Resposta Audível²⁴, um dos recursos muito comuns na telefonia mundial, opera com um sistema de *síntese de voz* muito mais apurado que o *Balabolka*, entre outros. A explicação é simples: quanto mais próximo da pronúncia humana estiver a voz do computador, mais se acentuam os conceitos de *usabilidade* e *ambiência amigável*. Em outras palavras, a irritação do cliente pode ser minimizada uma vez que ele consegue ao menos entender o que está sendo dito.

Nesse sentido, vale a pena avaliar o funcionamento do *Balabolka* - e outros como o *IBM Via Voice* ou *Dosvox*, em outras oportunidades - justamente para ter um parâmetro mínimo de comparação, tendo em vista que, comparar diretamente seu desempenho com o de um ser humano em termos de realização de fala, é ainda um tanto desfuncional. O Hall 9000 de Kubrick ainda é ficção, mas as pesquisas estão evoluindo e a contribuição desta pesquisa, ainda que pequena, certamente auxiliará na construção de dispositivos mais sofisticados e, de fato, úteis para o ser humano.

3.2 ASPECTOS FONÉTICOS NO *DOSVOX* ENQUANTO APLICATIVO TIPO TEXTO-FALA

A voz dos robôs em desenhos animados é inconfundível. De saída, o ritmo e a prosódia - deixando de lado deliberadamente muitos outros elementos analisáveis - são nítida

²⁴ Um dos sistemas que opera com a URA no Brasil é o Asterisk que, em pesquisas vindouras, merece também atenção dos linguistas.

e propositadamente artificiais. Os desenhos animados e/ou o cinema, no entanto, não criaram sozinhos este estereótipo de voz metálica artificial. Os resultados dos primeiros experimentos no século XX que envolvem a tentativa de proporcionar a capacidade de fala ao computador, com *síntese de voz*²⁵ ou não, auxiliaram em muito sua criação²⁶. Isaac Asimov, Arthur Charles Clarke e mesmo brasileiros, como André Carneiro, Otávio Aragão, Tibor Moricz, Fábio Fernandes, Bráulio Tavares, Jorge Luís Calife, Fausto Fawcet e muitos outros autores de ficção científica do orbe também deram sua contribuição. Textualmente, cada qual descreve a seu modo as características audíveis da voz dos autômatos. Desde os mais antigos livros, seriados e filmes de ficção até os mais recentes gibis, como os da *Turma da Mônica*, apresentam vez ou outra tais características em alguns personagens robóticos.

Essas nuances também podem ser encontradas em alguns aplicativos de computadores. Nesses, a similaridade com o estereótipo há pouco citado não é propositado, mas resultado de uma série de fatores de ordem técnica, absolutamente discutíveis no âmbito da Fonética e/ou Fonologia.

O *Balabolka*, estudado na seção anterior, é um entre vários aplicativos tipo *texto-fala* disponíveis gratuitamente na grande rede para instalação em computadores pessoais. Na presente seção, discute-se sobre o *DosVox*, que também se enquadra nessa categoria – tipo *texto-fala* – mas que foi desenvolvido para uma finalidade um pouco distinta em relação aos outros que desempenham funções similares.

Muito distante da realidade representada em filmes e/ou livros de ficção científica, tais softwares ainda não têm a capacidade de simular com grande êxito a presença de outro ser humano. Em outras palavras, ‘bater um papo’ - no sentido mais amplo encerrado por essa expressão - com um computador e, entremeios, solicitar-lhe serviços, daria aos conceitos de *usabilidade* e *ambiência amigável* um patamar altíssimo por ora inalcançável, ainda que absolutamente imaginável. Os aplicativos que convertem texto em fala, por mais aprimorada que seja a sua versão correspondente, estão muito aquém disso.

²⁵ *Síntese de voz* é o recurso pelo qual é possível conseguir a voz sintética, ou seja, a voz que um software ou aplicativo produz ao formar as palavras com partes de voz humana previamente gravadas. A expressão *síntese de voz*, juntamente com *reconhecimento de voz*, *autenticação de voz* ou *comandos e controle por voz* constituem as quatro áreas de processamento de sinais de voz (HUGO, 1996; DAMASCENO, 2006; TEVAH, 2007; YNOGUTI, 2000).

²⁶ O Vocoder, criado dos anos trinta do século XX, era um analisador e sintetizador para voz humana. Concebido inicialmente para permitir sintetização e criptografia da voz para transmissões sigilosas, foi posteriormente muito utilizado para fins artísticos, principalmente na música.

A fala, nesse caso, opera muitas vezes com o conceito de *síntese de voz*. Logo, não é de fato a voz do computador²⁷ que se ouve, mas partes de uma voz humana previamente gravadas que são montadas pelo programa numa sequência audível e interpretável pela audição humana, de acordo com o modo como o algoritmo foi disposto.

O *DosVox* também tem essa característica, porém, como foi desenvolvido especificamente para pessoas com necessidades especiais no quesito visão, traz também um recurso que é algo como um *menu audível*²⁸, ou seja, para cada função escolhida pelo usuário acompanha a pronúncia do item escolhido em *voz sintética*. Isso auxilia em muito uma pessoa que não enxerga - ou que enxerga de modo insuficiente - a usar um computador uma vez que não só ele pode explorar os menus e funções com segurança como também pode ouvir o texto que escreveu. A propósito, versões mais sofisticadas de Tical, podem também oferecer um menu audível.

Nesta seção, portanto, estão expostas algumas dessas características do *menu audível* do *DosVox* em termos fonéticos no sentido de procurar avaliar sua aplicabilidade a diversos brasileiros, uma vez que os fonemas do Português Brasileiro nele reproduzidos correspondem a realizações fonéticas típicas de determinadas variantes e não de outras. Mais do que isso, em alguns casos, fica difícil estabelecer correspondência significativa e a análise pode apresentar resultados inusitados.

3.2.2 O *DosVox*

O *DosVox* é um aplicativo para computadores pessoais desenvolvido especificamente para auxiliar pessoas, com pouca ou nenhuma visão, a criar e ler textos. Desenvolvido pelo Núcleo de Computação Eletrônica da Universidade Federal do Rio de Janeiro - UFRJ, que há alguns anos se empenha no aprimoramento desse aplicativo, o *DosVox* pode permitir a essas pessoas um alto nível de independência no estudo e no trabalho (PROJETO DOSVOX, *online*, 2012). Para tanto, é necessário que a pessoa com tal necessidade especial seja dotada do sentido da audição. Não é, portanto, exagero dizer que um computador pessoal cujo dispositivo de áudio esteja desconfigurado ou em mau funcionamento pode implicar o mau uso ou até mesmo a impossibilidade de utilização do aplicativo.

²⁷ Popularmente, costuma-se atribuir ao computador a capacidade de 'falar' quando este reproduz sequências sonoras inteligíveis e/ou familiares a um falante nativo ou não. O conceito que envolve a ação de falar, no entanto, é muito mais complexo do que isso.

²⁸ O menu audível pode ser considerado uma extensão do menu visual convencional, uma vez que é organizado mentalmente de modo similar e, ambos, estão entre as IHCs mais usuais quando se trata de PLN - ver seção 3.2.

De acordo com o próprio Projeto *DosVox* (PROJETO DOSVOX, *online*, 2012), o aplicativo proporciona uma utilização altamente *amigável* em função justamente do sistema de som ser feito por meio da *síntese de voz*, o que o diferencia de aplicativos mais antigos que fazem o uso de outro sistema de encadeamento de sons, em que a voz produzida é claramente robótica. A título de curiosidade, por ora, é válido lembrar que, em princípio, a apresentação textual do Projeto *DosVox* não parece preocupar-se com a distinção entre *síntese de voz* e *síntese de fala* – usam uma ou outra ao longo do texto. Essa distinção, pouco importante para os objetivos da presente seção, será discutida em pesquisas posteriores.

De qualquer forma, tal como enfatiza o Projeto, a maioria das mensagens sonoras no *DosVox* é produzida por voz humana previamente gravada, o que pode enquadrá-lo como “um sistema com baixo índice de estresse para o usuário, mesmo com uso prolongado²⁹” (PROJETO DOSVOX, *online*, 2012).

Outro recurso interessante elencado pelo texto do Projeto é a compatibilidade do *DosVox* com a maior parte dos sintetizadores existentes por usar a interface padronizada SAPI do Windows que, tranquilamente, é o sistema operacional mais conhecido e utilizado no mundo.

O Projeto também fala das vantagens da interface que, como será visto adiante, é de fato relativamente simples e de fácil utilização. Pessoas com deficiência visual, no entanto, e dependendo do grau dessa deficiência, jamais poderão ter essa mesma impressão, pois não podem vê-la. No máximo, conseguem imaginar a disposição dos itens na tela a partir do menu auditivo/sonoro que lhes é disposto quando da navegação a partir das funções alfa-numéricas do teclado do computador. Essa característica da interface, portanto, só pode ser apreciada e/ou avaliada por pesquisadores ou por monitores que previamente acompanham essas pessoas quando dos primeiros passos de utilização.

3.2.3 Estabelecendo Comparações

Embora o processo de avaliar aplicativos desse tipo, sob questões de ordem fonética, tenha tendência a suscitar resultados distintos em função das especificidades de cada um, foi inevitável estabelecer comparações com o *Balabolka*.

²⁹ Está claro que esse quesito estresse vinculado ao uso também não será avaliado nesta seção, pois, além de não se constituir da finalidade específica desta parte do trabalho, é bastante subjetivo e exigiria um estudo à parte bem mais aprofundado.

Primeiramente, não há como alterar a voz do *DosVox* de modo rápido e prático como se faz com o *Balabolka* – este nos oferece facilmente as vozes denominadas *Felipe* e *Raquel*, montadas com partes de vozes de um homem e de uma mulher, respectivamente. No *DosVox* também há mais de uma voz, mas o critério de distribuição não parece obedecer a alguma questão estética ou de usabilidade, mas de funcionalidade: ao escolher uma função do menu, a voz manifesta-se de imediato, independentemente de qual seja – homem ou mulher, esta ou aquela.

Isso não implica afirmar que a qualidade das vozes disponíveis do *DosVox* é sempre similar ou bem equilibrada. Pelo contrário, ao navegar pelos menus, pode-se notar de imediato que as diferenças são gritantes.

Ao abrir o programa, vozes femininas manifestam-se prontamente apresentando o programa. Para a versão analisada, a 4.3, a mensagem em áudio é a seguinte: “Sistema *DosVox*. Versão quatro ponto três ‘a’. Núcleo de Computação Eletrônica da UFRJ. Boa tarde. O que você deseja?”³⁰.

A qualidade das vozes - que não parecem ter a mesma matriz humana - nessa apresentação é significativamente boa se comparada ao *Balabolka*. Na verdade, não se trata de síntese de voz como é o caso deste último, mas de sintaxes gravadas na íntegra e não partes de palavras montadas pelo aplicativo. A título de breve esclarecimento, o fragmento em áudio “Versão quatro ponto três ‘a’” não só permite verificar voz humana diversa do restante da apresentação como também torna notório o modo como as partes textuais sofreram divisão. A tela de apresentação pode ser apreciada na Figura 2.

³⁰ Importante atentar para o detalhe de que a saudação – bom dia, boa tarde ou boa noite - alterna de acordo com o período do dia, desde que o relógio do sistema operacional disponível no computador em que o *DosVox* está instalado se encontre devidamente configurado e, tal detalhe alinha-se perfeitamente com conceitos de *usabilidade* e *ambiência amigável*, muito discutidos quando se trata de softwares de um modo geral.

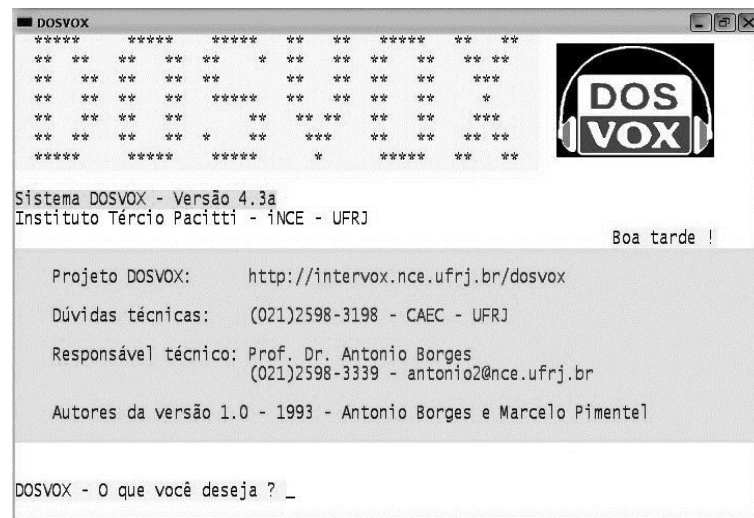


Figura 02 - Tela de apresentação do *DosVox*

Após a mensagem, basta que o usuário acesse o menu a partir do comando F1 - recurso ajuda – ou, se já tiver prévio conhecimento do aplicativo, a partir das teclas referência ‘t’, ‘e’, ‘l’ entre outras, como na Figura 03. Essa informação sobre a possibilidade da função F1, no entanto, não é apresentada ao usuário em forma de áudio. Para alguém com deficiência visual saber que é possível conseguir mais informações sobre o *DosVox* a partir do comando F1 é necessário que alguém, como um monitor, o acompanhe quando das primeiras utilizações. Ele será exibido em formato de texto na tela e, naturalmente, também em áudio. A imagem da tela do menu pode ser apreciada na Figura 3 a seguir.

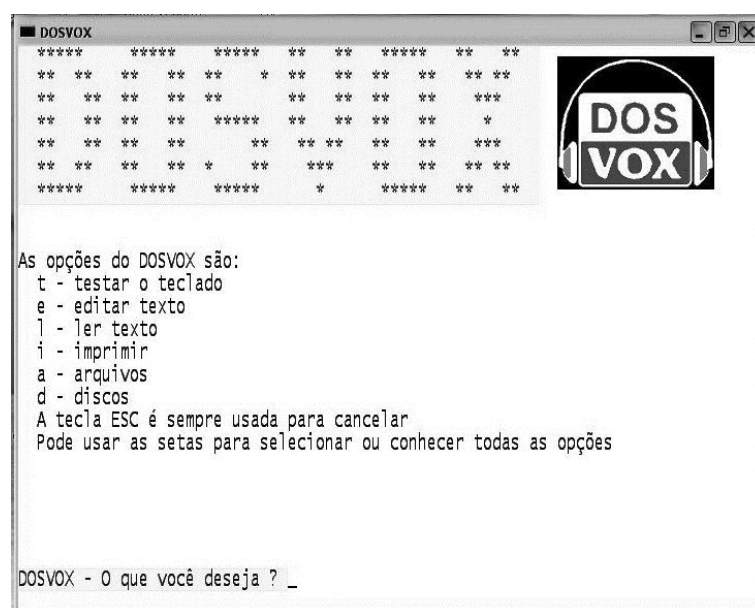


Figura 03 - Menu primário do *DosVox*

Vale lembrar, a título de esclarecimento, que todas as figuras relativas aos menus e telas do *DosVox* não foram aqui apresentadas com as cores originais, mas em negativo e monocromático, ou seja, a composição cromática foi editada. A escolha tem duas razões de ser: o fundo original é em preto, o que ocasiona certo desconforto à leitura para pessoas que não têm deficiência visual, principalmente do modo impresso que pode ser o caso desta seção; aqui as cores não estão sendo avaliadas – o que certamente seria bastante interessante para um trabalho que contemplasse a dimensão semiótica/semiológica.

Quanto ao menu inicial, basta pressionar a tecla F1 para que ele seja acionado imediatamente e, como ocorre na abertura do programa, ditado item por item por uma voz feminina. Até esse ponto a voz tem boa qualidade em termos de dicção e pronúncia e quase sempre é a mesma. Quando se faz a escolha por um dos itens, a avaliação fonética passa a ter mais trabalho e aumenta em complexidade.

O item ‘a- arquivos’, por exemplo, acionado pela tecla alfabética ‘a’, apresenta ao lado direito da tela uma lista de todos os arquivos pré-existentes e/ou criados e salvos pelo usuário, sempre obedecendo à ordem alfabética, tal como pode ser visto na Figura 4.

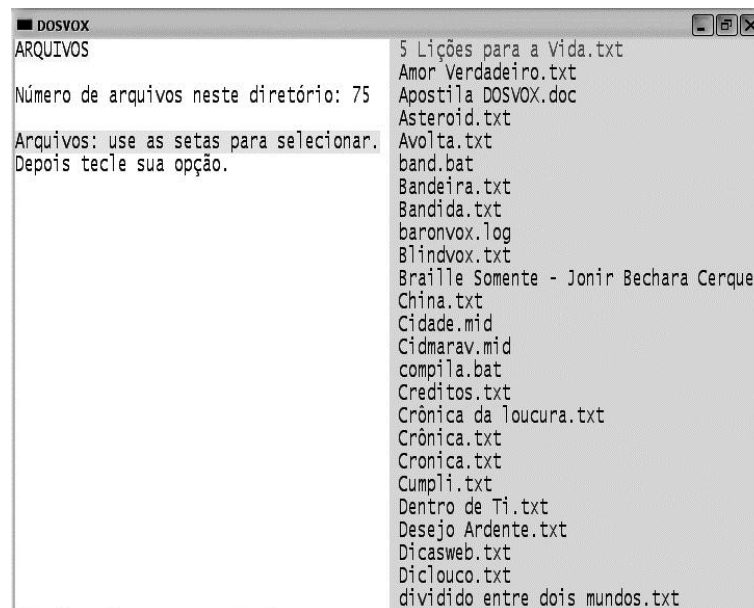


Figura 04 - Menu de arquivos do *DosVox*

Como de práxis, tudo o que aparece escrito na tela - incluindo o primeiro arquivo disposto na lista - é apresentado também em áudio. O nome no primeiro arquivo (documento), no entanto, mostra-se com uma voz quase ininteligível.

Esse ponto da exploração do aplicativo é um dos momentos em que se percebe claramente o contraste entre os diferentes modos e/ou critérios de operar com a *síntese de voz*.

Como dito anteriormente – e, claro, considerando-se a sequência de acionamentos dos respectivos menus adotada até o presente momento - a qualidade da voz desde o clique duplo no ícone do *DosVox* até imediatamente antes da ‘pronúncia’ do nome do primeiro documento tipo ‘txt’ presente na lista de arquivos é significativamente boa: são de fato sintaxes ou palavras gravadas na íntegra e não partes de palavras montadas pelo aplicativo.

A pronúncia do nome do documento denominado ‘5 Lições para a Vida ponto te xis te’ é resultante de um processo de síntese de voz humana bastante similar àquele encontrável no *Balabolka*, ou em outros da mesma natureza. O critério de formação das palavras, no entanto, é diverso. Provavelmente em função de ter sido desenvolvido em uma época na qual a capacidade de processamento dos computadores era muito inferior à de hoje, muitos quesitos atuais imprescindíveis eram impensáveis para a ocasião. Talvez seja por isso que as partes gravadas parecem obedecer a um critério obsoleto para os padrões mais atuais. No entanto, a comparação dessa voz que pronuncia o nome do documento com aquela da mensagem em áudio, quando da abertura do programa apresentada há pouco, pode desencorajar alguns linguistas a continuar com a análise, tamanha é a discrepância entre um padrão e outro.

Os dizeres “Sistema DosVox. Versão quatro ponto três ‘a’. Núcleo de Computação Eletrônica da UFRJ. Boa tarde. O que você deseja?” fazem lembrar logo em primeira mão um falante nativo do estado de São Paulo, mais especificamente da capital e de regiões com pronúncia similar. A transcrição³¹ aproximada para cada uma das sintaxes separadas por ponto de nossa transcrição grafemática seria:

[sis'tēmɐ dɔs'vɔks], [ver'sãw 'kwatru 'põtu 'trejʃ a],
 ['nukljɯ 'dʒi kõputa'sãw 'grafikɐ 'dɐ uefi'ehi'zɔtɐ],
 ['boɐ 'tardʒɪ], [ukivo'sede'zeʒɐ].

Note-se por essa transcrição que, apesar de a equipe responsável pelo aplicativo ser do estado do Rio de Janeiro, a pronúncia está bem longe do padrão carioca. Note-se também que esse comentário é apenas uma curiosidade uma vez que, se comparado à avaliação do *menu audível* disposta mais adiante, praticamente não causa estranheza ao usuário.

Por outro lado, a pronúncia do documento denominado “5 Lições para a Vida”, feita na íntegra incluindo a extensão (‘5 Lições para a Vida ponto te xis te’) quase que certamente soará muito estranha a qualquer falante do Português Brasileiro: risível, robótica, artificial e

³¹ Para todas as transcrições aqui realizadas foi usada a fonte SIL Doulos IPA, disponível gratuitamente na página do *Summer Institute of Linguistics - SIL International* (2012).

até ininteligível. A prosódia é estropiada, o ritmo é desengonçado e os fonemas incomuns para a vizinhança fônica em questão. Apenas esses motivos já tornam difícil a associação dessa pronúncia à realização fonética de alguma região específica do Brasil. A sua transcrição aproximada seria [ˈsĩku liˈsõjs ˈpara ˈvidɐ ˈpõtu ˈte ˈʃis ˈte] e ainda que essa realização³² se pareça com a de algum paulista ou norte-paranaense, basta avaliar outras pronúncias relativas à mesma lista de documentos tal como mostra a Figura 04 para que se verifique rapidamente a ausência de um padrão fonético facilmente identificável.

A seguir, está disposta a lista de algumas das transcrições para apreciação e análise. Como não é objetivo desta seção - e mesmo do trabalho - avaliar à exaustão todos os detalhes de todas as *realizações* presentes no *DosVox*, foram apresentadas apenas aquelas que se constituam em exemplos mais práticos à presente discussão. Para que as análises sejam mais objetivas e eficientes, serão suprimidas as transcrições das extensões dos documentos tais como .txt, .mid, .qst entre outras. Todos os títulos dos conteúdos dispostos na lista foram transcritos grafematicamente tal como se encontram no original – maiúsculas, minúsculas, caixa alta, espaços entre outros.

Tabela 05 - Nomes dos arquivos e respectivas transcrições

a)	Amor Verdadeiro.txt	[aˈmoxvɛdaˈdeirɔ]
b)	Noturno.mid	[noˈtuxnɔ]
c)	Avolta.txt	[aˈvɔltɐ]
d)	Luzes do arrebol.txt	[ˈluzizduaxɐˈbɔw]
e)	Nunca desista.wma	[ˈnũkadeˈzistɐ]
f)	História das Copas.qst	[istˈɔriɛdɛsˈkopɐs]
g)	Apostila DOSVOX.txt	[aposˈtʃilɛdɔsvˈɔks]

Fonte: Texto organizado pelo autor

Começando pela primeira opção, “Amor verdadeiro”, cuja transcrição apresenta por duas vezes o som velar desvozeado [x], primeiro próximo da vozeada [v] e depois próximo da também vozeada [d], em vez da esperada [ɣ] em função da vizinhança fônica. Lembrando que os sons tendem a ser modificados pelo ambiente em que se encontram (SILVA, 1999). Tal ocorrência repete-se em “Noturno”, momento em que novamente a velar [x] avizinha-se de [n].

³² A *realização de fala* é uma expressão comumente utilizada entre linguistas, mormente aqueles que trabalham especificamente com Fonética e Fonologia (SILVA, 1999). Quando se fala em *realização de fala* no campo da *síntese de voz* ou *síntese de fala*, trata-se evidentemente de uma analogia, uma vez que a máquina não é dotada de um aparelho fonador e necessita, ao menos, de partes de vozes humanas gravadas para montar palavras.

Em “A volta”, cuja transcrição é [a'voltɐ], há algo inusitado como a lateral [l], mais dificilmente encontrável em realizações da maior parte dos brasileiros. Essa lateral contrasta imediatamente com a transcrição de “Luzes do arrebol”, para a qual, tomando como base a decisão do sintetizador no exemplo anterior, seria presumível [ˈluzizdwaxeˈbol], também com uma lateral em final de palavra, em vez de [ˈluzizdwaxeˈbow] como se manifestou.

Para as três últimas opções, várias observações podem ser feitas. A fricativa [s] manifesta-se nas três opções de acordo com a norma paulistana ou norte-paranaense. No entanto, quando se encontra a velar [x] em construções como “Amor verdadeiro” e “Noturno”, ainda que, como discutido há pouco, seja um som também incomum tendo em vista a vizinhança fônica, talvez fosse também natural que se encontrasse algo um pouco mais carioca como [ʃ] em “Nunca desista”, “História das copas” e “Apostila DosVox” como [ˈnũkadeˈziʃtɐ], [iʃˈtɔriɐdɐʃˈkopɐʃ] e [apoʃˈtilɐdɔsvˈɔks] e isso não ocorre.

A vogal anterior semi-aberta [ɔ] também é realizada de modo diverso do esperado em “História das copas”, como pode ser observado na Tabela 05. Na palavra ‘história’ há o fonema [ɔ], meio aberto como em quase todo o Brasil. Porém, a palavra ‘copas’ apresenta [o], meio fechado e incomum entre as realizações de brasileiros em geral. Note-se que, nessas duas palavras, essa vogal é tônica. O [o] pretônico, no entanto, aparece em “Noturno” e “Apostila DosVox”, sem alçamento. Esse detalhe, contudo, serve apenas para lembrar que, em se tratando de falantes, o alçamento de [o] para [u] na condição pretônica comentada seria absolutamente possível para uma amostragem maior, principalmente se os dados fossem coletados de modo espontâneo³³.

Esses poucos e superficiais detalhes, analisados a partir da Tabela 05, são suficientes para causar estranheza ao usuário do aplicativo. Afinal, ainda que o interlocutor saiba que se trata de uma máquina com as suas limitações típicas, algumas vezes não é necessário alterar mais que um som para que a dificuldade de compreensão se estabeleça.

Sabe-se que apenas um traço do feixe de características fônicas de um único fonema pode alterar o significado de uma palavra (SILVA, 1999; BORBA, 1991; LOPES, 1999). No entanto, exemplos clássicos como ‘pato’ e ‘bato’, que se constituem de palavras com significado absolutamente diferente, não servem de comparação tendo em vista que para o

³³ Um interessante estudo sobre o alçamento da vogal [o] pode ser apreciado em Kailer (2012).

caso de [kopɐs], mesmo que pouco recorrente em Português Brasileiro, não gera grandes problemas uma vez que não há outra palavra em Português com similaridade fônica considerável que tenha significado diverso.

A dificuldade, portanto, está em o usuário estabelecer correspondência com alguma variedade do Português Brasileiro para que possa *calibrar* sua audição à nova condição linguística estabelecida pelos recursos do aplicativo. Em outras palavras, o *DosVox* em alguns momentos, pode fazer lembrar paulistas, paranaenses, gaúchos, cariocas, mineiros e nenhum deles ao mesmo tempo.

O problema, porém, não é apenas esse de os fonemas serem incomuns para a vizinhança fônica em questão e/ou por dificultar correspondência direta com alguma variedade brasileira. Como foi comentado ao início, em alguns dos *menus audíveis* do *DosVox*, a prosódia e o ritmo parecem pouco naturais.

Válido lembrar, no entanto, que são aplicativos como o *DosVox* e o *Balabolka* entre outros, que convidam o linguista a sempre refletir sobre como as sutilezas do ritmo - muitas vezes apenas passíveis de ser mensuradas em milissegundos - e da prosódia - com todas as suas qualidades tonais são importantes, carregadas de significado e difíceis de serem reproduzidas por máquinas de modo tão espontâneo como é feito naturalmente pelos seres humanos.

Por outro lado, quando ritmo e prosódia são minimamente estranhos à audição do falante, este automática e inconscientemente identifica de imediato a diferença, tenta a partir de seu conhecimento de mundo localizar a variante em alguma categoria que lhe seja familiar ou plausível. Isso ocorre, entre outros fatores, porque falar determinado idioma requer um conhecimento que normalmente ultrapassa uma dimensão puramente linguística (SILVA, 1999).

Daí, não havendo uma correspondência direta entre o que ouviu e o que conhece ou concebe como possível, pode chegar à conclusão de que tal manifestação linguística não é natural de sua região ou mesmo de seu país. Quando o grau de discrepância é elevado, tende a levar à desconfiança ou até a uma condição risível.

Para o caso do *DosVox* ou do *Balabolka* - este já discutido na seção 3.1 - basta fazer alguns testes e navegar pelos menus para que o ouvido extremamente crítico e implacável do falante nativo identifique alterações ou mesmo falhas ou anomalias.

Concluindo, aquilo que se convencionou chamar de voz robótica no século XX, como foi dito anteriormente, não é encontrável apenas em descrições detalhadas de livros de ficção

científica, filmes, novelas e seriados dessa mesma modalidade, além de gibis, desenhos animados e nas brincadeiras entre seres humanos. Essas nuanças de voz também podem ser verificadas em alguns aplicativos para computadores. Nesta seção foram discutidos alguns aspectos que levam o *DosVox* a ter também algumas dessas particularidades. Evidentemente que nele, assim como em outros aplicativos, a similaridade com o estereótipo veiculado pela ficção não é propositada, mas resultado de uma série de fatores de ordem técnica que foi discutida no âmbito da Fonética e/ou Fonologia.

Como foi visto, a breve transcrição feita na Tabela 05 demonstrou que vários fatores podem causar estranheza ao usuário em alguns dos submenus do *menu audível*, seja ele pertencente a qualquer região do país: sons permanecem inalterados quando deveriam mudar com a vizinhança fônica; algumas vogais e semivogais alternam entre si sem obedecer a um padrão discernível; a prosódia de modo geral é notadamente alheia a qualquer variante do Brasil.

Causar estranheza, no entanto, não tira o mérito do aplicativo que há muito auxilia pessoas com necessidades especiais de visão quanto à utilização básica de computadores. O hábito de uso, no entanto, é essencial para que o usuário possa acostumar-se ou, como colocado anteriormente, *calibrar* sua audição às particularidades fonéticas lá disponíveis.

Porém, isso apenas não é suficiente tendo em vista a evolução rápida por que passam todas as tecnologias nesse começo de século. Saltar de uma condição de voz robótica para algo bem mais próximo do real, como ocorre no *Balabolka*, pode ser a próxima etapa para desenvolvedores de aplicativos como este, tendo em vista que quanto mais proximidade houver entre a realização de um falante nativo e os sons reproduzidos pelo aplicativo, mais natural torna-se o uso e mais acentuado tornam-se os conceitos de usabilidade.

Um limite máximo de *usabilidade e ambiência amigável*, ainda que dificilmente possa ser mensurado, pode estar relacionado à condição em que um falante nativo terá muita dificuldade em apontar falhas de ordem fonética e/ou prosódica na realização da máquina, momento em que as vozes robóticas só serão bem vindas em obras de ficção e/ou brincadeiras domésticas.

3.3 PROCESSAMENTO DE LINGUAGEM NATURAL, PROCESSAMENTO DE SINAIS DA FALA, GEOLINGUÍSTICA

Lee de Forest, em 1906, talvez jamais tivesse imaginado que sua válvula termiônica seria substituída pelo transistor cinquenta anos depois. Mesmo que tivesse previsto algo tão

improvável, não imaginaria que o transistor pudesse ficar tão pequeno a ponto de poder ser impresso em uma pastilha de silício juntamente com outros milhares deles formando um microchip, capaz, dependendo de sua arquitetura interna, de processar muitos dados por segundo, que se conhece hoje como computador eletrônico.

Tão comum quanto os equipamentos e eletrodomésticos valvulados até o início dos anos 1950 é o computador nessa segunda década do século 21. Ele ainda talvez seja mais comum comparativamente aos antigos valvulares porque está de fato em todos os lugares: celulares, smartphones, vídeo games, televisores, equipamentos de som, automóveis, aviões, trens, navios e até mesmo dentro de nossos corpos, miniaturizados, escondidos e funcionais.

Com a válvula eletrônica, o poder de controlar dispositivos à distância passou da credence nas habilidades dos magos aos transmissores de ondas eletromagnéticas - equipamentos de rádio - já no início do século XX. Com o transistor, os autômatos que interagem e até conversam com seres humanos saltaram dos livros e filmes de ficção científica para a realidade, no mesmo século, um pouco mais tarde, no florescer da década de sessenta.

As áreas de PLN e de PSF evoluíram muito a partir dessa época e os computadores acompanharam essa evolução: estão bem menores, mais leves e portáteis, consomem menos energia e sua capacidade de processamento aumenta cada vez mais³⁴. A Linguística, por sua vez, não ficou atrás e faz parte desse conjunto de áreas em ascensão. Muitos conceitos e elementos já pesquisados e experimentados em cada uma dessas áreas costumam convergir algumas vezes em tecnologias que concentram os esforços de cada uma delas. Esta seção, portanto, expõe elementos que mostram como a Linguística de um modo geral tem marcado presença em diferentes esferas tecnológicas como a informática, as telecomunicações, a acessibilidade, a Domótica – de *domus* + robótica, relativa à automação residencial inteligente (BOLZANI, 2004) - entre outras.

3.3.2 PLN, PSF e Geolinguística

Todas as áreas de pesquisa têm um ponto de referência para o qual convergem todos os olhares quando se busca pelas origens. Na área de PLN, embora obsoleto, um dos programas de computador mais citados até hoje é o Eliza (WEIZENBAUM, 1966).

³⁴ A conhecida observação do Gordon Moore de que a capacidade de processamento dos computadores dobra a cada 18 meses (BAUER, WESTFALL, DIAS, 2013; HAWKING, 2009) é curiosamente interessante e assustadoramente comprovável.

Conhecido também como robô de conversação, tal como alguns outros da mesma natureza, tinha a função de simular diálogos com um interlocutor humano. Esses diálogos eram todos realizados por meio da linguagem escrita, ou seja, utilizando o teclado alfabético.

O curioso é que o Eliza foi implementado para assumir o papel de um psiquiatra e várias pessoas que interagiram com ele naquela época acreditaram estar de fato falando com um profissional humano. Assim, não seria exagero imaginar que, se em vez do teclado, o Eliza fosse equipado com um reconhecedor de voz e um sintetizador de voz, o experimento de Weizenbaum talvez tivesse sido considerado ainda mais polêmico do que foi na época. Atualmente, a versão mais incrementada de Tical com *síntese* e *comandos por voz* não causa a menor estranheza.

Como comentado no capítulo 3, esses dois recursos compõem duas das quatro áreas mais importantes do Processamento de Sinais na Fala. Além da *síntese de voz* e do *reconhecimento de voz*, há também a *autenticação de voz* e *controle por voz*. Das quatro, talvez o que menos interesse num primeiro momento nesta Tese seja a *autenticação de voz* em função de se tratar basicamente de um recurso voltado à biometria: um dispositivo só pode ser acionado pela voz de determinada pessoa, por exemplo. Os outros três podem ser tanto estudados quanto implementados ou aperfeiçoados por intervenção da Linguística. Vários aplicativos existentes no mercado são exemplos disso.

O descontinuado *IBM Via Voice*, por exemplo, é um do tipo fala-texto (*speech to text*), ou seja, converte palavras faladas em texto e enquadra-se como recurso de *reconhecimento de voz*. O *Balabolka* e o *DosVox*, cada qual desenvolvido para um fim específico, enquadram-se como recurso de *síntese de voz*: convertem textos escritos em textos audíveis, ou seja, reproduzem com maior ou menor acurácia a fala humana pertencendo ao tipo texto-fala ou *text to speech*. Para a Geolinguística, o PLN e o PSF talvez sejam apenas conhecidos de vista e, portanto, precisam ser apresentados. Os dados do ALiB podem tranquilamente estar presentes em um robô de conversação - como o Eliza ou o Lunar (WOODS, 1972) - e isso é, no mínimo, bastante interessante para pesquisadores, professores e alunos. Os modos como isso pode ser implementado são muitos: o aplicativo Tical presente nesta Tese é um deles.

A explicação dessa proposição é simples pois as diferentes maneiras como a mesma palavra é pronunciada no Brasil e a variação lexical para o mesmo conceito poderiam ser exploradas tanto pelo PLN quanto pelo PSF: no primeiro, incrementando sintaxes e vocabulário (BARBOSA, 2004); no segundo, aprimorando vozes sintéticas e reconhecimento (DAMASCENO, 2005; HUGO, 1995; TEVAH, 2007, YNOGUTI, 1999).

Com essas três áreas colaborando mutuamente - *síntese de voz e reconhecimento e controle por voz* - não fica difícil imaginar uma máquina recepcionista - em forma humanoide ou não - respondendo a perguntas básicas dos visitantes com o registro linguístico e vocabulário típicos da região onde está ocorrendo um congresso. Ainda que pareça pura ficção, é importante lembrar que as empresas de telefonia há muito utilizam reconhecedores e sintetizadores de voz dotados de um banco de dados que permite ao usuário fazer suas escolhas a partir das perguntas feitas pelo atendimento eletrônico. A diferença é que eles sempre soam aproximadamente paulistas, ou cariocas, ou gaúchos, ou baianos, dependendo da procedência do usuário que ouve. Isso ocorre porque, muitas vezes, as vozes utilizadas são peças de áudio gravadas em estúdio contendo palavras, expressões ou frases inteiras e os locutores contratados para tal função podem pertencer à mesma região de atuação da empresa em questão.

Trata-se da URA comentada na seção 3.1.4 que, normalmente, equipa dispositivos de telefonia como pequenos sistemas de PABX³⁵ ou outras centrais eletrônicas maiores e têm a função de identificar palavras-chave na fala do usuário para que este possa navegar pelos itens do menu disponíveis por determinada empresa de telefonia ou telecomunicações. Basicamente operam com um sintetizador de voz e um reconhecedor de voz atrelados a um banco de dados específico. Se o usuário diz, por exemplo, uma palavra que não está no banco de dados como ‘galinha’, ou um palavrão qualquer, justamente porque não acionam nenhum item ou subitem do menu, automaticamente o sistema ordena que o computador faça novamente uma pergunta ao usuário na tentativa de continuar o processo. Muitos usuários furiosos e outros simplesmente curiosos colocam o sistema à prova.

A URA e a IVR são quase a mesma coisa. IVR, do inglês, é sigla de *Interactive Voice Response* e tem a mesma função. Importante lembrar que o recurso URA - ou IVR - não é usado apenas para as telecomunicações, mas também em outras aplicações, entre elas nos automóveis mais modernos para possibilitar o famoso recurso *hands-free*, que permite atender ao telefone ou verificar um correio eletrônico sem tirar as mãos do volante e os olhos do trânsito, apenas falando e ouvindo.

A Linguística entra aqui. Esses tipos de equipamentos e recursos podem e devem ser avaliados em quesitos como funcionalidade, acessibilidade e *ambiência amigável*. Para simplificar, quanto menos robótica e mais próxima do natural parecer a voz sintética do equipamento, mais aceitável será entre os humanos. Para tanto ela deve atender a muitos

³⁵ *Private Automatic Branch Exchange* ou Comutador Automático de Ramais Privados.

parâmetros, entre eles a prosódia e a entonação, que devem preferencialmente atender bem ao gosto do interlocutor: uma pergunta deve soar como pergunta, uma afirmação como afirmação, etc. O fato é que isso não é tão simples assim. A maior parte das vozes sintéticas em Português Brasileiro que pode ser considerada de boa qualidade aproxima-se dos registros linguísticos típicos das cidades de São Paulo ou do Rio de Janeiro, ou seja, são essencialmente paulistanas ou cariocas em termos fonéticos.

Então, nesse quesito, a Geolinguística e o ALiB podem contribuir muito. Uma URA no ramo de telecomunicações que possa casar o registro linguístico da voz sintética com aquele que é típico da região do usuário seria algo fantástico em termos de *ambiência amigável*. Para um cidadão de Toledo, no Paraná, acessar o menu de uma empresa de telecomunicações e poder ser atendido com a voz sintética que corresponde à dele talvez seja vantajoso a ambos: usuário e empresa. Isso pode também se estender aos automóveis, afinal, uma montadora fornece o mesmo modelo a todo o Brasil e um sistema *hands free* adaptado ao registro linguístico de seu proprietário é bastante adequado. Nesse âmbito, o único trabalho da Geolinguística com base de dados suficientemente grande e precisa para dar suporte aos desenvolvedores é o ALiB, tanto no nível fonético quanto no lexical.

Seria muito estranho, no entanto, uma montadora situada no estado da Bahia fornecer um veículo com sistema *hands free* adaptado ao registro local para o Rio Grande do Sul, ou vice e versa: são 3100 quilômetros que pressupõem variação diatópica. Basta tentar imaginar um robô nitidamente ‘gaúcho’ interagindo com um soteropolitano ou, se preferir, um cidadão de Porto Alegre ouvindo uma voz sintética feminina parecida com aquelas de cantoras ou atrizes do Nordeste. Note-se que a questão não é ser inteligível ou não, mas simplesmente atender ao quesito de *usabilidade*.

Evidentemente que essas possibilidades já se encontram em outro âmbito de desenvolvimento, diferente daquele de Weizenbaum. O Eliza não possuía voz sintética e tudo era feito por meio do teclado e monitor, respectivamente, dispositivos de entrada e saída de dados. Mesmo assim foi confundido com um psicanalista. A parte ruim disso tudo talvez seja a evolução das máquinas nessas áreas tornar-se prejudicial a algumas pessoas em proporção muito maiores daquelas que envolveram o Eliza.

William Aaron Woods, em 1972, desenvolveu o Lunar (WOODS, 1972), outro robô de conversação especificamente desenvolvido para a NASA, cuja função era responder perguntas sobre composição de rochas lunares. Note-se que a finalidade é absolutamente diversa daquela do Eliza, mas o funcionamento é similar.

Outras implementações em softwares, no entanto, virão e certamente não contemplarão apenas as três áreas interdisciplinares citadas ao início e ao longo desta seção e, eventualmente, em todo o trabalho. Muitas outras esferas científicas trabalharão em conjunto para desenvolver novas tecnologias que envolvam competências aqui comentadas.

De modo geral, a Linguística, esteve presente em diferentes esferas tecnológicas como a informática, as telecomunicações, a Domótica, a acessibilidade, entre outras. As áreas de PLN, de PSF e da Geolinguística - especificamente - têm bastante a ver uma com a outra. As possibilidades são muitas e não se pode prever com exatidão até que ponto se pode chegar em termos tecnológicos. Resta aos pesquisadores envolvidos continuar a verificar mais especificidades de suas respectivas áreas, mas sem perder de vista a dimensão interdisciplinar, afinal, 'robô que fala' não é apenas coisa de ficção científica e os estudos geolinguísticos não são apenas para aqueles que fazem o curso de Letras.

4 DADOS DO ALIB EM FUNCIONAMENTO: PLN, ROBÔS DE CONVERSACÃO E INTELIGÊNCIA ARTIFICIAL

(...) nenhum computador é rápido o suficiente para superar a explosão combinatória gerada pela maioria dos problemas (RICH, p. 04)

4.1 O ALIB NO *IBM VIA VOICE*: PESQUISA GEOLINGUÍSTICA ASSOCIADA A APLICATIVO DE RECONHECIMENTO DE FALA

Esta seção tem por objetivo avaliar brevemente o funcionamento de um aplicativo³⁶ para reconhecimento de fala considerando-se variantes do Português Brasileiro. Para realizar esse estudo, foi usado o *IBM Via Voice*³⁷ versão 9, para o Português Brasileiro, que é um aplicativo do tipo *speech to text*, ou seja, converte texto falado em texto escrito. As variantes do Português Brasileiro foram extraídas do ALiB e consistem de fragmentos específicos das gravações originais.

A relevância do experimento está em aplicar de modo diferenciado um material coletado por muitos pesquisadores ao longo de vários anos a fim de obter mais informações que possam auxiliar em uma utilização mais precisa da máquina no que diz respeito à fala humana.

4.1.1 O ALiB e Algumas Possibilidades em Âmbito Tecnológico

O ALiB³⁸, conforme explicitado, é uma obra única na história da Geolinguística brasileira, tanto pela abrangência em termos territoriais quanto pela precisão dos dados coletados, que se originam de uma pesquisa de campo na qual foi coletada, por meio de gravadores, a fala de pessoas de diferentes regiões a partir de um questionário padronizado.

³⁶ Pensando em um uso terminológico mais adequado, o termo *aplicativo* é preferível a *software* ou *programa*. *Software* é muito geral e abarca muitas modalidades e categorias. *Programa*, embora usual no cotidiano, pode conotar diretamente as linguagens de programação que são acessíveis e manipuláveis apenas a iniciados no ramo de programação. Por isso, tal como fizemos quando estudamos o *Balabolka* e o *DosVox*, não chamaremos o *IBM Via Voice* - e outros de natureza similar - de *software* ou *programa*, mas sim de *aplicativo*.

³⁷ O *IBM Via Voice* começou a ser desenvolvido pela IBM na década de 1950. Atualmente, os direitos exclusivos de distribuição global para o *ViaVoice* pertencem à empresa Nuance, cujo produto *Dragon Naturally Speaking* também poderá ser estudado em outra ocasião. Essa informação pode ser obtida em IBM (2013) e Nuance (2013).

³⁸ Importante consultar Aguilera (1990) especificamente sobre o Paraná e Projeto Atlas Linguístico do Brasil (2013).

Para a palavra ‘luz’, por exemplo, resposta esperada com a questão número 009 do Questionário Fonético-Fonológico, são sempre necessários dois homens e duas mulheres em cada localidade - exceto nas capitais em que a coleta requer oito pessoas - e os indivíduos do mesmo sexo devem pertencer a duas faixas etárias diferentes. Essas quatro pronúncias podem, se for o caso, ser confrontadas com outras quatro pronúncias coletadas com o mesmo critério em outra cidade que fica a 400 quilômetros de distância (variação diatópica) daquela primeira e assim por diante.

Evidentemente, essa seria apenas uma das múltiplas aplicações com as quais os dados no ALiB podem contribuir. Entre elas, há a dimensão lexical, tendo em vista que nem sempre a mesma ideia/conceito envolvida em um item de questionário dá como resposta a mesma palavra. Para ilustrar, na carta 197 do Atlas Linguístico do Paraná – ALPR II – há oito acepções diferentes para ‘montanha’ (ALTINO, 2012), o que faz cair por terra a superficial concepção do senso comum de que algo tem apenas uma designação. ‘Morro’, ‘serra’, ‘lomba’, ‘cerro’, ‘monte’, ‘pico’, ‘barranco’ e ‘perau’ são itens lexicais usuais que recobrem a mesma ideia/conceito e estão distribuídos no Estado nas dimensões diatópica, diafásica, diassexual e diageracional.

Nesta seção, no entanto, a variação lexical não está em discussão, mas as diferentes realizações fonéticas da mesma palavra. ‘Luz’, citada anteriormente, pode ser pronunciada de maneiras diferentes, tanto fonética quanto prosodicamente.

Quando há menção sobre aplicativos de reconhecimento de fala, sejam eles resultados dos avanços tecnológicos ou da necessidade e desejo de tentar materializar a imaginação humana, é exatamente essa parte que nos interessa. A máquina que pode ouvir, entender e realizar a tarefa requerida pelo homem apenas por meio de um comando por voz depende desse fenômeno físico das ondas que se propagam pelo ar - o som - e o ALiB nos proporciona esse material.

Surge então um questionamento bastante comum em relação ao que foi afirmado no parágrafo anterior que é relativo à não necessidade de levar em conta o modo como a mesma palavra é pronunciada nas diferentes regiões do país, uma vez que após um período de amostragem e treinamento³⁹, determinado aplicativo pode armazenar diferentes prosódias e timbres de voz em seu banco de dados e isso pode ser suficiente para fazê-lo funcionar.

³⁹ O treinamento diz respeito, muito basicamente, ao procedimento de dispor ao computador uma quantidade mínima de amostras que permita o processamento e um resultado alinhado à finalidade para a qual o aplicativo, software ou equipamento foi desenvolvido (YNOGUTI, 1999; HUGO 1995; TEVAH, 2006).

Considere-se, então, a seguinte situação hipotética: um carioca – nativo da capital e que viaja apenas eventualmente⁴⁰, o que é muito comum - desenvolve um reconhecedor de voz para uma plataforma específica e ele mesmo carrega o banco de dados com a palavra ‘luz’ utilizando sua própria fala e a de outros nove colegas de seu departamento totalizando cem amostras fonéticas previamente gravadas - dez de cada. Tal amostragem pode ser considerada significativa para alguns aplicativos, considerando-se um item lexical apenas. O aplicativo, no entanto, será embarcado em um equipamento que deverá funcionar a partir de comandos por voz em todo o território nacional, ou seja, para quaisquer pessoas independentemente de idade, sexo ou escolaridade. Como esse equipamento - elevador, cadeira de rodas, automóvel - é dotado de um recurso para acionamento de luzes ele requer, pela lógica, um comando vocal, como a palavra ‘luz’.

Estatisticamente falando, o banco de dados tem cem amostras e isso é significativo (YNOGUTI, 1999; DAMASCENO, 2005; HUGO 1995; TEVAH, 2006). Linguisticamente falando, tal banco de dados, no entanto, dispõe apenas da pronúncia do programador carioca e seus nove colegas: todos homens, da mesma faixa etária, similar grau de escolaridade - desenvolvedores têm, em sua maioria, nível superior - e da mesma região. Isso pode ser ruim.

Note-se que, de saída, o critério para o teste do equipamento quase que certamente não está seguindo um parâmetro minimamente científico no que diz respeito à aplicação da amostragem em termos geolinguísticos. O método de coleta de dados adotado pelo ALiB notoriamente tem muito a contribuir nesse quesito. Supondo que o desenvolvedor carioca conhecesse tal método e tivesse à sua disposição uma equipe e patrocínio para coletar seus próprios dados, ainda assim levaria bastante tempo para treiná-la e ainda muito mais para desenvolver um questionário significativamente eficiente a ponto de proporcionar como resultados as palavras de que precisasse.

Esses são apenas alguns dos problemas em termos aplicativos pelos quais alguns desenvolvedores têm passado quando trabalham com máquinas comandadas pela voz humana.

Retorne-se à palavra ‘luz’. [ˈlus], [ˈluis], [ˈluʃ] e [ˈluiʃ] são algumas das mais prováveis pronúncias encontráveis em território nacional. A considerar novamente de modo hipotético que há apenas essas quatro em todo o Brasil e que são necessárias todas elas no banco de dados do desenvolvedor citado há pouco, conclui-se que seja preciso substituir as cem amostras coletadas em seu departamento por um número maior. Caso o suposto

⁴⁰ Ser ‘pouco viajado’ é um dos critérios extralinguísticos necessários a um informante no ALiB uma vez que, por possuir pouquíssimas influências linguísticas externas - às quais estão expostas os viajantes, por exemplo - apresenta sempre uma consistência confiável em termos de registro linguístico.

desenvolvedor carioca adote o método utilizado pelo ALiB para o qual há necessidade de dois homens e duas mulheres, provavelmente terá de conseguir bem mais que apenas cem amostras - considerando-se dez de cada indivíduo.

Há, no entanto, para essa situação hipotética, grande possibilidade de serem encontradas outras variantes, o que começa a inviabilizar o funcionamento em geral do equipamento tendo em vista que, quanto maior o banco de dados, maior deve ser a capacidade de processamento do computador envolvido no processo. Não se deve esquecer também de que ainda está sendo considerada apenas uma palavra - 'luz' - e o equipamento hipotético não tem somente esse recurso. Note-se, portanto, que o processo de desenvolvimento não é simples.

4.1.2 O *IBM Via Voice* e Outras Possibilidades

Basicamente, para que o *IBM Via Voice*⁴¹ funcione de modo minimamente adequado, basta instalá-lo no computador, configurar o microfone e fazer alguns ditados. Na prática, porém, há alguns obstáculos a serem vencidos. Primeiramente, o microfone que necessita ter ao menos qualidade mediana, deve estar próximo da boca em uma distância bastante regular, ou seja, segurar um microfone com a mão não é a melhor opção: um pedestal ou suporte auricular ajudam bastante; o ruído sonoro de ambiência deve ser o menor possível: pessoas conversando ou ruído de máquinas tornam-no pouco funcional; os ditados são longos e mesmo assim são necessárias constantes correções no banco de dados por conta do usuário: o próprio suporte ao usuário nos avisa o tempo todo.

Após isso tudo, é possível experimentá-lo com olhos e ouvidos mais habituados às variações geolinguísticas, dialetais, fonéticas ou fonológicas. Daí o ALiB no *IBM Via Voice* surgir como uma proposta bastante inusitada, interessante e produtiva, ainda que bastante incomum.

Em outras palavras, esse recurso tem pleno potencial de tornar-se uma maneira eficiente de processar e digitalizar as respostas dos informantes de modo diverso do tradicional, em que um linguista precisar ouvir e transcrever cada um dos fonemas em observação.

⁴¹ O *IBM Via Voice* começou a ser desenvolvido pela IBM na década de 1950. Atualmente, os direitos exclusivos de distribuição global para o ViaVoice pertencem à empresa Nuance, cujo produto *Dragon Naturally Speaking* também poderá ser estudado em outra ocasião. Essa informação pode ser obtida em IBM (2013) e Nuance (2013).

A propósito, qualquer pesquisa geolinguística - ou sociolinguística, ou dialetológica – que se sirva de um aplicativo de reconhecimento de fala pode, de fato, parecer pouco comum a muitos pesquisadores da área, embora necessária, como também será visto adiante.

Retornando à palavra ‘luz’. No *IBM Via Voice* ela está lá, no banco de dados e no ditado de treinamento. Ao pronunciá-la - evidentemente com o registro típico do pesquisador em questão - ela é prontamente escrita na folha do editor de texto eletrônico do aplicativo e é o que ele tem mesmo de fazer, afinal, foi para tal fim que o desenvolveram.

Surge então outro questionamento, dessa vez incomum - e formulado deliberadamente durante o desenvolvimento deste estudo - sobre como fazer o *IBM Via Voice* funcionar com a fala dos informantes do ALiB: eles moram em muitos lugares diferentes e distantes do país e encontrar todos eles novamente para testar o aplicativo é improvável. Isso não será feito.

Como já comentado, o ALiB proporciona esse material. Basta operar com ele de modo a buscar o que interessa à aplicação. A relevância do estudo, já apresentada na Introdução, está justamente em aplicar de modo diferenciado esse material riquíssimo coletado por muitos pesquisadores ao longo de vários anos a fim de obter mais informações que possam auxiliar em uma utilização mais precisa da máquina no que diz respeito à fala humana.

Então a palavra ‘luz’ foi extraída das gravações do ALiB e o *IBM Via Voice* foi preparado para ‘ouvi-la’. Quando ele ‘entende’, há como resultado a palavra ‘luz’ escrita em seu editor. Quando ele ‘não entende’, há como resultado ‘um absurdo qualquer’ ou outra palavra diferente de ‘luz’ escrita por ele - esta, porém, com proximidade fônica significativa.

Para realizar tal processo, é necessário algum aplicativo editor de som que permita visualizar as ondas e copiar a parte que nos interessa. Pode ser o Audacity, o Sonar, o Sound Forge, entre muitos outros. Aqui foi utilizado o Sound Forge⁴² versão 8 que, embora muito antiga e até obsoleta para muitos profissionais da área de gravação de áudio, sonoplastia e/ou sonorização, serve-nos sobremaneira.

Ao abrir o arquivo de áudio com a entrevista, é possível ir direto à parte onde há a palavra ‘luz’ pronunciada pelo informante, copiá-la e gerar outro arquivo de áudio - bem pequeno em termos de megabites - apenas com ela. O procedimento pode ser repetido quantas vezes for necessário até que o pesquisador julgue suficientes a qualidade e a quantidade de amostras a serem extraídas.

Tendo as amostras editadas já em mãos, resta fazer o *IBM Via Voice* ouvi-las. Para tanto, há algumas maneiras diferentes de fazê-lo.

⁴² Dados sobre a nova versão podem ser encontrados em Sony (2013)

Uma delas, tal como mostra a Figura 05, seria colocar um microfone próximo aos autofalantes do computador que reproduzem os arquivos de áudio ‘luz’. Esse microfone, devidamente conectado a outro computador no qual funciona o *IBM Via Voice*, captaria o som ‘luz’ e o enviaria para o processamento que nos daria um possível resultado: a palavra ‘luz’ no editor do aplicativo.



Figura 05 - Utilização de sistema de áudio e captação por microfone externo

Outra maneira, tal como mostra a Figura 06, seria conectar um computador a outro utilizando um cabo de áudio. Esse cabo ligaria a saída de áudio daquele onde estão os arquivos ‘luz’ à entrada de áudio do outro computador no qual funciona o *IBM Via Voice*. O sistema de áudio do segundo computador enviaria o som ‘luz’ para o processamento e nos daria o mesmo possível resultado: a palavra ‘luz’ no editor do aplicativo.

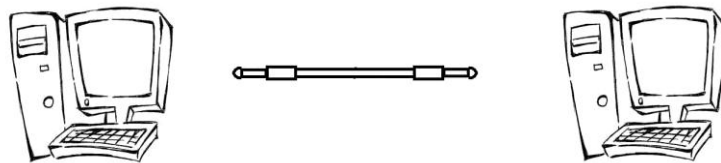


Figura 06 - Utilização de um cabo de áudio

Ainda que experimentais, ambas as maneiras demonstradas parecem funcionar. A segunda, no entanto, representada pela Figura 6 ficaria livre da eventual captação de ruídos de ambiência e, por não utilizar um microfone, requer menos ajustes e mostrou-se mais adequada à aplicação aqui proposta.

Para colocar em prática, foram editados 16 arquivos de áudio do ALiB contendo a palavra ‘luz’. Essas cópias foram feitas com a utilização do aplicativo para edição de áudio *Sound Forge* versão 8. Dele, foram utilizados apenas os processos de cópia e colagem do fragmento de áudio requisitado, ajuste de volume em decibéis e inserção de início e fim - *fade in* e *fade out* - para cada uma das pronúncias. A Figura 07 ilustra a cópia da palavra ‘luz’ sendo feita a partir do registro 213-1, da cidade de Cândido de Abreu.

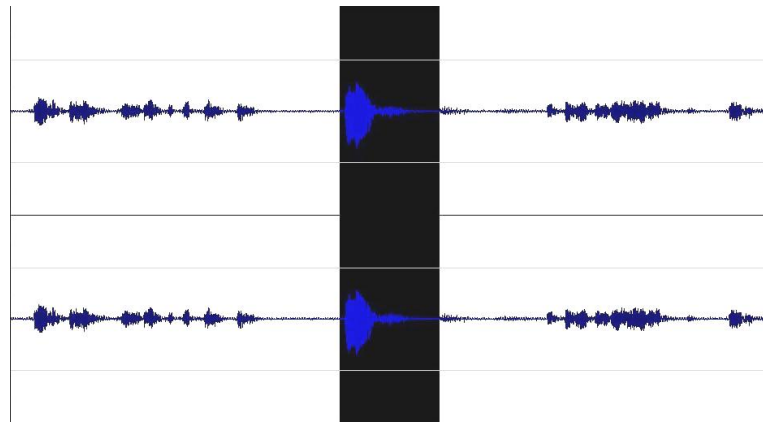


Figura 07 - Fragmento de áudio com a palavra luz sendo copiado

Importante salientar que um dos critérios do ALiB para coleta de dados estabelece que sempre haja quatro informantes de cada cidade do interior. Como o critério para este experimento, especificamente, estabeleceu um número de quatro cidades do interior, há à disposição 16 pronúncias. A escolha por quatro cidades equidistantes do interior do Estado do Paraná deve-se à disponibilidade de um instrumental bastante satisfatório; à proximidade geográfica da Universidade Estadual de Londrina; às características das variáveis /R/ e /S/.

Feitas as edições das 16 pronúncias, os computadores foram conectados tal como proposto na Figura 6 e, assim, pôde-se colocar em prática a proposta inicial deste estudo: o *IBM Via Voice* ouviu o ALiB.

Seria uma hipocrisia por parte do pesquisador dizer que qualquer resultado serviria, independentemente do que fosse, afinal, é pesquisa. O luxo - *ecce veritas* - seria o *IBM Via Voice* reconhecer a maior parte das pronúncias do ALiB, mas isso não aconteceu.

Foram necessários vários penosos, demorados e enfadonhos ajustes: o de **vocabulário**, o de **pronúncia** e o de **nível de sinal**.

O de **vocabulário** e o de **pronúncia** são feitos diretamente junto ao assistente do *IBM Via Voice*. Ele próprio coloca o usuário para ler um bom trecho do *Dom Casmurro*⁴³. Isso permite ao aplicativo processar o nível de sinal do microfone, a qualidade de áudio, o timbre, a prosódia e a potência da voz do usuário - entre muitos outros - além do ruído de ambiência provenientes de automóveis, máquinas, murmúrios etc. Às vezes, são necessárias várias horas e não é possível fazer isso em um único dia.

⁴³ A escolha por fragmentos de textos da Literatura Brasileira por parte dos desenvolvedores é interessante uma vez que há menor recorrência de termos estranhos ao português do Brasil. Por outro lado, como a obra já é antiga, muitos dos novos itens lexicais recorrentes a partir da segunda metade do século XX não ocorrem.

O de **nível de sinal** requer um cuidado técnico específico. As entradas de microfone ou linha - *line in* - têm sensibilidades e impedâncias diferentes e os diferentes computadores possuem adaptadores de áudio de fabricantes diversos. Feitos os ajustes entre as tecnologias coexistentes, tudo começa a funcionar.

Algumas vezes ele decodifica e escreve exatamente a palavra. Outras não. Quando não escreve exatamente a palavra ‘luz’, grafa algo com proximidade fônica significativa que, a julgar pela maior precisão do ouvido do pesquisador que é falante nativo e pela incapacidade de a máquina avaliar fatores linguísticos e mesmo extralinguísticos, é satisfatório.

Importante ressaltar que o conceito de ‘satisfatório’ aqui diz respeito à proximidade fônica de um ou mais fonemas na mesma palavra. Considerando que a palavra ‘luz’ pode ter de três a quatro fones⁴⁴ para a maioria dos falantes brasileiros, se o aplicativo conseguir identificar um de três, já se tem um terço. Pode parecer pouco para um ouvinte nativo - e, no trato cotidiano, é - mas para uma máquina nessas condições é um bom começo. As realizações [ˈluis] de ‘luz’ e [luˈiz] de ‘Louise’ são muito próximas em algumas das 16 amostras utilizadas no experimento e, embora não sejam idênticas, apresentam ambas basicamente o mesmo número de fones: quatro.

A Tabela 06, a propósito, foi elaborada especificamente para trazer mais informações que colaboram com essa colocação. Nela é prático apreciar todos os dados do experimento tabulados do modo mais visual possível.

As colunas contendo o número do registro em áudio, a cidade de origem desse registro, o Estado correspondente e o sexo do informante são os dados de referência do ALiB. As colunas contendo a palavra utilizada, a proximidade fônica e o índice básico de acerto - indicadores *sim*, *não* ou *eventualmente* - são as informações e resultados específicos do experimento.

⁴⁴ O *fone* é uma unidade sonora verificável na produção da fala, um segmento vocálico ou consonantal passível de transcrição fonética e difere-se do *fonema*, que é uma unidade sonora que se distingue **funcionalmente** de outras unidades da língua (SILVA, 1999) – grifo nosso.

Tabela 06: Resultado do experimento

Registro	Cidade	Estado	Sexo	Palavra	Proximidade Fônica	Acerto
210-1	Umuarama	Paraná	homem	luz	louise	event.
210-2	Umuarama	Paraná	mulher	luz	x	não
210-3	Umuarama	Paraná	homem	luz	duas, nos	event.
210-4	Umuarama	Paraná	mulher	luz	x	não
212-1	Campo Mourão	Paraná	homem	luz	x	não
212-2	Campo Mourão	Paraná	mulher	luz	onze	event.
212-3	Campo Mourão	Paraná	homem	luz	os	event.
212-4	Campo Mourão	Paraná	mulher	luz	onze	event.
213-1	Cândido de Abreu	Paraná	homem	luz	luz	sim
213-2	Cândido de Abreu	Paraná	mulher	luz	x	não
213-3	Cândido de Abreu	Paraná	homem	luz	doze	event.
213-4	Cândido de Abreu	Paraná	mulher	luz	x	não
215-1	Toledo	Paraná	homem	luz	x	não
215-2	Toledo	Paraná	mulher	luz	doze, onze	event.
215-3	Toledo	Paraná	homem	luz	x	não
215-4	Toledo	Paraná	mulher	luz	ruas, onze	event.

Fonte: Texto organizado pelo autor

O *IBM Via Voice* ouviu 30 vezes cada um dos 16 registros de áudio ‘luz’ elencados na Tabela 06, sendo 10 vezes em volume baixo, 10 vezes em volume médio e 10 vezes em volume alto⁴⁵. Os resultados dessas três categorias foram somados e apenas a média foi considerada porque mostraram resultados similares. Em termos de volume, apenas a quase ausência de sinal - muitíssimo baixo - e o excesso - saturação de áudio - trouxeram resultados muito discrepantes e essas grandezas não foram utilizadas no experimento. O critério para os indicadores *sim*, *não* e *eventualmente* foi estabelecido tal como indicado na Tabela 07:

Tabela 07: Critério para os indicadores

	Mais de quinze vezes	Menos que duas vezes	Entre duas e dez vezes
SIM	X		
NÃO		X	
EVENTUALMENTE			X

Fonte: Texto organizado pelo autor

A nona linha da Tabela 06 está grifada em negrito porque ela é o ponto de partida das colocações a seguir. Pode-se notar que o registro ‘luz’ obtido junto ao informante identificado como 1 (homem, Faixa etária I, do nível fundamental) no ponto 213-1 de Cândido de Abreu foi o que conseguiu melhor resultado, ou seja, maior quantidade do indicador *sim*. Isso ocorreu não em função do sexo, da idade ou do grau de escolaridade, mas da qualidade de

⁴⁵ As grandezas baixo, médio e alto foram baseadas no indicador de sinal de áudio do próprio *IBM Via Voice*, em que verde escuro é baixo, verde claro é médio - ou ideal - e vermelho é alto. Para o conceito de *volume* aqui abordado, vale lembrar que está vinculado diretamente ao nível de sinal mensurável em decibéis e/ou milivolts, muito comum em engenharia de áudio e engenharia elétrica.

áudio envolvida na gravação original. Esse registro estava praticamente isento de estalos, estática, murmúrios ou ruídos de máquinas ou automóveis. Outro detalhe importante é que a palavra ‘luz’ foi pronunciada sem quaisquer vizinhanças fônicas: após a pergunta, o informante respondeu apenas ‘luz’, e não ‘é a luz!’ ou ‘ah, luz, né?’.

Especificamente sobre vizinhança fônica, relevante lembrar que ela ocorre sempre levando-se em consideração a cadeia da fala, momento em que alguns fonemas se alteram, são suprimidos ou substituídos por outros, dependendo do ambiente em que se encontram (SILVA 1999, LOPES, 1999, BORBA, 1991, MUSSALIM & BENTES, 2001). Na seção 3.1, quando o aplicativo *Balabolka* tipo texto-fala foi analisado, foram utilizados, entre outros exemplos, as frases ‘Ele não comia **desde** ontem’ e ‘Ele não comia **deste** ontem’. Lá, as palavras *desde* e *deste* foram realizadas de modos diferentes e muito próximas das realizações de falantes nativos do Português Brasileiro. Em outras palavras, o *Balabolka*, para esse caso de vizinhança fônica estava muito próximo da performance humana.

Para todos os resultados com o indicador *eventual*, pode-se verificar que a maior dificuldade do aplicativo foi identificar e processar o fone /λ/ das gravações, tendo em vista que em todos há uma fricativa: *Louise*, *duas*, *nos*, *onze*, *os*, *doze*. Além disso, com exceção de *onze* e *doze*, as amostras exibem ou o fone /u/ presente em ‘luz’, seja naturalmente como em *Louise* [lu'is], *duas* [duəs] e *ruas* [xuəs], seja pelo alçamento⁴⁶ como em *nos* e *os*, [nus] e [us] respectivamente. Os resultados *onze* e *doze*, ainda que pareçam muito distantes foneticamente de ‘luz’, têm as fricativas, embora vozeadas. Possuem também o alçamento do [e] para [i], fone também presente na realização de ‘luz’ para alguns casos no Brasil e em algumas das 16 amostras: [lujs] e/ou [luis]. Importantíssimo relembrar ainda para esse indicador - *eventual* - que, embora a Tabela 06 exiba apenas as palavras com proximidade fônica justamente porque a função dela é apresentar os resultados mais numerosos tal como proposto no critério da Tabela 07, não quer dizer que o *IBM Via Voice* não tenha escrito uma ou outra vez exatamente ‘luz’.

Finalmente, para os resultados com o indicador *não* foi decidido que seria irrelevante para esta ocasião discuti-los todos. Primeiro porque não possuíam proximidade fônica significativa uma vez que, em muitos casos, era apenas a fricativa, em outros, apenas uma

⁴⁶ Mesmo que aqui não se trate de um alçamento da vogal [o] em posição pretônica, é importante frisar que as gravações desse experimento são dados coletados no Estado do Paraná tão somente e, portanto, é interessante contar com a contribuição de Kailer (2012).

vogal. Em segundo, porque foram em grande quantidade e aleatórios, provavelmente em função da qualidade da gravação.

Considerando o exposto, embora bastante incomum, como dito anteriormente, a proposta de fazer um aplicativo tipo fala-texto (*speech-to-text*) ouvir o ALiB é necessário. Outros experimentos serão feitos nesse sentido e com aplicativos ou softwares também diferentes. Outras máquinas serão empregadas e novos profissionais estarão envolvidos. Um Atlas Linguístico, seja ele qual for, mas principalmente o ALiB, com toda a sua riqueza, não pode ficar restrito aos compêndios de papel tampouco apenas à apreciação dos profissionais da área de humanas. Eles precisam ser estudados, processados e utilizados em mais e mais experimentos.

Duas observações importantes que já podem ser feitas a partir desse experimento é que (i) o procedimento aqui proposto é de fato passível de ser feito e que (ii) a qualidade de áudio interfere em muito no reconhecimento automático.

Mesmo considerando que todas as gravações do ALiB são suficientes para a análise quando se pode contar com o ouvido apuradíssimo do linguista, a variação da qualidade de áudio captada ainda é insuportável para a máquina. Enquanto não há aplicativos ou softwares em dispositivos que ouvem exatamente como humanos - o que pode ser impossível - outras medidas podem ser tomadas.

Estalos, estática, automóveis, máquinas, murmúrios, excesso de reverberação, saturação, entre muitos outros, podem, como foi dito, não ser obstáculo ao linguista, mas são elementos que geram sinais extras que são gravados na íntegra quando da entrevista. Transformados em sinais elétricos por qualquer microfone de qualquer sistema, acabam por compor um só corpo de áudio com muitas faixas de frequência audíveis.

O questionário do ALiB já é bem adaptado à realidade nacional - entre outras de suas características, as palavras envolvidas nas entrevistas pertencem ao léxico do vernáculo. Os critérios de entrevista também atingiram grau significativo de precisão pois todos os inquiridores são minimamente treinados e seguem os critérios básicos para a obtenção de êxito na coleta de dados. Para coletas futuras, vale a pena desde já arriscar uma previsão: atingir também um merecido grau de excelência na captação de áudio pode significar um salto gigantesco para o ALiB, no Brasil do século XXI: recepcionistas automáticos, reconhecedores de voz, analisadores de fones e até mesmo jogos digitais que operem com variação linguística são apenas alguns dos exemplos que poderiam se servir dos estudos geolinguísticos em termos de utilização de nossa língua.

A palavra ‘luz’ - em Português Brasileiro - gravada em um estúdio com os mais apurados critérios técnicos pode ser identificada por pessoas de diversas nacionalidades e, provavelmente, também pelo *IBM Via Voice*. A mesma palavra, gravada juntamente com os sons de um ônibus e uma motocicleta acelerando ao mesmo tempo, na maior parte das vezes, só pode ser interpretada por um falante nativo - seria uma tarefa muito penosa a um francês ou a um norte americano, por exemplo. Nesse contexto, a máquina ainda deixa a desejar: o *IBM Via Voice* ‘ouve’ o ALiB, mas com restrições.

4.2 TECNOLOGIAS ON-LINE COM COMANDOS POR VOZ

As tecnologias on-line são basicamente aquelas que dependem de uma conexão com a internet para operar. Muitos jogos, sistemas de busca e de consulta/cadastro funcionam assim e basta que algum problema prejudique ou interrompa a conexão para que se tornem inoperantes. Em se tratando de tecnologias on-line com comandos por voz, as condições não são diferentes. Como em outros sistemas de mesma natureza, o comando por voz – sinal de áudio – precisa ser enviado para processamento a um sistema que se encontra implementado em um servidor – ou vários - a partir do qual opera o aplicativo responsável por analisar a voz. Esse servidor pode estar a alguns quarteirões ou a milhares de quilômetros. Próximo ou distante, quando há falha na transmissão, o conjunto deixa de funcionar e não há resposta.

Excetuando-se essa grande desvantagem, e considerando-se apenas as conexões estáveis e suficientemente velozes para comportar o aplicativo em uso, essas tecnologias costumam operar relativamente bem tendo em vista que, cada vez mais, aumentam as instituições, empresas e usuários que as utilizam. Mais que isso, normalmente dispõem de sistemas bastante robustos com muitos computadores, cada um dos quais com diversos núcleos de processamento, quantidades gigantescas de memória e banco de dados extremamente sofisticados. Nesse sentido, esta seção discute sobre o funcionamento de uma dessas tecnologias on-line e que opera a partir de computadores domésticos ou dispositivos móveis como smartphones e similares.

4.2.1 As Muitas Tecnologias Disponíveis

Seria muita pretensão tentar falar sobre ou descrever todas as tecnologias que estão disponíveis em termos de comandos por voz da atualidade. Ao restringir essa tarefa apenas a smartphones, ainda assim seria algo quase impossível de ser feito por dois motivos: (i) a

quantidade de marcas e modelos é imensa, (ii) a proliferação de novas marcas e novos modelos, além daqueles já consagrados, é quase exponencial. A Tabela 08 apresenta apenas seis entre os mais conhecidos smartphones que operam com essas funcionalidades – provavelmente, até a publicação deste texto, alguns já estarão fora de linha:

Tabela 08 – Smartphones

aparelho	marca	SO	aplicativo
iPhone	Aple	iOS	Siri
Galaxi	Samsung	Android	Google Now
Nexus	LG	Android	Google Now
MotoG	Motorola	Android	Google Now
Lumia	Nokia	Windows Phone	Cortana
Xperia	Sony	Android	Google Now

Fonte: Starmundo (*online*, 2015)

Esta Tabela, meramente ilustrativa e muito pouco eficaz quanto à capacidade de mostrar um panorama significativo do que estava sendo utilizado quando esse texto estava em construção, já apresenta um universo difícil de ser explorado em termos de avaliação linguística. Em outras palavras, avaliar a eficiência de sistemas que operam com comandos por voz a partir desses seis aparelhos seria ainda uma obra extensa e pesada. Estão fora da Tabela 08 outros Sistemas Operacionais como Firefox OS (LG), Tizen (Samsung) e o Ubuntu toch (vários aparelhos).

Considerando que três deles – Galaxi, Nexus e MotoG – operam com o mesmo sistema operacional e, assim sendo, podem servir-se de um aplicativo em comum - Google Now - por ser o mais procurado num universo de dezenas, a tarefa ficaria um pouco mais fácil, mas ainda complexa. É necessário também levar em conta que todas as empresas desenvolvedoras de Sistemas Operacionais disponibilizam seus produtos não apenas para smartphones, mas a todos os aparelhos da categoria móvel - *mobile* - como notebooks, ultrabooks, netbook, laptop, tablets, PDAs, console portátil, Ultra Mobile PC, coletores de dados entre outros.

O universo de tecnologias, marcas, modelos e sistemas é vasto e mutante e, por esses motivos, não há como acompanhar todas as mudanças e atualizações de modo eficiente sem que fique nada pra trás. Qualquer escolha do objeto de estudo nesse campo não deve prescindir do fato de que há um terreno movediço a ser transposto. Por isso, uma das saídas talvez seja avaliar aquilo que mais está sendo utilizado: encontra-se em maior evidência, é seguido pelos concorrentes e abrange a maior quantidade e variedade de equipamentos disponíveis na ocasião da pesquisa.

4.2.2 O Buscador por Voz da Google

Se o critério para escolha de uma tecnologia on-line com comandos por voz aponta para os quesitos: *mais utilizado, em maior evidência, seguido pelos concorrentes e abrangente em termos de quantidade e variedade de equipamentos disponíveis quando da pesquisa*, então talvez a melhor opção seja o buscador por voz da Google. De acordo com a Serasa Experian, uma das maiores agências de inteligência em marketing digital da atualidade, o Google é responsável por mais de 90% de participação na internet no ano de 2014 entre os buscadores (SERASA Experian, *online*, 2014). Embora não esteja sozinho, muitos usuários sequer sabem da existência dos outros⁴⁷ que, juntos, correspondem a menos de 10% das preferências (EXAME, *online*, 2014). O motivo da escolha, portanto, é relevante e estatisticamente justificável.

Mesmo considerando que o tradicional buscador Google por escrito e o moderno buscador Google por voz sejam aplicativos diferentes, ambos operam em conjunto e basta abrir o sistema de busca em qualquer navegador para que todos os recursos estejam prontos para uso: o conhecido campo para digitar e o ícone do ‘microfone’ para acionar o processador de voz, indicado pela seta na Figura 08.



Figura 08 - O ícone do ‘microfone’ em destaque

Ao clicar no ícone do buscador por voz - ‘microfone’ - imediatamente o sistema anuncia por escrito a frase ‘fale agora’, o ícone do microfone começa a pulsar com um halo vermelho. Então, o usuário pronuncia o que quer que o buscador encontre. Nomes conhecidos como Youtube, Facebook, Tweeter, entre outros, não geram problemas e ele identifica de imediato. Outros, como celular, smartphone e tablet, também são prontamente identificados pelo processador e devolvem uma resposta. Dessa forma, a opção foi fazer testes com itens

⁴⁷ Bing Brasil, Ask Brasil, Google.com, Yahoo! Brasil, respectivamente 2º, 3º, 4º e 5º lugares na referida pesquisa da Revista Exame.

menos comuns, mas de categoria similar tal como sites de popularização de dados e/ou de relacionamento (Youtube, Facebook, Tweeter) ou computadores portáteis (celular, smartphone e tablet).

Importante salientar que o critério mais importante adotado para os testes foi evitar, o quanto possível, frases ou orações inteiras. No máximo, algumas expressões nominais como ‘Nuvem de Oort’ ou ‘caixa de satélites’ foram utilizadas. Isso impede, de certa forma, que o buscador possa se valer de seu banco de dados em linguagem escrita para deduzir uma palavra em determinada sintaxe, recurso típico em qualquer buscador moderno. Outro critério foi apagar todo o histórico do navegador e reiniciá-lo a cada procedimento periodicamente – ao menos a cada bloco de palavras – para que páginas encontradas não pudessem ser resgatadas. A voz e seu grau de monitoramento, a sala, o computador, o microfone e respectivos ajustes de ganho e volume utilizados na captação dos comandos foram sempre os mesmos. Tomados esses cuidados mínimos, os testes iniciaram-se.

4.2.3 Testes e Resultados

A palavra ‘eletrônica’, talvez não tão comum quanto aquelas citadas anteriormente, passou no teste. Algumas das grandezas elétricas a ela relacionadas bem como os nomes de alguns componentes eletrônicos listados na Tabela 09 também foram testados. Os componentes geraram 100% de acerto. As grandezas elétricas, no entanto, não. ‘Ohm’, por sua vez, resultou várias vezes na expressão “Não entendi. Tente novamente”⁴⁸ ou na palavra ‘homem’, impressos na tela do aplicativo. ‘Ampère’ resultou em ‘Ambev’ - conhecidíssimo fabricante de bebidas - várias vezes, ‘watt’ em ‘Avast’ - popular anti-vírus - em repetidas tentativas, ‘coulomb’ e ‘farad’ em uma série de disparates de toda ordem - foneticamente similares, aproximados ou diversos. Os itens marcados com asterisco são aqueles que não resultaram em sucesso de busca e esse critério serve para as tabelas seguintes.

⁴⁸ Aviso por escrito do aplicativo quando não ocorre identificação alguma

Tabela 09 – grandezas elétricas e componentes eletrônicos

grandezas	componentes
Volt	resistor
Hertz	transistor
Ampére*	capacitor
Watt*	varistor
Ohm*	disjuntor
Farad*	capacitor variável
Coulomb*	

Fonte: Texto organizado pelo autor

A escolha pelo assunto ‘Eletrônica’ - assim como ‘Linguística’ e outros bastante diferentes entre si - nesta seção, não é aleatória. A tendência dos buscadores mais modernos é sempre ter em sua implementação o conceito de ‘perfil de usuário’ levado muito a sério. Assim, preferências, crenças e experiências da pessoa que usa o buscador podem ser registradas a cada acesso, ou seja, o fato de alguém apreciar frequentemente ‘Eletrônica’ ou ‘Antigomobilismo’, permite ao buscador associar tal assunto ao perfil que gradativamente vai sendo incrementado.

O uso de ontologias também tem sido cada vez mais frequente nesses casos, procedimento que não só permite maior especificidade ao perfil individual, como também colabora para a criação de perfis de grupos (ALMEIDA; BAX, 2003), bastante usual em algumas redes sociais ou aplicativos específicos para aparelhos móveis – *mobile*. Note-se que, embora o objetivo nesta seção não seja versar sobre o conceito de ‘perfis de usuários’, ‘ontologias’ ou ‘aprendizado de máquina’ – *machine learnig*, alguns dos resultados dos testes aqui realizados podem ser explicados a partir desses conceitos.

Partindo para outro ramo de especialidade, a palavra ‘antigomobilismo’, presumidamente menos comum que ‘eletrônica’, também passou no teste. Algumas subcategorias como os nomes de alguns automóveis, como os listados na Tabela 10 também. Carros muito conhecidos no Brasil, principalmente por aficionados, não foram listados por serem muito óbvios em um campo lexical como esse. Fusca, Kombi, Opala, Maverick, Landau, Galaxie, Dodge Dart foram todos identificados pelo sistema, mas ficaram fora da listagem. ‘Ponta de eixo’, salvo sob uma condição de monitoramento absurdo em que a pronúncia torna-se absolutamente teatral ou forçadamente didática (CAMARA JR, 2006; 2004) como [‘põtə’dʒejʃu], resultava sempre em ‘ponto de exu’, provavelmente mais buscado que seu ‘parceiro’ fônico. ‘Pistão’, também altamente dependente da pronúncia teatral como [pis’tãw] em vez de [ps’tãw], resultou em “Não entendi. Tente novamente”.

Tabela 10 – automóveis e peças de reposição

automóveis	peças de reposição
Vemaghet	graxeta
Aero Willys	trizeta
Gordini	homocinética
Romi-Isetta	caixa de satélite
Simca Chambord	virabrequim
DKV	torquímetro
	almotolia
	ponta de eixo*
	pistão*

Fonte: Texto organizado pelo autor

Saindo da especialidade automobilística e migrando para a Linguística, termos técnicos da área foram adicionados aos testes. A Tabela 11 apresenta todos os que foram experimentados com o sistema de processamento da Google. ‘Geossociolinguística’ foi processado como ‘gil sociolinguística’, ‘g1 sociolinguística’ e, interessante, ‘gel sociolinguística’, relacionando o domínio Sociolinguística ao conhecido evento de Linguística no Brasil, o GEL - Grupo de Estudos Linguísticos do Estado de São Paulo. De qualquer forma, não gerou bons resultados. Quanto às denominações das variáveis extralinguísticas, salvo para o trio diastrática/diatópica/diafásica, respectivamente associadas às variações relativas ao extrato social/localização geográfica/grau de formalidade, muito comentado em sites educacionais e/ou voltados à Linguística, todos os outros termos que se iniciam com o prefixo grego *dia* levaram a resultados inconsistentes quase na mesma proporção que ‘farad’ da Tabela 09.

Tabela 11 – domínios da Linguística e variações

Domínio	Variações
Linguística	Idioleto
Dialetologia	Socioleto
Sociolinguística	Diastrática
Linguística textual	Diatópica
Análise do discurso	Diafásica
Geolinguística*	Diamésica*
Geossociolinguística*	Diagenérica*
	Diageracional*
	Diassexual*

Fonte: Texto organizado pelo autor

Utilizar dados do ALiB (CARDOSO et al., 2014a; 2014b, PROJETO Atlas, 2014; COMITÊ; 2001) também pareceu muito conveniente e produtivo para a análise, por isso algumas das cartas semântico-lexicais presentes no Atlas também foram utilizadas para os

testes. ‘Tangerina’ e todas as suas variantes presentes nas cartas L05a a L05e foram pronunciadas.

Tabela 12 – variações para tangerina - cartas L05a a L05e

variações de tangerina	presentes nas entrevistas
mexerica	vergamota
poncã	(mexerica) do rio
maricote	margote*
laranja cravo	mixirique*
tanja	
carioquinha	
bergamota	
mimosa	

Fonte: ALiB (CARDOSO, 2014b)

Com o intuito de estabelecer mais algumas comparações e sanar mais algumas dúvidas, a mesma lista de palavras testadas no experimento com o processador de áudio V2 foi testada no buscador por voz. As cinco cores, que lá foram confundidas pelo V2 com outras palavras de considerável proximidade fônica, ganharam notável distinção com esse processador on-line. A Tabela 13 ilustra a assertiva:

Tabela 13 – cores e palavras com proximidade fônica

Cores	palavras com proximidade fônica	
vermelho	pentelho	fedelho
verde	lerdis*	ver-te*
azul	anu	tatu
amarelo	chinelo	marmelo
branco	manco	banco

Fonte: Texto organizado pelo autor

A palavra ‘verde’, no entanto, merece uma ressalva. Ela parece ser uma espécie de caixa de Pandora para os processadores de áudio quando se trata de Português Brasileiro. Como comentado na seção 4.3, embora existam pouquíssimas palavras em nosso idioma que mantenham significativa proximidade fônica com ela e disso pudesse resultar pouco interesse em estudar suas particularidades por questões mesmo estatísticas, as razões pelas quais isso ocorre poderiam, por extensão, resolver alguns problemas ainda encontráveis em termos de comandos por voz.

Conforme se apresentou, as tecnologias on-line com comandos por voz proliferam na grande rede e são acessadas a partir de aplicativos cada vez mais numerosos. Para avaliá-los sob o crivo da Linguística, é necessário um elenco de critérios e isso se deve à melhora considerável ocorrida nos algoritmos desde aplicativos como o *IBM Via Voice* até o buscador

por voz da Google avaliado aqui. Nesse novo sistema, a sintaxe auxilia muito na identificação dos verbetes duvidosos ao sistema uma vez que se vale do banco de dados em Linguagem Natural Escrita para preencher as possíveis e eventuais lacunas. Por esse motivo, o principal dos critérios foi o de utilizar apenas palavras isoladas nos testes.

Embora esses algoritmos venham sendo atualizados e incrementados constantemente para atender às novas necessidades de mercado e os sistemas que deles se servem pareçam impecáveis aos olhos de alguns usuários, ainda apresentam falhas básicas muito similares às daquelas de aplicativos mais antigos. A substituição da palavra buscada por outra de maior ou menor proximidade fônica permanece recorrente, embora o número de acertos, como demonstrado nos testes, seja bem menor, tal como é o caso específico ilustrado pela Figura 09.

No entanto, ainda é cedo para afirmar que o sistema reconhece com maior precisão as palavras mais conhecidas. Isso implicaria aceitar que termos vinculados a especialidades como Eletrônica, Mecânica e Linguística fossem do domínio de todos. Mesmo assim, é possível esboçar uma teoria mais aceitável baseada nas observações: ele reconhece com maior precisão as palavras mais buscadas e rejeita, mesmo que foneticamente prováveis, aquelas que estatisticamente figuram menos na fala dos usuários. Entre outros exemplos das tabelas, pouquíssimas pessoas sabem o que é uma ‘caixa de satélites’, no entanto, ele a encontrou. O mesmo aconteceu com ‘albedo’ ou ‘paralaxe’, típicos da área de Astronomia, também encontrados. Daí, no outro extremo, ‘ver-te’ e ‘lerdes’ serem preteridos, mesmo com pronúncias monitoradas, assim como ‘Geolinguística’ ou ‘Geossociolinguística’, não por serem desconhecidos de todos os usuários, mas, ao que tudo indica, por serem pouco procuradas em relação a outras.

4.3 RELAÇÃO ENTRE RECONHECIMENTO E COMPREENSÃO DE VOZ: EXPERIMENTO PARA ANÁLISE LINGUÍSTICA

Pedir a uma máquina que realize tarefas apenas com comando por voz é algo presente em muitas obras de ficção científica. No clássico do cinema *Blade Runner* (BLADE Runner, 1982), por exemplo, há uma cena em que o caçador de replicantes Rick Deckard, interpretado por Harrison Ford, dá ordens orais ao computador para que amplie, centralize e enquadre determinada imagem da qual necessita extrair informações relevantes à sua investigação. O computador atende prontamente a todas as suas solicitações até que Deckard consegue finalmente visualizar o que procurava. Embora isso ainda pareça para muitas pessoas um

recurso inexistente, mesmo em se tratando de século XXI, muito do que foi propagandeado em filmes e livros do gênero há algumas décadas no que diz respeito a comandos por voz⁴⁹ já é realidade.

O comando por voz evoluiu tanto nas últimas duas décadas que não são mais necessários equipamentos caríssimos e softwares dedicados e específicos para que ele seja possível. A exemplo disso, há uma infinidade de eletrônicos que operam com esse recurso no dia a dia como os corriqueiros smartphones, sistemas embarcados em automóveis, dispositivos de Domótica em residências e os buscadores na internet.

No entanto, embora ambos dependam do processamento de áudio, *reconhecimento de voz e compreensão de voz* não são a mesma coisa e, na maior parte das vezes, são confundidos pelos usuários pelo fato de estarem intimamente correlacionados. Nesse âmbito, esta seção apresenta um experimento linguístico com um desses dispositivos procurando demonstrar diferenças existentes entre *reconhecimento de voz e compreensão de voz*.

Na época em que o longa-metragem foi lançado, a máquina fictícia utilizada por Deckard poderia ser classificada por linguistas e desenvolvedores como dotada de *reconhecimento e compreensão de voz* em função da quantidade de comandos que ela reconhecia e as sub-rotinas de cada um deles. Em outras palavras, enquanto ligar e desligar uma lâmpada de determinado cômodo da casa no âmbito da Domótica requer no máximo dois comandos como ‘liga’ e ‘desliga’, *ampliar, centralizar e enquadrar* determinada imagem, como solicitou o personagem no clássico do cinema, implica especificar cada vez mais um comando dado em primeiro plano: primeiro *amplia*, depois, cada qual ao seu tempo, *centraliza e enquadra*. Em outras palavras, os itens lexicais envolvidos multiplicam-se e a complexidade aumenta.

Isso não quer dizer que, fora da ficção científica, sub-rotinas por comandos por voz também não possam ser adaptáveis a uma simples lâmpada do mundo real. Pode-se, por exemplo, operar com nuances como aquilo que se chama popularmente de ‘dimerização’⁵⁰ que consiste em dar intensidades diferentes ao brilho da lâmpada em funcionamento. Então, em vez de apenas o insosso e dicotômico ‘liga-desliga’, há ‘liga-menos-menos-menos-

⁴⁹ Os comandos por voz podem ser atendidos por *reconhecimento de voz* ou por *compreensão de voz*, dependendo da finalidade e/ou robustez dos sistemas empregados. Para o usuário, pouco importa se é um ou outro, bastando apenas que a ordem seja cumprida pela máquina, ou seja, tem que funcionar bem.

⁵⁰ A ‘dimerização’, como é popularmente conhecida no âmbito da eletrotécnica diz respeito ao uso do *dimmer*, dispositivo elétrico-eletrônico capaz de controlar a intensidade de uma lâmpada. O homógrafo ‘dimerização’ - *dimerization* - encontrável nos dicionários diz respeito a um fenômeno químico em que uma reação que forma um dímero (DICIONÁRIO Aurélio, 2014).

desliga’. Esse exemplo pode também se aplicar a uma série de equipamentos elétricos em uma residência. Um chuveiro elétrico pode vir equipado com o inusitado ‘liga-menos-menos-menos-desliga’ que, por uma questão de praticidade, conforto e precisão de comandos, poderia ser programado para ‘liga-morno-quente-maisquente-desliga’ – Tabela 14. Note-se que agora, tanto para a lâmpada quanto para o chuveiro pode-se operar com cinco comandos reconhecíveis e a quantidade de fonemas e respectivas combinações a serem processados multiplica-se:

Tabela 14 – Conjunto de cinco possíveis comandos por voz

lâmpada	chuveiro
liga	liga
menos	morno
menos	quente
menos	mais quente
desliga	desliga

Fonte: Texto organizado pelo autor

É adequado, portanto, explicar o processamento de áudio de um ponto de vista mais voltado à Linguística, uma vez que o objetivo desta seção é apresentar um experimento que procure demonstrar diferenças entre *reconhecimento* e *compreensão de voz* e, não necessariamente, extinguir dúvidas sobre o processamento de áudio, cuja complexidade envolvida é extremamente grande e as áreas de conhecimento que se inter-relacionam multiplicam-se com o passar dos anos.

Por mais esse motivo, é importante explicar que *reconhecimento* e *compreensão de voz* são ações diferentes não apenas em termos linguísticos, mas também no âmbito do desenvolvimento tecnológico. Em algumas áreas da Engenharia, enquanto este último fundamenta-se no entendimento do sentido, do significado da mensagem como um todo (HUGO, 1995), *reconhecimento* está relacionado basicamente a um único fonema ou conjuntos deles - sílabas e palavras. Dos dois, será dada aqui muito mais ênfase ao conceito de *reconhecimento* e o equipamento utilizado para o experimento será o *Voice Recognition Module V2* (doravante apenas V2). O motivo de sua escolha é o fato de ser utilizado amplamente em aplicações básicas de robótica e Domótica, oferecer soluções práticas e baratas para aplicações gerais, não depender de conexões de rede como o buscador da Google e, diferentemente, do *IBM Via Voice*, não necessitar de Sistema operacional, além de oferecer suporte para voz humana em geral – não depende de um único idioma. No entanto, também possui grandes desvantagens em relação aos outros dois como se vê a seguir.

A propósito, uma afirmação contida do manual do V2 sobre o *reconhecimento de voz* poderia fazer parte dos critérios de análise de equipamentos que operam com *comandos por voz*. Lá consta a informação de que *reconhecimento de voz* se dá apenas quando o dispositivo ‘sabe exatamente’ o que o usuário está dizendo - “Voice recognition is something that knows exactly what you were saying” (VOICE Recognition, *online*, 2014). Tal informação, no entanto, estaria muito mais vinculada à *compreensão de voz* uma vez que a máquina precisa ser bem mais robusta em termos de software e hardware para fazer isso.

Para dar uma ideia, seria necessário implementar, além de aplicações sobre Fonética, Fonologia, Morfologia e Sintaxe, também outras sobre Semântica lexical, Semântica Composicional e Pragmática e padrões discursivos (JURAFSKY; MARTIN, 1999). Um dos exemplos mais modernos desse universo seria *Watson* da IBM, que opera a partir de linguagem natural - evidentemente -, aprendizado dinâmico, geração de hipóteses e é considerado pelos próprios desenvolvedores como ‘uma extensão natural daquilo que os humanos podem fazer em seu melhor’ (IBM Watson, *online*, 2015). O Watson talvez seja o mais próximo daquilo que pode ser considerado hoje como *compreensão por voz* – vai além do *reconhecimento de voz*.

No outro extremo da linha evolutiva de equipamentos que reagem aos sons produzidos por humanos, existem aqueles acionados por som de palmas - *clappers*⁵¹ - que, certamente, não foram projetados para operar com *reconhecimento e compreensão de voz* e, do ponto de vista linguístico, interessariam pouco neste momento. Com custo significativamente baixo, é possível dotar persianas com esse recurso: uma palma, abrem; duas palmas, fecham - ou vice-versa. Mesmo esses também dependem minimamente de processamento de áudio, ainda que bastante primitivo e simplificado. O circuito eletrônico, nesse caso, deve distinguir entre os estampidos do choque entre as mãos e outros sons pertencentes a uma faixa de frequência e pressão sonora similares, como eventuais ruídos, sons de alguns animais, gritos de crianças, gargalhadas ou mesmo enfáticas e espontâneas interjeições como ‘ah!’, ‘eh!’, ‘uh!’ entre outras. Porém, esses dispositivos, embora atendam a comandos por sons, são muito primitivos e não podem ser classificados como dotados dos recursos de *reconhecimento e compreensão de voz*⁵².

⁵¹ O *clapper* talvez seja o dispositivo comercial mais conhecido dessa categoria, embora não seja o primeiro da história. Trata-se de um comutador elétrico comandado por som patenteado nos EUA, em 1986, que serve para ligar eletrodomésticos: http://en.wikipedia.org/wiki/The_Clapper

⁵² Importante salientar que, nesse momento, algumas vezes, não há clara distinção entre ‘voz’ e ‘fala’ na literatura científica de um modo geral. Em se tratando de Fonética e Fonologia, esses dois termos manifestam-se cada qual em contexto específico, mas, algumas vezes, são equivalentes. Em expressões como ‘voz humana’ e

Migrando para sistemas mais sofisticados em relação aos que funcionam com palmas, temos aqueles que atendem a comandos básicos por voz como ‘liga’ e ‘desliga’ e que são aplicáveis a quase tudo o que funciona com eletricidade em uma residência: lâmpadas, televisores, ventiladores, condicionadores de ar, exaustores etc. Para esses, o processamento de áudio precisa ser bem mais elaborado porque não são apenas sons que os comandam, mas sons minimamente organizados: palavras. É válido lembrar que em termos de sofisticação e precisão eles são bem diferentes, mas em quantidade de comandos são análogos, conforme se demonstra na Tabela 15:

Tabela 15 – Comparativo entre comandos

	comando por palmas	comando por voz
acionar	uma palma	palavra ‘liga’
Desacionar	duas palmas	palavra ‘desliga’

Fonte: Texto organizado pelo autor

Buscando por mais recursos e por uma funcionalidade um pouco mais adequada, qualquer projetista da área da Domótica agora saberia que, caso seu cliente optasse por dotar vários de seus eletrodomésticos com o recurso de comandos por voz, ele teria um problema. Ao entrar em sua sala de estar e dizer ‘liga’, seriam acionados ao mesmo tempo lâmpada, televisor, ventilador e condicionador de ar. Isso ocorreria porque os sistemas citados estariam dotados de apenas dois comandos programados: ‘liga’ e ‘desliga’.

O passo seguinte seria então especificar: ‘ventilador liga’ e ‘ventilador desliga’ ou ‘televisor liga’ e ‘televisor desliga’. Dessa forma, os sistemas, integrados ou não, saberiam qual aparelho ligar. Surge então o primeiro problema que diz respeito ao volume de informação a ser processado. ‘Liga’ possui bem menos fonemas que ‘ventilador liga’. Além disso, necessita de pouco tempo de execução e, por conseguinte, menor espaço para gravação digital, algo bastante crítico quando se trata de processamento de dados de um modo geral. Note-se, por esse pequeno detalhe, que o processamento de áudio depende de uma série de fatores. A um usuário que não se importa em empregar mais e mais palavras, a fim de solicitar uma tarefa simples, pode ser indiferente dizer “por favor computador, pode por gentileza

‘fala humana’ pode-se dizer que se trata quase do mesmo fenômeno e alguns linguistas até diriam que não há ‘fala’ que não seja humana. Quando se distingue, por exemplo, entre um fonema vozeado e outro não-vozeado, é clara a presença das vibrações das cordas – ou pregas - vocais em um e não em outro, respectivamente. Não se discute, portanto, sobre fonema ‘falado’ e fonema ‘não-falado’, justamente porque o ‘fonema’ é abstrato, enquanto os ‘fones’ e ‘alofones’ são fenômenos físicos. Em Processamento de Sinais da Fala, ‘reconhecimento de voz’ e ‘reconhecimento de fala’ equiparam-se, assim como ‘comandos por voz’ e ‘comandos por fala’, ainda que esse último seja bem menos recorrente. De qualquer forma, um mínimo de cuidado com a terminologia deve ser tomado.

colocar o ventilador em funcionamento” em vez de “ventilador liga” tão somente. Aos linguistas, projetistas e desenvolvedores de sistemas isso é um problema.

Considerando o exemplo do ventilador e o dotando de recursos ‘dimerizados’ como o do chuveiro e o da lâmpada citados há pouco, haveria ainda mais trabalho e muito mais variáveis precisariam ser consideradas se o ventilador tivesse que operar juntamente com o condicionador de ar e a iluminação:

Tabela 16 – Possíveis comandos por voz em uma sala de estar

ventilador	condicionador de ar	iluminação
ventilador liga	ar liga	luz liga
ventilador menos	ar menos	luz menos
ventilador menos	ar menos	luz menos
ventilador menos	ar menos	luz menos
ventilador desliga	ar desliga	luz desliga

Fonte: Texto organizado pelo autor

Tomando então a Tabela 16 como uma nova realidade de Domótica, haveria a possibilidade ou de um sistema independente para cada um dos aparelhos ou de um único sistema integrado capaz de controlar a todos por meio de um processamento de áudio mais sofisticado que a opção primeira. Note-se que para ‘ventilador desliga’ - [vẽtʃila'doʁdiz'ligə] - há uma complexidade fonética bastante diversa daquela encontrada em ‘desliga’ - [dez'ligə]. Além disso, como comentado anteriormente, é necessário mais memória digital e maior capacidade de processamento. Para esses casos, o *reconhecimento de voz* já basta, ou seja, não é necessária a *compreensão de voz*.

Ainda sobre a Tabela 16, considere-se que, no Brasil, o advérbio ‘menos’ varia para ‘menas’ em diversas localidades, tanto na distribuição por nível de escolaridade quanto por faixa etária, tal como indicam as cartas linguísticas do ALiB, M03E e M03G, respectivamente (CARDOSO et al., 2014b). O equipamento tem também de dar conta dessa realidade morfossintática, ou seja, não se trata apenas de variação diageracional (idade), diasssexual (vozes masculina ou feminina) e/ou diatópica (local de origem, regional), mas de uma palavra que apresenta, ao menos, um fonema diferente. Importante lembrar que ‘menas’ ocorreria mais provavelmente em situações em que o nome a ser alterado é feminino como *potência*, *velocidade*, *luz*, *altura*, abertura e raramente para *vento*, *volume*, *frio*, *brilho* entre outros – pouco possível que ocorra mentalmente uma construção como ‘menas frio’ para um aparelho de ar condicionado, por exemplo.

Deve-se salientar ademais que, até esse ponto da discussão, ainda não foram expostos os problemas dos ruídos de ambiência, responsáveis por muitas falhas em dispositivos

comandados por voz, além das diferenças existentes entre as vozes das pessoas, como timbre, potência sonora, entonação, prosódia entre outros, assuntos que requerem pesquisas à parte.

Os exemplos de Domótica, elencados até então, representam apenas alguns dos muitos empregos do recurso de comando por voz desde sua gênese até a atualidade, mas já possibilitam visualizar uma parte da problemática envolvida entre a idealização de um sistema - prancheta - e sua plena operacionalidade - usuário final e variantes linguísticas correspondentes.

O experimento proposto para o presente estudo tem a ver com Domótica, mas pode ser ampliado para outros fins, como disposto adiante. Trata-se basicamente da utilização e aplicação de um dispositivo facilmente encontrável à venda em lojas do ramo - virtuais ou físicas - e é destinado a pesquisadores, desenvolvedores, projetistas e hobistas.

Distribuído pela Elechouse, o V2 é dotado de um processador específico para esse fim que funciona por interface USB, consome pouca energia e apresenta relativa facilidade quanto à programação e configuração. A Figura 09 dá uma noção de sua aparência física.

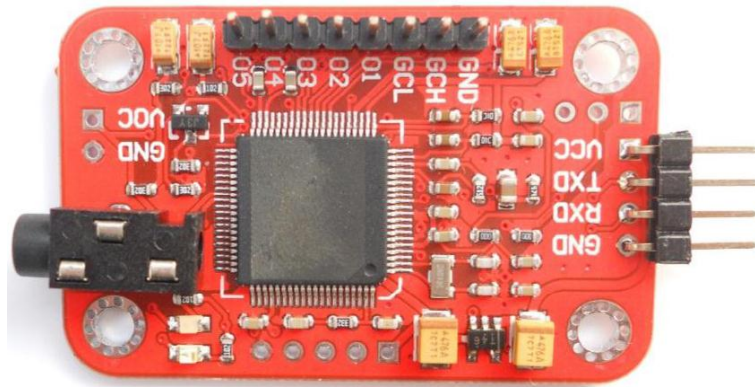


Figura 09- *Voice Recognition Module V2*

O que ele faz é bastante simples ao observador, embora seja sabidamente a junção de uma série de tecnologias que vem sendo desenvolvidas nas últimas décadas. Sua função é gravar e armazenar três grupos de comandos por voz, cada qual com cinco peças⁵³ de até 1.3 segundo de extensão em tempo de execução. Após gravados e armazenados, os comandos podem acionar praticamente qualquer dispositivo elétrico por meio de suas portas lógicas.

⁵³ Para o caso desse equipamento, o comando por voz diz respeito a uma palavra ou expressão no idioma correspondente. No manual do V2, cada comando por voz é denominado 'peça' e o que mais importa é o tempo de duração. A palavra 'geossociolinguística', por exemplo, pronunciada de modo espontâneo, pode não ser uma 'peça' válida por ultrapassar o máximo de 1.3 segundo estabelecido pelo fabricante.

Tabela 17 – Amostra de possíveis grupos com 5 peças (comandos) cada

grupo 01	grupo 02	grupo 03
Amarelo	primeiro	direita
Vermelho	segundo	frente
Verde	terceiro	para
Azul	quarto	trás
Branco	quinto	esquerda

Fonte: Texto organizado pelo autor

Por limitações técnicas do próprio V2 e pela necessidade de vários circuitos adicionais, a utilização de comandos como os exibidos na Tabela 16 serão feitos em outro experimento. Neste, a opção foi por algo mais simples - eletronicamente falando – pois, quando se trata do acionamento de equipamentos que operam com nuances ajustáveis como ventiladores, chuveiros, cadeiras ou camas reclináveis, entre outros, o projeto fica mais complexo e oneroso.

Portanto, para testá-lo, em um dos grupos, foram gravados os nomes de cinco cores: vermelho, verde, azul, amarelo e branco. Essas palavras têm algumas vantagens: são relativamente curtas, conhecidas por todas as pessoas - mesmo em tenra idade - e acionam no experimento lâmpadas com as cores respectivas, o que visualmente é incontestável – salvo para algumas pessoas com problemas de visão.

Além disso, a complexidade linguística envolvida em ‘vermelho, verde, azul, amarelo e branco’ - cinco palavras - é maior que em ‘liga-menos-menos-menos-desliga’ - apenas três palavras/comandos. Para esse último caso, é válido notar que, embora para dimerizar seja necessário repetir sequencialmente a palavra ‘menos’ mais de uma vez, a função do reconhecedor é apenas distinguir entre ‘liga’ e ‘menos’. Quem vai decidir se é ou não para dimerizar é o software, responsável por calcular o intervalo de tempo entre as palavras pronunciadas.

É claro que, fora de um contexto experimental e científico como esse, acionar cores pode ainda ser considerado pouco útil em termos de Domótica, embora a Cromoterapia⁵⁴ ocupe espaço considerável entre as terapias alternativas. Porém, como foi dito, as palavras poderiam ser outras e, em vez de lâmpadas coloridas, elevadores - (grupo 02/Tabela 17), veículos sobre rodas (grupo 03/Tabela 17) ou eletrodomésticos (Tabela 16) poderiam ser acionados, substituições relativamente fáceis de serem feitas no âmbito da eletroeletrônica. Em outras palavras, a eletricidade não questiona o que ela está acionando, ela simplesmente é encaminhada até seu destino por meio de condutores e semicondutores e lá faz o seu trabalho.

⁵⁴ A título de curiosidade, a Cromoterapia foi reconhecida como terapia alternativa pela Organização Mundial de Saúde em 1976 (BRASIL, 2009).

Os inconvenientes desse experimento são os mesmos elencados há pouco quando se discutiu sobre equipamentos que funcionam por comandos por voz: o ruído de ambiência e as diferenças existentes entre as vozes. O próprio manual (VOICE Recognition, *online*, 2014) do V2 deixa claro que possivelmente ele não atenderá aos comandos de um amigo, mesmo pronunciando as mesmas palavras. Quanto aos ruídos de ambiência, eles devem ser minimizados ao máximo durante a gravação e execução das palavras, caso contrário o processamento de áudio fica claramente comprometido.

Além do V2, as lâmpadas também contaram com alguns cuidados técnicos e todas foram montadas em um gabinete - Figura 10 - com revestimento opaco e branco que permite visualizar suas cores apenas quando estão ligadas. Cada uma delas tem uma baia própria e isolada que impede a contaminação entre as cores quando estão acionadas simultaneamente e potência suficiente apenas para permitir boa resolução cromática mesmo em ambiente iluminado. O V2 também foi alojado no interior do gabinete e as conexões com o computador foram feitas por meio de aberturas na tampa traseira.



Figura 10 - Gabinete cromático

Tomadas todas as precauções básicas que pudessem permitir melhor funcionamento do experimento, os testes começaram a ser feitos. Primeiramente, procedeu-se com a gravação

de cada uma das peças, como são denominadas no manual do V2 (VOICE Recognition, *online*, 2014). Então vermelho, verde, azul, amarelo e branco foram pronunciadas ao microfone no V2 do modo mais natural possível, embora seja um fato que o grau de automonitoramento é sempre presente e significativo quando a pessoa sabe que está sendo gravada, independentemente de ser um informante ou o próprio pesquisador.

De qualquer forma, o software que controla as gravações está programado para solicitar sempre duas amostras de cada peça, então as compara e, caso sejam equivalentes, finaliza o registro e automaticamente permite a gravação da peça seguinte. Analogamente, é quase como o linguista durante uma transcrição: sempre ouve mais de uma vez. Note-se por esse detalhe que o V2 tem critérios interessantes quanto à sua concepção no quesito funcionalidade uma vez que não permite continuar a gravação se o usuário também não adotar critérios mínimos quanto à pronúncia.

Gravadas todas as peças, o V2 foi colocado à prova. Devidamente conectado ao gabinete com as lâmpadas, funcionou muito adequadamente para cada uma das cores e as respectivas lâmpadas foram acionadas e desacionadas com bastante êxito. As transcrições das peças são [veɾ'melʝu], ['veɾ'dʒɪ], [a'zuw], [ama'rɛlu] e ['brãku]. A Tabela 18 apresenta o modo como os comandos se distribuem. Pode-se perceber que, embora o comando para ligar e desligar seja o mesmo - a própria palavra - pode-se, tranquilamente, dizer que há cinco comandos para o gabinete, do mesmo modo como há para cada um dos eletrodomésticos listados na Tabela 15.

Tabela 18 – Relação dos comandos e cores utilizados no experimento

cores	para ligar	para desligar
vermelho	vermelho	vermelho
verde	verde	verde
azul	azul	azul
amarelo	amarelo	amarelo
branco	branco	branco

Fonte: Texto organizado pelo autor

Algumas variantes também foram experimentadas, principalmente das palavras ‘vermelho’ e ‘verde’ por possuírem maior incidência de variação no Brasil: o /R/ em coda silábica em interior de palavra tem essa característica – Cartas Fonéticas F04 C5 e C6 (CARDOSO et al., 2014b). Mesmo com as típicas limitações de um falante que não é nativo da região correspondente, o sistema respondeu bem a ponto de as falhas serem desconsideradas em análise final. Isso não quer dizer que o sistema opere com *compreensão de voz*, mas comprova que o método de processamento de sinal certamente leva em conta o

tempo de gravação e pressão sonora de cada peça. Algumas das variações simuladas que obtiveram êxito como comandos estão na Tabela 19.

Tabela 19 – Variantes linguísticas dos comandos utilizados no experimento

peça gravada	variantes simuladas como comandos				
[veɾ'melju]	[veχ'melju]	[veh'melju]	[veχ'melju]	[veh'melju]	[ver'melju]
[^l veɾdʒɪ]	[^l veχdɪ]	[^l vehdɪ]	[^l veχdʒɪ]	[^l vehdʒɪ]	[^l verdʒɪ]

Fonte: Texto organizado pelo autor

No entanto, embora tenha acionado a lâmpada vermelha com a peça de áudio [veɾ'melju] assim como ocorreu com todas as outras lâmpadas e respectivas peças, também acionou a lâmpada vermelha com as execuções [pẽ'telju] e [fe'delju]. O mesmo se aplicou à lâmpada amarela, ou seja, foi facilmente acionada com a peça de áudio [ama'relu], mas quase igualmente com as execuções [ʃi'nelu] e [ma'ɾmelu]. Da mesma forma, o V2 ‘entendeu’ [a'zuw] com [ta'tu] e [^lbrãku] com [^lmãku], tal como demonstra Tabela 20. Está claro que, adotados os critérios adequados, ele funciona bem com as peças previamente gravadas, mas também aceita os comandos quando a execução tem fonemas, prosódia e pressão sonora similares ou aproximados. Em termos linguísticos mais técnicos, quando há pares mínimos como ‘faca/vaca’ ou ‘sumir/zunir’ em que estão em jogo Contraste em Ambiente Idêntico – CAI – ou Contraste em Ambiente Análogo - CAA (SILVA, 1999) – respectivamente, o sistema vai mal.

Tabela 20 – Comparativos entre as peças de áudio e execuções aproximadas ou similares

Peça	execuções aproximadas	
[veɾ'melju]	[pẽ'telju]	[fe'delju]
[^l veɾdʒɪ]	[^l leɾdʒɪs]	[^l veɾtʃɪ]
[a'zuw]	[ã'nu]	[ta'tu]
[ama'relu]	[ʃi'nelu]	[maɾ'melu]
[^l brãku]	[^l mãku]	[^l bãku]

Fonte: Texto organizado pelo autor

A peça [^lveɾdʒɪ], no entanto, tal como mostra a Tabela 19, sofreu menores problemas porque palavras em Português com disposição de fonemas, prosódia e pressão sonora similares ou aproximados são bastante raras – elas muitas vezes coincidem com palavras que rimam. Porém, isso não significa que o V2 também não possa entender [^lveɾdʒɪ] com alguma

outra eventual execução inusitada como [ˈleɾdʒis]⁵⁵ ou a locução [ˈveɾtʃi]⁵⁶, pouquíssimo utilizada na fala atualmente.

Originalmente, o gabinete cromático foi idealizado e montado tão somente para demonstrar a funcionalidade do V2 em termos linguísticos. Isso seria feito apenas junto às bancas de professores em algumas versões do SEDATA – Seminário de Dissertações e Teses em Andamento na UEL como extensão do projeto de pesquisa e/ou na presença de outros pesquisadores interessados. No entanto, devido à sua versatilidade nos quesitos visual - não há como não vê-lo - e portabilidade - dimensões compatíveis com porta-malas de vários automóveis – passou a integrar aulas, apresentações e palestras.

Ao longo de todo o segundo semestre de 2014, eventos científicos, palestras, reuniões, divulgação de vestibular a até uma feira de inovação tecnológica cederam espaço ao experimento que acabou surpreendendo - o pesquisador - pela capacidade de prender a atenção das pessoas de um modo geral. Além disso, o experimento proporcionou, concomitantemente, a divulgação da presente pesquisa, o projeto de Regime de Jornada Integral junto à Fatec Garça (São Paulo), o PPGEL e, claro, o ALiB.

Durante suas exposições, muitas dúvidas foram tiradas, muita curiosidade foi suscitada, várias ideias surgiram, um significativo número de pessoas testou o experimento emprestando suas vozes e certa quantidade de conclusões que agora fazem parte do corpo desta seção puderam ser tiradas a partir das observações linguísticas. Entre os fatos que mais chamaram atenção está o de que ficou evidente o cuidado que os desenvolvedores desse tipo de equipamento devem ter com as variantes diassexuais: quando as peças eram gravadas com vozes masculinas, ele funcionava mal com as mulheres e vice-versa. Outro fato interessante tem a ver com o resultado do equipamento quando exposto a uma variedade do Português que não foi estudado pelo ALiB. Durante a *V Feira de Ideias e Inovação* promovida pela AINTEC-UEL, o experimento foi testado por um angolano, que, interessadamente, teve êxito com todas as cores, exceto o verde – foi o único homem em todo o semestre que não conseguiu acionar este comando.

Embora possam parecer detalhes pouco significativos a outras áreas de pesquisa, esses dois fatos certamente têm potencial de gerar boas pesquisas nas quais duas respostas podem ser buscadas: (i) se os equipamentos estão preparados para dar conta das diferenças entre vozes masculinas e femininas (ii) se há de fato, tecnicamente falando, algo de peculiar na palavra *verde* considerando-se todas as distribuições de variantes no Brasil (diassexual,

⁵⁵ Trata-se da 2ª pessoa do plural verbo ‘ler’ no infinitivo pessoal e futuro do subjuntivo.

⁵⁶ ‘Ver-te’.

diageracional, etc). Para essa última dúvida teórica, é válido lembrar que o buscador da Google, inversamente, só identifica ‘verde’ para palavras com proximidade fônica.

Em síntese, o propósito de apresentar este experimento linguístico com o V2 foi atingido com bastante êxito. Em outras palavras, demonstrar diferenças existentes entre *reconhecimento* e *compreensão de voz* por meio do referido dispositivo, dotado da capacidade de processamento de áudio, tornou-se bastante produtivo, mesmo considerando que tal fenômeno também é perfeitamente observável com o *IBM Via Voice* ou o buscador do Google analisados nesta Tese.

Verificou-se, entre outros detalhes que fica clara a diferença entre *reconhecimento de voz* e *compreensão de voz* e não há dúvidas de que entre as tarefas do V2 não se encontra a de *compreensão* - ele não ‘sabe’ o que foi dito. Em outras palavras, comporta-se – na melhor das hipóteses - como um estrangeiro tentando identificar com seu ouvido ainda pouco treinado às nuances entre palavras com proximidade fônica. Desconsiderando-se a melhor das hipóteses, seu resultado está bastante próximo daquele de personagens surdos de quadros humorísticos. Sistemas mais sofisticados que operam com linguagem natural escrita identificam, há muito, os campos semânticos – entre outros critérios – situação em que ‘chinelo’ é facilmente descartado por não pertencer ao grupo ‘cores’. Os buscadores na internet, entre muitos, fazem o uso desse recurso paralelamente a outros (MANFIO; MORENO; BARBOSA, 2014a; 2014b; BARBOSA, 2004).

Rick Deckard, em *Blade Runner*, provavelmente ficaria decepcionado com seu equipamento se obtivesse esses resultados, pois, além do transtorno ocasionado pelo mau funcionamento, sua investigação ficaria prejudicada. No filme, como comentado anteriormente, a máquina parece não ter dúvidas quanto ao que o detetive pronuncia, pois atende ao comando por voz com a respectiva ação, mesmo considerando que há apenas uma fração de segundo entre uma e outra. Considerando-se um *continuum* evolutivo que parte dos equipamentos acionados por som - como os *clappers* - até aqueles que realizam a *compreensão de voz*, o equipamento fictício do filme está claramente mais próximo deste último recurso: a *compreensão*.

Nesse mesmo *continuum*, o experimento aqui apresentado parece deixar claro que, embora tenha ocorrido o *reconhecimento de voz* seguido por frações de segundo da execução do *comando por voz*, a *compreensão de voz* é inexistente. O próprio *reconhecimento* é precário uma vez que há identidade entre palavras que possuem certa similaridade fonética, mas bastante divergentes entre si em quesitos semânticos. Ainda sugere que, possivelmente, dentre os muitos equipamentos disponíveis no mercado que operam com *reconhecimento* e

comando por voz, a grande maioria pode apresentar a mesma falha, pois os métodos utilizados no processamento do sinal de áudio são similares.

4.4 PROFESSOR TICAL EM TRÊS ESTÁGIOS

Os oráculos são personagens bastante comuns na ficção, seja na literatura ou no cinema. Consultar um deles sempre traz ao interlocutor a possibilidade de alterar radicalmente o curso dos acontecimentos, normalmente de modo a favorecer quem o consulta. A ficção científica por vezes apresentou várias versões de oráculos cibernéticos ou robóticos que exercem esse papel. O filme “A.I. Artificial Intelligence” (A.I., 2001), de Steven Spielberg, por exemplo, mostra um oráculo holográfico denominado Dr. Saber (“Dr. Know”) que responde sobre muitos temas. A ciência, embora ainda muito distante do que o cinema tem a oferecer, tem procurado reproduzi-los.

Os *chatbots*, por exemplo, são máquinas que interagem com humanos por linguagem escrita. A diferença entre esses e os *chats* comuns (*Conversational Hypertext Access Technology*) em que há humanos interagindo com humanos por meio do computador, é que os *chatbots*, ou robôs de conversação⁵⁷, como também são conhecidos, simulam um interlocutor em conversação e têm o diferencial de operar com um banco de dados mais preciso quando se trata de um assunto específico. Em ambos os casos, a proliferação de sua existência na grande rede está associada a vários fatores ligados à interação.

Pensando no Eliza (WEIZENBAUM, 1966), que é uma grande referência nessa categoria e em vários outros robôs existentes que interagem e/ou respondem perguntas sobre diversos assuntos, verificou-se a possibilidade de criar um de tipo similar que pudesse ser atrelado a um banco de dados com informações sobre Dialetolegia e Geolinguística, especificamente o ALiB - Atlas Linguístico do Brasil (PROJETO Atlas, 2013; COMITÊ, 2001). Assim, esta seção apresenta um protótipo de robô de conversação, Professor Tical, baseado em conceitos de PLN que responde perguntas da área de Linguística, especificamente sobre Geossociolinguística no Brasil.

O robô foi implementado na linguagem C++, utilizando o Visual Studio 2010 da Microsoft e esse exemplar do protótipo funciona na plataforma Windows 32 e 64 bits, sendo os sistemas operacionais XP, Seven e Windows 8 com o Microsoft .NET Framework 4

⁵⁷ Daremos preferência à expressão ‘robô de conversação’ ou simplesmente ‘robô’ neste trabalho, embora eventualmente possamos nos referir também aos anglicismos *chatbot* ou *chatterbots*, mesmo que não sejam sinônimos perfeitos.

devidamente instalado. O banco de dados foi elaborado com base em perguntas e respostas bastante básicas e em um número bastante reduzido apenas para atender a testes iniciais e ao conteúdo deste estudo. O PLN será abordado nos quesitos mais elementares quando da abordagem sobre as tabelas *hash*⁵⁸.

4.4.1 Robô de Conversação sobre Dialetologia e Geossociolinguística

O Professor TICAL - Tecnologia Interativa Conversacional sobre Assuntos Linguísticos (doravante apenas ‘Professor Tical’) é um robô de conversação que tem por função responder perguntas sobre Dialetologia, Geossociolinguística e teorias afins.

A área de conhecimento temática escolhida para o Professor Tical é relevante por três principais motivos: (i) praticamente não há robôs de conversação que versem sobre Linguística e, muitos menos, em Português; (ii) ele pode operar, no mínimo, como agente de difusão cultural da área envolvida e de todas as teorias a ela associadas; (iii) é grande a necessidade de procurar manter o idioma ativo na Sociedade da Informação e o protótipo tem potencial de contribuir para a proliferação de interfaces de acesso a banco de dados baseadas em PLN.

Por esse prisma, as estatísticas são bastante favoráveis ao grau de ineditismo da proposta, pois um robô de conversação que responda sobre tais temas e em Língua Portuguesa certamente é novidade.

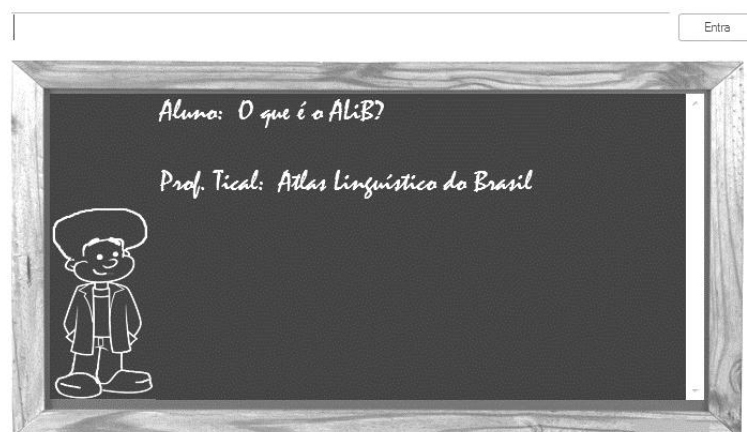


Figura 11 - Impressão de tela inteira do Professor Tical

⁵⁸ O termo *hash* significa basicamente ‘picadinho’ e, por extensão, ‘bagunça’ e relaciona-se ao método de transformação de chave, denominado *hashing*. Trata-se uma estrutura de dados que associa palavras-chave – ou chaves de pesquisa - a elementos correlacionados - ou valores - e tem como diferencial a capacidade de realizar uma busca rápida e obter um valor a partir de uma entrada simples como uma palavra (ZIVIANI, 1999).

Tal como pode ser observado na Figura 11, esse protótipo apresenta bastante simplicidade em sua aparência e funcionalidade, ou seja, o ambiente é minimamente *amigável*: há um personagem à esquerda que representa o Professor Tical e um fundo imitando uma lousa - que pressupõe uma sala de aula. Existem também nos diálogos as indicações ‘Professor Tical’ e ‘aluno’ distinguindo os respectivos turnos dos interlocutores.

O processo de personalização e caracterização de elementos temáticos pode a alguns pesquisadores parecer ‘perfumaria’ à primeira vista, mas constitui-se num procedimento necessário quando se trata de aprimorar quesitos voltados ao *design* de interação (PREECE, 2005).

As necessidades de criar robôs que versem sobre os mais diversos tipos de assuntos têm também justificativas das mais diversas. Interessante salientar que os robôs de conversação costumam atender quase às mesmas funções básicas da maior parte dos jogos. Dentre essas funções básicas, podem-se destacar as mais relevantes que são: entretenimento, treinamento, educação e difusão cultural (HUIZINGA, 2007).

Entre muitos, um exemplo disso talvez seja o Robô Ed (CONPET, 2014), disponibilizado na página do Conpet - Programa Nacional da Racionalização do uso dos derivados do petróleo e do gás natural - que, embora tenha sido idealizado, mormente para dar informações sobre assuntos ligados ao meio ambiente, educação e difusão cultural, não deixa de ter um viés de entretenimento, para aqueles que apenas querem se divertir com as possibilidades de respostas, e treinamento, para quem está disposto a estudar, avaliar e promover o conteúdo disponibilizado.

Evidentemente, não são todos os robôs de conversação que atendem aos quatro requisitos ao mesmo tempo e, quando atendem, nem sempre abarcam todos na mesma proporção. Aquele que é criado estritamente para fins educativos, também pode ser objeto de estudo de pedagogos ou profissionais da área de educação. Outro, cuja gênese remeta ao entretenimento, pode conter informações interessantes e educativas a algumas faixas etárias específicas.

O próprio clássico Eliza, embora tenha sido criado experimentalmente pelo seu desenvolvedor para aplicação de algumas teorias e tecnologias da época, mesmo hoje continua sendo utilizado como objeto de estudo e/ou para diversão, ainda que também impulse, de modo indireto, algumas discussões sobre psicanálise: o robô ‘ouve pacientemente’ o que a pessoa tem a dizer e a deixa à vontade para falar sobre o que quiser. Não é exagero frisar, portanto, apenas a título de curiosidade, que ter alguém que ‘ouça pacientemente’, há muito, é uma das grandes necessidades dos seres humanos ao redor do

mundo. Esse talvez seja um dos motivos pelos quais o Eliza ficou tão famoso: pessoas acreditaram e ‘apreciaram’ serem consultadas.

Embora o Professor Tical não tenha sido criado para portar-se como um psicanalista, tem mais ou menos a mesma funcionalidade básica do Eliza - guardadas as devidas proporções e condições de produção - que é a de aplicar determinados tipos de tecnologias e teorias e associá-las à possibilidade de impulsionar certa esfera de conhecimento. Para o Eliza, não necessariamente a Psicanálise, mas o PLN de modo geral. Para o Tical, não apenas a Dialetologia e a Geossociolinguística no Brasil, mas também o PLN entre outros. Assim, o que se espera dele, num primeiro momento em que estiver disponibilizado em rede, é nada menos que aquilo que foi dito há pouco, ou seja, que as diferentes pessoas utilizem o aplicativo do modo como melhor lhes convier. Em outras palavras, que explorem uma das vantagens do sistema, que é interagir com o robô de forma dialógica em vez de fazer a escolha entre elementos de uma lista ou palavras-chaves sublinhadas em um parágrafo (ROTHERMEL, 2007).

Como todo robô de conversação, o Professor Tical tem um banco de dados. Esse banco de dados vem sendo carregado manualmente com informações relacionadas a perguntas básicas sobre a Linguística Geral, Dialetologia, Geossociolinguística, ALiB e alguns conceitos transversais. A Tabela 21 ilustra o paralelo entre algumas potenciais perguntas e suas respostas.

Tabela 21: Paralelo demonstrativo de perguntas e respostas

PERGUNTAS	RESPOSTAS
O que é o ALiB? O que é ALiB? De que se trata o ALiB?	Atlas Linguístico do Brasil
Desde quando existe o ALiB? Que idade tem o ALiB? Qual o tempo de existência do ALiB? Quando foi criado o ALiB?	Surgiu em novembro de 1996 durante o <i>Seminário Caminhos e Perspectivas para a Geolinguística no Brasil</i> na cidade de Salvador
Quem criou o ALiB? Quem inventou o ALiB? Quem concebeu o ALiB? Quem idealizou o ALiB?	Ele foi criado por vários pesquisadores que fazem parte do Comitê Nacional do Projeto ALiB, entre eles, a Diretora Presidente do Comitê Suzana Alice Marcelino da Silva Cardoso. A sede de seu lançamento é UFBA
A qual área pertence o ALiB? De que área é o ALiB? De qual área é o ALiB?	Geossociolinguística ou Geolinguística Pluridimensional.
O que é a Geossociolinguística? O que é Geossociolinguística? De que se trata a Geossociolinguística?	Associação da Geografia Linguística com a Sociolinguística
O que é Sociolinguística? Do que se trata a Sociolinguística? De que se trata a Sociolinguística?	Sociolinguística seria um ramo da Linguística que leva em conta o aspecto social das línguas. Porém, como não é possível distinguir entre uma linguística geral para as línguas e uma Sociolinguística apenas para os aspectos sociais, “a Sociolinguística é a Linguística” (CALVET, 2002).
O que é Geolinguística? Do que se trata a Geolinguística? De que se trata a Geolinguística?	É um método dialetológico e comparativo pelo qual se traçam mapas de formas linguísticas encontráveis no âmbito geográfico relativo à língua ou dialeto estudado (COSERIU, 1982). Em outras palavras, busca-se a delimitação de zonas isoglóssicas de alguns vocábulos e sons com o intuito de descrever a circunscrição e a irradiação de formas linguísticas (AGUILERA, 2005). Consideram-se elementos fônicos, lexicais, gramaticais, etnográficos e relativos às variações diastráticas (BRANDÃO, 1991).

Fonte: Calvet, (2002), Coseriu (1982), Aguilera (2005), Brandão (1991), texto organizado pelo autor.

Todas as respostas, e não apenas essas que estão ilustradas Tabela 21, são baseadas em literatura específica que versa sobre as teorias relacionadas e tal literatura encontra-se também listada nas Referências deste trabalho. Quando se fala em ALiB, portanto, não há como não falar em Dialetologia, Sociolinguística, Geolinguística, Geossociolinguística, Linguística Geral, Fonética e Fonologia, entre outras e vice-versa. Assim sendo, mesmo que as respostas pareçam incompletas ou insuficientes do ponto de vista teórico, em função da complexidade inerente em explicar de modo sucinto cada uma das áreas relacionadas e os respectivos fenômenos linguísticos, elas estão lá e cumprem com a sua finalidade básica.

Porém, ainda que as respostas estejam completas no banco de dados, o sistema não é carregado com as perguntas na íntegra, ou seja, com a sintaxe tradicional tal como mostradas na Tabela 21. Ele é programado para identificar e combinar palavras presentes nas perguntas do interlocutor e selecionar a resposta que talvez atenda ao questionamento elaborado. Não

há, por exemplo, no banco de dados a pergunta “O que é o ALiB?”. Há apenas “alib” e “o que é”. Da mesma forma não há a pergunta “O que é Geolinguística?”. Há apenas “geolinguística”. A expressão “o que” serve para as duas e para muitas outras. Sendo assim, toda vez que alguém pergunta ‘o que é’, o sistema busca basicamente duas palavras-chave.

O robô funciona a partir do conceito de PLN e, para tanto, utiliza tabelas tipo *hash*. Em uma versão posterior e bem mais aprimorada que essa foi possível incrementar o sistema com recursos de PSF especificamente *síntese e reconhecimento de voz*⁵⁹. Nessa condição, os diversos modos como a mesma palavra é pronunciada no Brasil e a variedade lexical relativa à mesma lexia podem ser explorados paralelamente tanto pelo PLN quanto pelo PSF: respectivamente, incrementando sintaxes e vocabulário e aprimorando vozes sintéticas e reconhecimento (MANFIO, 2014c). Aqui estarão em evidência apenas PLN e tabelas *hash*.

O PLN diz respeito quase sempre ao processamento de linguagem escrita, mormente fazendo o uso de conhecimentos léxicos, sintáticos e semânticos. Estão, portanto, incluídos no conceito de PLN a compreensão, geração e outras tarefas como tradução de idiomas, cuja precisão depende em muito do processamento de linguagens naturais (RICH, 1993; SCHILDT, 1989).

O *hash*, também conhecido como ‘tabela de dispersão’ ou ‘tabela de espalhamento’ tem como diferencial realizar uma busca rápida e obter o valor correlacionado a partir de uma entrada simples como ‘ALiB’. Como todo este processo ocorre em memória durante a execução do Professor Tical, o armazenamento físico dos dados é feito através de arquivos XML (eXtensible Markup Language).

Tical opera com duas tabelas *hash*: uma para entradas/respostas e outra para sinônimos - substantivos e verbos - tempos verbais . Também possui uma lista encadeada e ordenada para as entradas do usuário e cadastro da tabela *hash* - entradas/respostas. Um exemplo de tabela entradas/respostas está disposto na Tabela 22:

⁵⁹ Dois estudos que relacionam essas duas áreas do PSF com Dialetologia e Geossociolinguística são “Aspectos fonéticos no *DosVox* enquanto aplicativo tipo texto-fala” (MANFIO, 2014a) e “Como funcionam alguns fonemas no aplicativo Balabolka” (MANFIO, 2014b) incorporados a esta Tese respectivamente nas seções 2.2 e 2.1.

Tabela 22: Paralelo demonstrativo de entradas e respostas x possibilidades

	POSSIBILIDADE 'X'	POSSIBILIDADE 'Y'
Entrada	Como se denomina macaxeira no Brasil?	Como é denominada macaxeira?
Resposta	Em quase toda a Região Norte, as pessoas chamam-na de 'macaxeira'. No Centro-Oeste e cidade de São Paulo predomina o termo 'mandioca'. Na Bahia e nas capitais Vitória, Rio de Janeiro, Florianópolis e Porto Alegre é conhecida como 'aipim'	

Fonte: Texto organizado pelo autor

A tabela de sinônimos, por sua vez, não atende necessariamente a estudos de semântica formal uma vez que, em princípio, o robô pode ser dotado de um banco de dados baseado em conhecimentos empíricos de pessoas pertencentes a diversas áreas de atuação. Se a existência da sinonímia perfeita já é bastante questionada por muitos linguistas (BORBA, 1991), para essa aplicação ela se torna uma discussão desnecessária e passa a interessar pouco. Considerem-se, portanto, como sinônimos aqui as palavras que, quando substituídas uma pela outra, não alteram a significação da sequência de modo a comprometê-la (BORBA, 1991). Sendo assim, o que se leva em conta nesse momento de desenvolvimento do protótipo é a experiência do falante. Eis um exemplo do próprio Tical na Tabela 23:

Tabela 23: Paralelo demonstrativo entre palavra e sinônimo

	PALAVRA	SINÔNIMO
Verbo	denominado	denomina
Substantivo	macaxeira	mandioca

Fonte: Texto organizado pelo autor

A maneira como o sistema irá gravar na tabela entradas/respostas não obedece, como dito anteriormente, à norma gramatical em termos de sintaxe. O modo ordenado para o sistema prevê algo como demonstrado nas Tabelas 22 e 23: 'denomina/denominado'. A tabela de sinônimos serve para alterar as perguntas do usuário e das respostas cadastradas. Levando-se em conta as respostas já cadastradas, há na Tabela 24, como exemplos, algumas perguntas equivalentes entre si.

Tabela 24: Bloco de perguntas equivalentes já cadastrado e sua resposta

ENTRADAS	RESPOSTA
Para que serve o ALiB?	Pesquisas e estudos sobre os modos de pronúncia e uso de palavras nas diferentes regiões do Brasil
Qual a função do ALiB?	
Qual a finalidade do ALiB?	
Qual o objetivo do ALiB?	

Fonte: Texto organizado pelo autor

A partir daí pode-se proceder com o cadastro dos sinônimos da seguinte maneira tal como organizado na Tabela 25.

Tabela 25: Sinônimos para perguntas sobre o ALiB

PALAVRA	SINÔNIMO
Finalidade	função
Objetivo	
Serve	

Fonte: Texto organizado pelo autor

Nas frases cadastradas para entrada/reposta, poderia ser feito seguindo o padrão demonstrado na Tabela 26.

Tabela 26: Entradas e respostas sobre o ALiB

FRASE (entrada)	FRASE (transformada)	RESPOSTA
finalidade ALiB	ALiB função	Pesquisas e estudos sobre os modos de pronúncia e uso de palavras nas diferentes regiões do Brasil

Fonte: Texto organizado pelo autor

Nesse caso, o sistema grava só uma combinação para a frase. Desse modo, qualquer uma das quatro entradas indicadas na Tabela 24 será respondida. Há, no entanto, um problema que necessita ser resolvido paralelamente ao processo de incrementação do banco de dados pois, caso haja alguma palavra idêntica em outra frase, esta também será afetada. Para resolver isso, são necessários alguns cuidados básicos como testes de utilização por diferentes pessoas e maior diversidade vocabular em contraste do banco de dados.

Outro dado relevante é que o programa deve estar preparado para operar bem quando ocorrem falhas de digitação ou algum descuido do interlocutor quanto à grafia e/ou pontuação. A Figura 12 mostra um flagrante bastante típico quando se trata do uso de chats no cotidiano.

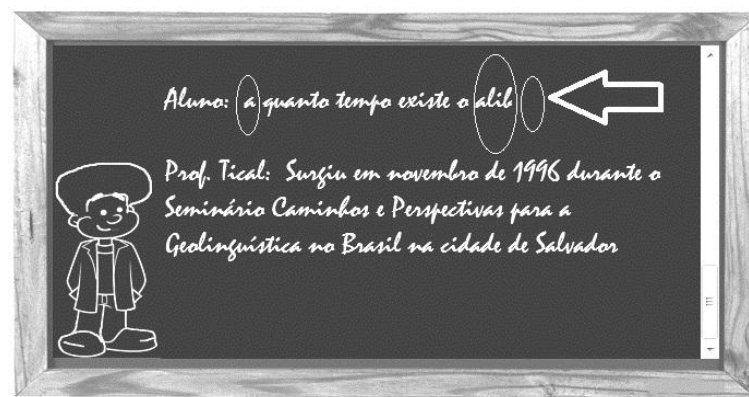


Figura 12 - Falhas de digitação, grafia e pontuação na pergunta.

É possível notar, nesse caso, ao menos três deles: (i) ‘há’ do verbo haver por ‘a’; (ii) supressão das maiúsculas necessárias para grafar a sigla ALiB; (iii) inexistência do ponto de

interrogação que pressupõe pergunta. Todos, no entanto, ocorrendo ao mesmo tempo, não interferiram na possibilidade de resposta, ou seja, o Professor Tical atendeu à solicitação.

Isso deve ocorrer porque, tal como comentado anteriormente, o programa analisará o conjunto ‘alib’/‘há quanto tempo’/‘existe’ diferente do conjunto ‘alib’/‘o que’/‘o que é’. Em versões futuras, poderão ser incluídos recursos que identifiquem a presença ou ausência do ponto de interrogação, momento em que o programa poderá questionar o interlocutor se este fez uma pergunta ou uma afirmação, recurso bastante sofisticado quando se pretende aproximar o comportamento do robô ao do humano.

Importante lembrar que, sob o ponto de vista de alguns desenvolvedores na área de PLN, um sistema que pretende processar linguagem natural normalmente deve abarcar módulos associados aos níveis linguísticos de processamento da linguagem. Esses níveis distribuem-se praticamente em todas as áreas de conhecimento linguístico como Fonética e Fonologia - incluindo-se a Prosódia - Morfologia, Sintaxe, Semântica e Pragmática. Além disso, jamais abre mão de utilizar um recurso essencial: o léxico (BARBOSA, 2004). Não é de espantar, portanto, que Tical se pareça em vários aspectos com muitos sistemas em operação no cotidiano e, entre eles, todos os que são baseados em ILNDB - *Interface em Linguagem Natural para Banco de Dados*, ou seja, sistemas que permitem ao usuário acessar informações armazenadas em um banco de dados escrevendo pedidos em qualquer linguagem natural (ANDROUTSOPOULOS; RITCHIE; THANISCH, 1994). Para ficar em apenas dois exemplos, qualquer sistema de busca disponível na internet ou em uma grande biblioteca física.

Assim sendo, o Professor Tical aproxima-se daquilo que alguns desenvolvedores chamam de analisador léxico-morfológico, responsável por identificar palavras e/ou expressões isoladas em sentenças (BARBOSA, 2004), embora, como já comentado, esse protótipo apenas se baseie nesse processo, isto já é suficiente para enquadrá-lo como um analisador em Processamento de Linguagem Natural.

Além disso, tal como comentado anteriormente, é apenas mais um entre muitos robôs similares existentes funcionando na rede ou disponíveis para serem baixados. O seu diferencial encontra-se nos três pontos citados logo ao início deste texto: (i) praticamente não há robôs de conversação que versem sobre Linguística e que operem em Português; (ii) pode operar como agente de difusão cultural da área envolvida e de todas as teorias a ela associadas; (iii) é necessário manter o idioma ativo na Sociedade da Informação.

O protótipo aqui apresentado certamente estará em outro estágio de desenvolvimento quando este trabalho for publicado em função de se encontrar sempre em processo constante

de incrementação que o dotará de maior precisão, acurácia e relativa proximidade com a interação entre pessoas.

Além disso, por aproximar-se do conceito de analisador léxico-morfológico pelo fato de operar com identificação de palavras e expressões presentes em construções sintáticas, o Professor Tical tem grande potencial de, não só dar conta de responder a algumas perguntas sobre Dialectologia, Geossociolinguística e teorias afins, mas também navegar por um mar de cartas lexicais também relacionadas ao Atlas Linguístico do Brasil.

4.4.2 Tical no WhatsApp

O WhatsApp Messenger - conhecido apenas por WhatsApp - é um dos mais populares aplicativos de mensagem multiplataforma para smartphones no Brasil e, entre suas grandes vantagens, além do alto grau de *usabilidade* e *ambiência amigável*, está o fato de que os usuários podem enviar mensagens ilimitadas sem pagar por SMS - *Short Message Service* ou Serviço de Mensagens Curtas.

Nesse âmbito, o propósito desta seção é apresentar o funcionamento do robô de conversação Tical no mensageiro on-line WhatsApp. A relevância do estudo e do projeto está no fato de que são poucos os *chatbots* que respondem perguntas por meio de aplicativos desse tipo e ainda apresenta o diferencial de possibilitar consultas em âmbito global e sem custo adicional aos alunos e pesquisadores. Tical, que dá respostas a questões da área de Linguística (MANFIO; MORENO; BARBOSA, 2014b), opera no WhatsApp por intermédio de um recurso criado com o aplicativo *WhatsAppAPI.net* e de uma linha de telefonia móvel particular. Para funcionar 24 horas por dia, requer também um servidor simples conectado à rede da Universidade com a qual os desenvolvedores mantêm vínculo em nível de pós-graduação.

A propósito, para desenvolvedores, ele também é bastante viável uma vez que há gradativamente mais ferramentas que tornam o aplicativo compatível com diferentes linguagens de programação e softwares. Esta seção apresenta justamente um desses exemplos, pois foi possível utilizar os recursos de um robô de conversação por meio do WhatsApp. A partir de então, vários conceitos vinculados ao ALiB desde a sua gênese até perguntas específicas sobre a existência de uma carta lexical com as variantes para *mexerica* podem ser consultados.

Tical é um projeto com um grau de interdisciplinaridade bastante alto. Em um primeiro momento, contava com seu desenvolvimento e implementação em Visual Studio e

C++, tabelas *hash* ou *hashing*, banco de dados da área de Linguística, além de promover a difusão cultural do ALiB, que é o maior trabalho de pesquisa realizado na área de Geossociolinguística do Brasil até a atualidade (CARDOSO et al., 2014a; 2014b). Fazê-lo operar também no WhatsApp, portanto, foi apenas mais um passo de sua evolução.

O professor robô tornou-se oficialmente operacional durante o III CIDS - Congresso Internacional sobre Dialectologia e Sociolinguística - realizado na UEL em Londrina no ano de 2014 (MANFIO; MORENO; BARBOSA, 2014b). Lá, esteve disponível no site do evento enquanto protótipo para que os participantes pudessem testá-lo. Na ocasião, pensou-se na possibilidade de Tical estar disponível também para dispositivos móveis e então a ideia de não fazer o pesquisador esperar por um ponto fixo de rede para poder consultar dados sobre o Atlas Linguístico do Brasil pareceu pertinente.

Algumas modificações, no entanto, foram necessárias para que Tical se tornasse compatível com o *WhatsAppAPI*. Para tanto, ele foi reescrito em C# e algumas adequações relativas às filas - perguntas simultâneas enviadas pelos interlocutores - tiveram de ser realizadas, uma vez que as respostas de Tical normalmente são mais longas que aquelas dos usuários comuns no WhatsApp. Feito isso, bastaram uma linha funcional de telefonia celular e um acesso à grande rede a partir de um computador tipo desktop básico.

No entanto, nunca é demais lembrar que Tical ainda se encontra na categoria de protótipo pelo fato de ainda não possuir um banco de dados significativamente amplo que possa atender estatisticamente à maior parte das potenciais perguntas na área, considerando que, eventualmente, elas possam ser feitas apenas por alunos, professores, pesquisadores e desenvolvedores. Esse universo de interlocutores já é bastante grande e caso seja considerada também a interação com o público em geral, seriam necessários outros cuidados específicos com o banco de dados de um modo geral.

As perguntas e respostas em Tical são cadastradas manualmente. Palavras e expressões equivalentes entre si são cadastradas como sinônimos no sistema e esse procedimento tem sido feito por pessoas pertencentes à área envolvida do tema – ALiB – uma vez que, muitas das vezes, é o contexto que determina as relações sinonímicas ou de equivalência de significado (MANFIO; MORENO; BARBOSA, 2014a). Tical também conta com um sistema automático de registro para perguntas não respondidas, ou seja, quando ele não encontra a resposta, avisa ao interlocutor que não a possui e faz o registro para que os desenvolvedores possam incrementar o banco de dados posteriormente, do qual certamente farão parte todas as perguntas existentes nos questionários (COMITÊ, 2001) elaborados e

utilizados como guia pelo Comitê Nacional do Projeto ALiB durante as entrevistas com os 1100 informantes de todo o Brasil.

Como explicitado, a interface de Tical, embora simples, apresenta alguns elementos relacionados à concepção de professor: uma representação do docente por meio de um personagem trajando jaleco ao lado do quadro, a lousa com fundo negro e os dizeres em branco simulando o giz. Esses estereótipos do magistério colaboram grandemente com o conceito de *ambiência amigável*, cujos quesitos são muito discutidos e respeitados por alguns desenvolvedores atentos ao *design* de interação (PREECE, 2005). A Figura 13 apresenta uma amostra dessas características e contempla uma versão mais atualizada que aquela das Figuras 11 e 12.

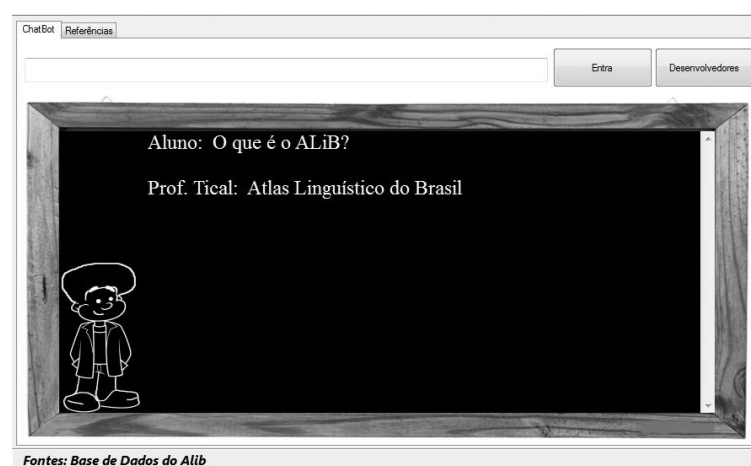


Figura 13 - Interface convencional de Tical com uma amostra de pergunta e resposta

O conceito de *usabilidade* também foi levado em consideração. Basicamente, há apenas o campo para digitar as perguntas e a lousa onde perguntas e respostas podem ser visualizadas. A dimensão dos caracteres é satisfatória para a maior parte das pessoas e o alto contraste – preto e branco – facilita bastante a leitura, mesmo com excesso de luz ambiente ou monitores com problemas cromáticos. O cursor retorna automaticamente ao campo de digitação a cada processo e a utilização do mouse para clicar e acionar o botão ‘entra’ a cada formulação é opcional: ao interlocutor, que já está com as mãos posicionadas, basta pressionar ‘entra’ no teclado para que a busca seja feita.

Quase todas essas características, no entanto, não estão presentes na interface de Tical no WhatsApp. Como pode ser verificado na Figura 14, ela é quase exatamente a mesma do próprio aplicativo WhatsApp, apenas com as particularidades da identificação ‘Tical’ no cabeçalho e da sua capacidade em responder algumas perguntas sobre Linguística.

Evidentemente, a *usabilidade* e a *ambiência amigável* já foram estudadas e aplicadas pelos desenvolvedores do WhatsApp e, assim sendo, Tical faz o uso delas diretamente quando funciona a partir dele.

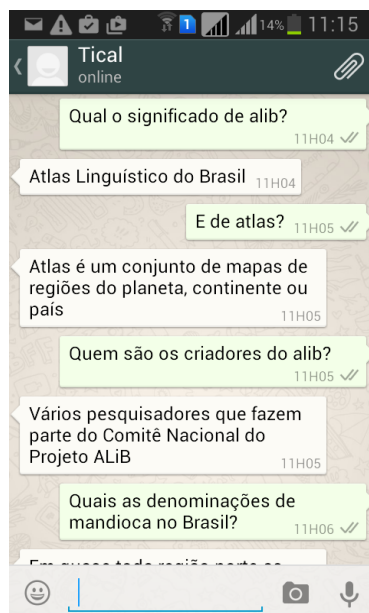


Figura 14 - Exemplo de interface com consultas feitas a Tical pelo WhatsApp

Entretanto, em ambas as interfaces - Tical convencional e Tical no WhatsApp – o que se vê em uso é a Linguagem Natural, é uma das quatro interfaces mais utilizadas em sistemas, juntamente com os *menus*, *form-fills* e *linguagem de comandos*. Em outras palavras, Linguagem Natural para esse caso é utilização de texto escrito em idioma preestabelecido - Português Brasileiro - a partir de um *prompt* (BARBOSA, 2004).

Para fazer o Tical operar com o WhatsApp, foi necessário utilizar a ferramenta *WhatsAPIInet* que é desenvolvida em C# e tem o código aberto, ou seja, *open source*. Por esse motivo, Tical foi reescrito também em C# para gerar compatibilidade com a API - *Application Programming Interface* ou Interface de Programação de Aplicativos. Isso foi necessário porque não havia outra maneira mais rápida de fazer funcionar, no WhatsApp, um *chatbot* como Tical, sem uma ferramenta que fornecesse os recursos para os processos de envio e recebimento de mensagens geradas por um robô de conversação em interação com um humano.

Importante lembrar que, embora neste estudo a API *WhatsAPIInet* tenha sido utilizada para um robô de conversação, ou seja, para uma aplicação específica, ela pode também ser

usada para desenvolvimento de todo e qualquer tipo de aplicativo desktop, móvel ou web⁶⁰ em que esteja envolvido o WhatsApp. Exemplos disso são potenciais utilizações em Domótica e âmbito industrial, situação em que qualquer equipamento pode, não só ser monitorado pelo seu responsável como também informá-lo sobre seu status, o que permite ao interlocutor decidir e enviar à máquina o comando adequado à situação/condição informada.

A tarefa de casar as funções e recursos do WhatsApp com aqueles típicos de um robô de conversação, no entanto, não se deram de modo automático. Textos escritos de maior extensão geraram problema em função do modo como o *hash* ordena suas listas. Mais que isso, ocasionam lentidão também por causa do modo como a rede móvel opera nas diferentes cidades e regiões. Então, decidiu-se limitar a quantidade de palavras a ser consultada por Tical a apenas quinze, uma vez que a função dos *chatbots* de um modo geral é dar conta de perguntas mais sucintas e objetivas. A Figura 15 a seguir representa muito bem isso.

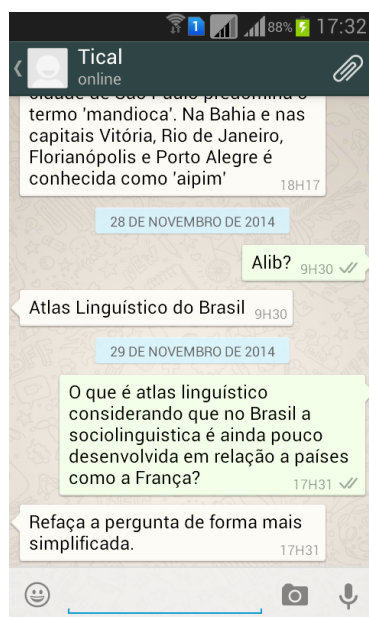


Figura 15 - Exemplo de consulta pelo WhatsApp com mais de 15 palavras

Caso um interlocutor elabore uma pergunta contendo um número de palavras maior que o previamente especificado, como está representado na Figura 15, a resposta será “Refaça a pergunta de forma mais simplificada”. Note-se, apenas a título de curiosidade, que a pergunta que deu origem a essa resposta possui 22 palavras: “O que é atlas linguístico considerando que no Brasil a Sociolinguística é ainda pouco desenvolvida em relação a países como a França?”. O número de 15 pode parecer um pouco arbitrário, porém resolve o

⁶⁰ São denominados ‘aplicativo desktop’, ‘aplicativo móvel’ e ‘aplicativo web’ aqueles que, respectivamente, são implementados para computadores domésticos ou portáteis, smartphones e servidores de redes em geral.

problema e força o interlocutor a ser mais objetivo, uma vez que a função do Tical – ao menos por enquanto – não é discutir à exaustão determinado assunto ou elemento teórico.

Outra alteração feita, mesmo após a constatação de funcionamento de Tical no WhatsApp, foi para resolver o problema da quantidade de mensagens que Tical poderia receber simultaneamente. Para solucionar isso, Tical precisaria colocar todas as perguntas em uma fila e responder a cada uma delas paralelamente. Esse procedimento só foi possível por meio da utilização de outro processo denominado *thread*, cuja função é criar um recurso que, embora opere dentro do Tical, funciona de modo independente buscando as respostas disponíveis no banco de dados.

De um modo global, o objetivo desta seção foi apresentar o funcionamento do robô de conversação Tical no mensageiro on-line WhatsApp. Pôde-se, com a demonstração, verificar dois fatos que colaboram com a relevância do projeto: são poucos os *chatbots* que respondem perguntas por meio de aplicativos desse tipo; possibilita consultas em âmbito global e sem custo adicional aos alunos, pesquisadores e/ou entusiastas. Tical, que até então possuía o potencial de funcionar on-line e off-line, ou seja, a partir um servidor conectado à rede ou instalado diretamente em computador pessoal, ganha mais versatilidade, pois pode responder a perguntas da sua área de especialidade - a Linguística - por meio de dispositivos móveis dos mais diversos.

4.4.3 Tical por Comandos e Síntese de Voz

Joseph Weizenbaum provavelmente teria ficado fascinado com a possibilidade de disponibilizar sua criação na internet para que mais e mais pessoas pudessem testá-la. Mais que isso, dotar o Eliza com recurso de comandos e síntese de voz seria algo sem precedentes. Porém, na década de 1960, a tecnologia disponível não tornava isso possível. O Eliza, capaz de portar-se como um psicanalista ao simular diálogos com um interlocutor humano, foi desenvolvido para testar e aplicar conceitos teóricos de diferentes áreas. Toda a conversa, porém, tinha que ser realizada por meio da linguagem escrita, ou seja, utilizando o teclado alfabético.

Da mesma forma, como quase todos os projetos de cunho científico incluindo o Eliza, o Professor Tical (MANFIO; MORENO; BARBOSA, 2014a e 2014b) tinha como objetivo básico em sua gênese dar maior aplicabilidade a alguns elementos de ordem mais teórica, tanto no que diz respeito à área de Processamento de Dados quanto à de Linguística – especificamente, o ALiB (PROJETO Atlas, 2013; COMITÊ, 2001, CARDOSO et al., 2014a;

2014b). Em outras palavras, como sua concepção envolve desde sempre um procedimento interdisciplinar, interdepartamental e interinstitucional, o grau de aplicação das diversas teorias envolvidas tinham que, de alguma forma, apresentar resultados, ora no formato de pequenas soluções, ora como dados tabulados em diagramas e tabelas.

Em outra fase de desenvolvimento e pesquisa, Tical passou de um robô de conversação convencional para um *chatbot* capaz de operar com o aplicativo para troca de mensagens WhatsApp (WHATSAPP, *online*, 2014), tal como apresentado na seção 4.4.2. O novo recurso proporcionou ao robô maior flexibilidade uma vez que se tornou disponível a qualquer dispositivo móvel capaz de se conectar à rede, não estando mais limitado a computadores fixos.

Num momento ainda posterior e mais atual, considerando a proposta de ter à disposição um projeto de aplicação mais completo possível e ao mesmo tempo um objeto de pesquisa, foram associadas ao Tical as capacidades de operar com comandos e síntese de voz. A partir desse ponto da trajetória evolutiva, uniram-se no projeto *processamento de linguagem escrita, processamento de linguagem falada, linguística geral, inteligência artificial, desenvolvimento de sistemas, tabelas hash, redes sociais, educação, entretenimento, difusão cultural* entre outros.

Assim, nessa etapa, o objetivo foi apresentar o robô de conversação Tical numa dimensão mais ampla e com todas as funcionalidades que a implementação tornou possíveis até o fechamento desta seção. Nessa versão, o Tical mostra-se como um projeto relevante em termos científicos justamente por concentrar vários diferenciais: (a) é um dos poucos robôs que ‘versam’ sobre linguística disponíveis, (b) está todo em Português Brasileiro, (c) opera em dispositivos móveis a partir de aplicativo de mensagem instantânea, (d) possui síntese de voz e (e) atende a comandos por voz. Por esses motivos, seu grau de aplicabilidade é bastante alto e esse quesito colabora grandemente para que continue a ser incrementado com o intuito de evoluir e manter-se compatível com tecnologias futuras.

O Tical tornou-se operacional a partir conceitos básicos de PLN. Importante lembrar que o PLN divide-se em *Processamento de Linguagem Escrita* e *Processamento de Linguagem Falada* (RICH, 1993; BARBOSA, 2004). Para que Tical se tornasse operacional também em *Processamento de Linguagem Falada*, uma vez que operava apenas com texto escrito por meio do teclado alfabético, foi necessário lançar mão de outros programas e aplicativos, além daqueles já utilizados para a implementação básica de Tical.

Como há vários modos de fazer comandos se tornarem efetivamente funcionais em termos de Interação Homem Máquina (IHC), foi necessário escolher uma das opções

disponíveis para realizar tal tarefa. Entre todas as possibilidades de escolha, utilizamos o Coruja (LAPS, *online*, 2014a).

O Coruja, de acordo com os próprios desenvolvedores, “é um sistema de reconhecimento de voz com suporte a ditado” (LAPS, *online*, 2014a). Comparado com sistemas menos complexos, fundamentados em gramáticas, tem a vantagem de permitir o desenvolvimento de aplicativos como transcrição forense, criação automática de legenda entre outros. O motivo da utilização do Coruja é que ele se mostra eficaz para a aplicação em questão, opera em Português e está disponível para uso em pesquisa científica.

O Loquendo (NUANCE, *online*, 2014) foi empregado para teste em função da qualidade e acurácia da síntese de voz, tal como foi feito com o *Balabolka* - seção 3.1. Outros sintetizadores estão disponíveis, mas sua qualidade é baixa e isso implica diretamente o grau de *ambiência amigável* do aplicativo. Válido lembrar que ele não é gratuito e para versões futuras do Tical, em que estejam implicadas outras finalidades que não apenas a pesquisa científica, provavelmente será necessário usar um recurso próprio desenvolvido especificamente para esse fim.

O Coruja funciona a partir do Tical, ou seja, ao acionar o executável do robô clicando no ícone, automaticamente são acionados o banco de dados básico do Professor - contendo todas as respostas por escrito cadastradas - e o banco de dados de reconhecimento de voz - contendo as palavras-chave que compõem possíveis perguntas a Tical.

Para incrementar o banco de dados de reconhecimento de voz do Tical basta incluir a palavra, expressão ou frase junto à gramática do Coruja que, de imediato, passa a identificar os fonemas falados a partir dos elementos grafemáticos digitados. A Figura 16 a seguir apresenta alguns dos comandos inseridos e prontos para funcionar. Note-se que há, além das perguntas típicas feitas ao Tical como “qual o significado de atlas” ou “quem criou o alib”, elementos tematicamente isolados como as indicações vetoriais “frente”, “direita”, “esquerda”, “parar” e mesmo cores como “vermelho” entre outros.

```

grammar.xml - Bloco de notas
Arquivo  Editar  Formatar  Exibir  Ajuda
<?xml version="1.0" encoding="utf-8" ?>
<GRAMMAR>
  <RULE>
    <LIST>
      <P>FRENTE</P>
      <P>DIREITA</P>
      <P>ESQUERDA</P>
      <P>PARAR</P>
      <P>VERMELHO</P>
      <P>TICAL</P>
      <P>BUSCAR</P>
      <P>O PROFESSOR</P>
      <P>PROFESSOR</P>
      <P>QUAL O SIGNIFICADO DE ATLAS</P>
      <P>QUEM CRIOU O ALIB</P>
      <P>QUANDO FOI CRIADO O ALIB</P>
      <P>QUAL A FINALIDADE DO ALIB</P>
      <P>O OVO OU A GALINHA</P>
      <P>COMO CHAMAM A TANGERINA</P>
      <P>GEOSSOCIOLINGUISTICA</P>
      <P>SOCIOLINGUISTICA</P>
    </LIST>
  </RULE>
</GRAMMAR>

```

Figura 16 - Tela exibindo uma lista disponível no banco de dados

Após incluídos os itens na gramática, ele passa a reconhecer algumas pronúncias equivalentes. Para o caso de palavras específicas e diretamente relacionadas à área de atuação do Tical como ‘sociolinguística’ ou ‘geossociolinguística’, o que se espera é que sejam reconhecidas imediatamente e é o que acontece. Quando elas ocorrem na expressão que compõe a pergunta como “o que é sociolinguística?” ou “o que é geossociolinguística?”, cuidados como pronúncia mais empostada e pausada são necessárias para melhor o funcionamento.

Palavras não existentes oficialmente no léxico do Português Brasileiro também podem ser reconhecidas. Para tanto, basta proceder da mesma forma como é feito com palavras existentes adicionando-as na gramática do sistema tal como ilustra a Figura 16. ‘Paculafife’ e ‘mocolotate’, por exemplo, criadas especificamente para testar isso, apresentaram bons resultados.

Esse tipo de teste com palavras não existentes oficialmente não é fortuito. Nas diferentes áreas de conhecimento são recorrentes os verbetes que ainda não estão incorporados no léxico e que, portanto, não figuram nos dicionários de Português Brasileiro. As palavras “geolinguística” e “geossociolinguística”, por exemplo, até o fechamento deste texto, não estavam presentes nos dicionários institucionalizados mais conhecidos⁶¹. Isso não significa que não possam ser cadastradas no banco de dados de Tical ou que não possam ser reconhecidas grafematicamente e foneticamente pelas duas frentes do sistema – *processamento de linguagem escrita e processamento de linguagem falada*.

⁶¹ Mesmo considerando que praticamente não há dicionários cujos discursos não sejam institucionalizados, ou seja, reconhecidos como legítimos por nossas instituições (NUNES, 2002), salvo alguns glossários e listas vocabulares alternativos, todos têm limitações, entre elas, a impossibilidade de incorporar rapidamente todas as palavras novas mais utilizadas.

Os problemas dos ruídos de ambiência, entretanto, responsáveis por muitas falhas em dispositivos comandados por voz, precisam ser levados em consideração. Lugares com muita reverberação ou naturalmente ruidosos certamente acarretarão mau funcionamento do sistema, seja utilizando o microfone interno do computador ou um sistema externo acoplado aos fones de ouvido. É necessário também considerar as diferenças fisiológicas existentes entre pessoas: timbre, potência sonora, entonação e prosódia, entre outros, podem ser empecilhos ao bom funcionamento.

Desconsiderados esses obstáculos, justamente porque, para o teste inicial foram utilizadas apenas as vozes dos pesquisadores⁶² num ambiente relativamente silencioso, e ajustados os parâmetros comentados há pouco, o aplicativo funciona bastante bem.

Observando os resultados gerais dos primeiros testes, foi possível pensar em algumas soluções para que os comandos por voz pudessem operar também de modo parcial, com um pouco mais de flexibilidade. Nesse sentido, verificou-se que seria interessante o usuário poder escolher entre fazer a pergunta por escrito - utilizando o Processamento de Linguagem Escrita - ou falada - utilizando o Processamento de Linguagem Falada. Para tanto, foram consideradas quatro opções de utilização, descritas a seguir.

A primeira opção seria um ambiente naturalmente ruidoso como uma avenida, uma cantina, grupo de pesquisas ou uma oficina em que o usuário, impossibilitado de fazer a pergunta oralmente, escreve-a utilizando-se do teclado e recebe a resposta por escrito na tela. A primeira versão do Tical foi projetada para isso e não estão envolvidos aqui os comandos por voz.

A segunda opção leva em conta as mesmas condições da primeira, com a diferença de o usuário poder escolher o recurso de resposta ditada, ou seja, por síntese de voz. Há aí a conversão de linguagem escrita para falada ou conversão texto-fala - o TTS ou *text to speech* - e também não estão envolvidos aqui os comandos por voz.

Como terceira opção, o usuário está em sua residência pesquisando e, portanto, livre de quase todos os ruídos externos presentes em primeira e segunda opções. Então escolhe perguntar oralmente ao Tical, mas receber a resposta por escrito para que não tenha o trabalho de digitar ou transcrever. Há aí a conversão de linguagem falada para escrita ou conversão fala-texto - o STT ou *speech to text*.

⁶² O aporte linguístico para todas as versões de Tical foi dado pelo próprio autor da presente tese, mas o desenvolvimento ficou sempre a cargo do Professor Fabio Carlos Moreno, que também contribuiu com sua voz em muitos experimentos.

Assim, a quarta opção desconsidera o texto escrito para pergunta ou resposta. O usuário faz a pergunta oralmente e recebe a resposta também oralmente. O que ocorre aqui é o reconhecimento de voz capaz de atender ao comando e trazer como resposta um texto em síntese de voz. Guardadas as devidas proporções e o profundo fosso existente entre a realidade e conceitos imagísticos, os recursos do aplicativo chegam bem próximos àqueles dos filmes de ficção científica com a vantagem de lidar tão somente com o Português Brasileiro e falar do nosso próprio atlas linguístico, o ALiB.

Essas quatro opções, no entanto, ainda não estão disponíveis em conjunto e serão observadas com mais critério em momento oportuno, quando do desenvolvimento de um estudo especificamente para isso.

No panorama apresentado, atingir o objetivo de apresentar o robô de conversação Tical numa dimensão mais ampla, ou seja, com todas as funcionalidades que a implementação tornou possíveis até o fechamento deste texto, foi uma tarefa árdua, não porque tenha sido necessário começar do zero ou porque faltassem ferramentas ou conteúdo, mas pelo fato de ter sido realizada basicamente a quatro mãos: não havia uma equipe numerosa.

O esforço, no entanto, foi recompensado. Nessa versão, Tical mostra-se como um projeto bastante relevante em termos científicos justamente por concentrar vários diferenciais já comentados aqui: é um dos únicos robôs que ‘versam’ sobre Linguística, está todo em Português Brasileiro, opera em dispositivos móveis a partir de aplicativo de mensagem instantânea, possui síntese de voz e atende a comandos por voz – esses dois últimos recursos ainda não operam no modo on-line.

Por esses motivos, seu grau de aplicabilidade é bastante alto e o trabalho tornou-se, ao mesmo tempo, projeto e objeto de pesquisa. Em outras palavras, o trabalho serviu não apenas para aplicar teorias das áreas relacionadas mas também para testar a funcionalidade de comandos por voz com Português Brasileiro e conteúdo afim: o ALiB.

Reconhecidas plenamente as funcionalidades de Tical após essa fase prototipal, resta dar ênfase e continuidade ao projeto. Para tanto será necessária uma equipe técnica que pode constituir-se de um grupo de pesquisa interdisciplinar - ou vários deles - e muito trabalho. Grupos de pesquisa devidamente cadastrados junto a órgãos de fomento vislumbram possibilidade de apoio financeiro às pesquisas, sem o qual, historicamente, muitas boas ideias tiveram sua viabilidade comprometida e permanecem apenas no papel.

Independentemente da existência de grupos ou equipes vindouros, o estudo já serviu para comprovar a dimensão funcional de Tical enquanto conceito, para permitir a utilização de conhecimentos teóricos adquiridos e para projetar a partir dos resultados obtidos alguns

outros caminhos possíveis: Tical pode ainda operar com recursos de Realidade Virtual e/ou Aumentada e uma versão holográfica do docente cibernético seria apenas mais alguns passos adiante.

5 CONSIDERAÇÕES FINAIS

A “Avaliação de dispositivos acionados por voz e texto para o Português Brasileiro” pôde ser verificada linguisticamente e esse foi o objetivo geral do trabalho.

Pode-se afirmar também que os objetivos específicos foram atingidos uma vez que a Tese (i) traz informações que podem auxiliar profissionais e pesquisadores da área de Linguística Geral a aplicar de modo mais pontual conhecimentos sobre Fonética e Fonologia às tecnologias vindouras de comando e/ou reconhecimento de voz. Da mesma forma (ii) possibilita aproximação de pesquisadores das áreas de Humanas com os da área das Exatas – Tical só foi possível devido às parcerias interdisciplinares já comentadas.

Critérios mais apurados levam a resultados mais consistentes - como em toda e qualquer avaliação - e, para conseguir estabelecer tais critérios, recorreu-se a uma base teórica adequada, pertinente e interdisciplinar, tal como comentado no capítulo 1.

A hipótese era de que apenas uma pequena parte de todos os equipamentos existentes ou disponíveis poderia ser acionada de modo eficiente por qualquer falante do Português Brasileiro. Tal hipótese foi parcialmente comprovada porque restaria testar equipamentos aos quais não há acesso, como os de uso corporativo ou militar.

As perguntas geradas por essa hipótese puderam ser respondidas com uma boa margem de precisão.

Para a pergunta (I), “os equipamentos que operam com comandos por voz funcionam bem para qualquer variedade linguística no Brasil?”, verificou-se que equipamentos com esse recurso não funcionam bem para qualquer variedade. A maior parte dos falantes do Português Brasileiro não tem como acessar tais equipamentos de modo legitimamente espontâneo, natural e/ou sem monitoramento na fala. A variante do próprio pesquisador não lhe deu acesso preciso a vários comandos durante os experimentos. As simulações de variantes também falharam e o mesmo pode ser dito quando da utilização das gravações do ALiB, cujos resultados foram tabulados na Tabela 06.

Quanto ao questionamento (II), “quais são os critérios linguísticos adotados pelos desenvolvedores desses equipamentos?”, pode-se afirmar que há uma quantidade muito pequena de desenvolvedores que trabalha com assessoria de linguistas, ou seja, a maioria esmagadora opera com empirismos e critérios discutíveis. O principal motivo que leva a essa resposta parece bastante óbvio: tal como apontaram os experimentos ao longo de todo capítulo 4, qualquer linguista minimamente atento teria identificado as falhas ainda na etapa

de desenvolvimento. Outro motivo é que citações sobre elementos de Sociolinguística ou Fonética/Fonologia são raras nas obras de desenvolvedores, tanto no Brasil como no exterior.

Em relação à pergunta (III), “o que ocorre de fato é reconhecimento de voz ou outro tipo de processamento de sinais de áudio?”, não restaram dúvidas de que, praticamente todos os equipamentos comerciais conhecidos e disponíveis à população operam com reconhecimento de voz básico e sistemas de processamento de áudio similares entre si. Em outras palavras, eles falham em contextos muito parecidos. Alguns sistemas mais modernos que contam, não apenas com um banco de dados lexical bastante amplo, mas também com um recurso de atualização de dados constante, como é o caso do buscador por voz da Google, parecem operar muito bem, mesmo com expressões inteiras, mas isso não significa que tenha capacidade de *compreensão de voz*.

Considerando que não há como oferecer respostas curtas para perguntas complexas, embora os resultados dos experimentos estejam todos à disposição, ora diluídos, ora concentrados, ao longo do trabalho, as respostas contidas nestas Considerações Finais, ainda assim, sintetizam a essência de todas as discussões feitas até o presente momento.

Tal como divulgado no capítulo 1, a essência do trabalho consiste em experimentos feitos com três modalidades distintas de dispositivos. A Tabela 26, baseada na Tabela 01, atualiza essa informação e adiciona Tical como o quarto dispositivo experimentado.

Tabela 27 - Equipamentos analisados e Tical

	Crítérios	IBM Via Voice	V2 Voice Recognition Module	Buscador por voz da Google	Tical
(i)	compatibilidade com o idioma	possui suporte para Português Brasileiro	suporte para voz humana em geral	‘reage’ bem ao Português Brasileiro	suporte para a fonologia do Português
(ii)	grau de difusão entre usuários	é um clássico (ocidental) para edição de textos	bastante difundido na área técnica	mundialmente conhecido por usuários na internet	ainda não avaliável
(iii)	banco de dados atualizável	possui léxico básico integrado manipulável	não possui banco de dados integrado mas é manipulável	possui amplo léxico integrado não manipulável	capacidade de, teoricamente, operar com léxico ilimitado
(iv)	grau de utilização atual	ainda utilizado por projetistas e estudantes para Domótica e robótica	amplamente utilizado por projetistas e hobistas para Domótica e robótica	abrange a maior quantidade e variedade de equipamentos disponíveis na ocasião da pesquisa	ainda não avaliável

Fonte: Texto organizado pelo autor

Tal como anunciado no capítulo 1, abriu-se a oportunidade de incrementar todos esses experimentos com dados do ALiB de diversas ordens e esse procedimento suplantou as expectativas linguísticas da abordagem: diminuiu-se a aleatoriedade da aplicação e mostrou-se uma nova dimensão para utilização de dados já pesquisados.

Como previsto na Introdução, a comprovação da hipótese naturalmente imprime uma poderosa inércia de movimento em que divulgar e aplicar outras áreas de conhecimento aos estudos já existentes sobre a Geolinguística do Brasil torna-se inevitável. A criação em dezembro de 2014 do Grupo de Pesquisas interdisciplinar denominado *Estudos dos Sinais da Fala do Português Brasileiro* junto ao CNPq e vinculado ao PPGEL-UEL é, tranquilamente, mais uma amostra de que isso é tanto possível quanto salutar à condução e à continuidade de todas as pesquisas relacionadas. Uma previsão nesse contexto é de que, entre outras coisas, o protótipo Tical não só evolua para algo mais sofisticado e abrangente em termos interdisciplinares, como também sirva de guia e base para pesquisadores vindouros.

REFERÊNCIAS

- 2001, UMA ODISSÉIA no Espaço. Direção: Stanley Kubrick. Produção: Stanley Kubrick. Intérpretes: Keir Dullea, Gary Lockwood, William Sylvester, Douglas Rain e outros. Roteiro: Stanley Kubrick e Arthur C. Clarke. Música: Johann Strauss II e Richard Strauss. 1968. Califórnia.
- AGUILERA, Vanderci de Andrade (org.) *A Geolingüística no Brasil: trilhas seguidas, caminhos a percorrer*. Londrina: EDUEL, 2005.
- AGUILERA, Vanderci de Andrade. *Aspectos Linguísticos da Fala Londrinense*. Vol. 1. Dissertação (Mestrado em Letras). 229 f. Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP). Assis: 1987.
- AGUILERA, Vanderci de Andrade. *Atlas Lingüístico do Paraná*. 1990. v. 1. Tese (Doutorado em Letras) – Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Assis: 1990.
- AGUILERA, Vanderci de Andrade. *Atlas Lingüístico do Paraná*. 1990. v. 2. Tese (Doutorado em Letras) – Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Assis: 1990.
- AGUILERA, Vanderci de Andrade; ALTINO, Fabiane Cristina. *Por um atlas pruridimensional: pesquisas e pesquisadores*. *Revista Alfa*, n. 56. São Paulo, 2012.
- ALKMIM, Tânia. *Sociolingüística* (parte I). In: MUSSALIM, Fernanda, BENTES, Anna Christina. (orgs.). *Introdução à lingüística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.
- ALMEIDA, Maurício B. ; BAX, Marcelo P. *Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção*. In *Revista Ci. Inf.*, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019.pdf>>. Acesso em: 10 jan. 2016.
- ALTENHOFEN, C. V.; KLASSMANN, M. S. (orgs.) *Atlas Linguístico-Etnográfico da Região Sul do Brasil – ALERS: cartas fonéticas e morfossintáticas*. Porto Alegre: Editora da UFRGS; Florianópolis: Ed. UFSC, 2011.
- ALTINO, Fabiane Cristina (org.) *Múltiplos Olhares sobre a diversidade lingüística: nos caminhos de Vanderci Andrade Aguilera*. Londrina: Mioograf, 2012.
- ALTINO, Fabiane Cristina. *Atlas Linguístico do Paraná – ALPR II*. Tese (Doutorado em Estudos da Linguagem). Centro de Letras e Ciências Humanas. Programa de Pós-Graduação em Estudos da Linguagem. 182 f. Universidade Estadual de Londrina, Londrina, 2007.
- ANDROUTSOPOULOS, I.; RITCHIE, G. D.; THANISCH, P. *Natural Language Interfaces to Databases: An Introduction*. *Journal of Natural Language Engineering*, Cambridge University Press. 1994.
- ARAGÃO, Maria do Socorro Silva de. *A Neutralização Dos Fonemas / v – z - Z / no falar de Fortaleza*. In PROFALA - Variação e Processamento da Fala e do Discurso Análises e Aplicações. Disponível em: <<http://www.profala.ufc.br/index.html>>. Acesso em: 22 mar. 2015.

ARARIBÓIA, G. *Inteligência Artificial*. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda., 1988.

BABINI, Maurizio; MARRANGHELLO, Norian. *Introdução às redes neurais artificiais*. São Paulo: Cultura Acadêmica; São José do Rio Preto: IBILCE, UNESP, 2007.

BALABOLKA. Disponível em: <<http://www.cross-plus-a.com/br/balabolka.htm>>. Acesso em: 13 jan. 2012.

BALIEIRO Jr, Ari Pedro. Psicolinguística. In MUSSALIM, Fernanda, BENTES, Anna Christina. (orgs.). *Introdução à lingüística II: domínios e fronteiras*, v.2. São Paulo: Cortez, 2001.

BARBOSA, Cinthyan Renata Sachs Camerlengo de. *Técnicas de parsing para gramática livre de contexto lexicalizada da língua Portuguesa*. (Doutorado em Engenharia Eletrônica e Computação) – Curso de Engenharia Eletrônica e Computação – Instituto Tecnológico de Aeronáutica. São José dos Campos, 2004. 171 f.

BAUER, Wolfgang; WESTFALL, Gary D.; DIAS, Helio. *Física para Universitários*. São Paulo: AMGH, 2013.

BLADE Runner. Direção: Ridley Scott. Produção: Michael Deeley. Intérpretes: Harrison Ford, Rutger Hauer, Sean Young, Edward James Olmos, Daryl Hannah e Brion James J. J. Johnson. Roteiro: Hampton Fancher, David Peoples. Música: Vangelis. 1982. EUA.

BOLZANI, Caio Augustus Morais. *Residências inteligentes: Domótica, redes domésticas, automação residencial*. São Paulo: Editora Livraria da Física, 2004.

BORBA, Francisco da Silva. *Introdução aos estudos linguísticos*. 11 ed. Campinas, SP: Pontes, 1991.

BRANDÃO, Silvia F. *A geografia lingüística no Brasil*. São Paulo: Ática, 1991.

BRASIL. Congresso. Senado. Projeto de lei no 64, de 04 de março de 2009. **Diário do Senado Federal**, Poder Executivo, Brasília, DF, 04 março. 2009. Seção 1, pp. 03547-03550.

CAGLIARI, Gladis Massini; CAGLIARI, Luiz Carlos. Fonética In MUSSALIM, Fernanda; BENTES, Anna Christina. (orgs.). *Introdução à lingüística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.

CALLOU, D.; LEITE, Y. *Iniciação à fonética e à fonologia*. Rio de Janeiro: Jorge Zahar, 2001.

CALVET, Louis-Jean. *Sociolinguística: uma introdução crítica*. Tradução Marcos Marcionilo. São Paulo: Parábola, 2002.

CAMACHO, Roberto Gomes. *Sociolinguística (parte II)* in MUSSALIM, Fernanda; BENTES, Anna Christina. (orgs.). *Introdução à lingüística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.

CAMARA JR, Joaquim Mattoso. *Dicionário de Língua e Gramática: referente à Língua Portuguesa*. 10 ed. Petrópolis: Vozes, 1981.

CAMARA JR, Joaquim Mattoso. *Estrutura da língua portuguesa*. 13 ed. Petrópolis, J: Vozes, 2004.

CAMARA JR, Joaquim Mattoso. *Para o estudo da fonêmica Portuguesa*. Rio de Janeiro: Organização Simões, 2006.

CARDOSO, Suzana Alice Marcelino da Silva et al. *Atlas Linguístico do Brasil: Introdução*. Vol. 1. Londrina: Eduel, 2014a.

CARDOSO, Suzana Alice Marcelino da Silva et al. *Atlas Linguístico do Brasil: Cartas Linguísticas I*. Vol. 2. Londrina: Eduel, 2014b.

CARDOSO, Suzana Alice Marcelino. O Atlas Linguístico do Brasil: um projeto nacional. In: AGUILERA, Vanderci de Andrade (org.). *A Geolinguística no Brasil: caminhos e perspectivas*. Londrina: EDUEL, 1998.

CARUSO, Pedro. *Atlas Linguístico do Estado de São Paulo*. Assis: Unesp, 1983.

CASTILHO, Ataliba Teixeira de. *Variação dialetal e ensino Institucionalizado da língua portuguesa*. In: BAGNO, Marcos (org.). *Linguística da Norma*. São Paulo: Loyola, 2004.

COMITÊ Nacional do Projeto ALiB. *Atlas Linguístico do Brasil: questionário 2001/Comitê Nacional do Projeto ALiB*. - Londrina: Editora UEL, 2001.

CONPET - Programa Nacional da Racionalização do uso dos derivados do petróleo e do gás natural. *Robô Ed*. Disponível em: <<http://www.ed.conpet.gov.br/br/converse.php>>. Acesso em: 24 abr. 2014.

COSERIU, Eugênio. A geografia linguística. In: COSERIU, Eugênio. *O homem e a sua imagem*. Rio de Janeiro: Presença; São Paulo: USP, 1982.

CRYSTAL, David. *Dicionário de Linguística e Fonética*. Tradução Maria Carmelita Pádua Dias. Rio de Janeiro: Jorge Zahar Editor, 1988.

CUNHA, C.; CINTRA I. *Nova gramática do Português*. Rio de Janeiro: Nova Fronteira, 2001.

CUNHA, Celso. *Língua Portuguesa e realidade brasileira*. 3. Ed. Rio de Janeiro, Tempo Brasileiro, 1972.

DAMASCENO, Eduardo Filgueiras. *Avaliação das bibliotecas de reconhecimento e síntese de fala em ambientes virtuais*. 2005. 129 f. Dissertação (Mestrado em Ciência da Computação) – Centro Universitário Eurípedes de Marília - UNIVEM. Marília, 2006. Disponível em: <http://www.fundanet.br/servico/aplicativos/mestrado_dir/dissertacoes/Avalia%E7%E3o_da_bibliotecas_de_reconhecimento_e_s%EDntese_de_f_52_pt.pdf>. Acesso em 12 fev. 2012.

DAMASCENO, Eduardo Filgueiras; PEREIRA, Tatiane Valau; BREGA, José Remo Ferreira. *Implementação de serviços de voz em ambientes virtuais*. Artigo. Disponível em: <<http://www.dcc.ufla.br/infocomp/artigos/v4.3/art09.pdf>>. Acesso em 12 fev. 2012.

DICIONÁRIO Aurélio Eletrônico da Língua Portuguesa. 1 CD. 2014.

DUBOIS, Jean et al. *Dicionário de Linguística*. ed. 9. São Paulo: Cultrix, 2004.

DUCROT, Oswald; TODOROV, Tzvetan. *Dicionário Enciclopédico das Ciências da Linguagem*. 3 ed. São Paulo: Perspectiva, 2001.

ELECHOUSE. Disponível em: <http://www.elechouse.com/elechouse/index.php?main_page=product_info&cPath=168_170&products_id=2151>. Acesso em: 02 jul. 2014.

FARACO, Carlos Alberto. *Norma-padrão brasileira: desembaraçando alguns nós*. In BAGNO, Marcos (org.). *Linguística da Norma*. São Paulo: Loyola, 2004.

FILGUEIRAS, Miguel. Interfaces de Linguagem Natural. In: MATEUS, Maria Helena; BRANCO, António Horta (orgs). *Engenharia da Linguagem*. Lisboa: Edições Colibri, 1995.

HAGÈGE, Caroline; DUARTE, Inês. Construção de gramáticas formais para o Processamento de Linguagem Natural. In: MATEUS, Maria Helena; BRANCO, Antonio Horta (orgs). *Engenharia da Linguagem*. Lisboa: Edição Colibri, 1995.

HAWKING, Stephen. *O universo numa casca de noz*. Tradução Monica Gagliotti Fotunato Friaça. São Paulo: Ediouro, 2009.

HUGO, Marcel. *Uma interface de reconhecimento de voz para o sistema de gerenciamento de central de informação de fretes*. 1995. 60 f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Estadual de Santa Catarina. 1996. Disponível em: <<http://www.tede.ufsc.br/teses/PEPS0413-D.pdf>>. Acesso em 14 fev. 2012.

HUIZINGA, Johan. *Homo Ludens: o jogo como elemento da cultura*. 4 ed. São Paulo: Perspectiva, 2007.

IBM Watson. *What is Watson*. Disponível em: <<http://www.ibm.com/smarterplanet/us/en/ibmwatson/what-is-watson.html>>. Acesso em: 13 jan. 2015.

IBM. Disponível em: <<http://www-01.ibm.com/software/pervasive/viavoice.html>>. Acesso em: 03 jun. 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Estimativas da população residente no Brasil e unidades da Federação com data de referência em 1º de julho de 2014*. Disponível em: <ftp://ftp.ibge.gov.br/Estimativas_de_Populacao/Estimativas_2014/estimativa_dou_2014.pdf>. Acesso em: 03 jan. 2015.

JOTA, Zélio dos Santos. *Dicionário de Linguística*. 2 ed. Rio de Janeiro: Presença, 1981.

JURAFSKY, Daniel; MARTIN, James H. *Speech and language processing: an introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Englewood Cliffs, New Jersey, 1999.

KAILER, Dircel. Alçamento da vogal pretônica /o/ em duas regiões paranaenses. In *Revista Signum: Estudos Linguísticos*. Londrina n. 15/1, p. 201-221, jun. 2012.

KOCH, Walter; KLASSMANN, M. S.; ALTENHOFEN, Cléo Vilson. *Atlas Linguístico-etnográfico da Região sul do Brasil – ALERS*. Porto Alegre/Florianópolis/Curitiba: UFRGS/Ed. UFSC/Ed. UFPR, 2002.

LABOV, William. *Padrões sociolingüísticos*. São Paulo: Parábola Editorial, 2008.

LAPS - Laboratório de Processamento de Sinais. Disponível em: <<http://www.laps.ufpa.br/falabrasil/descricao.php>>. Acesso em: 11 out. 2014a.

LOPES, Edward. *Fundamentos da Linguística contemporânea*. 17 ed. São Paulo: Cultrix, 1999.

LYONS, J. *Língua(gem) e lingüística*. (Trad. de Marilda Averborg Winckler et alii). Rio de Janeiro: Pontes, 1987.

MANFIO, Edio Roberto. *Aspectos fonéticos no DosVox enquanto aplicativo tipo texto-fala*. In: *XXI Seminário do CELLIP - Paranaguá*, 2013. Disponível em: <<http://cellip.files.wordpress.com/2013/08/anais-do-xxi-cellip.pdf>>. Acesso em: 21 jun. 2014a.

MANFIO, Edio Roberto. Como funcionam alguns fonemas no aplicativo *Balabolka*. In: *Revista Via Litterae*. Anápolis, v. 4, n. 2, p. 191-204, jul./dez. 2012. Disponível em: <www2.unucseh.ueg.br/vialitterae>. Acesso em: 21 jun. 2014b.

MANFIO, Edio Roberto. Processamento de Linguagem Natural, Processamento de Sinais da Fala, Geolinguística e um Naco de Humor. In: *Anais do X Seminário de Iniciação Científica Estudos Linguísticos e Literários - Sóletras*. UENP – Jacarezinho, 2013. Disponível em: <http://www.cj.uenp.edu.br/index.php/institucional/eventos/1-soletras/event_details>. Acesso em: 21 jun. 2014c.

MANFIO, Edio Roberto; MORENO, Fabio Carlos; BARBOSA, Cinthyan Renata Sachs Camerlengo de. Tecnologia Interativa Conversacional sobre Assuntos Linguísticos - Tical: Linguagem e Significação. In: IX Seminário de Estudos sobre Linguagem e Significação e X Simpósio de Leitura da UEL "Convenções e Ousadias da Linguagem". **Caderno de Resumos**. Londrina: UEL, 2014a, p. 54-55.

MANFIO, Edio Roberto; MORENO, Fabio Carlos; BARBOSA, Cinthyan Renata Sachs Camerlengo de. Professor Tical: Robô de Conversação sobre Dialetologia e Geossociolinguística. In: III CIDS - Congresso Internacional de Dialetologia e Sociolinguística – Variação, Atitudes linguísticas e Ensino. 2014. Londrina, UEL. **Caderno de Resumos**. ISBN: 978 85 7846 297 0. Londrina, 2014b. p. 48.

MUSSALIM, Fernanda, BENTES, Anna Christina. (orgs.). *Introdução à lingüística I: domínios e fronteiras*, v.1. São Paulo: Cortez, 2001.

MUSSALIM, Fernanda, BENTES, Anna Christina. (orgs.). *Introdução à lingüística II: domínios e fronteiras*, v.2. São Paulo: Cortez, 2001.

NASCENTES, Antenor. *O idioma nacional*. 3 ed. Rio de Janeiro: Livraria Acadêmica, 1960.

NUANCE – Loquendo. Disponível em: <<http://www.nuance.com/for-business/by-solution/customer-service-solutions/solutions-services/inbound-solutions/loquendo-small-business-bundle/interactive-tts-demo/index.htm>>. Acesso em: 18 out. 2014.

NUANCE. Disponível em: <<http://www.nuance.com/>>. Acesso em 03 jun. 2013.

NUNES, José Horta. *Dicionarização no Brasil: condições e processos*. In: NUNES, José Horta; PETTER, Margarida (orgs.). *História do saber lexical e a construção de um léxico brasileiro*. São Paulo: Humanitas / FFLCH / USP: Pontes, 2002.

ORLANDI, E. P. *Língua e conhecimento lingüístico: para uma história das idéias no Brasil*. São Paulo: Cortez, 2002.

PREECE, Jennifer; YVONNE, Rogers; SHARP, Helen. *Design de interação: além da interação homem-computador*. Tradução de Viviane Possamai. Porto Alegre: Bookman, 2005.

PROJETO Atlas Linguístico do Brasil. Disponível em: <<http://twiki.ufba.br/twiki/bin/view/Alib/WebHome>>. Acesso em: 02 fev. 2013.

PROJETO DOSVOX. Disponível em: <<http://intervox.nce.ufrj.br/dosvox/>>. Acesso em: 28 mar. 2012.

RECTOR, Mônica. *A linguagem dos jovens*. Petrópolis: Vozes. 1975.

RICH, Elaine. *Inteligência Artificial*. Tradução Maria Cláudia Santos Ribeiro Ratto. São Paulo: Makron Books, 1993.

ROMANO, Valter Pereira. Balanço crítico da Geolinguística brasileira e a proposição de uma divisão. In: *Revista Entretextos*. Londrina, v. 13. jul/dez 2013. Disponível em: <<http://www.uel.br/revistas/uel/index.php/entretextos/article/view/16388/13897>>. Acesso em 23 dez. 2015.

ROTHERMEL Alessandra. *Maria: Um chatterbot desenvolvido para os estudantes da disciplina "Métodos e Técnicas de Pesquisa em Administração"*. In: Simpósio de Excelência em Gestão e Tecnologia. Disponível em: <http://www.aedb.br/seget/artigos07/1429_artigos2007eget.pdf>. Acesso em: 21 jun. 2014.

SCHILDT, Herbert. *Inteligência Artificial utilizando linguagem C*. Tradução Cláudio Gaiger Silveira e Mônica Soares Rufino. São Paulo: McGraw-Hill, 1989.

SERASA Experian. *Google Brasil lidera buscas em junho, segundo Hitwise*. Disponível em: <<http://noticias.serasaexperian.com.br/google-brasil-lidera-buscas-em-junho-segundo-hitwise/>>. Acesso em: 29 jan. 2015.

SIL INTERNATIONAL. Disponível em: <<http://www.sil.org/>>. Acesso em: 03 jun. 2012.

SILVA NETO, Serafim da. *Guia para estudos dialetológicos*. 2 ed. Belém: Conselho Nacional de Pesquisa: Instituto Nacional da Amazônia, 1958.

SILVA, Thaís Cristófar. *Fonética e Fonologia do Português: roteiro de estudos e guia de exercícios*. São Paulo: Contexto, 1999.

SILVEIRA, Sérgio Amadeu da. *Software livre: a luta pela liberdade do conhecimento*. São Paulo: Editora Fundação Perseu Abramo, 2004.

SONY (Recort. Edit. Produce. The complete Sound Forge Family of applications). Disponível em: <<http://www.sonycreativesoftware.com/soundforgesoftware/>>. Acesso em 10 fev. 2013.

STAR Trek. Série. 50 min. Criador: Gene Roddenberry. Produtores: Gene Roddenberry, Gene L. Coon, John Meredyth Lucas, Fred Freiberger. Estados Unidos, Setembro de 1966.

STARMUNDO. *Sistemas operacionais que existem hoje para smartphones*. Disponível em: <<http://smartmundo.com/sistemas-operacionais-para-smartphones/>>. Acesso em: 29 jan. 2015.

TARALLO, Fernando. *A pesquisa sociolinguística*. 7. ed. Série Princípios. São Paulo: Ática, 2001.

TEVAH, Rafael Teruszkin. *Implementação de um sistema de reconhecimento de fala contínua com amplo vocabulário para o Português Brasileiro*. 2006. 91 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2007. Disponível em: <<http://www.pee.ufrj.br/teses/textocompleto/2006053001.pdf>>. Acesso em 22 mar. 2012.

TRUDGILL Peter. *Sociolinguistics: an introduction*. Penguin Books: 1974.

VICCARI, Rosa Maria. *Tutor inteligente para a programação em lógica: idealização, projeto e desenvolvimento*. 1989. 462 f. Tese (Doutoramento em Engenharia Electrotécnica) Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Coimbra, Portugal, 1989. Disponível em: <http://www.lume.ufrgs.br/bitstream/handle/10183/17971/000031255.pdf?sequence=1&locale=pt_BR>. Acesso em: 12 fev. 2015.

VOICE Recognition Module V2 Manual. Disponível em: <<http://www.elehouse.com/elehouse/images/product/Voice%20Recognition%20Module/Manual.pdf>>. Acesso em: 02 jul. 2014.

ZIVIANI, Nivio. *Projeto de algoritmos: com implementação em Pascal e C*. 4. ed. São Paulo: Pioneira, 1999.

WEIZENBAUM J. *ELIZA - A computer program for the study of natural language communication between man and machine*. Communications of the ACM, v.9, n.1, p. 36-45, jan. 1966.

WHATSAPP. Disponível em: <<http://www.WhatsApp.com/>>. Acesso em: 17 out. 2014.

WOODS, Willian Aaron. *Progress in natural language understanding: An application to lunar geology*. Bolt Beranek and Newman Inc. Cambridge, Massachussets, 1972.

YNOGUTI, Carlos Alberto. *Reconhecimento de Fala Contínua Usando Modelos Ocultos de Markov*. 1999. 138 f. Tese (Doutorado em Engenharia Elétrica) – Unicamp. Campinas, 2000. Disponível em: <http://www.decom.fee.unicamp.br/lpdf/teses_pdf/Tese-Doutorado-Carlos_Alberto_Ynoguti.pdf>. Acesso em 16 mar. 2012.

ANEXOS

ANEXO A

Endereços Eletrônicos dos Arquivos de Áudio Reproduzidos pelo *Balabolka*

I- ‘Nós estamos utilizando o Balabolka hoje’:

<https://www.dropbox.com/s/0kn3zpixhmljcmv/n%C3%B3s%20estamos.wav?m>

II- ‘O *Balabolka*, que é um programa de fácil utilização, mostra-se bastante útil a diversas pessoas’:

<https://www.dropbox.com/s/0p0vrcbtai65fts/o%20Balabolka.wav?m>

III- ‘Um usuário comum (...)’:

<https://www.dropbox.com/s/5fnzticqivxte1t/um%20usu%C3%A1rio%20comum.wav?m>

IV- ‘Ele virá aqui hoje’:

<https://www.dropbox.com/s/uuoagk7vxjci5kx/Ele%20vir%C3%A1%20aqui%20hoje%20%282x%29.wav?m>

V- Arara (...):

<https://www.dropbox.com/s/bv0wcrub0ugr1i7/arara.wav?m>

VI- Barato (...):

<https://www.dropbox.com/s/urnb6rigtyv1o3z/barato.wav?m>

VII- Rata (...):

<https://www.dropbox.com/s/xdrx8zp3yyhevu3/rata.wav?m>

VIII- Celular (...):

<https://www.dropbox.com/s/66hwy9s5bgs8sf4/celular.wav?m>

VIII- ‘O mar está calmo’: