



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM LINGUÍSTICA

O USO DE REDES NEURAIS AUTO-ORGANIZÁVEIS PARA A
ANÁLISE DO CONHECIMENTO ACENTUAL EM
APRENDIZES BRASILEIROS DE LÍNGUA INGLESA

ANA CRISTINA CUNHA DA SILVA

FORTALEZA - CEARÁ
Outubro/2010

ANA CRISTINA CUNHA DA SILVA

O USO DE REDES NEURAIS AUTO-ORGANIZÁVEIS PARA A
ANÁLISE DO CONHECIMENTO ACENTUAL EM
APRENDIZES BRASILEIROS DE LÍNGUA INGLESA

Tese submetida à Coordenação do Curso de Pós-Graduação em Linguística da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Doutor em Linguística.

Orientadora: Prof^a. Dr^a. Ana Cristina Pelosi Silva de Macedo.

Co-Orientador: Prof. Dr. Guilherme de Alencar Barreto

FORTALEZA
2010


Esta tese foi submetida ao Programa de Pós-Graduação em Linguística como parte dos requisitos necessários para a obtenção do grau de Doutor em Linguística, outorgado pela Universidade Federal do Ceará, e encontra-se à disposição dos interessados na biblioteca de humanidades da referida Universidade.

A citação de qualquer trecho desta tese é permitida, desde que seja feita de acordo com as normas científicas.

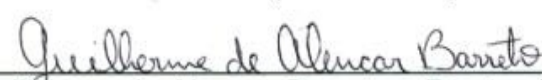
Tese aprovada em ___ / ___ / ____.

Ana Cristina Cunha da Silva

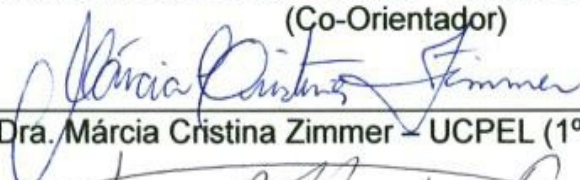
BANCA EXAMINADORA



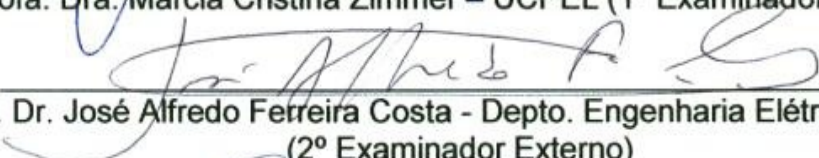
Profa. Dra. Ana Cristina Pelosi S. de Macedo – Depto. de Letras Vernáculas - UFC
(Orientadora)



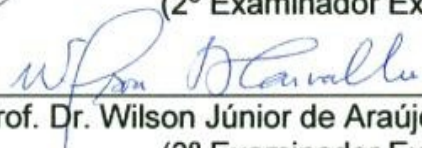
Prof. Dr. Guilherme de Alencar Barreto – Depto. de Teleinformática – UFC
(Co-Orientador)



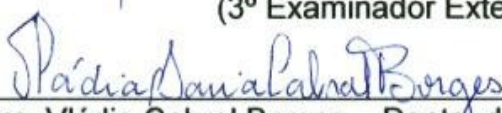
Profa. Dra. Márcia Cristina Zimmer – UCPEL (1º Examinador Externo)



Prof. Dr. José Alfredo Ferreira Costa - Depto. Engenharia Elétrica – UFRN
(2º Examinador Externo)



Prof. Dr. Wilson Júnior de Araújo Carvalho – UECE
(3º Examinador Externo)



Profa. Dra. Vlândia Cabral Borges – Depto. de Letras Estrangeiras - UFC
(Examinador Interno)

Profa. Dra. Emília Maria Peixoto Farias – Depto. de Letras Estrangeiras - UFC
(Suplente Interna)

Profa. Dra. Paula Lenz Costa Lima - UECE (Suplente Externo)

*Dedico este trabalho a todos os meus alunos,
por eles serem a minha maior motivação!*

AGRADECIMENTOS

A Deus, por ter me dado saúde, fé e coragem para eu alcançar mais este objetivo em minha jornada.

À orientadora e também amiga Ana Cristina Pelosi Silva de Macedo, por sua paciência, motivação, pelos conselhos brilhantes e por sempre ter acreditado no meu potencial e determinação.

Ao meu co-orientador, Guilherme de Alencar Barreto, pelo espaço de estudo concedido, por sua atenção, supervisão, visão inteligente e objetiva, e por sua total disposição em sempre me ajudar.

À FUNCAP e à CAPES-REUNI, pelas bolsas de estudo concedidas ao longo do percurso de três anos e meio de curso.

À Universidade Estadual do Piauí, instituição onde leciono, pelo tempo de afastamento concedido para a finalização deste trabalho.

Ao Programa de Pós-Graduação em Lingüística da Universidade Federal do Ceará, representado pela dedicada coordenadora Mônica Cavalcante e pelos funcionários Antônia e Eduardo. Muito obrigada pela atenção e carinho.

Aos alunos da Casa de Cultura Britânica – UFC, por gentilmente concederem entrevistas para a composição do banco de dados para esta tese.

A todos os meus colegas do Programa de Pós-Graduação em Lingüística que estiveram ao meu lado nos momentos difíceis, especialmente a Geórgia Maria Feitosa. Obrigada pela amizade sincera.

A José Álber Campos Uchoa, pelas valiosas discussões sobre Fonética e Fonologia e também a Otaciano Noronha Filho pela contribuição nas transcrições fonéticas.

Aos colegas do Laboratório de Teleinformática (LATIN), especialmente a Amauri Holanda, pelas dicas e suporte técnico.

À Hylnara Salatiel Bezerra de Menezes, pela presença marcante na reta final deste trabalho.

A Raimundo Isídio de Sousa e Joselena Dourado, pela revisão do texto final.

E por último, todavia não menos importante, agradeço a todos que contribuíram de forma direta ou indireta para que eu pudesse realizar esta pesquisa de forma motivada e com paz de espírito suficiente. Agradeço-lhes sinceramente.

“Assim como são os fenômenos lingüísticos são também as relações humanas: caóticas e, ao mesmo tempo, auto-organizáveis.” (Ana Cristina Cunha da Silva)

“...over and over again nature uses the same principles of self-organization to produce dynamic patterns on all scales.” (J. A. Scott Kelso)

RESUMO

O objetivo desta tese foi investigar como o conhecimento prosódico está organizado em um estágio inicial de aquisição de L2 em aprendizes brasileiros de inglês com a ajuda de uma rede neural conexionista. A abordagem proposta neste trabalho consiste primeiramente em "quantificar" as elocuições dos aprendizes de L2 na forma de coeficientes LPC e outras características linguísticas/fonéticas que possam representar o fenômeno aqui estudado (Transferência do Conhecimento Prosódico do Português para o inglês). A este processo dá-se o nome de "extração de características" da fala (*feature extraction*), uma importante etapa na abordagem conexionista do processamento da fala. Em segundo lugar, uma vez determinadas as características do item lexical ou da frase produzida por cada aprendiz, são inseridos esses dados na rede neural a fim de analisar as propriedades (regularidades) estatísticas do conjunto de falantes como um todo. Em terceiro, utiliza-se ferramentas de visualização para analisar como a rede organiza os falantes e quais informações são mais relevantes para este processo de formação de grupos (e.g. nível de proficiência, uma certa característica ou propriedade da fala, entre outros). A rede utilizada é conhecida como Mapa Auto-Organizável (*Self-Organizing Map*, SOM). A rede SOM organiza os falantes por grau de similaridade em grupos bem definidos (clusters). A aplicação da rede SOM neste contexto é, portanto, inovadora. A rede SOM é implementada no ambiente Matlab usando o pacote Som toolbox, que é um conjunto de rotinas de programação desenvolvidas pelo grupo de pesquisa da Finlândia, também inventores da rede SOM. Os resultados das simulações apontam que a rede SOM pode vir a ser usada mais frequentemente para avaliar o grau de distância a que um grupo de aprendizes está do grupo de falantes nativos. Dessa forma, uma rede neural pode vir a ser aplicada como ferramenta no contexto de determinação de nível de proficiência em língua estrangeira.

Palavras-chave: Aquisição de L2. Mapas auto-organizáveis. Análise de agrupamentos. Análise Multi-Rótulo.

ABSTRACT

The objective of this dissertation was to investigate how the prosodic knowledge is organized in an early stage of L2 acquisition in Brazilian learners of English with the help of a connectionist neural network. The approach proposed in this research is first, to quantify the utterances of L2 learners in the form of LPC coefficients and other linguistic/phonetics features that can represent the phenomenon studied here (Transfer of the prosodic knowledge from Portuguese to English). This process is called speech feature extraction, an important step in the connectionist approach to speech processing. Second, since certain features of the lexical item or sentence produced by each learner are determined, these data are entered into the neural network to analyze the statistical properties (regularities) of the set of speakers as a whole. Third, visualization tools are used to analyze how the network organizes speakers and what information is most relevant to this process of group formation (e.g. proficiency level, a certain characteristic or property of speech, among others). The network is known as Self-Organizing Map (Self-Organizing Map, SOM). The SOM organizes speakers for similarity degree in well-defined groups (clusters). Application of SOM in this context is therefore innovative. The SOM network is implemented in Matlab environment using the SOMtoolbox package, which is a set of programming routines developed by the research group in Finland, also the inventors of the SOM. The simulation results indicate that SOM might be used more frequently to assess the degree of distance that a group of learners is to the group of native speakers. Thus, a neural network might be used as a tool in the context of determining the level of foreign language proficiency.

Keywords: L2 Acquisition. Self-organizing maps. Cluster analysis. Multi-Label Analysis.

RESUMEN

El objetivo de esta tesis ha sido investigar cómo el conocimiento prosódico se organiza en una fase temprana de la adquisición de L2 en los estudiantes de Inglés con la ayuda de una red neuronal conexionista. El enfoque propuesto en este trabajo es primero "cuantificar" las declaraciones de los alumnos de L2 en forma de coeficientes LPC y otras características lingüísticas y fonéticas que pueden representar el fenómeno estudiado aquí (Transferencia de Conocimiento prosódico de Portugués para Inglés). Este proceso le da el nombre de "extracción de características del habla (extracción de características), un paso importante en el enfoque conexionista para el procesamiento del habla. En segundo lugar, ya que ciertas características del elemento léxico o la sentencia producida por cada alumno, estos datos se introducen en la red neuronal para analizar las propiedades (regularidades) estadísticas del conjunto de datos en su conjunto. En tercer lugar, utilizamos herramientas de visualización para analizar cómo se organiza la red de altavoces y cuál es la información más relevante de este proceso de formación de grupo (nivel de habilidad, por ejemplo, una determinada característica o propiedad de expresión, entre otros). La red se conoce como Self-Organizing Map (SOM). La SOM organiza los datos por el grado de similitud en grupos bien definidos (clusters). La aplicación de SOM en este contexto es por lo tanto innovadora. El SOM se implementa mediante la caja de herramientas de Matlab (toolbox), que es un conjunto de rutinas de programación desarrollado por el grupo de investigación en Finlandia, también los inventores de la SOM. El programa de simulación de resultados que SOM puede ser utilizado con mayor frecuencia para evaluar el grado de distancia desde la cual un grupo de alumnos es el grupo de hablantes nativos. Por lo tanto, una red neuronal puede ser utilizado como una herramienta en el contexto de determinar el nivel de dominio del idioma extranjero.

Palabras clave: Adquisición de la L2. Auto-organización de los mapas. El análisis de conglomerados. Análisis de multi-etiqueta.

LISTA DE SÍMBOLOS FONÉTICOS

Os símbolos fonéticos utilizados para a transcrição das emissões deste trabalho foram extraídos do IPA, International Phonetic Alphabet, e encontram listados abaixo:

Vogais

Símbolo	Palavra-chave
i	beat, feed
ɪ	bit, did
eɪ	date, paid
ɛ	bet, bed
æ	bat, bad
ɑ	box, odd, father
ɔ	bought, dog
oʊ	boat, road
ʊ	book, good
u	boot, food, student
ʌ	but, mud, mother
ə	banana, among
ɜr	shirt, murder
aɪ	bite, cry, buy, eye
aʊ	about, how
ɔɪ	voice, boy
ɪr	beer
ɛr	bare
ɑr	bar
ɔr	door
ʊr	tour

Consoantes

Símbolo	Palavra-chave
p	pack, happy
b	back, rubber
t	tie
d	die
k	came, key, quick

g	game, guest
tʃ	church, nature, watch
dʒ	judge, general, major
f	fan, photograph
v	van
θ	thing, breath
ð	then, breathe
s	sip, city, psychology
z	zip, please, goes
ʃ	ship, machine, station, special
ʒ	measure, vision
h	hot, who
m	men, some
n	sun, know, pneumonia
ŋ	sung, ringing
w	wet, while
l	light, long
r	right, wrong
y	yes, use, music

Suprasegmentos

- ' ➡ acento primário (e.g.: ['dɛzɜrt])
- ˌ ➡ acento secundário (e.g.: ['sɛpəˌrɛrt])
- :
- ː ➡ longo (e.g.: [ɔː])
- .
- ➡ divisão silábica
- ◌
- ➡ link (ausência de pausa entre palavras)

LISTA DE SÍMBOLOS MATEMÁTICOS

K	número de grupos ou clusters
N	número de entradas de dados
i^*	índice do protótipo mais próximo
t	iteração atual
$\ \cdot \ $	distância ou norma euclidiana
\mathbb{R}^n	conjunto dos números reais no espaço de n dimensões
$x(t)$	um vetor de entrada da rede na iteração t
n	indicativo do número de dimensões do espaço amostral
$i^*(t)$	índice do neurônio vencedor na rede
$w_i(t)$	o vetor de pesos associado ao neurônio i
$w_{i^*}(t)$	neurônio vencedor
$h(i^*, i; t)$	função de vizinhança da rede SOM
$eq(t)$	vetor de erros de quantização
$eq(x(t))$	erro de quantização associado ao vetor $x(t)$
X	uma matriz $n \times n$
P	dimensão da quantização vetorial

LISTA DE FIGURAS

Figura 2.1: Representação dos pés do item lexical 'casa'	28
Figura 2.2: Hierarquia prosódica proposta por Nespor e Vogel	31
Figura 2.3: Hierarquia dos constituintes prosódicos.....	32
Figura 2.4: Os constituintes silábicos.....	33
Figura 3.1: Aparelho fonador	56
Figura 3.2: Digrama em bloco da produção da fala humana, também conhecido como sistema fonte-filtro	57
Figura 3.3: Anatomia do ouvido humano	60
Figura 3.4: Visualização de uma onda sonora na tela de um computador usando o software PRAAT	61
Figura 3.5: Sinal de tempo contínuo e amplitude contínua	62
Figura 3.6: Sinal de tempo contínuo e amplitude discreta	63
Figura 3.7: Processo de conversão de um sinal de tempo contínuo para tempo discreto	64
Figura 3.8: Sinal de tempo discreto e 16 níveis de quantização	66
Figura 3.9: Sinal quantizado	66
Figura 3.10: Exemplo de som surdo	69
Figura 3.11: Exemplo de som sonoro	70
Figura 3.12: Uso de LPC na produção de voz	70
Figura 4.1: Representação de um neurônio biológico em (a) e um neurônio artificial em (b)	77
Figura 4.2: Ilustração das várias etapas que compõem o potencial de ação em um neurônio	78
Figura 4.3: Modelo do neurônio artificial de McCulloch & Pitts.....	81
Figura 4.4: Neurônios organizados em um mapa de Kohonen bidimensional	87
Figura 4.5: Atualização do neurônio vencedor (best matching unit – BMU) e seus vizinhos em direção à amostra indicada com X.	89
Figura 4.6: Vizinhos (0, 1 e 2) da unidade mais central: (a) grade hexagonal e (b) grade retangular.	89
Figura 4.7: Exemplo de matriz U - Matriz visualizada com representação por superfície topológica	92
Figura 4.8: Exemplo de matriz-U matriz bidimensional com tons de cinza.....	92
Figura 4.9: Análise de dados a partir do SOM	93
Figura 5.1: Input da rede	109
Figura 5.2: Representação da interface gráfica do ambiente Matlab durante o processo de inicialização e treinamento da rede.	112
Figura 5.3: Arquitetura da rede neural	117
Figura 6.1: Gráfico típico do erro de quantização por épocas	124
Figura 6.2: Gráfico típico do erro topológico	124
Figura 6.3: Gráfico típico do erro topológico.....	125
Figura 6.4: Mapa rotulado com as características lexicais e de identificação dos aprendizes	126
Figura 6.5: Matriz-U gerada para a rede 10 x 10 da palavra object (verbo) após treinamento de 250 épocas (ainda sem marcação dos grupos)	130

Figura 6.6: Matriz-U com entrada formada por 10 coeficientes LP. Índice DB indica 3 clusters e o erro de quantização final é igual a 0,427	131
Figura 6.7: Matriz-U com entrada formada por 10 coeficientes MFC. Índice DB indica 2 clusters e o erro de quantização final é igual a 1,636	131
Figura 6.8: Mapa rotulado de uma rede SOM bidimensional 10 x 10 com estrutura de vizinhança hexagonal	133
Figura 6.9: Gráfico para o índice Davies-Bouldin em função do número k de agrupamentos em uma rodada de treinamento típica	132
Figura 6.10: Representação dos grupos no mapa treinado para k=3.....	132
Figura 6.11: Representação dos grupos no mapa treinado para k=3.	133
Figura 6.12: Matriz-U marcada com a representação dos grupos sugeridos pelo índice DB	134
Figura 6.13: Mapa rotulado de acordo com o nível de proficiência (1, 2, 3, 4, 6) .	135
Figura 6.14: Matriz-U evidenciando a formação de 2 grupos; um que contém indivíduos que transferem o acento primário e outro que não transfere	138
Figura 6.15: Mapa rotulado com “erro” para os conjuntos de dados referentes à palavra object (verbo), usando LPC 10, treinados com dados dos 30 aprendizes	139
Figura 6.16: Mapa rotulado com informações sobre a frequência das classes (rótulos) para cada neurônio vencedor	140
Figura 6.17: Mapa colorido sugerindo a formação de dois grupos, conforme indicação do índice Davies-Bouldin	141
Figura 6.18: Mapa rotulado com codificação relativa aos fenômenos de inserção de vogal em posição de coda	142
Figura 6.19: Mapa colorido sugerindo a formação de clusters	143
Figura 6.20: Matriz-U sem nenhuma representação significativa de clusters	143
Figura 6.21: Marcação da formação de grupos na Matriz-U	144
Figura 6.22: Mapa rotulado com identificação de indivíduos sem boa formação de agrupamento	145
Figura 6.23: Divisão do mapa rotulado sugerida pela rede em quatro grupos	145

LISTA DE QUADROS

Quadro 2.1: Parâmetros de Dresher e Kaye	31
Quadro 2.2: Lista de alguns pares mínimos em língua portuguesa	40
Quadro 2.3: Padrão acentual dos itens lexicais organizados por categoria lexical	42
Quadro 2.4: Determinação de Parâmetros Métricos do PB e inglês.....	45
Quadro 5.1: Rótulos dos vetores de características com seus respectivos significados	112
Quadro 6.1: Parâmetros de treinamento da rede SOM para o experimento 1	124
Quadro 6.2: Parâmetros de treinamento da rede SOM para o experimento 2	129
Quadro 6.3: Parâmetros de treinamento da rede SOM para o experimento 3	141

LISTA DE TABELAS

Tabela 5.1: Classificação de nível de proficiência de língua inglesa de acordo com horas-aula acumuladas (distribuição da estrutura curricular da disciplina de língua inglesa em 2004.1 nas escolas particulares de Fortaleza)	104
Tabela 5.2: Estrutura curricular do curso de extensão em língua inglesa no semestre 2004.1	105
Tabela 5.3: Divisão dos participantes por horas-aula	106
Tabela 6.1: Resultados para algoritmo SOM – teste de desempenho da rede.....	123
Tabela 6.2: Número de clusters sugerido pelo índice DB.....	129
Tabela 6.3: Número de agrupamentos para cada palavra em cada algoritmo de extração de parâmetros da fala e ordem de predição	129

LISTA DE ABREVIATURAS E SIGLAS

ANN - Artificial Neural Networks

HMM - Hidden Markov Models

LPC – Linear Predictive Coding

CCB – Casa de Cultura Britânica

IA – Inteligência Artificial

L1 – Língua Materna

L2 – Segunda Língua

LE – Língua Estrangeira

PB – Português Brasileiro

PE – Português Europeu

PDP – Processamento distribuído em paralelo

RNA – Redes Neurais Artificiais

SOM – Self-Organizing Map (Mapa auto-organizável)

UFC – Universidade Federal do Ceará

DB - Davies-Bouldin

AR - Autorregressivo

EQ - Erro de Quantização

ET - Erro Topológico

SUMÁRIO

1 INTRODUÇÃO	19
2 FUNDAMENTOS LINGÜÍSTICOS	27
2.1 As teorias prosódicas	27
2.1.1 A sílaba	32
2.1.2 O acento.....	34
2.1.3 Níveis de acento.....	35
2.1.4 O acento em PB	37
2.1.5 O Acento em Inglês.....	42
2.1.6 Relação entre vogal reduzida e acento	45
2.2 Tom e Entonação	47
2.2.1 Padrões entoacionais do PB	49
2.2.2 Padrões entoacionais do inglês (norte-americano)	50
2.3 Resumo do Capítulo.....	53
3 FUNDAMENTOS DE PROCESSAMENTO DA FALA	54
3.1 O Processo de produção natural da fala	55
3.2 A onda sonora	58
3.3 Análise do sinal de fala.....	60
3.3.1 Sinais de tempo discreto	63
3.4 Parametrização do sinal da fala e Extração de características	67
3.4.1 Coeficientes de Predição Linear.....	68
3.4.2 Coeficientes Cepstrais.....	72
3.5 Resumo do capítulo	75
4 MODELOS CONEXIONISTAS PARA O PROCESSAMENTO DE LINGUAGEM ..	76
4.1 Redes Neurais Biológicas Vs. Redes Neurais Artificiais	76
4.2 Mapas auto-organizáveis e a rede de Kohonen	83
4.2.1 Os princípios de auto-organização.....	84
4.2.2 O Algoritmo de Treinamento da Rede de Kohonen.....	87
4.2.3 Busca pelo neurônio vencedor	88
4.2.4 A matriz de distância unificada (Matriz-U)	91
4.3 As vantagens de utilização da Rede Auto-Organizável.....	94
4.4 Aplicações dos Mapas Auto-Organizáveis ao Processamento de Fala Natural ..	96
4.5 Resumo do capítulo	99
5 METODOLOGIA DE PESQUISA	100
5.1 Caracterização dos participantes da pesquisa, o <i>corpus</i> e o método de coleta de dados	100
5.2 As tarefas e os testes	101
5.3 A organização dos participantes da pesquisa	102
5.3.1 Etapa de Aquisição dos dados e segmentação das palavras	107
5.3.2 Etapa de Processamento ou Parametrização do Sinal de Fala	107
5.3.3 Etapa de configuração da rede neural	112
5.3.4 Determinação dos parâmetros da rede	113
5.3.5 A simulação e suas fases.....	113

5.4 O uso da SOM Toolbox.....	118
5.5 Os experimentos	120
5.6 Resumo do Capítulo.....	120
6 RESULTADOS DAS SIMULAÇÕES	122
6.1 Simulação da representação de categorias lexicais.....	122
6.1.1 Metodologia da simulação.....	122
6.1.2 Resultados da simulação	123
6.2 Análise da formação de grupos em função do conhecimento prosódico do aprendiz	127
6.2.1 Metodologia da simulação.....	127
6.3 Simulação da transferência de padrões acentuais (acento primário) para substantivos, verbos e adjetivos	139
6.3.1 Metodologia da simulação.....	139
6.4 Simulação da Transferência de Padrões Vocálicos (substituição, apagamento, inserção e redução vocálica).....	143
6.5 Sumarização dos resultados	151
CONCLUSÃO.....	153
REFERÊNCIAS.....	159
ANEXOS	170
APÊNDICES.....	175

1 INTRODUÇÃO

Apesar de se afirmar muito frequentemente que a prosódia é uma das maiores fontes de dificuldades do aprendiz brasileiro de língua inglesa, não há dados suficientes na literatura de aquisição de L2 que mostrem a extensão dessas dificuldades.

O acento lexical, um dos elementos prosódicos mais importantes, é o maior responsável pelos casos de transferência linguística, sotaque predominante de L1 e fossilização em L2, de acordo com os estudos da área de fonologia da interlíngua. Modelos formais de aprendizado prosódico/acental (ARCHIBALD, 1994, 1998, 2004; MAIRS, 1989; SILVA, 2005) propõem como os aprendizes constroem seus sistemas entoacionais e acenais. Há duas hipóteses principais: a transferência do padrão de L1 ou a construção de um novo sistema.

Muitos estudos empíricos têm se dedicado a responder a essa questão, todavia não conseguiram de forma tão bem sucedida, muito provavelmente por não se utilizarem de um método de simulação eficiente. Os modelos conexionistas, por sua vez, mostram-se como uma ótima alternativa para a análise das regularidades estruturais do sistema prosódico dos aprendizes em fase inicial de aquisição de língua estrangeira (doravante LE), devido à sua grande capacidade de generalização, além de oferecer uma variedade de métodos para responder a tais questões, ou seja, as simulações conexionistas em redes neurais artificiais.

As redes auto-organizáveis (*self-organizing neural nets*), modelo de rede neural artificial de um novo conexionismo, têm-se mostrado úteis na caracterização do mecanismo subjacente a várias tarefas de aprendizagem de linguagem e na resolução de problemas de reconhecimento estatístico de padrões.

Vale ressaltar que ainda há uma escassez de estudos sobre a aplicação de redes neurais artificiais para dar conta dos processos relacionados à aquisição do acento de L2. Com isso, torna-se crucial preencher essa lacuna a partir da análise do conhecimento acental de determinados itens lexicais complexos em aprendizes brasileiros de língua inglesa em vários níveis utilizando-se de uma rede neural auto-organizável, tendo-se em mente que os resultados futuros dessas pesquisas irão

auxiliar de forma direta e indireta na determinação de níveis de proficiência linguística de aprendizes de L2.

A área de aquisição¹ da linguagem tem evoluído bastante desde o trabalho pioneiro de Rumelhart e McClelland (1986) sobre a aquisição do passado simples na flexão de verbos de língua inglesa. Ao longo desses anos, esse paradigma do conhecimento tem-se mostrado bem sucedido na simulação de comportamento linguístico, desde o reconhecimento de padrões fonológicos, ambiguidade semântica, compreensão leitora e aprendizagem de uma segunda língua (doravante L2) (MACEDO, 2008; POERSCH, 2004; PLUNKETT, 1998).

A necessidade de modelos dinâmicos de aquisição de representações linguísticas faz com que os cientistas de tendência emergentista escolham o conexionismo. Isto porque ele oferece um conjunto de ferramentas computacionais para explorar as condições sob as quais algumas propriedades linguísticas se consolidam, além de possuir várias vantagens, como, por exemplo, inspiração neural, representação e controle distribuído, processamento indutivo com representações prototípicas emergindo mais do que sendo inatamente pré-especificada, aprendizado sensível à estrutura e ao conteúdo, representações não-estáticas, distribuídas e graduais; generalização e transferência² como produtos naturais de aprendizado (ELLIS, 2003).

Um modelo específico de rede auto-organizável, os mapas auto-organizáveis (SOM) - comumente chamados de rede de Kohonen - apresentam-se como uma arquitetura conexionista moderna e são vistos como um modelo de interpretação de processos de transferência de conhecimento linguístico utilizados usualmente como ferramenta de visualização de dados, permitindo que se percebam mais facilmente inter-relações existentes em conjuntos complexos de dados.

¹ Apesar de a maioria dos conexionistas acreditar que a linguagem é aprendida e não adquirida, ainda mais no que se refere à L2, optamos aqui por não fazer distinção entre os termos 'aquisição' e 'aprendizagem', seja em língua materna ou estrangeira. Tais termos são usados aqui de forma livremente substituível. Também não faremos distinção entre os termos de língua estrangeira (LE) e segunda língua (L2), devido ao fato de ainda não haver um consenso entre os autores quanto aos domínios de atuação de ambos os tipos de convenção (Cf. Ellis, 1999).

² Segundo Gasser (1990 *apud* Zimmer, 2006), o fenômeno cognitivo em que os modelos conexionistas são extraordinariamente hábeis em simular é o da transferência de conhecimento – linguístico ou extralingüístico. Entende-se aqui por *transferência linguística* o processo de aprendizagem de uma dada língua estrangeira em que os aprendizes baseiam-se no conhecimento que têm de sua língua materna para compreender a forma de estruturação da língua alvo, seja no nível fonológico, sintático, semântico ou pragmático, bem como produzi-la (ZIMMER, 2006).

A despeito de as redes auto-organizáveis de Kohonen deterem uma importância reconhecida em outras áreas de compreensão da cognição humana, elas foram pouco exploradas para explicar fenômenos da linguagem. As redes de Kohonen já foram aplicadas em muitas áreas tecnológicas, a saber, reconhecimento automático de locutores (MAFRA, 2002; SOUZA JR, 2009; MÜLLER, 2006; LIMA, 2000), classificação bibliográfica, sistemas de navegação para imagens, diagnóstico médico, interpretação de atividade sísmica, compressão de dados, separação de fontes de som, modelagem ambiental. Na linguística, essa rede neural tem-se mostrado ideal para o processamento de linguagem natural, modelagem de aprendizado perceptual de categorias fonéticas (GAUTHIER; SHI; XU, 2007b) e particularmente para modelar a aquisição do léxico (LI; FARKAS; MCWHINNEY, 2004; LI, 2009). Alguns trabalhos já propuseram modelos computacionais prosódicos utilizando redes neurais auto-organizáveis para fins de transcrição texto-fala (do inglês, *text-to-speech* - TTS); no Chinês, há o trabalho de Tao e Cai (1998); no Finlandês, Vainio (2001); no tcheco, Adámek (2002) e na língua Nguni, o trabalho de Govender (s/d).

Outros estudos investigaram a plausibilidade da aplicação de redes neurais artificiais à modelagem da aquisição prosódica/acental. Gauthier et al. (2009) usaram modelos conexionistas para explorar se e como as crianças poderiam aprender o foco prosódico³ diretamente de *input* de fala contínua. Em três simulações utilizando redes neurais auto-organizáveis, os autores exploraram como o foco poderia ser aprendido a partir de sinais acústicos contínuos em Mandarim, que foram produzidos com tons lexicais co-ocorrentes e por vários falantes. Os resultados desse estudo mostraram que redes neurais não-supervisionadas podem desenvolver agrupamentos específicos de foco a partir de sinais de fala dinâmicos contínuos produzidos por vários falantes em várias condições de tom lexical, o que pode eventualmente conduzir à aquisição do foco.

Zimmer (2006) investigou a transferência do conhecimento fonético-fonológico do português brasileiro para o inglês (variedade norte-americana) durante a recodificação leitora de palavras sob uma perspectiva conexionista. Seu estudo analisou os dados coletados transversalmente junto a 156 adultos, falantes monolíngues de português e estudantes de inglês como língua estrangeira e

³ O foco é uma função comunicativa, que serve para colocar ênfase em uma determinada parte de um enunciado.

investigou a incidência de nove processos de transferência sobre as elocuições dos participantes divididos em quatro grupos de proficiência (básico, intermediário, pré-avançado e avançado) durante sessões de leitura de não-palavras e de palavras regulares. Foi apresentada, também, uma simulação computacional do fenômeno investigado em redes conexionistas do tipo *feedforward*, *i.e.*, cada unidade intermediária recebendo uma conexão de cada unidade de *input* e enviando, sucessivamente, uma conexão para cada unidade de *output*.

Nos resultados das simulações das leituras de não-palavras, a autora indica como direcionamento futuro a reformulação do *corpus* utilizado em sua tese, de forma que inclua um número maior de palavras que figurem como determinantes no treinamento da rede e que esse *corpus* reformulado fosse trabalhado na perspectiva da aprendizagem hebbiana⁴ em redes do tipo Kohonen. A autora também chamou a atenção para o desenvolvimento de um segundo tipo de modelo conexionista, considerado por ela e por seu orientador associado mais apropriado para simular a leitura em voz alta: uma rede sequencial para a exploração da temporalidade na fala. Segundo eles, a exploração de aspectos da temporalidade na fala⁵, tais como a taxa de velocidade de fala e a variação na duração das pausas, ajudaria a explicitar padrões de transferência do conhecimento suprasegmental entre o português e o inglês.

Li, Farkas, MacWhinney (2004) apresentaram um modelo de rede neural auto-organizável que simula o desenvolvimento lexical inicial, o DevLex. A rede consistiu de dois mapas auto-organizáveis (um mapa semântico e um mapa fonológico, ambos crescentes) que estavam conectados por meio de *links* associativos e foram treinados por aprendizagem hebbiana. O modelo captou uma série de fenômenos importantes que ocorriam na aquisição lexical inicial das crianças, uma vez que permitiu a simulação da representação de um ambiente em mudança linguística dinâmica na aprendizagem de línguas. Nas simulações de Li et al. (2004), o DevLex desenvolveu topograficamente representações organizadas por categorias linguísticas ao longo do tempo, modelou a confusão lexical como uma

⁴ Referente ao Biólogo canadense Donald Olding Hebb. O postulado de Hebb é uma das mais antigas e famosas regras de aprendizagem e é amplamente utilizado nas computações de redes neurais artificiais.

⁵ Albano (2002) compara a fala humana a uma orquestração de pulsações que podem comprimir-se, expandir-se, sobrepor-se temporalmente. O gesto articulatório é uma unidade de análise abstrata, ao mesmo tempo fonética e fonológica, que difere do segmento fônico do tamanho do fonema por ter uma natureza temporal e, portanto, uma duração necessária.

função da densidade de palavra e da semelhança semântica e ainda mostrou os efeitos da idade de aquisição no decurso da aprendizagem em um léxico crescente. Todos esses resultados promissores demonstram a competência das redes neurais competitivas não-supervisionadas (rede de Kohonen) nas simulações de modelagem de aquisição lexical.

A despeito de todas as análises exploratórias arroladas acima e suas áreas de aplicação, ainda não houve a investigação, tampouco uma análise exploratória dos fenômenos de transferência de L1 para L2 relacionados ao acento de palavra por meio de uma rede neural auto-organizável. Ressalta-se que a natureza auto-organizável da rede aliada ao *input* e ao *output* não supervisionado consolida-se como uma proposta válida e extremamente útil na determinação de nível de proficiência em língua inglesa, com base em categorias linguísticas. Motivados pela escassez de estudos na área, pela ânsia de consolidação de uma área de conhecimento que lide de forma profícua com o objeto de estudo e seus objetivos, as seguintes questões de pesquisa são apresentadas:

1. O sinal de fala das elocuições dos aprendizes e os coeficientes cepstrais extraídos dele constituem uma forma de codificação válida para a entrada da rede neural e seu posterior treinamento?

2. Quais elementos e características linguísticas associadas aos coeficientes são mais significativas/determinantes para a formação de agrupamentos no mapa auto-organizável/rede neural?

3. Partindo-se do pressuposto de que a rede neural organizará os aprendizes em função da pronúncia de determinadas palavras, os aprendizes serão organizados a partir de critérios de similaridade de pronúncia ou com base em outras informações que estão subjacentes aos dados?

4. A rede neural seria capaz de fazer generalizações sobre os processos fonológicos relacionados às vogais, consoantes e acentuais de L1 para L2 com base na formação de agrupamentos do mapa auto-organizável?

Considerando essas questões, as seguintes hipóteses de pesquisa são apresentadas:

1. Os coeficientes LP e os coeficientes cepstrais (MFC) condensam muita energia, frequência fundamental (F_0), banda de frequência, ou seja, informações

genuinamente ligadas à prosódia da fala e principalmente ao acento, o que se torna de fundamental importância para a codificação do *input* da rede;

2. A parametrização do sinal de fala por coeficientes LP e MFC é eficiente na categorização dos falantes por características prosódicas já que estes – os coeficientes – capturam informação prosódica necessária a ser utilizada posteriormente para a classificação dos aprendizes;

3. A rede neural analisará as propriedades (regularidades) estatísticas dos atributos relativos ao conjunto de dados dos falantes como um todo e organizará as informações por grau de similaridade em grupos bem definidos. Como os dados serão rotulados de forma diferenciada (*Multi-label analysis*), a emergência de certas categorias no mapa auto-organizável será identificável mais facilmente;

4. A rede neural é capaz de discriminar os coeficientes LP e MFC dos aprendizes e segregar os grupos de aprendizes usando como base os processos de transferência, um sinal claro de indicação de nível de proficiência linguística ou, no mínimo, de descrição de aquisição de L2.

O objetivo desta pesquisa é investigar como o conhecimento prosódico (acentual) de aprendizes brasileiros de inglês está organizado em um estágio inicial de aquisição de L2 com a ajuda de uma rede neural conexionista. A presente pesquisa utilizou-se do sinal de fala numericamente codificado como *input* para uma rede neural não-supervisionada. A partir daí, têm-se como objetivos específicos:

1. Determinar a codificação adequada para a entrada da rede neural, a fim de que esta organize os falantes por características linguísticas semelhantes em grupos homogêneos;

2. Analisar como a rede de Kohonen organiza os falantes em grupos (*clusters*) e se esta organização segue um padrão coerente com o nível de proficiência pré-determinado (tempo de exposição ao idioma);

3. Verificar quais informações/características são mais relevantes para este processo de formação de grupos (e.g. nível de proficiência, uma certa característica ou propriedade da fala, entre outros);

4. Determinar os parâmetros da rede de Kohonen e os rótulos que podem oferecer os melhores resultados para a visualização dos agrupamentos na matriz-U: dimensão da rede, tipo de neurônio utilizado, número de épocas, repetições do treinamento da rede etc.;

5. Avaliar o desempenho da rede neural durante as simulações que analisem a aquisição inicial do conhecimento acentual em vários aprendizes de vários níveis distintos (acento primário, acento secundário, inserção vocálica em final de sílaba) e representação de categorias lexicais.

Espera-se contribuir com este trabalho para os estudos de aquisição de L2, mais especificamente, os estudos de fonologia da interlíngua, teorias de aquisição de L2 e metodologia do ensino de língua inglesa. Os resultados desta pesquisa também virão a contribuir para a (re)formulação de teorias sobre a avaliação de desempenho e a determinação de nível de proficiência linguística em LE, dada a competência da rede de organização de grupos de indivíduos por características fônicas semelhantes além de ser uma ferramenta de visualização razoável para analisar a formação de agrupamentos a partir de simulação de representação de categorias lexicais. O presente estudo também oferece uma revisão do estado da arte sobre a utilização de redes neurais artificiais utilizadas na investigação de fenômenos de aquisição de L2.

Sinaliza-se a importância dessa pesquisa para a linha de lingüística aplicada pelo fato de os resultados contidos aqui contribuir para ampliar o uso de ferramentas tecnológicas envolvendo a fala para a melhoria do aprendizado de pronúncia de língua inglesa. Pesquisas sobre o *Ensino de Pronúncia Mediado por Computador (Computer-assisted Pronunciation Teaching)* (LEVIS, 2008; CHUN, 1998; MOLHOLT, 1988) nas últimas décadas indicam que tanto pesquisadores quanto professores de pronúncia vêm fazendo um maior uso da tecnologia computacional para desenvolver novas teorias e práticas metodológicas que se alinhem com a realidade acústica da fala. A teoria sobre processamento do sinal acústico da fala e as ferramentas de análise da fala são de fundamental importância para a resolução de muitas questões relacionadas ao ensino e à aprendizagem de línguas estrangeiras.

A pesquisa se situa dentro das atividades do grupo de pesquisa (GELP-COLIN) liderado pela orientadora Prof^a. Dra. Ana Cristina Pelosi Silva de Macedo, cujos aportes teóricos da psicolingüística e lingüística cognitiva contribuíram diretamente para o desenvolvimento da idéia inicial sobre cognição e linguagem aqui apresentada. A pesquisa ainda recebeu contribuições dos conhecimentos da Engenharia (Elétrica e de Teleinformática), por exemplo, a área de Processamento

de Sinais, de suma importância para a etapa de extração das características numéricas para o *input* da rede, bem como a área de Reconhecimento de Padrões e a utilização de seus métodos não-supervisionados. A tese apresenta uma nova proposta teórico-metodológica: a Análise Multi-rótulo (*Multi-label Analysis*), que se utiliza de rótulos criados a partir de processos fonético-fonológicos relacionados a vogais, acento e nível de proficiência.

A tese está organizada da seguinte forma: o capítulo 2 apresenta os fundamentos teórico-linguísticos que embasam a pesquisa, ou seja, as principais teorias linguísticas relacionados à prosódia; o capítulo 3 traz os principais fundamentos de processamento do sinal de fala; o capítulo 4 trata dos modelos conexionistas para o processamento da linguagem, entre elas a rede de Kohonen com as suas respectivas propriedades; o capítulo 5 explana os procedimentos metodológicos utilizados na pesquisa; e o capítulo 6 apresenta as simulações na rede neural com os respectivos resultados. Este trabalho é concluído com indicações possíveis para pesquisas futuras na área de aprendizagem de LE.

2 FUNDAMENTOS LINGUÍSTICOS

A prosódia, o nível fonético da língua que não é segmentável em unidades discretas como são os fones, acrescenta espontaneidade e significado pragmático à fala. Ela engloba as características de entoação, de intensidade e de duração fonética e, em português, está na base de distinções fonológicas importantes, como a alternância entre sílabas tônicas e átonas, ou como a diferença entre frases declarativas, interrogativas ou exclamativas. É também responsável pela distinção perceptiva entre informação nova e importante e informação conhecida e menos relevante em um texto/discurso (BRAGA; FREITAS; FERREIRA, 2003).

2.1 As teorias prosódicas

A teoria prosódica da Fonologia, também conhecida como Escola Linguística de Londres, permite que tanto elementos fonêmicos (segmentais) quanto prosódicos (suprasegmentais) sejam identificados e tenham um *status* igual, mas complementar, no inventário fonológico dos sistemas linguísticos (LYONS, 1987).

A maioria dos linguistas emprega o termo suprasegmental como referência a acento, tom e duração. O acento e a entoação são dois elementos suprasegmentais importantes para a concretização da cadeia da fala. Alguns autores pós-gerativos formularam teorias fonológicas que tomam por base a sobreposição desses elementos ao segmento.

Variações quanto ao uso do acento podem causar diferenças no ritmo das línguas. Até bem pouco tempo, costumava-se adotar a classificação de línguas como ritmo silábico ou tempo silábico (*syllable-timing*) (MAJOR, 1985). O francês é um exemplo: as sílabas tendem a repetir-se em intervalos regulares de tempo, em contraste com línguas, como o inglês e o português, chamadas línguas de ritmo acentual ou tempo acentual (*stress-timing*), em que a duração entre os intervalos das sílabas acentuadas é classificada como isocrônica⁶. Hoje em dia, já se sabe que

⁶ Referente ao termo isocronia (fenômeno em que as sílabas não-acentuadas diminuem sua duração de acordo com o número delas ocorrente entre duas sílabas acentuadas).

essa classificação é inviável, já que uma língua e seus dialetos podem apresentar características de ambos os tipos de ritmo (BARBOSA, 2000).

A Fonologia Métrica, iniciada por Liberman e Prince (1977), é uma teoria fonológica cujos segmentos estão organizados hierarquicamente (subordinados) em outros constituintes de ordem rítmica, como sílabas, pés e palavras. Crystal (2000, p. 117) apresenta a seguinte definição para essa teoria fonológica:

A fonologia métrica é uma abordagem que enfatiza a relação entre os segmentos, as sílabas, e os processos prosódicos, como ritmo e acento: as regras atribuem uma proeminência relativa aos nódulos de tipo irmão (binários) de um diagrama de árvore, de modo que um nóculo seja forte e o outro fraco, a fim de produzir uma árvore métrica que represente a estrutura métrica subjacente de uma palavra.

As palavras de Bisol (1994, p. 25), destacadas abaixo acerca do acento na teoria métrica, não pretendem ser exaustivas no que tange às suas dimensões, mas têm o mérito de fornecer alguns pressupostos básicos sobre essa teoria:

A Teoria Métrica afirma que o acento não se localiza diretamente na vogal, mas provém de uma relação que se estabelece entre as sílabas, formando um contorno de proeminência. É preciso, pois, para estabelecer o algoritmo acentual, apenas verificar como a língua organiza as suas sílabas em constituintes prosódicos, pés métricos propriamente ditos.

A Fonologia Métrica tem o pé como unidade mínima representativa. Xavier e Mateus (1990) definem pé acentual da seguinte forma:

Unidade da fonologia métrica, normalmente dissilábica, que tem dominância à esquerda, i. e., que se inicia com uma sílaba terminal acentuada, seguida por uma (ou mais) sílabas não acentuadas. O pé acentual é representado por Σ .

Exemplo: Σ
 /
 \
 σ σ
 c a s a

Figura 2.1: Representação dos pés do item lexical 'casa'.
Fonte: Hogg e McCully (1987).

Uma das regras principais, dentro da teoria métrica, que garante exceção à mobilidade do acento é a regra da Extrametricalidade⁷ (ou Extrametricidade), conceito introduzido por Libermann e Prince em 1977 que ganhou importância dentro da Fonologia Métrica com Hayes (1992).

Uma das funções da Extrametricalidade, que é restrita a regularidades idiossincráticas e que também tem o poder de tornar invisíveis certos segmentos, é ajustar a palavra prosódica aos domínios das regras gerais de atribuição do acento. Na proposta de Hayes (1981), ela incide sobre a sílaba final. Os pares mínimos do tipo dúvida/duvida e dívida/divida são garantidos pelo asterisco lexical que incide sobre todas as proparoxítonas. É o que assevera Bisol (1994), ao assumir que a regra de acento é *categorial blind*, ou seja, é insensível à categoria lexical. Lee (1994), por sua vez, não concorda com esse argumento e considera, em concordância com Mateus (2005), como se verá oportunamente, que há duas regras de acento: uma para verbo e outra para não-verbo. Sobre isso, Lee (1994, p. 40) afirma:

[...] a existência de duas regras de acento parece mais plausível. Para diferenciar a regra de atribuição de acento do verbo e do não-verbo, é preciso observar a distinção entre essas duas categorias (verbo e não-verbo), do ponto de vista da Fonologia e da Morfologia.

i) O acento distingue o verbo e o não verbo em casos como:

(17) NV (Não-Verbo)	V (Verbo)
fórmula	formula
número	numero
válido	valido
contínuo	continuo

Esses exemplos mostram que a categoria lexical é definida pelo acento.

Os exemplos anteriores mostram que a categoria lexical é definida pelo acento. Sobre a regra de Extrametricalidade em língua inglesa, Archibald (1998, p. 179) argumenta que:

Em inglês, existe o princípio da Extrametricalidade Consonantal (...) que determina que cada consoante em final de palavra (em qualquer palavra) é extramétrica. Isso nos permite manter a generalização que o acento é

⁷ As sílabas que são excluídas temporariamente para fins de regras acentuais são chamadas de Extramétricas. Esse método tem a função de simplificar regras de acentuação em várias línguas sem ter que expandir o inventário de pés básicos e, ainda, eliminar diferenças de padrões acentuais entre diferentes classes de palavras.

atribuído à sílaba final de uma palavra se ela contém uma rima ramificada (vogal ou ditongo longo ou consoante na posição de coda) – caso contrário o acento é atribuído à penúltima sílaba. Este tipo de extrametricalidade funciona bem para a atribuição do acento em verbos, mas se torna problemática para substantivos. Além da regra acima, o inglês tem um princípio de extrametricalidade do substantivo que determina que a rima final de um substantivo seja ignorada na atribuição do acento. Em outras palavras, categorias gramaticais diferentes atribuem um status extramétrico a diferentes estruturas. Substantivos marcam uma sílaba completa como extramétrica, enquanto as outras categorias marcam somente o segmento⁸.

De acordo com Mateus (2005), na Fonologia Métrica, o acento principal (ou acento primário) é dominado, na árvore métrica por nós fortes em todos os níveis e o acento secundário incide sobre uma sílaba que, numa árvore métrica, é apenas dominada por nós fortes. A força métrica de uma sílaba, segundo Hogg e McCully (1987), é a sua relativa proeminência que se representa esquematicamente através das grelhas métricas.

Assim sendo, pode-se formalizar que os padrões (também chamados de parâmetros) acentuais são o resultado da estrutura métrica de constituintes hierarquizados pertencentes a cada língua. O quadro a seguir traz a reprodução, em termos mais gerais, dos parâmetros utilizados na atribuição do acento propostos por Dresher e Kaye (1990 *apud* ARCHIBALD, 1998, p. 178):

<i>P1</i>	<i>A palavra é forte na [esquerda/direita].</i>
<i>P2</i>	<i>Os pés são [binários/irrestrito].</i>
<i>P3</i>	<i>Os pés são construídos da [esquerda/direita].</i>
<i>P4</i>	<i>Os pés são fortes na [esquerda/direita].</i>
<i>P5</i>	<i>Os pés são sensíveis à quantidade [sim/não].</i>
<i>P6</i>	<i>Os pés são sensíveis à quantidade na [rima/núcleo].</i>
<i>P8</i>	<i>Existe uma sílaba extramétrica [sim/não].</i>
<i>P8A</i>	<i>É extramétrico na [esquerda/direita].</i>

Quadro 2.1: Parâmetros de Dresher e Kaye.
Fonte: Dresher e Kaye (1990 *apud* Archibald, 1998).

⁸ Trecho original: “In English, there is a principle of Consonant Extrametricality (Hayes, 1990) which states that every word-final consonant (in any word) is extrametrical. This allows us to maintain the generalization that stress is assigned to the final syllable of a word if it contains a branching rhyme (long vowel or diphthong, or consonant in the coda) – otherwise stress is assigned to the penult. This type of extrametricality works well for stress assignment in verbs but is problematic for nouns. In addition to the above rule, English has a principle of Noun Extrametricality which states that the final rhyme of a noun is ignored in stress assignment. In other words, different grammatical categories assign extrametrical status to different structures. Nouns mark a whole syllable as extrametrical, while other categories mark only a segment.”

A Fonologia Prosódica trabalha com descrições coerentes dos traços não-lineares, no que diz respeito ao acento e a suas implicações no sintagma. Para os seguidores dessa teoria, a prosódia determina funções sintáticas nos contextos enunciativos, e a interface entre Fonologia e Sintaxe é uma realidade permanente. A esse respeito, Nespor e Vogel (1986, p. 299) afirmam:

A fonologia prosódica explica como o fluxo da fala é organizado dentro de um grupo finito de unidades fonológicas. É também, no entanto, uma teoria de interações entre fonologia e outros componentes da gramática. Enquanto a divisão da cadeia da fala dentro das várias unidades fonológicas faz referência a estruturas encontradas nos outros componentes da gramática, um aspecto fundamental da teoria prosódica é que os constituintes fonológicos por si só não são isomórficos a qualquer constituinte encontrado em qualquer lugar da gramática.⁹

Na Fonologia Prosódica, a proeminência origina-se da relação entre sílaba, pé e palavra fonológica, estes três sendo os constituintes mais baixos na hierarquia prosódica, que está organizada dessa forma, de acordo com Nespor e Vogel (1986): Sílaba (símbolo: σ), Pé (Σ), Palavra fonológica (ω), Grupo Clítico (C), Sintagma Fonológico (ϕ), Sintagma Entoacional (I) e Enunciado Fonológico (U), como demonstrado na figura 2.2:

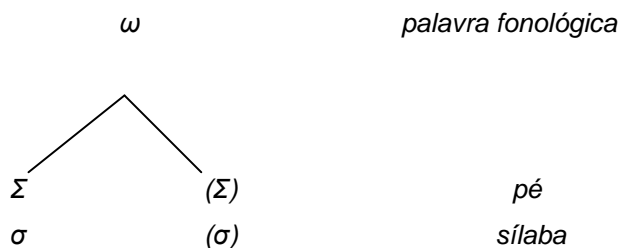


Figura 2.2: Hierarquia prosódica proposta por Nespor e Vogel (1986)¹⁰
 Fonte: Nespor e Vogel (1986).

⁹ Trecho original: "Prosodic phonology is a theory of the way in which the flow of speech is organized into a finite set of phonological units. It is also, however, a theory of interactions between phonology and other components of the grammar...While the division of the speech chain into various phonological units makes reference to structures found in the other components of the grammar, a fundamental aspect of prosodic theory is that the phonological constituents themselves are not isomorphic to any constituents found elsewhere in the grammar."

¹⁰ Por razões de economia de espaço, apresentamos somente os níveis hierárquico-prosódicos que são relevantes para nossa análise.

Já para Mateus (2005, p. 7), “O acento e os constituintes prosódicos – sílaba, palavra prosódica e sintagma entoacional – resultam das propriedades prosódicas dos sons e contribuem para a organização do enunciado”. Como se pode observar, Mateus não inclui o “pé” na sua hierarquização, conforme mostra a figura 2.3:

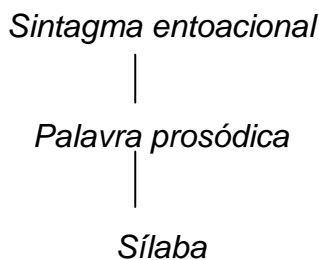


Figura 2.3: Hierarquia dos constituintes prosódicos.
Fonte: Mateus (2005).

2.1.1 A sílaba

A sílaba é uma unidade indispensável nos estudos de prosódia. Ela é a menor unidade nas quais as propriedades acústicas e estruturas temporais da fala estão integradas dentro de uma estrutura coerente com um gesto articulatorio subjacente. Uma unidade silábica possui propriedades relacionadas a outras unidades maiores (palavras) e menores (traços, fonemas) do que a sílaba; é a unidade central que tem descrições estruturais que se combinam ao longo da produção da fala.

Depois do aparecimento das Fonologias Não-Lineares¹¹, a sílaba passou a ter um tipo diferente de tratamento: o que antes era concebido como algo intuitivo e facilmente dividido conforme a quantidade de vogais passou a obedecer a regras e a padrões previamente estabelecidos por teóricos defensores de uma hipótese silábica. Do ponto de vista fonético, Callou e Leite (2003, p. 31) definem:

[...] A sílaba é um acréscimo da pressão do ar expelido dos pulmões pela atividade de pulsação dos músculos respiratórios que faz com que a saída do fluxo de ar não seja contínua, mas em jatos sucessivos. A cadeia sonora é composta por aclives, ápices e declives de sonoridade. O ápice da sílaba é ocupado por vogais enquanto os aclives e declives determinam

¹¹ Também chamada de fonologia autosegmental. A Fonologia Métrica e a Fonologia Prosódica são exemplos dessas fonologias de base gerativista. A referência à sílaba enquanto unidade linguística é praticamente inexistente nos modelos fonológicos que precedem à fonologia autosegmental.

as fronteiras silábicas, onde ficam as consoantes. Assim dizemos que as vogais ocorrem como núcleos silábicos e consoantes como margens.

Para a descrição da sílaba na língua portuguesa, Mateus (2005) utiliza o modelo europeu de Ataque-Rima. De acordo com a autora, “neste modelo, a unidade *sílaba* não domina diretamente as unidades segmentais: estas unidades segmentais agrupam-se em constituintes de planos intermédios, que são dominados pelo nó máximo, *i.e.*, o nó sílaba.” Observe-se a representação do modelo da estrutura interna da sílaba na figura 2.4:

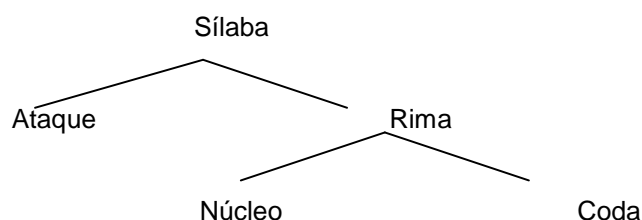


Figura 2.4: Os constituintes silábicos.
Fonte: Mateus, 2005.

Assim, de acordo com Mateus (2005, p. 77), no modelo de ‘Ataque-Rima’, a sílaba é uma unidade prosódica hierarquicamente organizada em constituintes silábicos:

- (i) no nível (ou fiada) da sílaba, o nó sílaba (σ) domina os constituintes Ataque (A) e Rima (R);
- (ii) no nível (ou fiada) da Rima, o Ataque e a Rima são nós irmãos, sendo que a Rima é constituída por um Núcleo (Nu) e por uma Coda (Cd)¹²;
- (iii) os constituintes terminais (Ataque, Núcleo e Coda) estão associados a posições rítmicas, ou seja, a posições de esqueleto, no nível (ou fiada) do esqueleto;
- (iv) os *constituintes podem ramificar* em duas posições; no caso de um constituinte terminal, a não ramificação corresponde a uma posição no nível do esqueleto e a ramificação corresponde a duas posições no nível do esqueleto;
- (v) cada posição rítmica ou de esqueleto pode ou não estar associada a um *nó Raiz* (que introduz a representação de um segmento em Geometria de Traços), no nível segmental.

Ladefoged (1993, p. 249) reconhece a relação do acento com a sílaba, já que “ele se aplica não só para indicar vogais e consoantes mas para a sílaba toda – o que quer que ela venha a ser. Uma sílaba acentuada é pronunciada com uma

¹² Nestes casos, diz-se que o constituinte Rima é ramificado.

quantidade maior de energia do que uma não-acentuada.”¹³

2.1.2 O acento

O acento é o único elemento prosódico responsável pela marcação de classe gramatical em língua portuguesa e em língua inglesa.

No aspecto fonético, acento se refere ao grau de intensidade empregado na produção de uma sílaba (CRYSTAL, 2000). Quanto ao aspecto fonológico, o acento é uma proeminência percebida como funcional, capaz de distinguir significados. Assim como existem sílabas átonas, existem palavras átonas, que são designadas como clíticas. Observe-se o que Xavier e Mateus (1990) afirmam sobre acento:

Grau de proeminência de uma vogal ou sílaba numa determinada sequência fonética. De um modo geral, faz-se a distinção entre vogais (ou sílabas) acentuadas e não acentuadas, considerando que as primeiras são mais proeminentes que as últimas. Esta proeminência pode ser devida a um aumento de intensidade (acento de intensidade), de duração (acento de quantidade) ou de altura (acento de altura), ou ainda a uma conjugação destas três propriedades. As sequências fonéticas portadoras de acento podem ser palavras, constituintes ou frases, distinguindo quaisquer destes níveis podem admitir-se diferentes graus de proeminência (graus de acento).

O acento pode ser computado no nível da palavra (acento lexical) e no nível da frase (acento frasal). Moraes (1995) explica que o acento lexical é inerente a cada vocábulo contendo mais de uma sílaba cuja localização pode eventualmente permitir a oposição entre palavras idênticas quanto a sua composição segmental.

Em línguas em que sua localização não é fixa, como o português, ele assume uma função semântica, responsável pela distinção de conteúdos referenciais, como em *sábida/sabida/sabiá*. No nível da frase, percebe-se que apenas algumas poucas sílabas são sentidas como efetivamente proeminentes. Esse acento frasal exerce “função sintática por dar uma maior coesão aos elementos por ele dominados, definindo e delimitando assim uma determinada extensão de cadeia sonora comumente designada por expressões como “grupo rítmico”, “grupo prosódico” (MORAES, 1995, p. 42). Albano et al. (1998, p. 1) atribuem

¹³ Trecho original: “It applies not to indicate vowels and consonants but to whole syllables – whatever they might be. A stressed syllable is pronounced with a greater amount of energy than an unstressed syllable.”

características acústicas específicas a cada um desses acentos: “Os acentos lexicais alongam não só a vogal, mas também a consoante da sílaba acentuada, enquanto os acentos frasais alongam também a consoante seguinte à vogal acentuada.”

Com base na afirmação acima, pode-se depreender que o componente acústico-temporal é de fundamental importância para a computação e manutenção do acento tanto no nível da palavra quanto no da frase. Quanto à intensidade, é conveniente levar em conta a extensão do segmento fonológico no contexto da fala a fim de se compreender a existência de níveis de acento relacionados à proeminência silábica, como tratará a próxima seção.

2.1.3 Níveis de acento

Convencionou-se criar denominações para os níveis de intensidade que uma proeminência desempenha na palavra ou na frase. Dessa forma, classificam-se essas intensidades da seguinte maneira, de acordo com Roach (1991, p. 78): “acento primário (o mais forte, com proeminência resultante de um movimento de *pitch*), acento secundário (mais fraco que o acento primário, mas mais forte que o não acentuado) e não acentuado (sem nenhuma proeminência notável)”. Ainda registra-se na literatura sobre tonicidade de língua inglesa o acento terciário, porém, não é dada muita atenção aos seus efeitos e funções na sentença.

Xavier e Mateus (1990, p.87) definem acento primário como “acento mais forte da palavra ou da frase; proeminência relativa apresentada pela sílaba acentuada (exemplo: CASas)”. Enquanto o acento primário de palavra (ou principal) é governado por princípios morfológicos e sua atribuição ocorre no nível lexical, o acento secundário se faz em um nível pós-lexical (sintático) “por ter em conta processos como a supressão de vogais átonas e permitir variação decorrente do contexto fônico” (MATEUS, 2005, p. 70). O acento primário não é dado metricamente, mas localmente, ou seja, especificamente, no âmbito da palavra. Enquanto a atribuição de acentos secundários deriva claramente do ritmo, a atribuição de acentos primários se dá localmente, no léxico. Assim, serão tratados o acento primário e o acento secundário como objetos de análise distintos.

Xavier e Mateus (1990, p. 88) definem acento secundário como o “acento que incide sobre uma vogal que não recebe o acento principal. Em português, só as

vogais pré-tônicas podem apresentar acento secundário no domínio da palavra. Cagliari (1995, p. 381) faz um resumo da trajetória que o acento secundário fez durante o último século:

Para a morfofonêmica, o acento era apenas um contexto. Para a fonologia gerativa padrão, um enfraquecimento na composição ou derivação de palavras. Para a fonologia lexical, trata-se de uma regra pós-lexical que não tem nada a ver com a regra de atribuição do acento primário. Para a fonologia métrica, trata-se de uma regra de eurrítmia que regula a saliência das sílabas, fora a sílaba tônica principal – ou seja, o acento secundário se mistura com outras saliências na onda rítmica, não tendo um status próprio.

Com relação ao acento secundário no PB, Collischon (1994) propõe que ele seja atribuído ao componente pós-lexical, através de um algoritmo¹⁴ que, sobre a estrutura previamente construída de acento primário (constituente ilimitado de cabeça à direita), constrói constituintes binários de cabeça à esquerda, obedecendo a uma direção (Direita-Esquerda), a partir do acento primário, e tomando como domínio de aplicação a palavra. O algoritmo de acento secundário seria, segundo a autora, insensível ao peso silábico¹⁵, cabendo-lhe, simplesmente, preencher a grade métrica e marcar o ritmo.

Sobre a inter-relação das propriedades rítmicas desencadeadas pelos níveis de acento, Abaurre e Galves (1998, p. 389) atestam:

O fato de que os acentos, primários ou secundários, são vistos como propriedades de domínios que podem ter apenas um acento, tem como consequência: 1) que a existência de um acento primário em uma sílaba implica sua ausência em todas as demais sílabas do mesmo domínio (i.e., no domínio da palavra, o que significa dizer que cada palavra é portadora de apenas um acento primário); 2) que a existência de um acento em uma sílaba implica sua ausência nas sílabas imediatamente adjacentes (ou seja, no interior do mesmo pé, o que significa dizer que cada pé é portador de apenas um acento secundário).

Geralmente mal interpretado por aprendizes brasileiros, por seu teor de intensidade em relação ao acento primário ser quase imperceptível, o acento

¹⁴ Algoritmo: sequência pré-definida de instruções que visam a alcançar sempre o mesmo resultado. Fonte: FERREIRA NETTO, Waldemar. Introdução à fonologia da língua portuguesa. São Paulo: Hedra, 2001.

¹⁵ O peso silábico é representado através da quantidade de moras. A mora (do latim, pausa, atraso). Em geral, uma sílaba formada por uma consoante e uma vogal breve contém uma mora somente, já que as consoantes em posição de ataque silábico não representam nenhuma mora. Já nas sílabas que contêm vogais longas ou ditongos, diz-se que essa sílaba possui duas moras.

secundário de língua inglesa recebeu pouca atenção nos estudos de Fonologia da Interlíngua. O acento secundário da língua inglesa, como na língua portuguesa, também contribui para a marcação do ritmo nas sentenças. Ladefoged (1993, p. 89) explica que, em longas palavras do inglês, existe mais de um nível de acento, que aparece mais comumente quando a palavra é dita de forma isolada ou em fim de sintagma. Isso se deve a fatores rítmicos, como é declarado por ele abaixo:

[...] a última sílaba acentuada num grupo tonal geralmente carrega o acento tônico. Em palavras longas que contêm dois graus de tonicidade, a aparente diferença nos níveis do acento primário e do acento secundário é realmente devido à superimposição de um padrão entoacional. Quando essas palavras ocorrem dentro de uma sentença em posição onde não existem efeitos entoacionais, então existirão diferenças nos níveis de acento. (p. 115)¹⁶

Como se pode perceber, na ausência de oposições entoacionais fortes o acento secundário desempenha uma função de assistente para a complementação do padrão rítmico-entoacional na cadeia da fala. Diante disso, é inegável a importância do acento secundário nas duas línguas aqui descritas. O que se espera quando da análise da aquisição do padrão acentual de inglês por aprendizes brasileiros é uma mínima transposição dos padrões de L1 para L2 no que concerne à computação do acento secundário. Todavia, alguns estudos de aquisição do padrão acentual do inglês (ARCHIBALD, 1994; SILVA, 2005; MAIRS, 1989) para citar como exemplos mais recentes não corroboram este fato.

Nas duas próximas seções serão apresentadas as principais diferenças entre os sistemas acentuais da língua portuguesa e da língua inglesa.

2.1.4 O acento em PB

Uma definição alternativa para acento ligada à Fonologia Não-linear é proposto por Mateus (2005, p. 58):

¹⁶ Trecho original: "...the last stressed syllable in a tone group usually carries the tonic accent. In longer words containing two stresses, the apparent difference in the levels of the first and the second stress is really due to the superimposition of an intonation pattern. When these words occur within a sentence in a position where there are no intonation effects, then there are no differences in the stress levels."

O acento em português resulta da combinação das propriedades de intensidade e duração do som vocálico que, em determinado ponto da sequência, apresentam valores relativamente mais elevados e marcam uma sílaba mais “forte” ou proeminente na palavra.

O PB é uma língua sensível à quantidade silábica na atribuição do acento principal de palavra e constrói seus constituintes prosódicos¹⁷ (pés métricos binários) de cabeça à esquerda, ou seja, o PB é uma língua de troqueau moraico (BISOL, 1994; MASSINI-CAGLIARI, 1999; MATEUS, 2005). Bisol (1994, p. 25), adotando o peso inerente da rima ramificada de Hayes (1981), admite que o acento em língua portuguesa seja atribuído da seguinte forma:

1. Regra do acento primário

Domínio: a palavra

- i. Atribua um asterisco (*) à sílaba pesada final, i.e., sílaba de rima ramificada;
- ii. Nos demais casos, forme um constituinte binário (não iterativamente) com proeminência à esquerda, do tipo (* .), junto à borda direita da palavra.

Bisol (1994) refere-se às sílabas pesadas como as que têm duas posições preenchidas na rima e leves apenas as que têm uma posição preenchida na rima. Bisol (1994, p. 25) faz uma ressalva quanto ao domínio de palavra:

[...] há uma diferença de estrutura morfológica entre verbos e não-verbos, em se tratando de nomes e adjetivos, a palavra fica entendida como radical + vogal temática. Em se tratando de verbos, a palavra fica entendida como radical + vogal temática + sufixo modo-temporal + sufixo número-pessoa, pois em qualquer um desses morfemas pode incidir o acento.

Contrariando as regras para atribuição acima, Lee (1994) apresenta uma alternativa para a regra do acento em português, ao afirmar que o acento só opera na raiz (radical), e não no domínio do nível de palavra, como afirmou Bisol (1994). Lee (1994) continua seu argumento afirmando, também, que, a partir de uma redução dos níveis de aplicação da regra de acento em PB e com o objetivo de minimizar o uso da Extrametricidade, existem somente duas regras de acento – uma para verbo e outra para não verbo. Lee (1994) ainda sustenta que a regra do acento do verbo aplica-se no domínio da palavra, enquanto a regra do não-verbo aplica-se no domínio do radical, e que ambas as regras não são sensíveis ao peso silábico,

¹⁷ Bisol (1994) afirma que o PB estrutura suas sílabas em pés métrico-binários de cabeça à esquerda, assim representados (* .), em que o asterisco representa a sílaba dominante e o ponto a sílaba dominada.

fato que vai de encontro a outros modelos de descrição do PB, como é o caso de Bisol (1994; 1999), Collischon (1999) e Mateus (2005).

O acento em PB, embora seja governado por algumas regras, tem posição relativamente livre e poderá recair em qualquer uma das três últimas sílabas do vocábulo, o que é chamado por Bisol (1994) de Restrição da Janela de Três Sílabas (RJT). Essa regra, que funciona como uma espécie de filtro acentual, admite que o acento só poderá recair em alguma das três últimas sílabas de palavra, fato esse enfatizado pela existência classificatória de tonicidade de palavras como “oxítona, paroxítona e proparoxítona”. As palavras do inglês que sofrem influência do padrão acentual do PB estão no apêndice E.

Muitas palavras possuem a mesma qualidade vocálica, quantidade e disposição linear de segmentos e suas categorias lexicais se distinguem apenas pela variação da posição do acento, que pode ser ou não representado graficamente, e elas são chamadas de pares mínimos, como se pode observar nos exemplos do quadro 2.2:

Palavra	Categoria lexical	Palavra	Categoria lexical
Referência	Substantivo Feminino	Referencia	3ª pessoa do singular do presente do indicativo; 2ª pessoa do singular do imperativo afirmativo do verbo Referenciar
Pronúncia	Substantivo Feminino	Pronuncia	3ª pessoa do singular, do presente do indicativo; 2ª pessoa do singular do imperativo afirmativo do verbo Pronunciar
Início	Substantivo Masculino	Inicio	1ª pessoa do singular do presente do indicativo do verbo Iniciar
Estudará	3ª pessoa do singular do Futuro do Presente do verbo Estudar	Estudara	1ª / 3ª pessoas do singular do pretérito mais-que-perfeito do verbo Estudar
Princípio	Substantivo Masculino	Principio	1ª pessoa do singular do presente do indicativo do verbo Principiar
Análise	Substantivo Feminino	Analise	1ª e 3ª pessoas do presente do subjuntivo; 2ª pessoa do singular do imperativo afirmativo e negativo do verbo Analisar
Empresário/a	Substantivo	Emprasarío/	1ª pessoa do singular e 3ª

	masculino/feminino	Empresaria	peessoa do singular do presente do indicativo, respectivamente, do verbo Empresariar
Ortógrafo	Substantivo Masculino	Ortografo	1ª pessoa do singular do presente do indicativo do verbo Ortografar
Prática	Substantivo Feminino	Pratica	3ª pessoa do singular do presente do indicativo; 2ª pessoa singular do imperativo afirmativo do verbo Praticar
Auxílio	Substantivo Masculino	Auxilio	1ª pessoa do singular do presente do indicativo do verbo Auxiliar
Número	Substantivo Masculino	Numero	1ª pessoa do presente do indicativo do verbo Numerar
Dúvida	Substantivo Feminino	Duvida	3ª pessoa do singular do presente do indicativo e 2ª pessoa do singular do imperativo afirmativo do verbo Duvidar
Última	Adjetivo Feminino (à exceção do caso em a palavra figura como substantivo no contexto coloquial sendo sinônimo de “novidade”, e.g., “Você sabe da última?”)	Ultima	3ª pessoa do singular do presente do indicativo e 2ª pessoa do singular do imperativo afirmativo do verbo Ultimar
Vômito	Substantivo Masculino	Vomito	1ª pessoa do singular do presente do indicativo do verbo Vomitar

Quadro 2.2: Lista de alguns pares mínimos em língua portuguesa.
Fonte: Elaboração da autora.

Ao se considerar que o acento nutre uma relação de dependência com a estrutura interna da palavra, tornando essa interface um fato de natureza morfológica, é possível formalizar algumas regras possuindo poucas exceções. Mateus (2005) e Lee (1994) formulam regras de acento cujo domínio é o radical, ou seja, é acentuada a última vogal do radical ou a vogal dos sufixos não marcados (CAGLIARI, 1999). Segundo Mateus (2005, p. 61), a regra para acentuação de nomes e adjetivos é a seguinte:

Acentuar a última vogal do radical. Restrição à regra geral de acento: Acentuar a penúltima vogal do radical nos itens com vogais marcadas no léxico como não-acentuáveis.

O que se percebe, através da formalização das regras acima, é que a vogal temática nunca interfere no acento, assim como o plural, que não tem peso algum sobre o acento.

Mateus (2005, p. 62-69) ainda fornece a regra de acentuação para verbos:

Nos tempos do presente em que a vogal temática foi mantida, o acento incide sobre a vogal temática (Infinitivos, Gerúndio e primeira pessoa do plural do Presente do Indicativo). Nos casos em que a vogal foi suprimida, o acento incide sobre a vogal que substitui a temática (primeira pessoa do plural do Presente do Conjuntivo); maioria dos casos, porém, o acento aplica-se na última vogal do radical; Nos Tempos do Passado o acento incide sempre na vogal temática; Nos Tempos do Futuro – Futuro do Indicativo e Condicional – o acento incide sempre na primeira vogal do sufixo.

Descritas as regras de acentuação de nomes e verbos na língua portuguesa, faz-se necessário expor um quadro dos itens lexicais utilizados nesta pesquisa, organizados por categoria lexical com seus respectivos padrões acentuais. O símbolo (') representa acento primário, conforme mostrado no quadro 2.3:

Substantivo	Verbo	Adjetivo
con'duta	condu'zir	--
ob'jeto	obje'tar	--
con'flito	conflitar (?)	confli'tante/ conflitu'oso
--	contes'tar	contestador / contestável
de'serto	deser'tar	desertante
separação	sepa'rar	sepa'rado / sepa'rável
advo'gado	advo'gar	--
apropri'ação	apropri'ar (-se)	a.pro.pri.'ado
estima'tiva/estima'ção	esti'mar	esti'mado
elabora'ção	elabo'rar	elabo'rado

Quadro 2.3: Padrão acentual dos itens lexicais organizados por categoria lexical.
Fonte: Elaboração da autora.

A observação do quadro 2.3 resulta nas seguintes conclusões: há uma tendência acentual dos substantivos trissílabos e polissílabos no PB serem

paroxítonos. Os verbos trissílabos e polissílabos no PB sofrem uma tendência a serem oxítonos. Há ainda outra tendência: a dos adjetivos trissílabos e polissílabos no PB serem paroxítonos.

Descritas as regras de acentuação de nomes e verbos no PB, passe-se a discutir a importância do acento em língua inglesa, com suas respectivas particularidades e contrastes com o PB.

2.1.5 O Acento em Inglês

O acento tem várias funções no inglês que oscilam entre um papel enfático, passando pelo contrastivo até poder indicar relações sintáticas entre palavras e partes de palavras, como é o caso de oposições de pares de palavras substantivo e verbo (*noun and verb*), tais como “*na insult* - /¹ɪnsʌlt/, *to insult* - /ɪm¹sʌlt/; *an overflow* - /¹oʊvərˌfləʊ/; *to overflow* - /¹oʊvərˈfləʊ/; *an increase* - /¹ɪnkriːs/; *to increase* - /ɪm¹kriːs/”.

O acento de palavra em inglês é livre, no sentido de gozar de mobilidade, e poderá cair em qualquer sílaba, diferentemente do que ocorre na língua portuguesa, em que acontece a Restrição da Janela de Três Sílabas (RJT). É comum haver a acentuação de itens na quarta sílaba, como é o caso de *comfortable*, *difficulty* e *favorite* (substantivos).

A língua inglesa tende a ser trocaica¹⁸ (*forte - fraco*) e sensível ao peso silábico. O pé geralmente é trocaico (o elemento mais forte à esquerda), mas algumas palavras possuem padrão iâmbico¹⁹ (o elemento mais forte à direita), o que permite separar palavras como *content/contént*, *object/objéct*, *permit/permít*.

Em sua discussão sobre acento do inglês, Katamba (1993) assume que este se trata de um elemento fonológico, que não é somente uma característica de consoantes ou vogais, mas uma propriedade da palavra em seu todo.

¹⁸ Troqueu (trochee) – definição: um pé métrico que contém uma sílaba acentuada seguida por uma sílaba não-acentuada, ou, em um metro quantitativo, de uma sílaba longa seguida por uma curta. Trecho original: definition: A metrical foot consisting of a stressed syllable followed by an unstressed syllable, or, in a quantitative metre, of a long syllable followed by a short. Fonte: TRASK, R. L. A dictionary of phonetics and phonology. London/New York: Routledge, 1996, p. 363.

¹⁹ Iâmbico – o fato de a proeminência principal localizar-se o mais à direita possível do domínio de aplicação do acento, seja ele o radical ou a palavra, sugere uma dominância à direita na formação dos constituintes prosódicos (pés) – ou seja, sugere um pé iâmbico.

Não existem regras infalíveis na determinação de sílabas que portarão o acento tônico em língua inglesa. De acordo com Prator e Robinett (1985) e Kreidler (1989), a regra geral para algumas palavras que funcionam como nomes e verbos é atribuir acento na primeira sílaba para os substantivos e adjetivos e na segunda sílaba para os verbos. Prator e Robinett (1985, p. 21-22) comentam acerca dos pares mínimos em língua inglesa:

[...] 3. Um grande grupo de palavras, que podem ser usadas tanto como substantivos ou verbos sem mudar a sua grafia, possui uma diferença na tonicidade para indicar a diferença em uso. Nesses casos, o substantivo possui o acento primário na primeira sílaba, o verbo na última... Os substantivos nesse grupo de palavras às vezes têm acento secundário na última sílaba: *increase* /'ɪnkriys/ e *overflow* /òwvɛrflòw/. Às vezes – como no caso de *concert* e *object* – o significado do substantivo tem pouca ou nenhuma relação com o verbo correspondente.²⁰

[...] 4. Generalizações úteis podem se feitas sobre o grande número de palavras polissilábicas em inglês que terminam em *-ate*. Algumas delas – tais como *duplicate* e *associate* – podem ser usadas como adjetivos, substantivos ou verbos; outras - tais como *consulate* (substantivo), *educate* (verbo), e *appropriate* (adjetivo ou verbo) – são usados somente de uma ou duas maneiras. No caso de todas as palavras desse grupo, entretanto, o uso como um adjetivo ou como um substantivo é indicado por desacentuar a vogal do fim de palavra e pronunciá-la como /ɪ / or / ə /. E.g.: *I must go to the consulate* /kánsəlɪt/. O uso como um verbo é indicado quando a vogal é pronunciada como /eɪ / com acento secundário final. E.g. *He'll duplicate* /dúwplɪkèyt/ the letter.²¹

Vale ressaltar um fato curioso em relação à acentuação secundária de um item da citação acima - *increase* /'ɪnkriys/, que gera conflito por estar em alguns dicionários (e.g.: *Longman Advanced American Dictionary*, 2000) como sendo uma lexia que não é portadora de acento secundário - /'ɪnkris/. Não se sabe ao certo a

²⁰ Trecho original “[...] 3. A large group of words, which may be used either as nouns or verbs without change in their spelling, have a difference in stress to indicate the difference in usage. In such cases, the noun has primary accent on the first syllable, the verb on the last.... The nouns in this group of words sometimes have secondary accent on the last syllable: *increase* /'ɪnkriys/ and *overflow* /òwvɛrflòw/. Sometimes – as in the case of *concert* and *object* – the meaning of the noun has little if any relationship to the meaning of the corresponding verb [...]”

²¹ Trecho original: “[...] 4. Helpful generalization can also be made about the large number of polysyllabic English words that end in *-ate*. Some of these – such as *duplicate* and *associate* – may be used as adjectives, nouns, or verbs; others – such as *consulate* (noun), *educate* (verb), and *appropriate* (adjective or verb) – are used in only one or two of these three ways. In the case of all words of this group, however, use as an adjective or a noun is indicated by leaving the vowel of the ending unstressed and pronouncing it as /ɪ / or / ə /. E.g. *I must go to the consulate* /kánsəlɪt/. Use as a verb is indicated by giving the ending secondary stress and pronouncing the vowel as /eɪ /. *He'll duplicate* /dúwplɪkèyt/ the letter.[...]”

que se deve tal fato. No entanto, o usuário da língua se adapta a essa idiossincrasia. Isto não se significa que tal informação também deva ser suprimida no processo de aprendizagem do falante de L2.

Pelo tratamento dado ao acento secundário em alguns dicionários de língua inglesa (às vezes, não existe nem a marcação dele em alguns vocábulos que portem o traço suprasegmental), observa-se uma leve negligência em relação a esse traço prosódico tão fundamental quanto o acento primário em alguns itens. O acento secundário desempenha uma função extremamente importante na determinação de componentes de par mínimo quando esses assumem a função de verbos, não podendo ser esquecido com tanta freqüência.

Em resumo, o posicionamento do acento primário ou secundário de um dado item lexical (componente de par mínimo) acontece em função de uma classificação morfológica que é gerada *ad hoc* no *continuum* da fala através do processamento sintático do restante dos componentes da sentença.

No PB, Massini-Cagliari (1992) aponta os principais correlatos acústicos do acento, no nível lexical e em ordem decrescente de importância, como duração, intensidade e qualidade vocálica.

No inglês, o termo 'acento' é a junção de três fatores perceptivos correlacionados a parâmetros acústicos: 1) quantidade/duração (ms) relacionada com o tamanho da sílaba; 2) intensidade (dB) relacionada à amplitude média alta; 3) altura (Hz), ou seja, o valor de F_0 mais elevado na elocução. A duração da vogal não é considerada no inglês; todavia, no PB, essa característica é fundamental para a determinação do acento de palavra.

Com base nas descrições dos parâmetros métricos do acento em PB e em inglês apresentados acima, o quadro 2.4 compara a atribuição do acento nas duas línguas:

Parâmetros	Inglês	Português
Parâmetro 1: A palavra é forte na:	esquerda	esquerda
Parâmetro 2: Os pés são:	binários	binários
Parâmetro 3: Os pés são construídos à:	direita	direita
Parâmetro 4: Os pés são fortes na:	esquerda	esquerda
Parâmetro 5: Os pés são sensíveis à quantidade?	sim	Sim
Parâmetro 6: Os pés são sensíveis à quantidade		

na:	rima	Rima
Parâmetro 8A: Existe uma sílaba extramétrica?	sim	Sim
P8: É extramétrico na:	direita	esquerda

Quadro 2.4: Determinação de Parâmetros Métricos do PB e inglês.
 Fonte: Adaptado de Archibald, 1998 e Massini-Cagliari, 1999, p. 97.

Como se pode observar, os padrões de atribuição do acento nas duas línguas são parecidos. Porém, ao se estudar a aquisição do acento de língua inglesa em aprendizes brasileiros, são verificadas evidências de transferência de parâmetros de L1 que influenciam na atribuição do acento de L2. Eventualmente, durante o processo de aquisição do sistema acentual de uma língua estrangeira, os erros cometidos pelos aprendizes podem referir-se à interferência de sua língua materna ou ser resultantes dos desvios possíveis em sua língua-alvo.

2.1.6 Relação entre vogal reduzida e acento

O conceito de vogal se revela de grande importância para a compreensão do fenômeno da atribuição do acento de palavra²². O acento determina a qualidade vocálica das sílabas ao seu redor por estabelecer uma relação íntima com as vogais da sílaba acentuada. Albano (2001) assevera que a tonicidade tende a atrair vogais baixas. Zimmer (2006) corrobora essa afirmação acrescentando que tanto no inglês como no PB quanto mais baixa a vogal mais longa ela tende a ser. Para Cunha e Cintra (1985, p. 36), antes de definir acento, é preciso ter em mente a noção do correlato físico intensidade e vogal relacionados:

A intensidade é a qualidade física da vogal que depende da força expiratória e, portanto, da amplitude da vibração das cordas vocais. As vogais que se encontram nas sílabas pronunciadas com maior intensidade chamam-se tônicas, porque sobre elas recai o acento tônico, que se caracteriza em português principalmente por um reforço de energia expiratória. As vogais que se encontram em sílabas não acentuadas denominam-se átonas.

Segundo Katamba (1993), entende-se por vogal reduzida uma vogal centralizada encontrada em posições não acentuadas de palavras e é geralmente

²² Acento de palavra (ou acento lexical): Acento que incide sobre a sílaba que, na palavra, apresenta proeminência relativamente às restantes. O lugar do acento de palavra (ou acento lexical) estabelece, em certas circunstâncias, distinção entre palavras. Exemplo: DUvida / duVida. Fonte: XAVIER, M. F. & MATEUS, M. H. (orgs.) Dicionário de termos lingüísticos, Vol I. Lisboa: Cosmos, 1990, p. 44).

representada por um *schwa* (vogal meio-central [ə] ou [ɪ]. Ladefoged (1993) destaca a importância do reconhecimento da vogal reduzida para a discriminação do acento, sob o ponto de vista do ouvinte. Segundo ele,

A coisa mais confiável para um ouvinte detectar se uma sílaba é tônica é se ela possui uma vogal mais longa do que ela seria se estivesse desacentuada. Mas isso não significa que todas as vogais longas são necessariamente acentuadas. A segunda e terceira vogal em *radio*, por exemplo, são comparativamente longas, mas elas não têm um impulso extra de ar dos pulmões que ocorre na primeira vogal. (p. 114)²³

A duração da vogal não é considerada no inglês; todavia, no PB, essa característica é fundamental para a determinação do acento de palavra. Segundo Cristófaros-Silva (2001, p. 77), as vogais acentuadas ou tônicas carregam o acento mais forte ou acento primário, e as vogais não acentuadas - átonas pretônicas ou postônicas - carregam acento secundário ou são completamente isentas de acento.

Prator e Robinett (1985) são específicos em relação à qualidade vocálica na língua inglesa e enumeram as vogais que podem ser pronunciadas nas sílabas tônicas: [i], [ɪ], [o], [ɔ], [e], [ɛ], [æ], [a], [ʊ], [u], [ə], [aʊ], [aɪ] e [ɔɪ], enquanto que a vogal na sílaba átona é quase sempre [ə] ou [ɪ], e alguns falantes ainda a pronunciam como [u]. No inglês, para deixar mais clara a distinção entre substantivos, adjetivos e verbos, no caso de palavras que contêm sufixo *-ate*, há a redução da vogal átona nos adjetivos e substantivos. Quando se muda de um adjetivo ou substantivo para um verbo, o acento permanece na mesma posição, mas a vogal “a” de *-ate* é reduzida completamente, enquanto que na pronúncia dos verbos não há a redução, mas sim a mudança para /eɪ/. Uma possível explicação para essa redução em direção ao som neutro /ə/ é consequência da aglutinação de sílabas não acentuadas: e.g.: *advocate* /'ædvəkət/ e *to advocate* /'ædvə,keɪt/, *animate* /'ænəməɪt/ e *to animate* /'ænə,meɪt/.

²³ Trecho original: “The most reliable thing for a listener to detect is that a stressed syllable frequently has a longer vowel than that same vowel would be if it were unstressed. But this does not mean that all long vowels are necessarily stressed. The second and third vowels in “radio”, for example, are comparatively long, but they do not have the extra push of air from the lungs that occurs on the first vowel.”

Um estudo sobre redução vocálica em preposições de língua inglesa foi desenvolvido por Watkins (2001). O autor focalizou a variabilidade no uso das vogais reduzidas em clíticos pronunciados em sentenças por aprendizes brasileiros de nível avançado de proficiência em língua inglesa. O autor aponta que os aprendizes, muito embora façam uso de vogais reduzidas ao falarem em inglês, não o fazem de maneira consistente. Em um prévio estudo piloto, o autor identificou que as palavras-função (*function words* – ver mais sobre isso na seção Tom e Entoação), as sílabas fechadas e as sílabas precedidas por uma sílaba métrica fraca eram mais propensas a resistir à redução vocálica. Esse resultado o levou a prever que fatores fonológicos, tais como a estrutura interna da sílaba bem como o ambiente segmental e métrico da palavra, poderia possivelmente influenciar na variação do uso das formas fracas das palavras-função. Sobre essa sistematicidade de emprego da vogal reduzida, Watkins conclui em seus achados que:

Ou os falantes não estão conscientes de quais formas eles estão usando (a forte ou a fraca), ou eles estão conscientes, mas possuem uma concepção mal desenvolvida do papel fonológico da redução vocálica, falhando sempre ao dar a distinção e a importância da qual ela tem para os falantes nativos.²⁴

A literatura sobre ensino de pronúncia de língua inglesa reporta que o mau uso da vogal reduzida e do acento secundário acarreta grandes prejuízos na cadeia da fala no que se refere à entoação, que constitui o próximo tópico a ser abordado.

2.2 Tom e Entoação

A entoação corresponde às variações de pitch (correlato perceptual frequência fundamental) percebidos por um ouvinte, que, por sua vez, correspondem às variações de F_0 . Uma análise da entoação é desempenhada ao se considerar padrões de *pitch* em termos de contorno, para o qual a média de *pitch*, a altura e a mudança de direção são geralmente caracterizadas (DELLER et al., 2000). Halliday

²⁴ Trecho original: “Either speakers are simply not aware of which form they are producing (the strong or the weak), or they are aware but have an underdeveloped concept of the phonological role of vowel reduction, failing to give the distinction the importance which it has for native speakers.”

(1961, p. 64) argumenta que “nós não podemos descrever totalmente a gramática do inglês falado sem fazer referência ao contraste demonstrado pela entoação.”

A entonação desempenha várias funções úteis na linguagem, sendo a mais importante a sinalização da estrutura gramatical. Um segundo papel da entonação é dar características secundárias, tais como atitude ou emoção, como a raiva, o sarcasmo, a ironia, entre outros. Seus correlatos acústicos são a frequência fundamental, a duração e a intensidade. Bolinger (1985) confere à entonação uma natureza gramatical e também uma emocional. Quanto ao caráter semântico da entonação no PB, Orsini (2005, p. 7) destaca:

Por ser um signo linguístico, associa forma e conteúdo tanto no nível linguístico (gramatical) quanto no extralinguístico (expressivo). Além disso, fornece tanto informações de natureza sintática quanto aquelas pertinentes à ordenação das unidades informacionais do enunciado. Seu domínio é a frase.

A entoação é ao mesmo tempo um traço universal e específico das línguas. É considerada universal por manifestar, em diferentes línguas, comportamentos entoacionais semelhantes. Algumas análises prosódicas translinguísticas mostram, por exemplo, que a presença *versus* a ausência da queda do valor da frequência fundamental ao longo do enunciado diferencia as sentenças declarativas das interrogativas. Podem ser considerados traços prosódicos universais: (a) elevação de F_0 nas primeiras sílabas da sentença; (b) repetição de padrões melódicos – ascendente ou descendente – ao longo da sentença; (c) tendência de o valor máximo da F_0 localizar-se na primeira palavra da sentença e (d) aumento da duração da última sílaba do grupo prosódico, bem como do primeiro fonema da sentença (VAISSIÈRRE, 1983 *apud* ORSINI, 2005).

Tratando-se de um traço específico, uma vez que a manifestação de uma curva ascendente ou descendente obedece a critérios que variam segundo a língua, a combinação de elementos agregados à entoação determina sua natureza. O grau de relevância de fatores tais como o limite de palavra, a estrutura morfológica e fonológica da palavra e os fenômenos de juntura²⁵ na determinação do contorno melódico variam entre as línguas. Para Vaissière, é possível que a inter-relação dos

²⁵ Segundo Crystal (1985, p. 100), “juntura é um termo usado na fonologia para indicar os traços fonéticos de fronteira que podem demarcar as unidades gramaticais, como os morfemas, palavras ou orações”.

parâmetros acústicos - F_0 , duração e intensidade - sejam responsáveis pela individualização das línguas. Hirst e Di Cristo (1998) corroboram essa afirmação explicando que:

A diferença prosódica entre as línguas origina-se do fato de os primitivos prosódicos poderem ser introduzidos dentro da representação fonológica em diferentes níveis. Quando as características prosódicas de uma palavra não são lexicalmente especificadas, elas precisarão ser introduzidas por regras que são adotadas para converter uma representação subjacente numa representação superficial (HIRST; DI CRISTO, 1998, p. 8).

Um tom ascendente, uma porção da entonação que caracteriza seu contorno, seguido de pausa, pode marcar toda uma sequência de palavras como sendo uma frase interrogativa. Da mesma forma, o tom descendente com pausa caracteriza a frase declarativa. Estabelecidos os padrões universais da entonação, apresentar-se-ão, na próxima seção, os padrões entoacionais das duas línguas aqui contrastadas, o português e o inglês.

2.2.1 Padrões entoacionais do PB

Moraes (1982) explica que a frequência fundamental é o traço mais significativo para a determinação do padrão entoacional de um enunciado em PB; trata-se de um parâmetro acústico, percebido, pelos interlocutores, como altura de voz (isto é, variações melódicas, na dimensão grave/agudo).

No PB, existe uma tendência à focalização no fim de sentença. Seus padrões entoacionais são similares aos do Português Europeu, com a exceção de que os falantes deste preferem uma curva entoacional descendente-ascendente nas perguntas afirmativas/negativas (*yes-no questions*), enquanto que os brasileiros utilizam um contorno ascendente sobre a sílaba acentuada e um contorno descendente na posição pós-tônica. Não se sabe ao certo o que motiva os dois usos. Alguns poucos estudos buscaram a elucidação deste fato e mesmo assim ainda falharam ao oferecer uma explicação plausível.

Em seus estudos descritivos do padrão entoacional do PB, Moraes (1999) advoga a favor de um modelo SV (sujeito-verbo) que privilegia o primeiro evento tonal alinhado com o início da tônica do verbo. Há uma tendência a atribuir

contornos entoacionais perceptivamente mais audíveis às primeiras palavras de conteúdo, em posição inicial da frase.

Estudos sobre a descrição perceptiva da entoação da frase interrogativa em português escolhem a variante mineira como a mais confusa em termos de interpretação semântica devido à mudança de contorno que diferencia do padrão do PB.

2.2.2 Padrões entoacionais do inglês (norte-americano)

A entoação da língua inglesa está intimamente ligada ao acento frasal. Geralmente, uma sílaba com acento frasal é produzida com uma nota musical mais alta que as outras sílabas não acentuadas. Os encadeadores de perfil melódico são as sílabas acentuadas, ao menos que a necessidade de ênfase faça com que o falante coloque mais força no enunciado. Nas palavras mais longas, os padrões de acento lexical podem desviar-se dos oficiais (posição isolada), especialmente porque a primeira sílaba tende a receber a tonicidade.

Em inglês, um tom descendente final pode indicar o fim de uma declaração (função sintática), porém uma queda alta na mesma posição pode transmitir surpresa ou indignação; uma queda baixa indiferença ou desgosto (FRY, 1970, p. 367).

Prator e Robinett (1985) explicam que o artifício favorito para indicar foco no inglês é a entoação. É possível se entoar uma nota alta para qualquer elemento na frase devido a uma grande liberdade que existe para o posicionamento do acento enfático. Porém, alguns gramáticos dividem todas as palavras da língua em palavras-conteúdo (*content words*²⁶) - que possuem significado por si só - e em palavras-função (*function words*), cujo significado pode ser pouco ou nem sequer ser transmitido. As palavras-conteúdo geralmente devem ser acentuadas numa frase e são elas: substantivos, verbos, adjetivos, advérbios, pronomes demonstrativos e pronomes interrogativos. As palavras-função são os artigos, as preposições simples,

²⁶ Desde o nascimento, a criança é sensível à estrutura prosódica da língua, o que lhe permite desempenhar uma primeira discriminação crucial entre palavras de conteúdo e palavras função ao se adquirir a estrutura da sua língua (SHI et al., 1999).

pronomes pessoais, possessivos adjetivos, pronomes relativos, conjunções comuns e alguns verbos auxiliares (PRATOR; ROBINETT, 1985).

As pausas são uma forma de dividir a sentença em grupos tonais. Um grupo tonal é uma parte da frase separada das outras por uma pausa, geralmente representada numa transcrição como uma barra (/). A inserção da pausa cumpre três funções primordiais: 1) esclarecer a frase; 2) enfatizar ou 3) simplesmente permitir que o falante possa respirar. A pausa dificilmente será colocada numa posição que comprometa a inteligibilidade da sentença, isto é, não é comum a inserção de pausas entre artigos e adjetivos ou entre os verbos auxiliares e os verbos principais que eles acompanham.

Bolinger (1985) utiliza um tipo de notação diferenciado das convenções já vistas para reproduzir os contornos entoacionais de língua inglesa. Seus parâmetros para detectar os padrões entoacionais são: interesse (informação nova), força e pausas. As perguntas tendem a ter um nível mais alto e as exclamações favorecem um nível ascendente seguido de queda do contorno melódico.

Para marcar a entoação, a maioria dos sistemas escolheu o sistema simplificado de notação que divide os tons em quatro tipos: normal, alto (*high - H*), baixo (*low - L*) e extra-alto (*extra-high - H**). Podem-se mostrar os movimentos da voz subindo e descendo, desenhando linhas em quatro níveis diferentes. Uma linha desenhada na base das letras de uma palavra indica que a palavra é pronunciada num tom normal; uma linha acima da palavra marca um tom alto; uma linha que se distancia um pouco para baixo indica que a palavra está num tom baixo, e uma linha que se levanta sobre a palavra marca um tom extra-alto.

Geralmente o movimento de um tom para outro acontece entre sílabas e é chamado de *shift* (mudança). Essa mudança é indicada por uma linha reta vertical (ex.: HOW ARE YOU?). Às vezes, no entanto, a voz se modifica de um tom para outro ao pronunciar uma sílaba; esse movimento dentro de uma sílaba é marcado por uma linha curva para cima ou para baixo, chamada de *slide* (deslize). No exemplo ALL DAY LONG, a pronúncia de LONG começa com uma nota mais alta que o normal, e então a voz desliza para baixo numa nota mais baixa que o normal antes do fim da sílaba.

É no final da frase que os nativos de língua inglesa usam a entoação mais uniformemente. Nessa posição, em certos tipos de frase, a voz geralmente sobe

acima do normal e cai também abaixo do normal. Esse é o chamado padrão entoacional ascendente-descendente. A nota mais alta normalmente coincide com a última sílaba acentuada. Observem-se os exemplos abaixo:

____ THE CÓFFEE IS HÓT.



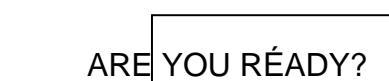
____ WHÁT TÍME DID YOU CÁLL?



O movimento deslizante de uma nota para outra dentro de uma única sílaba acentuada significa que a vogal da sílaba será tão alongada que poderá ser transformada em duas vogais levemente diferentes – um ditongo. Essas duas sílabas com tons diferentes e a ditongação resultante constituem uma das mais importantes diferenças entre o Inglês e muitas línguas. Neste exemplo, entoação e formação vocálica se encontram²⁷. A entoação ascendente-descendente é normalmente usada no fim de declarações simples sobre fatos (frases declarativas), comandos e perguntas que começam com uma palavra interrogativa (*what, who, which, when, etc*), ou seja, perguntas que começam com “*wh*” (*wh-questions*).

Outro padrão entoacional muito comum é o ascendente, usado normalmente no fim de frases que não começam com *wh-questions*, ou seja, perguntas de caráter geral que podem ser respondidas somente com sim ou não. Notem-se os exemplos a seguir:

____ ARE YOU RÉADY?



____ WILL YOU RÉAD IT FOR ME?



²⁷ Para fins de comparação, a ditongação no PB parece ser independente da tonicidade, podendo cada ditongo ser analisado como dois segmentos, pois a duração está associada com os graus de acentuação, sem ser lexical. Como mesmo diz Câmara Jr. (1977), o segmento é abstração, sendo discutíveis tanto quantos segmentos há em uma cadeia, quanto onde fica o limite entre os segmentos.

A voz normalmente é elevada para uma nota mais alta no último acento frasal, como acontece no padrão descendente-ascendente. A diferença entre esse dois padrões reside no fato de que, na entoação ascendente, as sílabas que seguem a subida são pronunciadas numa nota alta também. Ao fazer isso, o ouvinte tem uma sensação de não-finalização da mensagem, em oposição à sensação de finalização encontrada tipicamente nas frases de padrão ascendente-descendente.

2.3 Resumo do Capítulo

Este capítulo apresentou de forma sumarizada os principais fundamentos teóricos concernentes aos elementos prosódicos acento – primário e secundário - e entonação, sílaba e vogal reduzida nas duas línguas (PB e inglês).

O texto também trouxe um apanhado das teorias prosódicas, com seus elementos estruturais e a forma como estão organizados hierarquicamente. Demonstrou também a importância dos constituintes silábicos para a concretização da computação acentual – em seu nível primário e nível secundário - na cadeia entoacional da fala.

Algumas regras acentuais do PB e do inglês foram listadas a fim de gerar um contraste entre os dois sistemas acentuais para a compreensão dos erros de transferência de uma língua para outra.

O próximo capítulo apresentará os principais fundamentos de processamento da fala a fim de explicar como o sinal de fala, após ser numericamente codificado, pode ser usado como ferramenta de análise da linguagem e da prosódia.

3 FUNDAMENTOS DE PROCESSAMENTO DA FALA

O presente capítulo apresenta os principais tópicos para a compreensão do sinal acústico da fala, desde a sua geração, passando pelos principais elementos envolvidos na sua produção, até ele ser numericamente codificado e ser usado como ferramenta de análise da linguagem e da prosódia. Ainda apresenta-se neste capítulo noções sobre onda sonora, sinal com seus respectivos tempos (discreto e contínuo) e amplitudes, bem como os processos de digitalização de sinal analógico. O capítulo descreve também o processo de extração de características do sinal de fala e a parametrização deste e apresenta as técnicas de parametrização e processamento de sinal aqui utilizadas.

O sinal da fala é uma representação elétrica da pressão sonora acústica captada por um transdutor (microfone) que transforma a vibração sonora em ondas elétricas. Ao longo da história, todo o processamento de sinal da fala como filtragem, codificação, análise e síntese foi feito através de dispositivos analógicos, teoricamente concebidos e operados com grandezas contínuas, como tempo, voltagem, frequência etc. Hoje em dia, devido à evolução tecnológica, a maioria do processamento é feito através de equipamentos digitais de uso geral, como computadores pessoais, ou através de processadores digitais de sinais de uso específico (LILJENCANTS, 1997).

Após ser recebido pelo computador, o sinal da fala é submetido a diversas transformações até estar adequado para o uso da aplicação-alvo. A versão quantizada e em tempo discreto do sinal, após ter sido devidamente processada, serve como base para a extração das características e parametrização do sinal de fala que permitirá a sua identificação por sistemas de reconhecimento automático de voz (RAV). Há duas técnicas de parametrização bastante difundidas na extração de características da fala: codificação por predição linear (LPC) e coeficientes mel-cepstrais (MFCC).

As características do sinal extraídas por meio das técnicas LPC e MFCC podem ser utilizadas tanto para representação da energia e da frequência fundamental (F_0) - correlatos do acento – quanto para bandas de frequência. A análise da prosódia da fala só se torna possível graças aos dados provenientes da F_0 . As características extraídas do sinal de fala podem ser classificadas através de

modelos probabilísticos, a saber, Modelos Ocultos de Markov (HMMs - *Hidden Markov Models*) ou modelos não-paramétricos, tais como Redes Neurais Artificiais (ANNs - *Artificial Neural Networks*). Atualmente, as redes neurais artificiais substituem, satisfatoriamente e de forma subsimbólica (conexionista), os Modelos Ocultos de Markov e as árvores de *parser* na classificação de características do sinal de fala (Cf. MÜLLER, 2006).

3.1 O Processo de produção natural da fala

Fisiologicamente, a fala humana é gerada por meio de um complexo mecanismo de produção de ar oriundo da força dos músculos ao redor dos pulmões, músculos intercostais e diafragma, além da participação de estruturas anatômicas do trato vocal (faringe e cavidade oral), traqueia, laringe e cavidade nasal (narinas, fossas nasais, palato mole, véu palatino) e úvula. O ar que sai dos pulmões passa pelas pregas vocais, que se contraem ou se afrouxam emitindo sons, todo esse processo sendo subordinado a um comando cerebral.

O tecido mucoso ao redor das cordas vocais, ao ser acionado, transmite vibração às moléculas de ar que estão em contato com elas. Essa vibração é transmitida, de molécula em molécula, até atingir o nosso ouvido. A modulação, ou mudança de tom, inicia-se com as cordas vocais que propiciam a onda (ou frequência) fundamental (F_0) do sinal de fala. A F_0 é refletida no trato vocal, que vai da faringe até a cavidade nasal e ainda é modulada pela língua e pelos lábios. Como resultado, tem-se uma onda acústica complexa, com inúmeras informações acerca do locutor e da locução (MÜLLER, 2006).

Conforme mostrado na figura 3.1, as fossas nasais e o trato vocal constituem o principal filtro acústico da fala. O trato nasal começa no palato mole e termina nas narinas. A língua, os dentes (superiores e inferiores), os lábios (superiores e inferiores), o palato duro, os alvéolos e a úvula complementam a cavidade oral. Esses componentes anatômicos mais delicados se movem em diferentes posições para realizar os vários tipos de sons da fala e são conhecidos como articuladores. A mandíbula e o queixo também são considerados articuladores, já que são responsáveis pela movimentação dos outros articuladores (DELLER et al., 2000).

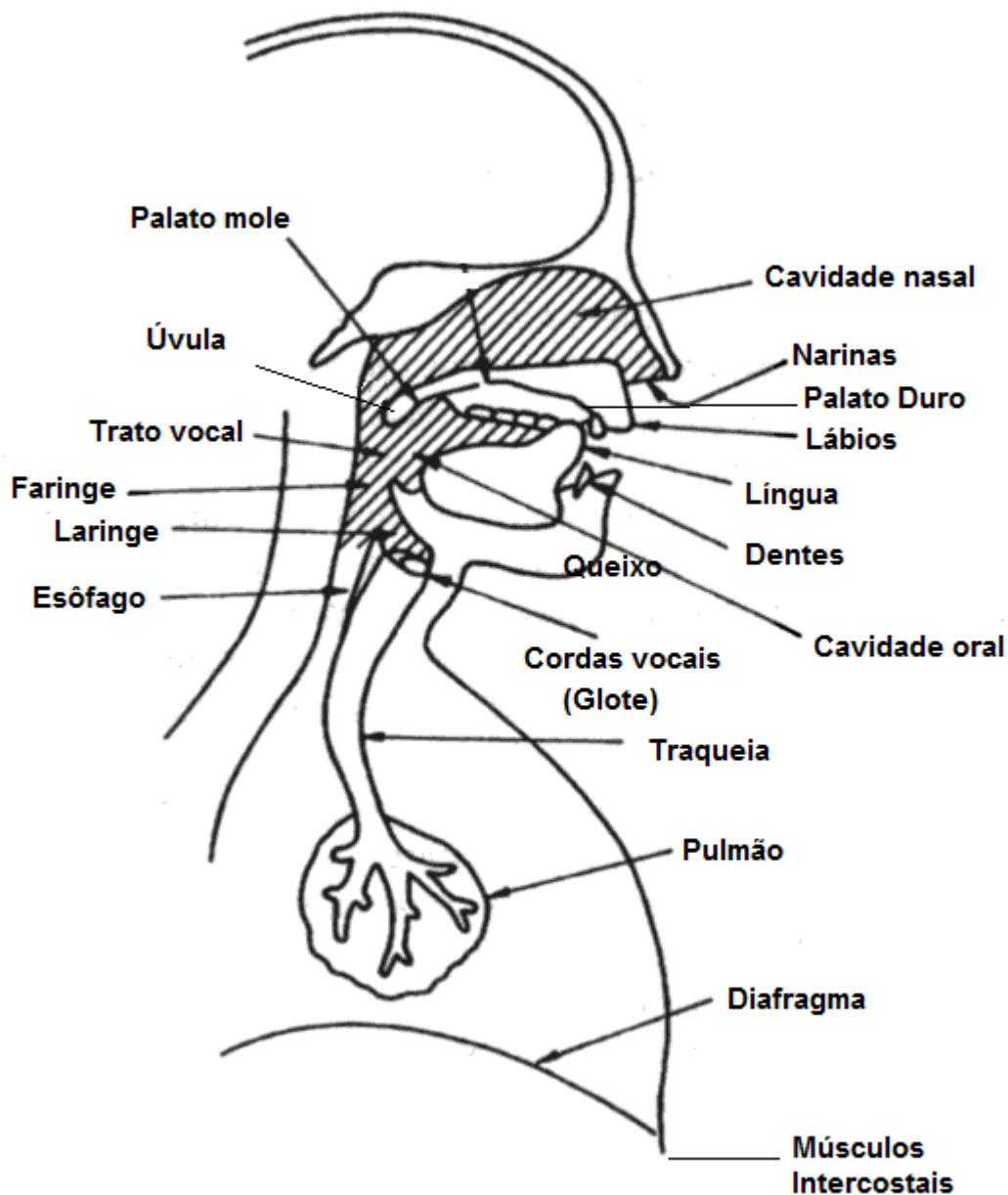


Figura 3.1: Aparelho fonador

Fonte: Adaptado de http://www.gta.ufrj.br/grad/09_1/versao-final/impvocal/propdosinal.html.

Neste momento, abandona-se a noção do trato vocal como um conjunto anatômico de órgãos e adota-se uma visão da engenharia que considera o trato vocal em termos de operação de filtragem acústica. Este filtro, constituído das fossas nasais e pelo trato vocal, é excitado pelo diafragma e é carregado em sua principal saída por uma radiação de impedância devido aos lábios. A boca e as fossas nasais agem como caixas de ressonância. Os articuladores, a maioria dos quais estão associados com o filtro, são usados para mudar as propriedades do sistema, sua

forma de excitação e seu carregamento na saída. Um modelo acústico simplificado é fornecido na figura 3.2:

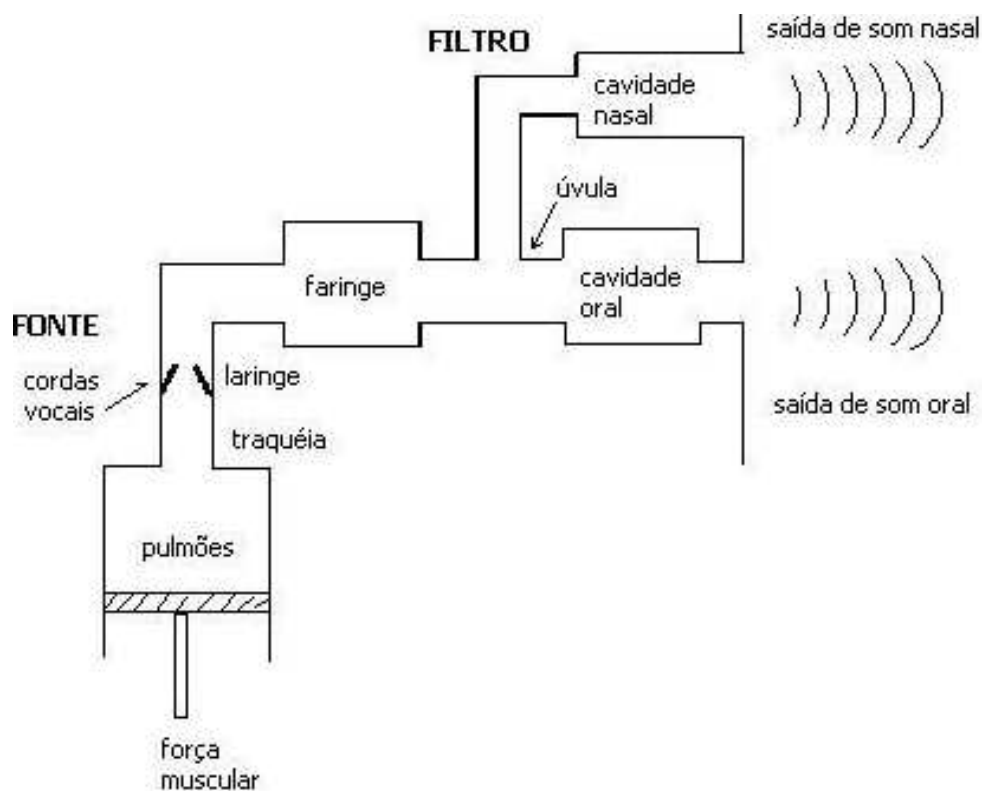


Figura 3.2: Digrama em bloco da produção da fala humana, também conhecido como sistema fonte-filtro.

Fonte: Deller et al. (2000, p. 103).

De acordo com Deller et al. (2000), o trato oral de um indivíduo adulto do sexo masculino mede, em geral, 14 cm. O reposicionamento dos articuladores do trato vocal pode variar ao longo de seu comprimento de zero (fechamento total) a uma área maior que 20 cm². Um comprimento típico do trato nasal em um adulto do sexo masculino mede aproximadamente 12 cm, e o acoplamento acústico entre os tratos vocais e nasais é controlado pelo tamanho da abertura no palato mole.

Em geral, o acoplamento nasal pode influenciar substancialmente as características de frequência do som irradiado pela boca. Se o palato mole estiver em posição baixa, o trato nasal será acusticamente acoplado para produzir os sons da fala. A abertura velar pode variar de zero a cerca de 5 cm² em um indivíduo adulto masculino. Para a produção de sons não-nasais, o palato mole fica

posicionado junto à cavidade da faringe, fechando a entrada da cavidade nasal e desacoplando-a do sistema de produção da fala.

Do ponto de vista técnico, a laringe com suas subdivisões (glote e pregas vocais) tem um papel significativo na produção da fala. A glote é o espaço, uma fenda entre as pregas vocais, que por sua vez estão na laringe. A função da laringe é fornecer uma excitação periódica ao sistema para produzir os sons vozeados (com vibração do tecido mucoso que recobre o músculo vocal²⁸) e os sons desvozeados (sem vibração do tecido mucoso). A corrente expiratória, ao sair da laringe, entra na cavidade da faringe que, por sua vez, lhe oferece duas vias de acesso ao exterior: a cavidade oral ou a cavidade nasal.

3.2 A onda sonora

Uma onda é uma perturbação, um impulso, que se propaga nos meios materiais (sólido, líquido e gasoso). A onda sonora é um tipo de onda mecânica longitudinal que transporta energia. É mecânica porque exige que as partes constituintes do equipamento que a produz devam mover-se sob atuação de forças específicas, e.g., cordas em um violão, martelo em um prego, teclas de um piano etc, e longitudinal porque os pontos do meio se deslocam na mesma direção de propagação da onda. Pode-se definir a frequência da onda sonora como:

$$T = \frac{1}{f} \quad (\text{eq. 1})$$

em que T é o período, e f a frequência. Vale ressaltar que quanto mais rápido é o movimento da fonte geradora de movimentação de partículas, menor é o seu período e maior é a sua frequência (e.g.: Experimente movimentar um objeto próximo ao seu ouvido para comprovar isso!).

O som possui algumas qualidades, a saber, intensidade e frequência, sendo esta a qualidade do som relacionada à sua altura, assim como a intensidade está ligada ao volume. Sons com grandes frequências são chamados de agudos e

²⁸ Ao contrário do que se costumava pensar, não é a contração do músculo vocal que por si só acarreta a vibração das pregas vocais. Na verdade o que vibra é o tecido mucoso que recobre o músculo vocal, chamado de tireoaritenóideo.

sons com baixa frequência, de graves. Em geral, as mulheres têm a voz mais aguda, isto é, emitem sons de maior frequência. A frequência do som é sempre medida em Hertz (Hz).

Para se medir a intensidade sonora, que está relacionada à energia transportada pela onda sonora, utiliza-se uma unidade conhecida como *bel*, que se relaciona com a amplitude da onda. É muito comum o uso de um submúltiplo do *bel*, o decibel (dB), que é um décimo de *bel*; assim $10 \text{ db} = 1 \text{ bel}$.

Para ser ouvida, a energia do som deve ser transportada de molécula a molécula do meio até chegar aos nossos ouvidos. Se o som é muito forte, isto é, se a quantidade de energia emitida pela fonte é alta, isso gera uma sensação desagradável ao ouvido, já que a grande quantidade de energia transportada exerce forte pressão sobre o tímpano. Considerando que energia sonora e vibração da fonte são diretamente proporcionais: quanto maior a vibração, maior a energia, portanto, a intensidade do som é maior quanto maior for a amplitude da onda.

No ouvido, existe uma membrana muito sensível, o tímpano, que vibra quando atingido pela energia sonora. A onda sonora, ao passar pelo ouvido externo, chega à membrana timpânica e faz vibrar todos os ossículos do ouvido médio (martelo, bigorna e estribo). Ao chegar ao estribo, que está acoplado a uma fina membrana chamada de janela oval, esse ossículo move-se fazendo a janela oval mexer-se com ele. No outro lado da janela oval, encontra-se a cóclea, um canal auditivo preenchido por líquidos e responsável por conversão de vibração provinda do meio externo em ondas de compressão, conforme figura 3.3.

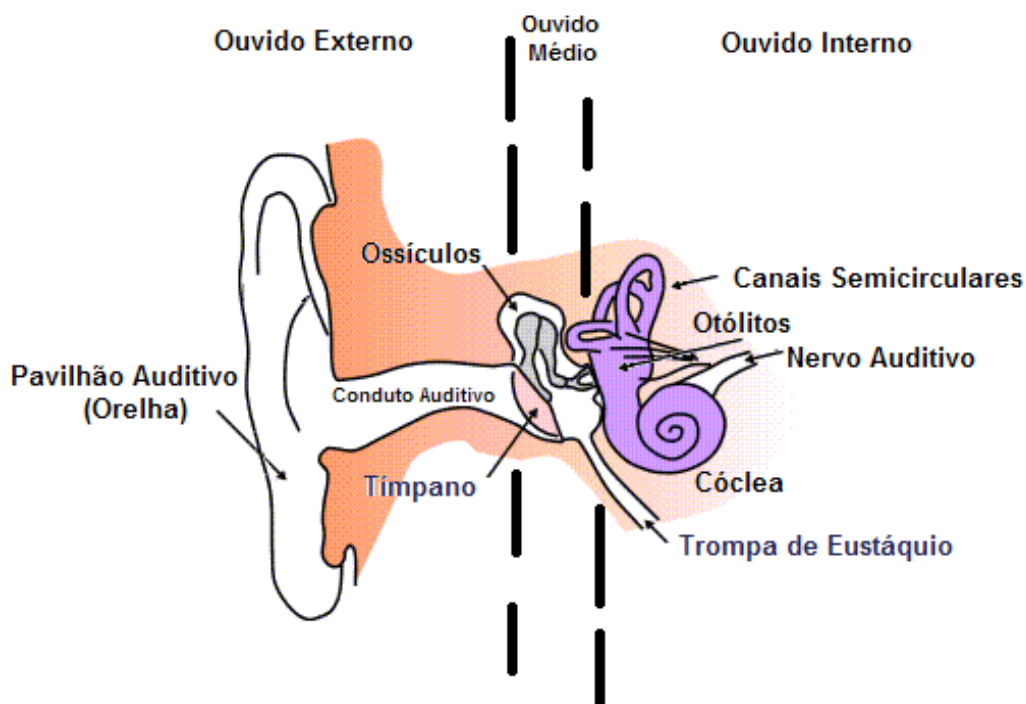


Figura 3.3: Anatomia do ouvido humano
 Fonte: Wikipedia

Quando as vibrações chegam à cóclea provenientes do ouvido interno, elas são convertidas em ondas de compressão que, por sua vez, ativam o órgão de Corti, responsável pela transformação das ondas de compressão em impulsos nervosos (também chamados de sinais elétricos) e que são enviados ao cérebro para interpretação posterior. Todo esse processo de captação, codificação da onda mecânica sonora e sua eventual decodificação em sinais elétricos, constitui-se na cadeia da fala.

3.3 Análise do sinal de fala

O aparato neurobiológico, juntamente com a cognição (percepção da fala) do ser humano, é competente o suficiente para fazer uma análise acústica do sinal de fala proveniente do meio exterior, como também construir a estrutura sintática dos conteúdos lexicais e definir as relações semânticas, dando conta de relações pragmático-prosódicas em milésimos de segundos. Uma máquina ou um *software* de compreensão de fala não seria capaz de executar tais tarefas tão rapidamente nem mesmo se ponderações matemáticas poderosas fossem feitas.

Grosso modo, chama-se de sinal a representação numérica da intensidade de uma certa grandeza física que está sendo medida/observada durante um dado intervalo de tempo, tais como tensão, corrente, onda sonora, temperatura força, deslocamento, aceleração etc. Por exemplo, ao se usar um microfone, o sinal de fala (onda de pressão sonora) é convertido em um sinal elétrico que pode ser medido e processado com o equipamento adequado e visualizado diretamente na tela do computador por meio de um *software* específico, conforme evidenciado na figura 3.4.

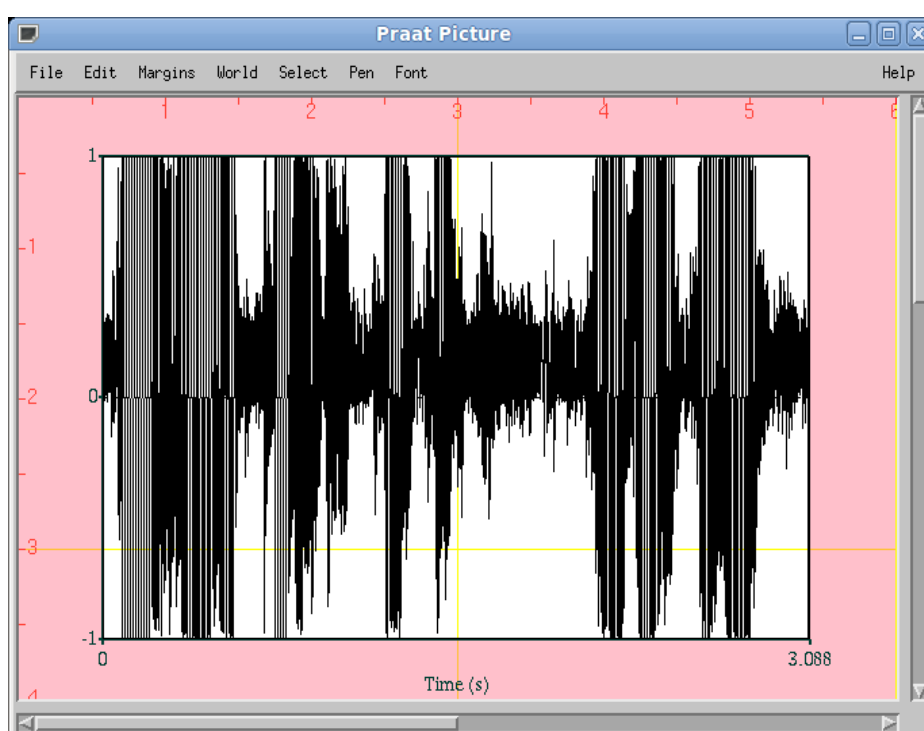


Figura 3.4: Visualização de uma onda sonora na tela de um computador usando o software PRAAT.
Fonte: Disponível em: <<http://www.fon.hum.uva.nl/praat/>>.

O gráfico mostrado na tela de um osciloscópio é geralmente chamado de forma de onda do sinal elétrico medido. Quando se usa um osciloscópio do tipo analógico para visualizar-se a forma de onda de um sinal elétrico, os valores de amplitude dessa onda estão definidos para todo e qualquer instante de tempo dentro do intervalo de medição. Nesse caso, o sinal é chamado de sinal de tempo contínuo ou sinal analógico.

Ao se denotar a amplitude do sinal pela variável x e a variável tempo pela letra t , $t \in \mathbb{R}$, então um sinal de tempo contínuo será representado matematicamente como $x(t)$ ou x_t , sendo a sua representação gráfica no plano cartesiano dado pela

curva $x(t)$ x t , em que a variável t é marcada no eixo das abscissas e a amplitude x no eixo das ordenadas. De maneira mais formal, um sinal é de tempo contínuo se o intervalo entre dois instantes consecutivos quaisquer t_1 e t_2 , $t_2 > t_1 > 0$, é contínuo, ou seja, não contém interrupções.

Um sinal de tempo contínuo pode ser classificado como sendo de amplitude contínua ou de amplitude discreta. A diferença fundamental entre esses dois tipos de sinais é estabelecida ao se verificar se a variável $x(t)$, que representa a amplitude do sinal, é uma grandeza contável ou não para dois instantes t_1 e t_2 quaisquer. Se for contável, então o sinal é de amplitude discreta; caso contrário, o sinal será de amplitude contínua. As figuras 3.5 e 3.6 ajudam a entender esta diferença.

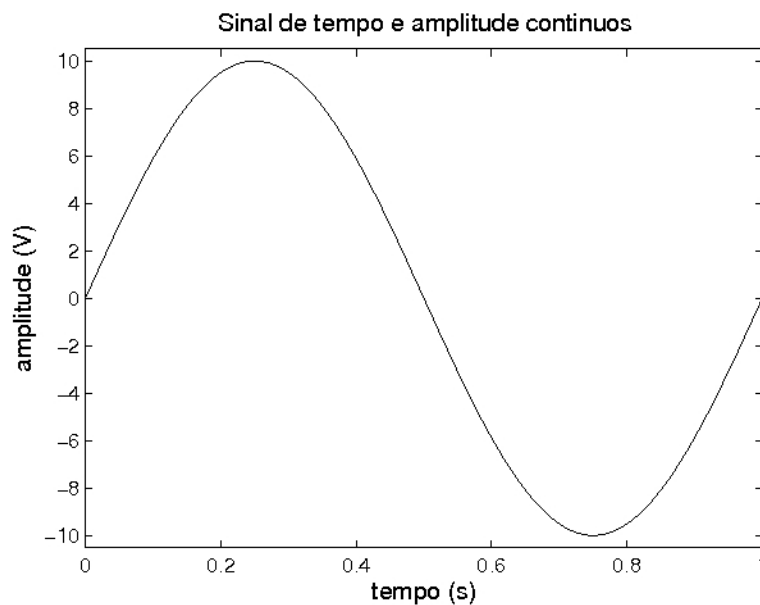


Figura 3.5: Sinal de tempo contínuo e amplitude contínua.
Fonte: Notas de aula de Barreto (2010)

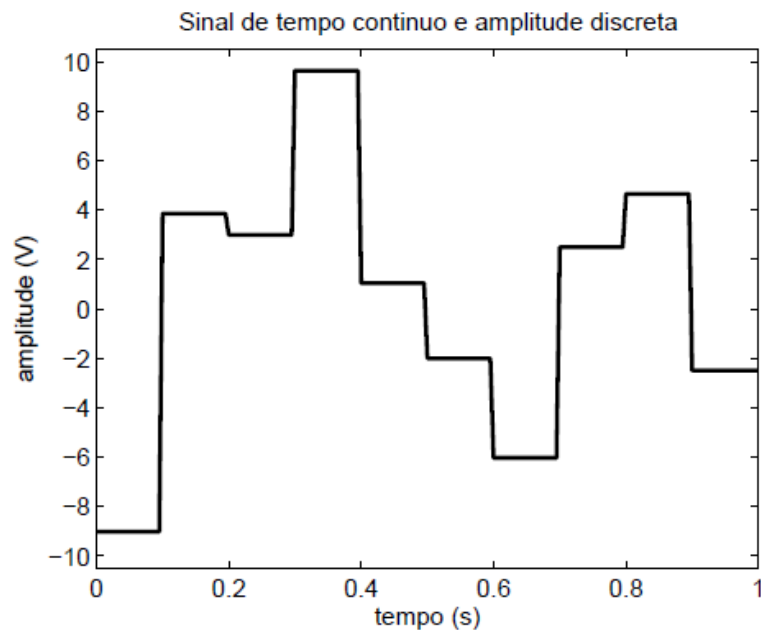


Figura 3.6: Sinal de tempo contínuo e amplitude discreta.
 Fonte: Notas de aula de Barreto (2010)

Nota-se que o sinal mostrado na figura 3.5 é de amplitude contínua, pois o intervalo $[x(t_1), x(t_2)]$ é contínuo, ou seja, tem infinitos pontos entre $x(t_1)$ e $x(t_2)$. Já o sinal da figura 3.6 é de amplitude discreta, visto que as amplitudes assumem uma quantidade finita de valores entre $[x(t_1)$ e $x(t_2)]$, sendo, portanto uma grandeza contável.

3.3.1 Sinais de tempo discreto

De acordo com Barreto (notas de aula, 2010), o uso de osciloscópios digitais e de osciloscópios virtuais está cada vez mais comum atualmente. Osciloscópios virtuais são aqueles implementados em *software* em computadores pessoais (PCs) ao contrário daqueles osciloscópios antigos, grandes e pesados que eram colocados geralmente em cima das mesas e lembravam um rádio antigo.

Em ambos os casos, o sinal de tempo contínuo é convertido em um sinal de tempo discreto por meio de conversores analógicos digitais (A/D). Um conversor A/D executa basicamente duas funções: amostragem e quantização. Por meio destas duas funções, um conversor A/D converte um sinal de tempo e de amplitude

contínuos em um sinal de tempo e de amplitude discretos. Eis a seguir, um exemplo de cada:

Amostragem: durante a amostragem, o conversor A/D captura, isto é, lê a cada ΔT unidades de tempo as amplitudes do sinal analógico $x(t)$, gerando uma sequência numérica dada por:

$$\{x(\Delta T), x(2\Delta T), \dots, x(n \cdot \Delta T), \dots, x(N \cdot \Delta T)\} \quad (\text{eq. 2})$$

em que a constante ΔT é chamada de período de amostragem, e o recíproco $f_s = 1 / \Delta T$ é chamado de frequência de amostragem (do inglês, *sampling frequency*) do sinal contínuo.

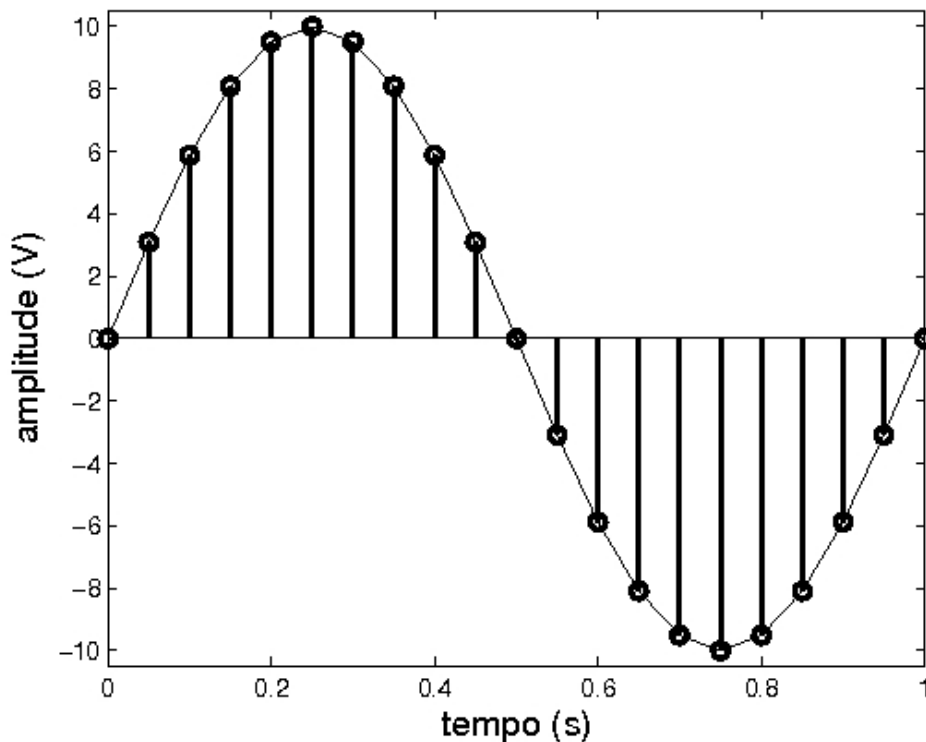


Figura 3.7: Processo de conversão de um sinal de tempo contínuo para tempo discreto. Fonte: Notas de aula de Barreto (2010)

Como mostrado na figura 3.7, para uma frequência de amostragem $f_s = 1$ KHz, significa dizer que em um segundo são capturados $N = 1000$ amostras do sinal

contínuo $x(t)$, cada uma separada de $\Delta T = 1$ ms da outra. A duração total do sinal é dada pelo produto $\Delta T \times N = 1$ s.

Em poucas palavras, o amostrador transforma o sinal de tempo contínuo em um sinal de tempo discreto, pois agora as amplitudes só existem para instantes de tempo múltiplos de ΔT . Nota-se que quanto menor ΔT , maior será a frequência de amostragem e, conseqüentemente, maior será a quantidade de pontos (ou amostras) a serem coletados do sinal analógico. A figura 3.7 ilustra este processo de conversão para um sinal senoidal.

Quantização – durante a quantização, cada um dos valores amostrados $\{x(\Delta T), x(2\Delta T), \dots, x(n \cdot \Delta T), \dots, x(N \cdot \Delta T)\}$ é convertido em um código binário, formado por um certo número de *bits*. Por exemplo, se o código é de 4 bits então cada valor amostrado é convertido para um dos 16 códigos possíveis, ou seja, 0000 0001 0011 0100 0101 0110 0111 1000 1001 1011 1100 1101 1110 1111.

Em outras palavras, estes códigos correspondem a 16 níveis discretos de amplitude entre o valor mínimo e máximo medido (ou permitido). Assim, a escolha do código a ser utilizado depende do valor da amostra no instante nT , $n = 1, \dots, N$. Em geral, o valor de $x(nT)$ é arredondado para o nível discreto mais próximo. A quantização é necessária, pois o computador só entende números binários. Observa-se que o processo de quantização introduz um erro de arredondamento, chamado de erro de quantização, que depende diretamente do número de bits do conversor A/D. O processo de quantização e seus efeitos sobre o sinal de tempo discreto estão ilustrados nas figuras 3.8 e 3.9:

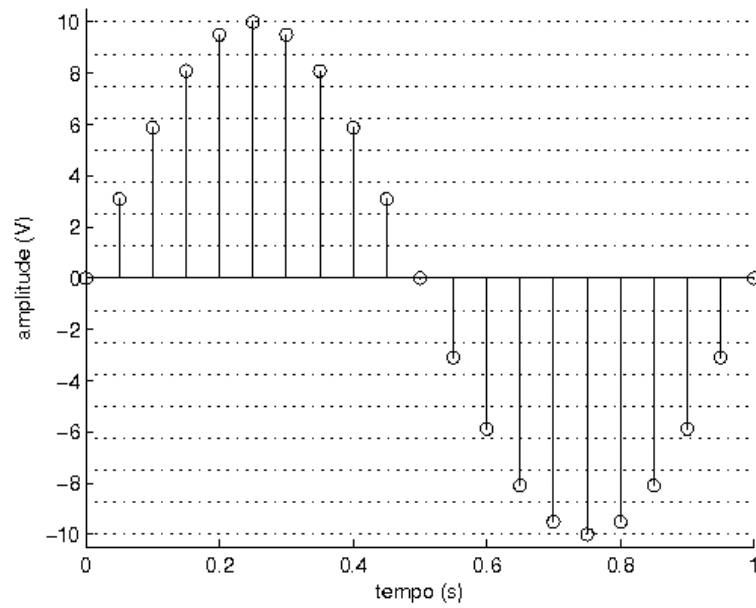


Figura 3.8: Sinal de tempo discreto e 16 níveis de quantização.
 Fonte: Notas de aula de Barreto (2010)

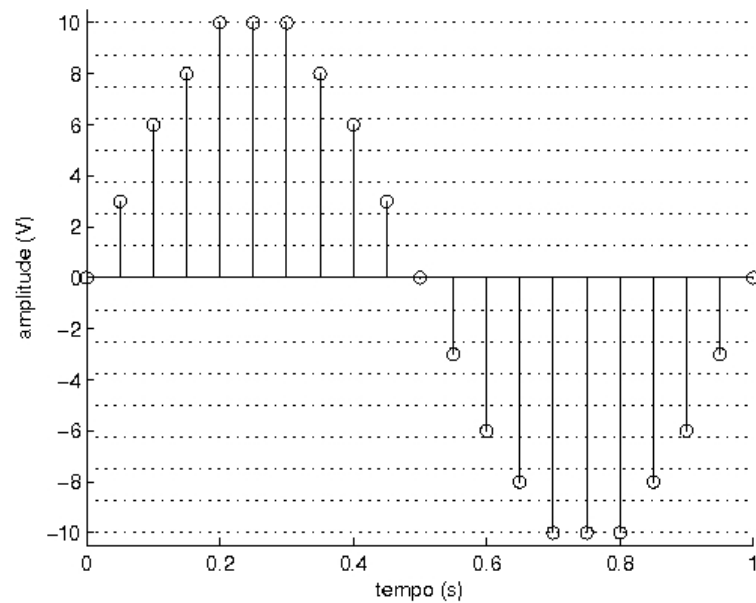


Figura 3.9: Sinal quantizado.
 Fonte: Notas de aula de Barreto (2010)

Em suma, o sinal digital é um sinal com valores discretos (descontínuos) no tempo e em amplitude. Isso significa que um sinal digital só é definido para determinados instantes de tempo e que o conjunto de valores que ele pode assumir é finito.

A digitalização de sinais analógicos é obtida por três processos:

- 1) **Amostragem** - discretização do sinal analógico original no tempo;
- 2) **Quantização** - discretização da amplitude do sinal amostrado;
- 3) **Codificação** - atribuição de códigos - geralmente binários - às amplitudes do sinal quantizado.

O número de *bits* e a taxa de amostragem (*sampling rate*) definem a qualidade de um sinal digital. Os sinais com boa qualidade possuem 16 *bits* e 44,1 kHz de taxa de amostragem. Todavia, atualmente, a taxa mínima de amostragem de sinais de fala estabelecida em sistemas computacionais é de 11kHz, valor suficiente para o processamento de fala em um nível básico. No capítulo 5, serão apresentados os valores de amostragem e a quantização os dados das elocuições utilizadas para a codificação do *input* da rede neural.

3.4 Parametrização do sinal da fala e Extração de características

O processo de extração de características é uma importante etapa na abordagem conexionista do processamento da fala e tomadas de decisão e classificação da rede neural. Essa etapa consiste na utilização de técnicas de transformação do sinal de fala original em uma representação matemática que permita a identificação de uma dada elocução, e é geralmente representado por um conjunto de vetores de características (SOUZA JR., 2009).

Analisar o sinal de fala por si só já é uma tarefa árdua quicá extrair suas características. O sinal de fala²⁹ é captado através de recepção física³⁰, e o seu processamento é normalmente realizado para a identificação da parte da voz referente aos elementos lexicais da fala. Feita a identificação do sinal, passa-se a

²⁹ O sinal pode ser gerado artificialmente por um circuito eletrônico (oscilador). Entretanto, na maioria das aplicações práticas, o sinal elétrico representa a variação de outra grandeza física no decorrer do tempo, convertida em eletricidade por um transdutor (um microfone, por exemplo).

³⁰ Qualquer informação útil para o circuito é considera-se como sinal. Qualquer informação indesejada, inútil, ou nociva introduzida involuntariamente no sistema é considerada ruído. Por exemplo, ao se falar em um microfone, este converte as ondas sonoras em uma tensão variável. Esta variação na tensão corresponde exatamente à frequência de oscilação das ondas sonoras emitidas. É o sinal elétrico puro. Há muitas outras fontes de ruídos que podem afetar o sinal elétrico, como descargas atmosféricas e raios cósmicos.

extrair o conteúdo identificador dos itens lexicais, no caso da presente pesquisa, somente para a representação de palavras, não da frase inteira.

Ao serem extraídos os parâmetros do sinal de fala, deve ocorrer, para fins de parcimônia, uma diminuição da sequência representativa do sinal. Toma-se, como exemplo, um segmento de sinal de fala de 20ms amostrado a 44 kHz sendo representado por uma sequência numérica de 880 pontos: sua representação paramétrica pode ser dada por uma sequência de 15 valores. Esta sequência de valores, ou coeficientes, é utilizada para representar o sinal de fala, pois, além de diminuir o tamanho do vetor que representa o segmento do sinal de fala, a sequência mantém a informação mínima necessária capaz de diferenciar os vários segmentos existentes (LIMA, 2000).

Normalmente, há quatro técnicas tradicionais para a representação do sinal linguístico, ou seja, a extração de parâmetros: 1) a transformada de Fourier; 2) transformadas ondeletas (*wavelets*); 3) Codificação Linear Preditiva ou Predição Linear (*Linear Predictive Coding* - LPC); 4) Coeficientes Mel-Cepstrais (*Mel-Frequency Cepstrum Coefficients*- MFCC) (MULLER, 2006).

As duas técnicas de extração de parâmetros que se mostraram mais convenientes para a análise dos dados da presente pesquisa foram as duas últimas técnicas: LPC e MFCC (Ver simulações apresentadas no capítulo 6).

3.4.1 Coeficientes de Predição Linear

A Codificação Linear Preditiva (*Linear Predictive Coding* - LPC) é uma técnica de parametrização e processamento do sinal da fala amplamente utilizada para a obtenção de coeficientes cepstrais nas áreas de reconhecimento automático de voz e sistemas de síntese texto-fala (CARICATTI; WEIGANG, 2001; MILONE; RUBIO, 2003; SCHÖTZ, 2002; SOUZA JR., 2009).

A técnica LPC é também utilizada para representar o envelope espectral de um sinal digital em forma compressa, usando a informação de um modelo preditivo linear. É uma das mais poderosas técnicas de análise e um dos métodos mais úteis para codificar a fala.

Com base na idéia de que a fala pode ser gerada através de um modelo de filtragem, no caso um modelo autorregressivo (AR), e também com base no fato

de que a fala pode ser representada por um processo estocástico estacionário por partes, a análise por predição linear assume que o sinal de fala é produzido por um zumbido no fim do tubo (garganta e boca), cuja ressonância resultante da excitação (ar vindo dos pulmões) forma as frequências de ressonância e que por sua vez origina os formantes ou bandas de frequência alteradas.

A técnica LPC considera que a fala é constituída de duas classes de sons: sons surdos (desvozeados) e sons sonoros (vozeados), que podem ser vistos nas figuras 3.10 e 3.11.

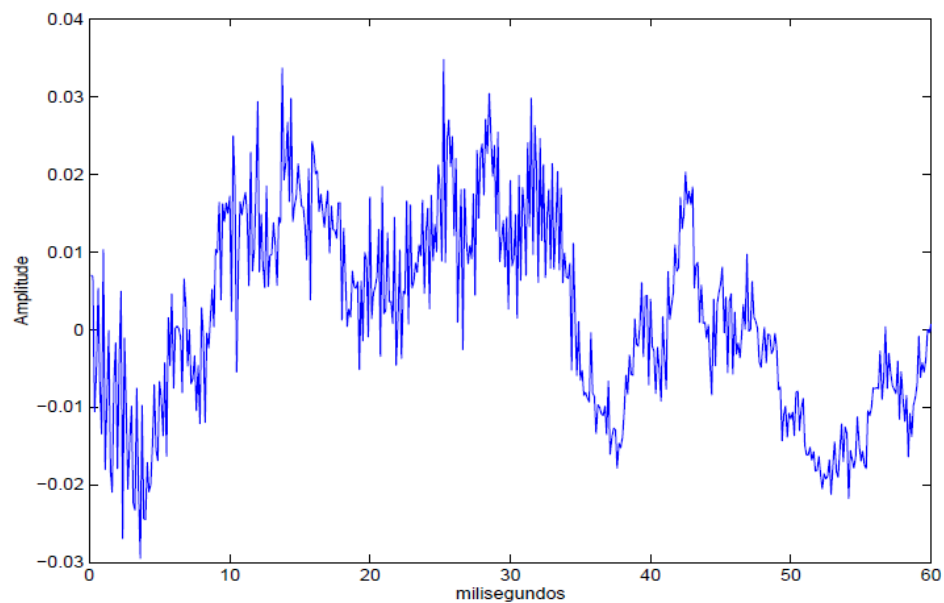


Figura 3.10: Exemplo de som surdo
Fonte: LIMA (2000).

Os sons surdos são tipicamente ruidosos, como é o caso das consoantes fricativas. Os sons sonoros apresentam uma forma de onda *quasi* periódica, como podem ser percebidos na onda das consoantes da figura 3.11.

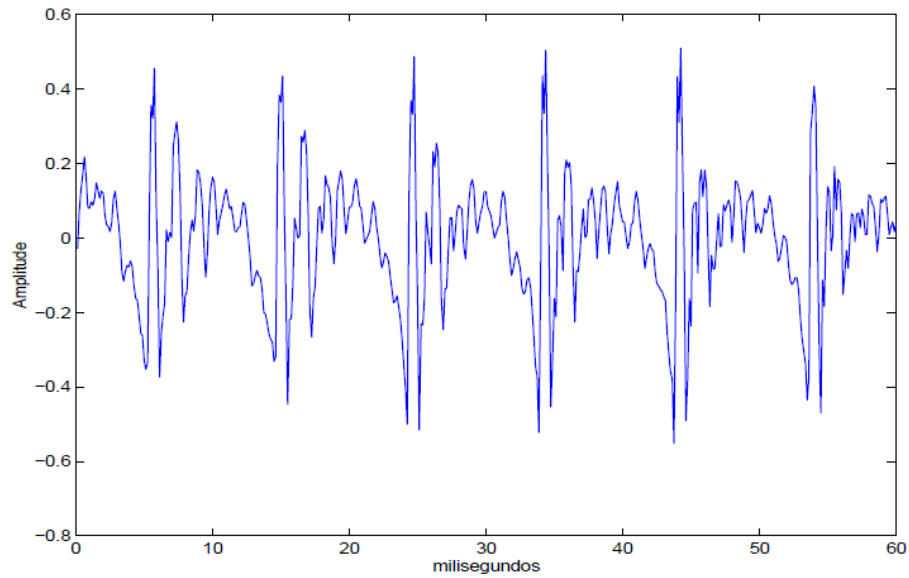


Figura 3.11: Exemplo de som sonoro
 Fonte: LIMA (2000).

No modelo LPC, esta distinção entre os sons ou fones ocorre em termos de excitação do filtro digital, sendo os sons surdos gerados por uma excitação do tipo ruído branco, e os sons sonoros a partir de uma excitação do tipo trem de pulsos (LIMA, 2000), como mostra a figura 3.12:

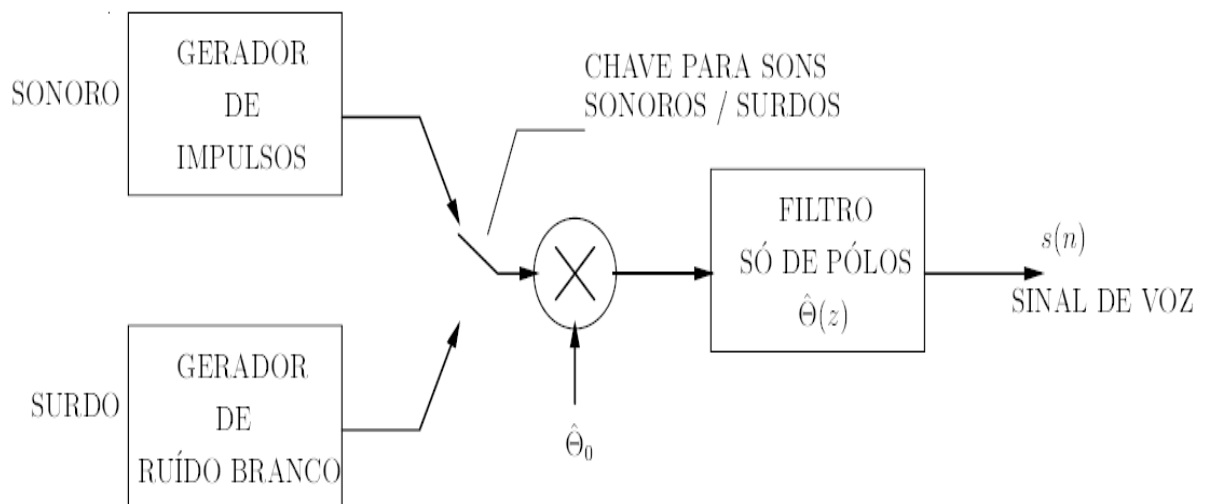


Figura 3.12: Uso de LPC na produção de voz.
 Fonte: LIMA (2000, p. 32).

A frequência do sinal de fala modifica-se à medida que o ar, vindo da excitação inicial dos pulmões, vai percorrendo o tubo (traqueia), modificando também seus formantes. Os sons produzidos pelas cordas vocais e amplificados no trato vocal e na cavidade nasal contribuem para a produção das principais características identificadoras do interlocutor: a intensidade (volume – *loudness*) e a frequência (altura – *pitch*), contidas na energia do sinal de fala³¹.

Os dois principais fatores responsáveis pelas variações dessas características são a pressão do ar subglotal e a tensão das cordas vocais. Esses fatores contribuem para variações na frequência fundamental, fonte do espectro e fonte de amplitude. O aumento na fonte de excitação (ar vindo dos pulmões) acarretará um aumento da pressão do ar subglotal, o que causará um aumento na média dos pulsos da corrente de ar produzidos na glote. Em resumo, a frequência e a intensidade são diretamente proporcionais à pressão do ar subglotal (DELLER et al., 2000, p. 140).

Por se acreditar na superposição de todos esses sinais advindos de fontes distintas (pulmões e glote), ou seja, na convolução formando um sinal composto, faz-se necessário separá-los ou deconvolucioná-los. Esse processo é chamado de extração do *cepstrum*, que será discutido, com mais detalhes, na próxima seção.

A técnica LPC analisa o sinal da fala estimando os formantes, removendo seus efeitos e estimando a intensidade e frequência do ruído restante. O processo de remover os formantes é chamado de filtragem inversa e o sinal remanescente após a subtração do sinal modelado filtrado, de resíduo.

Os números que descrevem a intensidade e a frequência do zumbido, ou seja, os formantes, bem como o sinal residual, podem ser armazenados. Pelo fato de o sinal de fala variar no tempo, esse processo é feito através de janelamento em pedaços do sinal de fala, chamados de *frames* (quadros).

Assim, a técnica LPC baseia-se na aproximação de um novo valor em uma série temporal por uma expressão recursiva linear de valores passados. Como justifica Kohonen (2001, p. 211):

³¹ O espaço entre as pregas vocais, chamado de glote, não produz ruídos ou zumbidos, como se costumava pensar. Estes, além dos estalos, são percebidos no fim do tubo, na boca e na garganta.

Os coeficientes LPC³² têm sido usados com frequência para a extração de características do sinal de fala, porque eles contêm a mesma informação do espectro de amplitude, mas são computacionalmente mais leves do que a transformada de Fourier (*Fast Fourier Transform* - FFT).

O sinal de fala com a combinação linear dos seus P valores passados e a excitação G(u) pode ser representado da seguinte forma:

$$s(n) = \sum_{i=1}^p a(i)s(n-i) + G(u), \quad (\text{eq. 3})$$

em que $s(n)$ é a saída do modelo (sinal de fala), $a(i)$ são os coeficientes de predição linear estimados e p é a ordem do modelo LPC utilizado (preditor).

A análise LPC parte do pressuposto de que o sinal de fala pode ser representado pela combinação linear das amostras atrasadas (SOUZA JR., 2009). Com isso, novos valores do sinal são estimados, janelados, bem como computados novos coeficientes para cada quadro do sinal de fala. Na prática, utilizam-se ordens de predição que variam de 10 a 20 coeficientes³³.

Em suma, o problema básico na análise LPC é determinar os coeficientes de predição, $a(i)$, considerados algoritmos de extração de parâmetros. O método utilizado para determinar os coeficientes foi o método de Burg³⁴.

3.4.2 Coeficientes Cepstrais

Ainda tendo em mente a ideia de que a fala é composta de uma excitação (ar que vem dos pulmões) em conjunto com uma resposta do trato vocal e que a separação dessas duas fontes não é tarefa fácil uma vez que os valores das duas fontes são multiplicados e representados no domínio da frequência, como mostra a equação abaixo,

$$s(n) = u(n) * h(u), \quad (\text{eq. 4})$$

³² O objetivo da análise preditiva linear é estimar parâmetros de um modelo autoregressivo, também conhecido como modelo *all-pole* do trato vocal. (DELLER, 2000, p. 471).

³³ Mais detalhes sobre a escolha da ordem de predição podem ser encontrados em Huang et al. (2001).

³⁴ Para mais detalhes sobre a complexidade computacional desses algoritmos, conferir Deller et al. (2000).

criaram-se os coeficientes cepstrais ou *cepstrum* para facilitar essa tarefa de separação de fontes. O *cepstrum* (pronunciado [kɛpstɹəm]) é definido como a Transformada de Fourier (*Fourier Transform*) do logaritmo da amplitude espectral de um sinal (SHAFER, 2008). Verbalizando a expressão matemática, o espectro de potência (ou espectro de força) de um sinal é a magnitude quadrada da transformada de Fourier do logaritmo da magnitude quadrada da transformada de Fourier de um sinal, ou seja:

$$|\mathcal{F}\{\log(|\mathcal{F}\{\text{sinal}\}|^2)}|^2, \quad (\text{eq. 5})$$

Há muitas formas de calcular o *cepstrum*. Tomando por base o espectro da potência, eis os passos a serem seguidos para consegui-lo:

1. Sinal digital: $\{x(n)\}$, $n = 1, \dots, N$;
2. Transformada de Fourier: $\mathcal{F}\{x(n)\}$;
3. Aplicar o valor absoluto: $|\mathcal{F}\{x(n)\}|$;
4. Elevar ao quadrado: $|\mathcal{F}\{x(n)\}|^2$;
5. Aplicar o logaritmo ao resultado do passo 4: $\log(|\mathcal{F}\{\text{sinal}\}|^2)$;
6. Aplicar Fourier ao resultado do passo 5: $\mathcal{F}\{\log(|\mathcal{F}\{\text{sinal}\}|^2)\}$;
7. Aplicar valor absoluto ao resultado do passo 6: $|\mathcal{F}\{\log(|\mathcal{F}\{\text{sinal}\}|^2)\}|$;
8. Elevar ao quadrado os resultados do passo 7: $|\mathcal{F}\{\log(|\mathcal{F}\{\text{sinal}\}|^2)\}|^2$.

Coeficientes mel-cepstrais consistem na representação de um espectro de potência de curta duração de um som, baseado na transformada de um cosseno linear de um espectro de força sobre uma escala mel de frequência não-linear. Para determinar o *mel-cepstrum*, isto é, aplicar a escala mel na extração de coeficientes cepstrais (MFCC - *Mel Frequency Cepstral Coefficient*), basta incluir um passo intermediário entre os passos 4 e 5 do algoritmo mostrado acima, em que se mapeia as frequências em Hertz (f) para a escala mel através da seguinte equação:

$$M = 2595 \log_{10} (f/700 + 1) \quad (\text{eq. 6})$$

Os coeficientes de frequência mel cepstral (MFCC) são provenientes de um tipo de representação cepstral. A diferença entre um *cepstrum* e um *mel-*

frequency cepstrum consiste em que, neste último, as bandas de frequência sejam igualmente espaçadas na escala mel, o que aproxima a resposta do sistema auditivo humano mais do que bandas de frequência linearmente espaçadas no *cepstrum* normal. Essa possibilidade gera uma melhor representação do som em termos de áudio compressão.

A amplitude do espectro do sinal de fala é constituída pela combinação do envelope espectral³⁵ que está associado ao filtro linear e da estrutura fina, em correspondência com a excitação glotal (FURUI, 2001 *apud* SOUZA JR., 2009). A propriedade básica do *cepstrum* consiste no fato de que ele permite uma separação entre o envelope espectral e a estrutura fina. Além disso, os *cepstra* têm-se mostrado uma representação eficiente do modelo de fala (DELLER et al., 2000).

Porém, Müller (2006) indica um problema no algoritmo para obtenção de coeficientes MFC: o fato da derivada do cosseno criar apenas uma representação para todas as faixas de frequência. Basta uma das faixas estar corrompida por ruído para afetar todos os coeficientes gerados. O autor aponta o uso de transformada ondeletas para corrigir o problema, ou seja, a análise do sinal por sub-bandas.

Não foi necessária a utilização da transformada ondeletas já que não importava se o espectro do fonema de alta frequência poderia dominar outro de baixa frequência, ocorrendo uma sobreposição. Todavia, todos os dados inseridos na rede passaram por uma fase de pré-teste com métodos de parametrização do sinal da fala por análise de predição linear e por extração das características da fala por coeficientes mel-cepstrais.

Optou-se pela análise de predição linear porque os coeficientes LP conseguem extrair a intensidade e a frequência do sinal de fala. Essas duas características são portadoras e indicadoras do elemento prosódico “acento”. No inglês, o acento é a junção de três fatores perceptivos correlacionados: 1) quantidade/duração (medida em ms) relacionada com o tamanho da sílaba; 2) intensidade (medida em dB) relacionada à amplitude média alta e 3) altura (medida em Hz), ou seja, o valor de F_0 mais elevado na elocução.

³⁵ Envelope espectral é uma função de frequência que combina amplitudes de partes individuais do espectro.

3.5 Resumo do capítulo

Este capítulo apresentou sucintamente os fundamentos do processamento de fala, o processo de produção natural da fala que definiu o trato vocal em termos de filtragem acústica, a descrição da onda sonora e a análise do sinal de fala e do sinal de tempo - contínuo e discreto, bem como sua amplitude – contínua e discreta. Viu-se também que o sinal digital é um sinal com valores discretos (descontínuos) no tempo e em amplitude.

Neste capítulo também foi apresentado o processo de extração de características como uma importante etapa na abordagem conexionista do processamento da fala e tomadas de decisão e classificação da rede neural. Nesta etapa, é de fundamental importância a utilização de técnicas de transformação do sinal de fala que permitam a identificação de uma dada elocução. As duas técnicas de extração de parâmetros apresentadas foram LPC e MFCC. Esses dois coeficientes conseguem extrair a intensidade e a frequência do sinal de fala, características indicadoras do elemento prosódico “acento”.

Após entender como o sinal de fala é numericamente codificado para servir como ferramenta de análise da linguagem e da prosódia, apresentaremos no próximo as redes neurais artificiais (RNA), mais precisamente dos mapas auto-organizáveis (self-organizing maps – SOM) de Teuvo Kohonen, como uma ferramenta para análise exploratória de dados aquisicionais linguísticos (*knowledge acquiring*), extração e mineração de dados multivariados. O propósito é apresentar e avaliar este tipo de rede como uma técnica possível para a detecção de padrões que emergem da atividade da rede, bem como para a visualização e a análise de componentes em bases de dados. Alguns conceitos sobre auto-organização serão também apresentados. Cabe ressaltar que o uso das redes de Kohonen para a análise exploratória de dados aquisicionais dessa magnitude é, portanto, inédita. Espera-se que os resultados das simulações tragam inúmeras contribuições para as teorias de aquisição de L2.

4 MODELOS CONEXIONISTAS PARA O PROCESSAMENTO DE LINGUAGEM

O presente capítulo tem o objetivo de descrever a rede neural artificial usada nesta pesquisa. Primeiramente, é feita uma pequena introdução às redes neurais - biológica e artificial - e posteriormente se segue à descrição da rede de Kohonen e os seus princípios básicos aplicados às simulações. O capítulo também conta com a justificativa do uso da rede de Kohonen em nossos experimentos bem como um breve relato das aplicações dos mapas auto-organizáveis ao processamento da fala.

4.1 Redes Neurais Biológicas Vs. Redes Neurais Artificiais

O paradigma conexionista se baseia na estrutura organizacional dos bilhões de neurônios que compõem o cérebro humano, tridimensionalmente conectados por meios de sinapses, e em como eles processam (trocam) informação entre si, não na forma de símbolos prontos como um todo, mas de padrões de atividade e de conectividade.

Biologicamente, os neurônios variam quanto ao tamanho, forma, função e relação entre eles e estão organizados em diferentes estruturas no cérebro humano. O cérebro tem cerca de cem bilhões de neurônios e uma quantidade ainda maior de células gliais (ou neuroglias, que fornecem sustentação essencial à manutenção do órgão) (AAMODT; WANG, 2009).

De acordo com Aamodt; Wang (2009), o envio de sinais entre as partes que compõem o neurônio é feito por meio eletroquímico. O neurônio possui uma membrana com carga elétrica distribuída que, ao entrar em contato com a membrana de outros neurônios, abre canais que permitem que íons a atravessem e gerem uma corrente que transmite o sinal elétrico pela membrana. Um neurônio típico é formado basicamente por dendritos, corpo celular (ou soma) e axônio. Este último é o meio pelo qual um neurônio se comunica com outros, ou seja, é a porta de saída da informação processada pela célula (AAMODT; WANG, 2009).

Os axônios no corpo humano podem conduzir sinais para longas distâncias, por exemplo, desde a coluna vertebral até a ponta dos dedos dos pés. Os dendritos são ramificações do corpo celular que funcionam como receptores de informação (estímulo) advinda de outros neurônios por meio dos axônios. O ponto de aproximação máxima entre um axônio e um dendrito é denominado de sinapse, componente fundamental de comunicação no cérebro. Na realidade, os neurônios nunca se tocam: a sinapse é, de fato, o espaço entre a extremidade do axônio e os dendritos (MACEDO, 2007, comunicação pessoal), como evidenciado na figura 4.1:

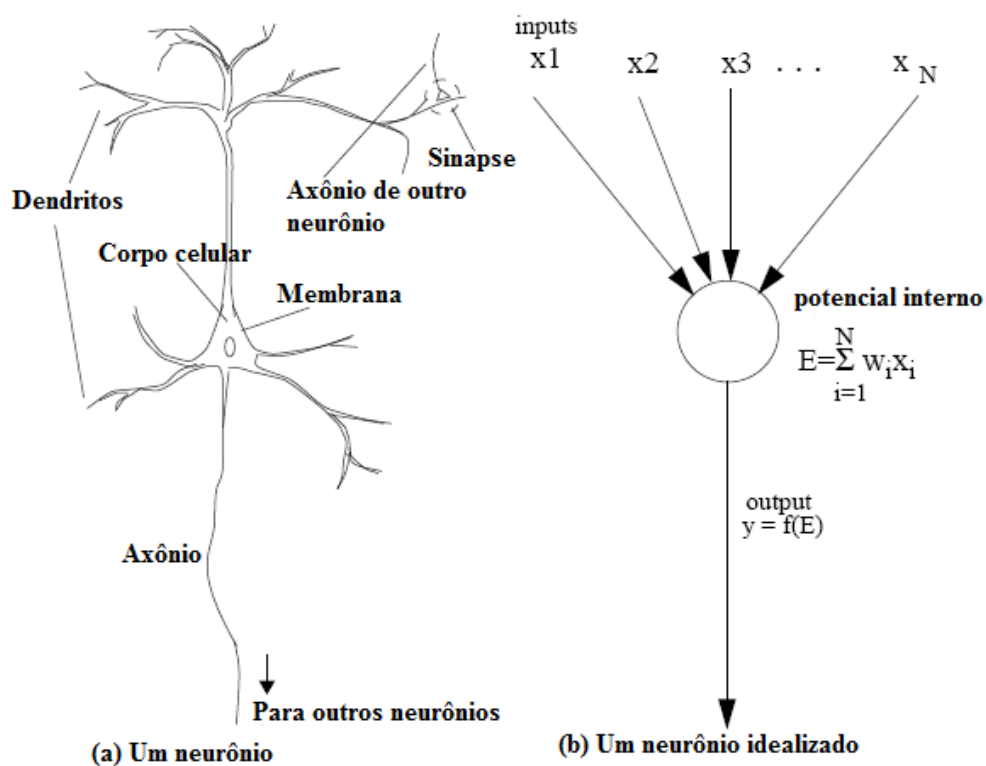


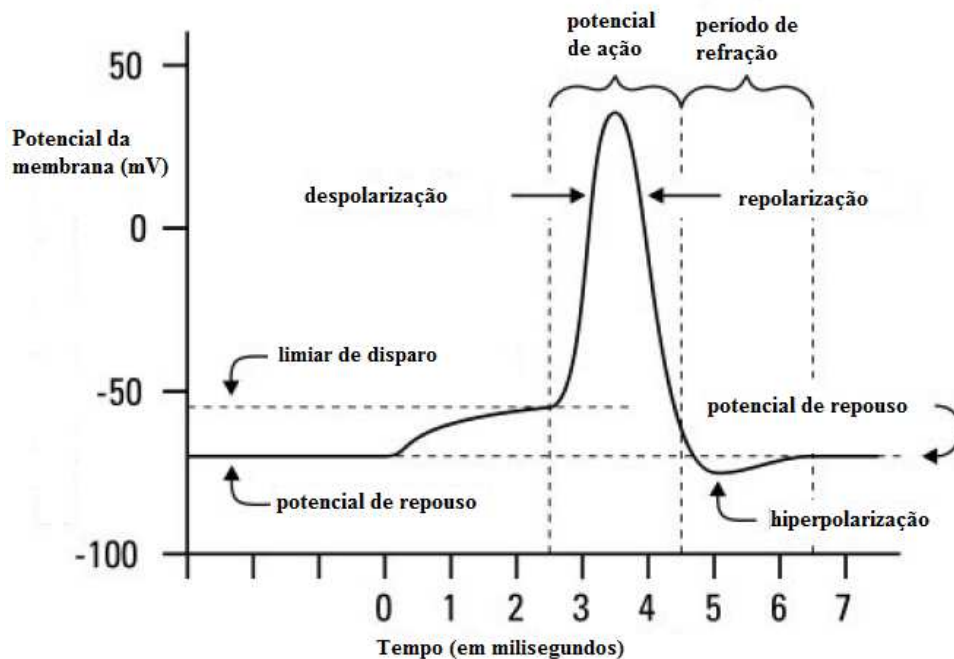
Figura 4.1: Representação de um neurônio biológico em (a) e um neurônio artificial em (b).
Fonte: Adaptado de Haykin (1999).

Os neurônios podem estabelecer milhares de conexões, ou seja, sinapses, com centenas de outros neurônios e se comunicar com músculos diretamente ou com órgãos sensoriais. As informações são transmitidas dos neurônios aos seus axônios via sinais elétricos que duram pequenas frações de segundos (AAMODT; WANG, 2009).

Esses sinais elétricos são também chamados de “picos” porque são o resultado de aumentos repentinos nas correntes elétricas cerebrais. Os picos, mais

conhecidos como potenciais de ação ou impulsos nervosos, percorrem o axônio a uma velocidade de vários metros por segundo a fim de transmitir sinais do cérebro para partes específicas do corpo realizarem uma determinada ação. A comunicação entre os neurônios depende de substâncias químicas denominadas neurotransmissores, que são liberadas a partir de pequenas áreas na terminação axônica por um potencial de ação.

O axônio emite um impulso elétrico (potencial de ação) somente se o balanço energético realizado no corpo celular for maior que um certo limiar. Neste caso, diz-se que o neurônio disparou ou está ativado. Vejamos na figura 4.2 o potencial de ação do neurônio em um dado instante de tempo:



Potencial de ação em um neurônio

Figura 4.2: Ilustração das várias etapas que compõem o potencial de ação em um neurônio. Fonte: Adaptado de Barreto (2008).

A informação de um neurônio devidamente estimulado é então codificada na frequência dos potenciais de ação. A chegada de um trem de pulso no botão sináptico localizado na região terminal do axônio provoca liberação de transmissores na fenda sináptica. Os neurotransmissores se ligam a receptores localizados nos dendritos ou corpos celulares de outros neurônios gerando sinais eletroquímicos.

De acordo com Cielo (1998), quando há atividade elétrica a ser propagada entre os neurônios, proteínas sintetizadas nas próprias células (neurotransmissores³⁶) preenchem o espaço sináptico e são recebidas pelos dendritos como sinais de *input*. Sendo assim, “não podemos considerar as sinapses como entidades fantasmas, já que são concretas, materiais” (CIELO, 1998, p. 44).

Pelo exposto, as Redes Neurais Biológicas (RNB) podem ser consideradas como padrões de atividade elétrica nas sinapses que constituem o código cerebral utilizado para armazenar o conhecimento. Para armazenar conhecimento, processá-lo e resgatá-lo necessita-se de estímulo exterior sensorial que irá desencadear processos de atividade mental à medida que atividades sinápticas reforcem as conexões já estabelecidas.

Partindo desse ponto de vista, conceitua-se o neurônio como um elemento biológico altamente adaptável³⁷, sendo responsável assim pela extrema plasticidade, flexibilidade, rapidez e capacidade do cérebro para operar com vários estímulos ao mesmo tempo. Dessa maneira, o armazenamento de informações está intimamente ligado a mudanças nas sinapses (conexões neurais). Podem ser percebidas respostas mais rápidas quando as ligações neurais são reforçadas. Quando o inverso ocorre, ou seja, quando não se apresenta repetição do estímulo, conseqüentemente há o enfraquecimento das sinapses.

Do ponto de vista do fluxo de informação, uma rede neural artificial pode conter três tipos de neurônios: (i) de entrada – análogos aos neurônios sensoriais dos seres vivos; (ii) de saída – análogos aos neurônios motores; e (iii) neurônios internos ou intermediários.

Uma vez apresentado o neurônio biológico, torna-se possível estudar as propriedades das redes de neurônios interconectados artificialmente. Uma rede neural artificial (RNA) pode ser definida como uma máquina planejada para imitar o modo como o cérebro realiza uma tarefa, via utilização de componentes eletrônicos

³⁶ Substâncias responsáveis pela transmissão neural compostas quimicamente de moléculas de proteínas. Existem aproximadamente uns 60 tipos de neurotransmissores e são classificados em quatro categorias: colinas, aminas biogênicas, aminoácidos e neuropeptídeos.

³⁷ A rede neural opera o processamento de conexões em paralelo, fato que pode vir a aumentar ou diminuir a sua atividade de acordo com condições específicas de outros neurônios que promovam um efeito excitatório (sinapses que facilitam a passagem do potencial de ação) ou inibitório (inibem a passagem do potencial de ação).

(*hardware*) ou simulação em programas de computador (*software*), a fim de promover a modificação dos pesos de suas sinapses resultando no armazenamento e recuperação do conhecimento experiencial (HAYKIN, 1999).

O sistema de arquitetura conexionista “genérico” funciona de forma paralela (PDP – *Parallel Distributed Processing*), em que todos os seus neurônios podem se comunicar entre si. Nós, nódulos, ou simplesmente neurônios, unidades básicas do sistema neural artificial e de inspiração biológica, são a base de todos os processos de treinamento e aprendizagem de uma rede neural artificial.

O modelo de neurônio artificial de McCulloch e Pitts (1943) é uma aproximação útil do neurônio real e serve até hoje como bloco construtivo básico de algoritmos de redes neurais artificiais. A proposta pretendia modelar aspectos ligados ao processamento da informação e, em última análise, a produção do pensamento³⁸.

A partir de então, modelos matemáticos têm tentado representar as sinapses, os dendritos, o corpo celular e o axônio. Com inspiração no neurônio biológico, cada ramo da árvore dendrítica é modelado como uma linha ou canal de transmissão por onde flui a informação de entrada x_j , $j=1, \dots, p$, em que p é a quantidade de canais de entrada. A força (ou eficiência) das conexões sinápticas de uma certa árvore dendrítica é modelada como um fator (peso sináptico), cujo papel é modular o fluxo de sinais passando por uma certa árvore dendrítica. A função do corpo celular de realizar o balanço ou acúmulo energético é modelada por uma operação de somatório sobre as entradas moduladas pelos pesos sinápticos. O axônio é modelado como uma chave ON/OFF que indica se o neurônio respondeu ao estímulo atual, ou seja, se houve ou não o envio de um potencial de ação (BARRETO, 2007).

³⁸ No conexionismo, o pensamento nada mais é do que o resultado da ordenação mecânica de uma série de representações ou símbolos, diferentemente do que era apregoado pela IA simbólica que, numa segunda fase, considerava a mente como um conjunto de proposições e o ato de pensar seria encadear tais proposições por meio de conectivos lógicos para solucionar problemas, de forma algorítmica.

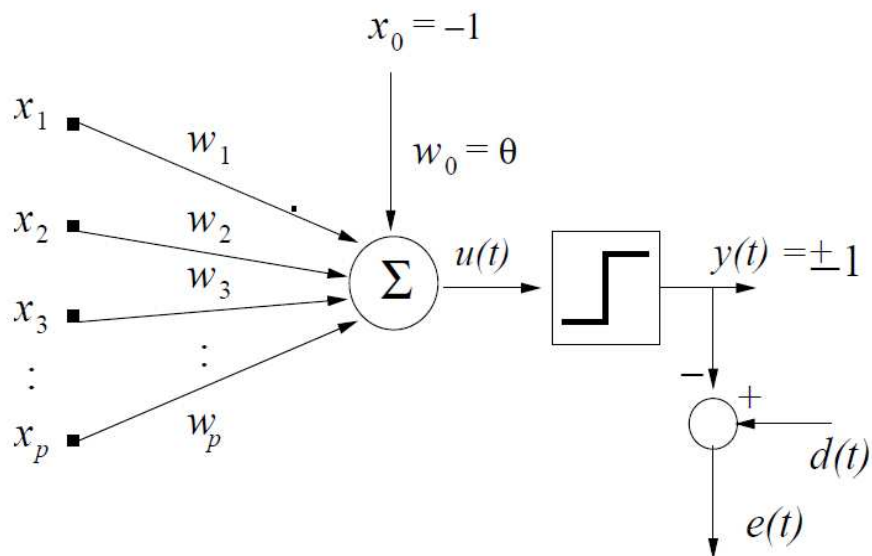


Figura 4.3: Modelo do neurônio artificial de McCulloch & Pitts.
 Fonte: Barreto (2007).

Conforme a figura 4.3, w_j são os pesos sinápticos, x_j são as entradas dos neurônios, $u(t)$ é a ativação do neurônio, $y(t)$ é a saída do neurônio, $d(t)$ é a saída desejada e $e(t)$ é o erro produzido na iteração t .

Em oposição ao sistema nervoso humano, constituído de aproximadamente 100 bilhões de neurônios, com cada neurônio mantendo cerca de 10.000 interconexões por meio de vias de transmissão (axônios) que podem atingir vários centímetros, a grande maioria das arquiteturas de redes neurais artificiais contém apenas algumas dezenas ou, no máximo, centenas de neurônios. Independentemente da topologia de uma rede neural, seu comportamento ao longo do tempo pode ser formalizado matematicamente através de ferramentas formais da álgebra linear³⁹ (BARRETO, 1998).

A cada instante de tempo, o estado de uma rede neural, com n neurônios, é representado por um vetor n -dimensional cujos componentes são os valores das atividades neurais, que são as saídas dos neurônios. As interconexões entre neurônios, com seus respectivos pesos, podem ser então representadas na forma de uma matriz de conexões sinápticas (alguns pesos podem ter valor zero, significando ausência de conexão) (HAYKIN, 1999). As matrizes de conexões

³⁹ Para mais informações sobre as propriedades dos modelos de álgebra linear no processamento distribuído em paralelo consultar Jordan (1986).

definem os padrões de conectividade entre as diferentes camadas de processamento de uma rede neural.

Do ponto de vista da topologia da interligação entre neurônios, uma rede pode ser de *alimentação para frente* (*feedforward*) ou *recorrente* (*recurrent*). Nas redes de alimentação para frente, os neurônios são organizados em camadas e a informação se desloca em um único sentido, entre camadas adjacentes. Nas redes recorrentes, não existe direção privilegiada para a propagação da informação, podendo haver retro-alimentação (*feedback*).

Embora exista uma ampla variedade de modelos de redes neurais artificiais, elas sempre seguem dois princípios básicos: a) o estado da rede neural com n neurônios pode ser descrito como um vetor de dimensão n contendo os valores de ativação dos neurônios; e b) a memória é criada ao se modificar a força das conexões entre as unidades neurais.

O primeiro modelo de rede neural, chamado de *Perceptron Simples*, foi proposto por Frank Rosenblatt, em 1957. Este modelo consiste em uma rede de duas camadas, sendo uma de entrada, utilizada apenas para redistribuição da informação de entrada, e outra camada de saída onde o processamento é realmente realizado, formada por neurônios binários. Por natureza, o *Perceptron*⁴⁰ se constitui como um classificador linear sem realimentação (*feedforward*).

É uma rede neural supervisionada, ou seja, tanto a entrada quanto a *saída desejada* são oferecidas durante o aprendizado e, caso a saída real sofra um desvio da *entrada desejada*, então os pesos são ajustados para reduzir o erro. É uma rede neural destinada a resolver problemas muito simples (linearmente separáveis). Embora esse modelo tenha dado o ponta-pé inicial nas redes neurais artificiais, foi contestado por cientistas da inteligência artificial (Papert e Minsky, por exemplo) no fim da década de 1960 por dar conta de uma classe limitada de problemas, relegando as redes neurais artificiais ao ostracismo ainda por algumas décadas.

A partir daí, diversos modelos neurais para classificação de padrões foram propostos, entre eles o Perceptron Multicamadas (*Multilayer Perceptron*-Rede MLP). Esse modelo supervisionado de rede neural, a rede MLP, faz uso do algoritmo

⁴⁰ A regra de aprendizagem do *Perceptron* permitiu a construção automática de regras de decisão com base em exemplos, de tal forma que ao final de uma fase de treinamento a rede *Perceptron* pudesse ser usada como um classificador de padrões.

de aprendizagem por retropropagação (*back-propagation*), proposto por Rumelhart, Hinton e Williams (1986). Este algoritmo se provou empiricamente útil por oferecer soluções práticas para uma variedade enorme de problemas e ensejou um novo conexionismo.

Um exemplo disso é a sua aplicação na resolução de problemas relacionados à sintaxe. Elman (1991) apresentou resultados promissores quando investigou como uma Rede Recorrente Simples (*Simple Recurrent Network – SRN – Cf. ELMAN, 1990 e 1991*) representava contexto livre e contexto sensível aos componentes gramaticais. Como se pode observar, o aprendizado supervisionado pode ser usado para modelar a aquisição linguística, todavia o treinamento requer um esquema mais elaborado para originar saídas desejáveis durante o aprendizado a partir dos dados de entrada.

Ainda sobre a aprendizagem via *Perceptron*, esta foi utilizada, sob uma perspectiva linguístico-morfológica, para simular o aprendizado do passado simples dos verbos da língua inglesa por Rumelhart & MacClelland (1986). Esse foi o primeiro trabalho a demonstrar que um único mecanismo poderia ser usado para derivar formas de tempo passado de verbos de seus radicais tanto nos casos de verbos regulares quanto verbos irregulares (NEGISHI, 1998).

A próxima seção traz mais uma rede neural artificial e seu funcionamento: os mapas auto-organizáveis, uma rede neural com algoritmo de aprendizagem não-supervisionado. Essa rede será utilizada nas simulações apresentadas no capítulo 6.

4.2 Mapas auto-organizáveis e a rede de Kohonen

Um mapa auto-organizável (SOM - *Self-Organizing Map*) é um tipo de rede neural artificial treinada por aprendizagem competitiva não-supervisionada baseada em princípios de auto-organização de sistemas que permite a representação de dados multidimensionais em espaços de dimensões menores (KOHONEN, 2001).

O modelo de Kohonen torna possível o agrupamento de padrões similares em certas áreas da saída da rede, como também a identificação de padrões com características comuns. O agrupamento acontece devido ao fato de o mapa ser formado por neurônios (unidades mínimas de processamento).

Ao se deparar com o modelo de funcionamento dessa rede, é necessário não se restringir demais a alguns dos conceitos vistos anteriormente, tais como neurônio artificial (*Perceptron* Simples, por exemplo), funções de ativação, conexões *feedforward*, redes recorrentes etc. Essa nova rede, embora ainda mantenha alguns termos de modelos conexionistas já vistos, não os utiliza da mesma forma e com o mesmo propósito, conforme será visto a seguir.

4.2.1 Os princípios de auto-organização

O desenvolvimento de mapas auto-organizáveis como modelos neurais foi motivado a partir da característica distintiva de organização cerebral local em mapas topologicamente ordenados (mapas corticais do cérebro humano).

O mapa auto-organizável foi criado para ser um sistema capaz de aprender por meio de auto-organização de forma neurobiologicamente inspirada⁴¹. Sob esta perspectiva, o aspecto mais relevante do sistema nervoso a ser levado em conta é a sua capacidade de formação topográfica de mapa, fato que o difere de outras redes neurais. Como dito por Kohonen (2001), “a localização espacial ou as coordenadas de uma célula na rede correspondem a um domínio particular de padrões de sinais de entrada”. O mapa computacional constitui-se, assim, como uma pedra fundamental básica na infraestrutura de processamento da informação do sistema nervoso.

Já que se espera que um modelo de rede não-supervisionada seja capaz de se auto-organizar, ou seja, de fornecer pontos entre as posições finais e iniciais dadas, faz-se necessário definir e conceituar “auto-organização”. Segundo Kelso (1995, p. 16-17 *apud* Barbosa, 2005), os conceitos elementares das condições de auto-organização são ⁴² :

⁴¹ Na realidade, os mapas auto-organizáveis de Kohonen possuem inspiração biológica, todavia não com uma plausibilidade biológica. Os mapas corticais nos quais eles se baseiam (mapa retinotópico, mapa tonotópico e mapa somatotópico) possuem uma inspiração totalmente biológica (BARRETO, 2009, comunicação pessoal).

⁴² Trecho original: “*Elementary Concepts of and Conditions for Self-Organization*”:

1. Patterns arise spontaneously as the result of large numbers of interacting components. If there aren't enough components or they are prevented from interacting, you won't see patterns emerge or evolve. The nature of the interactions must be nonlinear. This constitutes a major break with Sir Isaac Newton, who said in Definition II of the *Principia*: "The motion of the whole is the sum of the motion of all the parts." For us, the motion of the whole is not only greater than, but *different* than the sum of the motions of the parts, due to nonlinear interactions among the parts or between the parts and the environment.

1. Os padrões (combinação de qualidades ou parâmetros dependentes que forma um arranjo característico) emergem espontaneamente como resultado da interação de um grande número de componentes;
2. O sistema deve estar distante de seu estado de equilíbrio, que muitas vezes pode constituir um empecilho para mudança de estado a curto prazo;
3. O sistema tem um tom ou mais parâmetros de ordem que emerge(m) da própria coordenação entre as partes, mas que por sua vez influencia(m) o sistema como um todo;
4. A identificação dos parâmetros de ordem é mais fácil nos momentos de transição do estado do sistema, notadamente as bifurcações (a bifurcação é uma transição brusca no comportamento do sistema devido a pequena mudança num parâmetro de controle);
5. A perturbação do sistema (modificação de seu estado atual imposta exatamente) é a melhor maneira de determinar a relação funcional entre seus parâmetros de ordem (e de controle);
6. Os parâmetros de controle guiam o sistema para novos estados, mas não emergem da coordenação das partes do sistema;
7. A dinâmica dos parâmetros de ordem pode ser simples (atrator pontual ou cíclico) ou complexa.

O processo de auto-organização é interpretado como a emergência de um complexo padrão de organização a partir da desordem inicial, fruto da interação entre os componentes do sistema como um todo (BARRETO, 2002). Em suma, os conjuntos de neurônios, isto é, as redes neurais, sejam biológicas ou artificiais, apresentam uma capacidade de auto-organização que não é própria da lógica e ainda possuem propriedades emergentes de padrões globais.

2. The system must be *dissipative* and far from (thermal) equilibrium. Due to nonlinear interactions in the system, heat or energy doesn't diffuse uniformly but is concentrated into structural flows that transport the heat (dissipate it) more efficiently. As a result of dissipation, many of the system's degrees of freedom are suppressed and only a few contribute to the behavior. Intuitively, dissipation is equivalent to a kind of attraction that may take several forms (see 7 below).

3. *Relevant* degrees of freedom, those characterizing emerging patterns in complex systems, are called collective variables or order parameters in synergetics. An order parameter is created by the coordination between the parts, but in turn influences the behavior of the parts. This is what we mean by *circular causality*, which, incidentally, is not the same as tautology.

4. Order parameters are found near nonequilibrium phase transitions, where loss of stability gives rise to new or different patterns and/or switching between patterns. Order parameters may exist far from transitions as well, but it is difficult to identify them.

5. Fluctuations are continuously probing the system, allowing it to feel its stability and providing an opportunity to discover new patterns. Fluctuations are positive sources of noise, not just something to be rid of.

6. Parameters that lead the system through different patterns, but that (unlike order parameters) are not typically dependent on the patterns themselves, are called control parameters. Such control parameters may be quite *unspecific* in nature; that is, in no sense do they act as a code or a prescription for the emerging patterns. On the other hand, we will see that in biological systems boundary conditions may be quite *specific* in an informational sense, especially when a particular pattern is required (Cf. chapters 5 and 6).

7. The order parameter dynamics, the equation describing the coordinated motion of the system, may have simple (fixed point, limit cycle) or complicated solutions including deterministic chaos and stochastic (random) aspects thereby giving rise to enormous behavioral complexity.

Como Barreto (2002) afirma, “a limitação de recursos leva à competição entre sinapses, e em consequência, à seleção das sinapses mais vigorosas (mais adequadas) às expensas de outras menos adaptadas”. Os princípios de auto-organização seguem os mesmos princípios da evolução das espécies: a seleção natural, isto é, as sinapses que tendem a se enfraquecer são aquelas que não conseguem suceder em alcançar recursos limitados.

Em uma abordagem de aprendizado não-supervisionado (ou auto-organizado) não existe supervisor para avaliar o desempenho da rede em relação ao conjunto dos dados de entrada⁴³. A rede se adapta sozinha às regularidades dos dados de entrada, desenvolvendo uma habilidade para criar representações internas que codificam as características dos dados de entrada e, dessa forma, geram novas classes automaticamente (CASTRO, 2006).

Após o aprendizado não-supervisionado, ou seja, o aprendizado em que somente as entradas da rede são oferecidas, cada neurônio se torna um protótipo dos dados de entrada. Protótipos similares tendem a se agrupar próximos uns dos outros numa ordenação topológica à camada de saída, ou melhor, o mapa. Para um neurônio i ser o vencedor, a distância entre os vetores de peso⁴⁴ correspondentes (w_i) e um certo padrão de entrada x deve ser a menor medida (entre todos as unidades de saída da rede), geralmente aferida com a distância euclidiana. Consequentemente, a ideia é achar um neurônio de saída cujo peso de vetores é mais similar (isto é, que tenha a menor medida) ao padrão de entrada apresentado (CASTRO, 2006)

Durante o processo de auto-organização da rede, vários padrões de ordenação topológica emergem como resultado do processo de aprendizado dos padrões de entrada. Em suma, a rede neural aprende a distribuição estatística do padrão de entrada e oferece a saída que a caracteriza em relação à distribuição aprendida, *i.e.*, a saída corresponde à categoria a qual a entrada pertence, ou traços que são encontrados na entrada (NEGUISH, 1998). Basta apresentar os dados de treinamento à rede para que estes sejam classificados automaticamente, sem nenhuma supervisão externa (BITTENCOURT, 2006).

⁴³ Em um aprendizado não-supervisionado não existe informação de erro sendo retroalimentada dentro da rede; as classes dos dados são desconhecidas ou não rotuladas e a presença de um supervisor não existe mais.

⁴⁴ *Codebook vectors*: conjunto de vetores de peso da rede.

4.2.2 O Algoritmo de Treinamento da Rede de Kohonen

O objetivo de aprendizagem em um mapa auto-organizável é fazer com que diferentes partes da rede respondam de forma similar a certos padrões de entrada. Isto é parcialmente motivado pela forma como outras informações visuais, auditivas ou outras sensoriais são manipuladas em partes diferentes do córtex cerebral no cérebro humano (KOHONEN, 2001).

A maioria dos algoritmos de treinamento de redes neurais é inspirada, de forma direta ou indireta, pela lei de Hebb, que estipula que a intensidade de uma ligação sináptica entre dois neurônios aumenta se ambos são excitados simultaneamente. Numa rede neural baseada em aprendizagem Hebbiana, vários neurônios da camada de saída podem estar simultaneamente ativos.

Já na aprendizagem competitiva, os neurônios da camada de saída “competem” entre si pelo direito de responder, ou seja, de permanecerem ativos para um dado estímulo de entrada. Ao final desta competição, apenas uma unidade permanecerá ativa, como resposta àquela informação da entrada. Com um único neurônio de saída sendo ativado a cada iteração, essa propriedade torna o algoritmo apropriado para descobrir características estatísticas salientes dentro dos dados que podem ser usados para classificar um conjunto de padrões de entrada (CASTRO, 2006), de acordo com a figura 4.4.

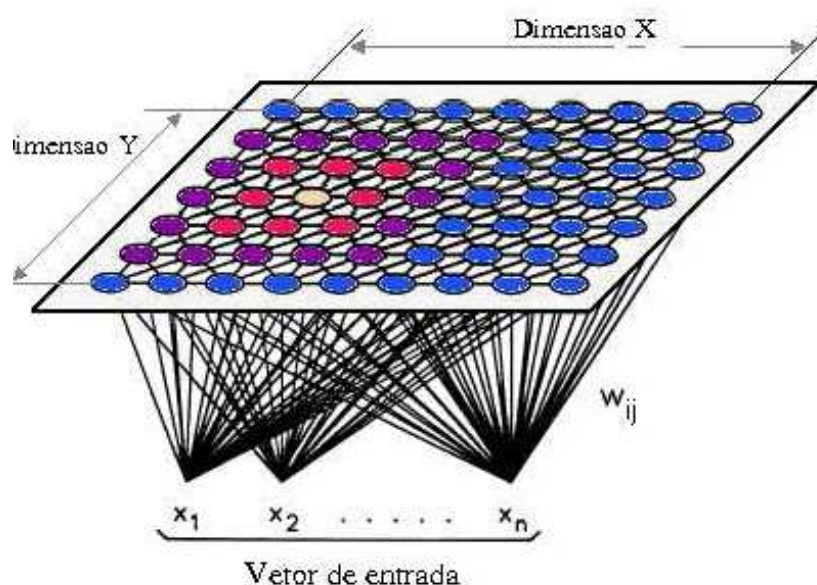


Figura 4.4: Neurônios organizados em um mapa de Kohonen bidimensional.
Fonte: Barreto (2007).

Os pesos dos neurônios são inicializados com valores aleatórios pequenos ou uniformemente amostrados do subespaço gerado pelos dois maiores autovetores (*eigenvectors*) principais. Com a última alternativa, a aprendizagem é muito mais rápida porque os pesos iniciais já estão relativamente bem ordenados. A cada época, ou seja, a cada apresentação dos dados à rede, todos os neurônios têm seus pesos ajustados.

A rede deve ser alimentada por um grande número de vetores de exemplo que representam, tanto quanto possível, os tipos de vetores esperados durante o mapeamento. Os exemplos são normalmente administrados várias vezes com iterações (também chamadas de épocas)⁴⁵.

Como o objetivo deste trabalho não é fazer um estudo exaustivo das propriedades matemáticas da rede de Kohonen, indica-se a leitura das obras de Kohonen (1985, 1998 e 2001, nesta ordem) para uma compreensão mais aprofundada dos princípios matemáticos que regem a teoria.

4.2.3 Busca pelo neurônio vencedor

Após a fase de treinamento da rede, os vetores protótipos são organizados no mapa. Em princípio, cada neurônio é um vetor de pesos de dimensão d , em que d é igual à dimensão dos vetores de entrada. Os neurônios estão conectados a neurônios vizinhos por meio de uma relação de vizinhança, que contribui posteriormente para a preservação topológica⁴⁶, estrutural, do mapa.

O processo de formação do mapa pode ser imaginado como uma rede, cujo arranjo usual de nós é um espaçamento regular em uma grade hexagonal ou retangular, que se espalha de acordo com a disposição da nuvem de dados. O algoritmo de treinamento move os vetores de peso de forma a espalhar a nuvem de dados na grade. Dessa forma, vetores com pesos similares ao final do treinamento serão neurônios vizinhos na grade (KOHONEN, 2001).

A função de vizinhança funciona como uma espécie de janela de ponderação (*weighting window*) que faz com que os neurônios mais próximos do

⁴⁵ Época (ou Iteração): cada vez que todo o conjunto de dados de entrada são apresentados à rede.

⁴⁶ Na rede SOM, os neurônios estão dispostos em arranjos geométricos fixos, i.e., em linha (SOM-1D) ou em grade (SOM-2D). Quando a rede SOM tenta fazer com que os protótipos de neurônios que são próximos no arranjo geométrico também estejam próximos no espaço de entrada é uma tentativa de obter um erro de quantização mínimo (CRUZ, 2007).

neurônio vencedor atual tenham seus vetores de pesos atualizados de forma mais intensa que aqueles neurônios que estão mais distantes do neurônio vencedor, ou seja, os neurônios vizinhos também possuem seus pesos atualizados, como mostra a figura 4.5:

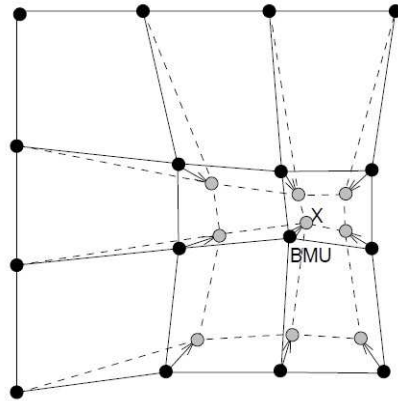


Figura 4.5: Atualização do neurônio vencedor⁴⁷ (*best matching unit* – BMU) e seus vizinhos em direção à amostra indicada com X.
 Fonte: Vesanto et al. (2000, p. 2).

A busca pelo neurônio vencedor acontece por meio da atualização dos pesos da unidade cujos pesos mais se aproximem dos valores dos dados de entrada. Em outras palavras, o neurônio vencedor tem seus pesos reajustados com maior intensidade, visto que para ele tem-se $h(i^*, i; t) = 1$. Para todos os outros neurônios, tem-se $h(i^*, i; t) < 1$, em que h é a função de vizinhança, i^* é o neurônio vencedor e i é o neurônio vizinho (CRUZ, 2007), de acordo com a figura 4.6.

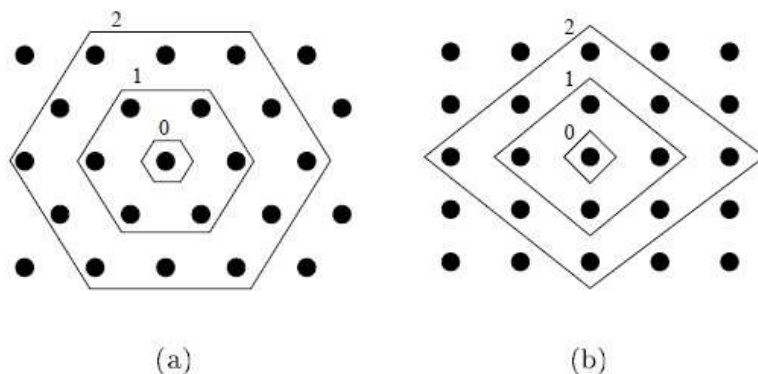


Figura 4.6: Vizinhos (0, 1 e 2) da unidade mais central: (a) grade hexagonal e (b) grade retangular.
 Fonte: VESANTO et al. (2000, p. 2).

⁴⁷ As linhas sólidas e pontilhadas correspondem à situação antes e depois da atualização, respectivamente.

A depender do tipo de algoritmo utilizado para o treinamento, as amostras podem ser apresentadas à rede por vez (treinamento seqüencial – *sequence training*) ou os dados podem ser apresentados à rede como um todo (treinamento em lote – *batch training*).

Durante o mapeamento, existirá um único neurônio vencedor (BMU), ou seja, o neurônio cujo vetor de peso permanece mais próximo do vetor de entrada. Isso pode ser simplesmente determinado, ao se calcular a distância euclidiana⁴⁸ entre o vetor de entrada e o vetor de peso.

Como explica Barreto (2007), o mapa $i^*(x) : X \rightarrow A$, definido pela matriz de pesos $W = (w_1, w_2, \dots, w_q)$, $w_i \in \mathbb{R}^p \subset X$, atribui a cada vetor de entrada $x(t) \in \mathbb{R}^p \subset X$ um neurônio vencedor $i^*(t) \subset X$, determinado por:

$$i^*(t) = \underset{i}{\operatorname{argmin}} \|x(t) - w_i(t)\|, \quad (\text{eq. 7})$$

em que $\| \cdot \|$ denota a distância euclidiana e t simboliza o tempo discreto associado às iterações do algoritmo.

Os vetores de peso, $w_i(t)$, também chamados de protótipos ou vetores de código, são treinados de acordo com uma regra de concorrência cooperativa de aprendizagem em que os vetores de pesos de um neurônio vencedor e seus vizinhos na matriz de saída são atualizados após a apresentação do vetor de entrada:

$$w_i(t+1) = w_i(t) + \alpha(t) h(i^*, i; t)[x(t) - w_i(t)] \quad (\text{eq. 8})$$

em que $0 < \alpha(t) < 1$ é a taxa de aprendizado, e $h(i^*, i; t)$ é a função de peso que limita a vizinhança do neurônio vencedor. Uma escolha usual por $h(i^*, i; t)$ é dada pela função gaussiana:

$$h(i^*, i; t) = \exp\left(-\frac{\|r_i(t) - r_{i^*}(t)\|^2}{2\sigma^2(t)}\right) \quad (\text{eq. 9})$$

⁴⁸ A distância euclidiana, uma das medidas de proximidade mais utilizadas entre agrupamento de dados, é utilizada para encontrar o vetor protótipo mais próximo dos vetores dos dados de entrada.

em que $r_i(t)$ e $r_{i^*}(t)$ são, respectivamente, as coordenadas dos neurônios i e i^* no mapa de saída A , e $\sigma(t) > 0$ define o raio da função de vizinhança num tempo t . As variáveis $\alpha(t)$ e $\sigma(t)$ devem decair com o tempo a fim de garantir a convergência dos vetores de peso para estados estáveis. Nesta tese, adota-se o decaimento exponencial, dado por:

$$\alpha(t) = \alpha_0 \frac{\alpha_T}{\alpha_0} e^{-t/\tau} \text{ e } \sigma(t) = \sigma_0 \frac{\sigma_T}{\sigma_0} e^{-t/\tau} \quad (\text{eq. 10})$$

em que α_0 (σ_0) e α_T (σ_T) são os valores iniciais e finais de $\alpha(t)$ ($\sigma(t)$), respectivamente. O ajuste dos pesos é realizado até que um estado global de ordenação e estabilidade dos vetores de peso seja alcançado. Essa fase é chamada de convergência do mapa (BARRETO, 2007).

4.2.4 A matriz de distância unificada (Matriz-U)

Os mapas de Kohonen são utilizados usualmente como ferramenta de visualização de dados, permitindo que inter-relações existentes em conjuntos complexos de dados possam ser percebidas e classificadas.

A matriz-U é a ferramenta canônica para a demonstração da distância de estruturas dos dados de entrada em um mapa auto-organizável. Seus métodos têm sido amplamente usados para agrupamento de conjuntos de dados de alta dimensão. Metaforicamente falando, ela apresenta uma “paisagem” das relações de distância dos dados de entrada no espaço dos dados. Eis algumas propriedades da Matriz-U (ULTSCH, 2003, p. 2):

- A posição das projeções dos pontos dos dados de entrada reflete a topologia do espaço de entrada. Essa característica é herdada diretamente do algoritmo SOM;
- Os vetores de peso de neurônios com grandes alturas U são muito distantes de outros vetores no espaço de dados;
- Os vetores de peso de neurônios com pequenas alturas U são circundados por outros vetores no espaço de dados;
- As projeções dos pontos dos dados de entrada são tipicamente encontradas em depressões (vales);
- O que sobra no espaço de entrada é encontrado nos funis do mapa;
- As cordilheiras sobre um ponto da matriz U apontam para fronteiras de agrupamento;

- Os vales sobre uma matriz U apontam para centros de agrupamentos.

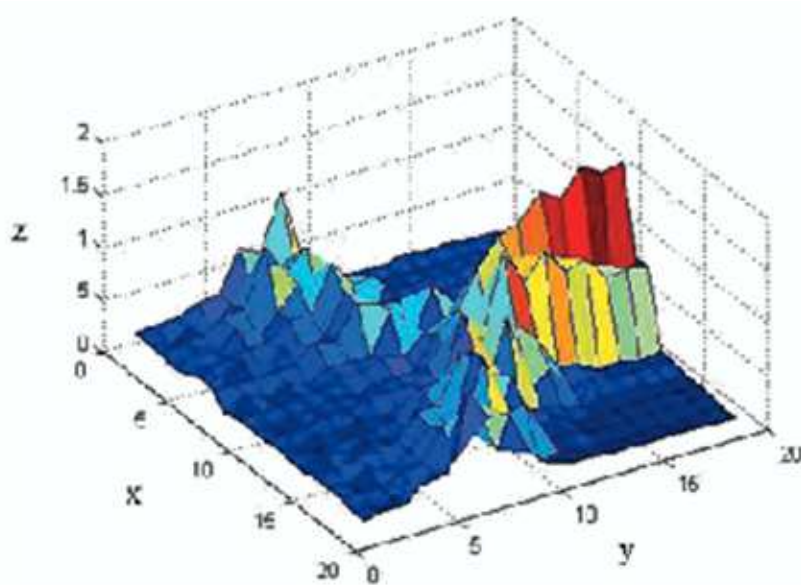


Figura 4.7: Exemplo de matriz U - Matriz visualizada com representação por superfície topológica.
Fonte: BOSCAROLI (2008, p. 68).

Após as duas fases da rede, a **formação** do mapa, usando-se exemplos de entrada por meio de quantização vetorial e o **mapeamento**, é possível visualizar a disposição dos neurônios vencedores na grade de neurônios, como mostra a Figura 4.8:

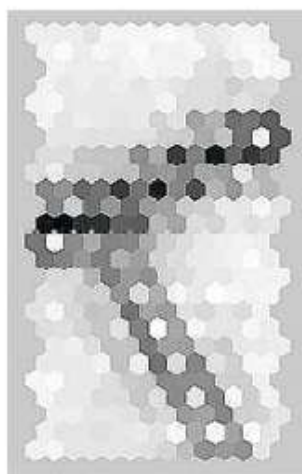


Figura 4.8: Exemplo de matriz- U matriz bidimensional com tons de cinza.
Fonte: BOSCAROLI (2008, p. 63).

Em uma matriz-U bidimensional, pode-se visualizar região de agrupamento e região de separação. Os valores representados na matriz-U para um nó particular correspondem a distâncias entre o vetor de pesos deste nó e os vetores de pesos de seus vizinhos mais próximos. Numa grade quadrada, por exemplo, podem-se considerar os 4 ou 8 nós próximos ao BMU ou seis nós em uma grade hexagonal, conforme figura 4.9.

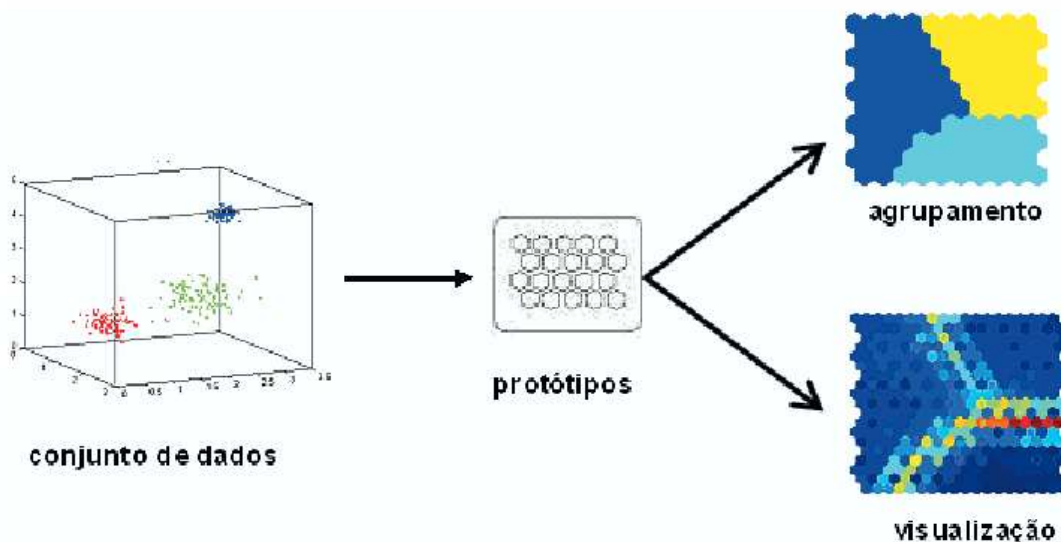


Figura 4.9: Análise de dados a partir do SOM.
Fonte: BOSCAROLI (2008, p. 70).

A despeito de a Matriz-U ser uma ferramenta poderosa na visualização de agrupamentos, Silva (2004) chama atenção para algumas restrições do método de identificação visual de agrupamentos desta ferramenta:

Para mapas pequenos a U-matriz gerada tende a ser complexa e de difícil identificação visual dos agrupamentos (...), além do que, a U-matriz só pode ser gerada a partir de mapas com grade bi-dimensional. Para mapas com dimensões de grade maior que 2 o processo de visualização da matriz de distância torna-se complexo.

Essa informação se tornou de grande valia para as decisões metodológicas apresentadas no capítulo 5. A partir de então, decidiu-se gerar mapas de médio porte e com dimensão bidimensional. Os resultados foram satisfatórios, como se poderá observar no capítulo 6.

4.3 As vantagens de utilização da Rede Auto-Organizável

Os sistemas conexionistas têm obtido êxito ao explicar o funcionamento de uma gama de fenômenos de interesse para a ciência cognitiva ao prover modelos mais adequados para a explicação de processos cognitivos, por serem biologicamente plausíveis, ao contrário da IA simbólica. De acordo com Vainio (2001, p. 3):

As redes neurais são conhecidas por sua capacidade de generalização de acordo com a similaridade de seus insumos, mas também por distinguir diferentes saídas de padrões de entrada que são similares apenas na superfície. Como consequência, as redes têm o poder de prever, depois de uma fase de aprendizagem adequada, mesmo os padrões nunca vistos antes.⁴⁹

Os mapas auto-organizáveis podem ser usados para categorizar o *input* de uma maneira não-supervisionada, o que é muito útil em vários níveis de categorização linguística (ÁGOSTON, 2005, p. 114). Um dos princípios fundamentais da teoria conexionista afirma que o aprendizado se dá na ausência de um *feedback* explícito. No entanto, há um *feedback* implícito a todo momento. As redes neurais não-supervisionadas apresentam uma visão dissidente dentro da própria teoria a qual fazem parte (conexionismo) por afirmarem que o aprendizado se dá sem a ajuda direta de um professor. No aprendizado não-supervisionado a rede tem de descobrir sozinha relações, padrões, regularidades ou categorias nos dados que lhe vão sendo apresentados e codificá-las nas saídas.

Desta forma, a maneira de categorização da rede neural auto-organizável de Kohonen se mostra como uma forma mais conveniente de responder a questões lingüísticas de maneira mais prática e rápida, já que exclui uma das variáveis que tornam o processamento da informação lingüística mais complexo, sou seja, o *feedback*.

Já os *Perceptrons* multicamadas obtêm seu conhecimento de maneira supervisionada, o que vai de encontro com o mundo real, em que a maioria do conhecimento é obtida sem supervisão (VAINIO, 2001).

⁴⁹ Trecho original: "Neural networks are known for their ability to generalize according to the similarity of their inputs but also to distinguish different outputs from input patterns that are similar only on the surface. As a consequence, networks have the power to predict, after an appropriate learning phase, even patterns they have never seen before."

Além disso, as redes de aprendizagem não-supervisionada são um auxiliar poderoso na explicação de diversos processos linguísticos, como a análise lexical, a aquisição do léxico e da sintaxe, a explicação de interferências linguísticas, a reavaliação de processos cognitivos, o código escrito e o redimensionamento da inteligência artificial (POERSCH; ROSSA, 2007).

As redes de aprendizado supervisionado, embora tenham sido capazes de modelar muitos aspectos do aprendizado linguístico da criança, apresentam falta de plausibilidade biológica e psicológica. No aprendizado linguístico real, a criança **não** recebe um retorno constante sobre o que está incorreto na sua fala (tipo de correção de erro oferecido à rede por algoritmos de aprendizagem supervisionada) (Cf. LI et al., 2004). Para os linguistas cognitivos que defendem uma cognição de forma situada, a aprendizagem supervisionada, por sua vez, não pode ser equiparada à forma como acontece na realidade, ou seja, a aprendizagem de maneira supervisionada não pode replicar o contexto do mundo real pelo fato de o sujeito não ser uma entidade isolada e estar o tempo inteiro recebendo *feedback* do contexto à sua volta.

Justifica-se a escolha dos mapas auto-organizáveis de Kohonen pelo fato de estes servirem ao propósito de interpretação de processos de classificação e transferência de conhecimento linguístico, mais especificamente o conhecimento acentual. Este tipo de rede é usado quando se deseja reconhecer diversos padrões para os quais não se tem conhecimento prévio das classes ou categorias e quando o que se deseja é justamente descobrir como os padrões estão organizados de forma rápida e precisa.

A vantagem de se usar o mapa auto-organizável para estudar processos de aprendizagem reside não somente em seus princípios auto-organizacionais (competição entre neurônios por recursos limitados e cooperação implementada pela função vizinhança), mas também em sua visualização e propriedades de abstração (KOHONEN, 1998). Além disso, a rede SOM é capaz de representar estruturas hierárquicas implícitas e modelar aspectos específicos de redes neurais biológicas. Constitui-se como um representante de um novo paradigma na inteligência artificial e na modelagem cognitiva; é uma ferramenta para análise estatística; é uma ferramenta para o desenvolvimento de aplicações complexas (HONKELA, 1997).

Para finalizar a justificativa para o uso da rede de Kohonen em tarefas que, a princípio, exigem o processamento de informação temporal, Barreto (2007) apresenta um apanhado das aplicações dos mapas auto-organizáveis de Kohonen na predição de séries temporais (TSP – *Time Series Prediction*). O autor deixa claro que, mesmo a despeito de as redes terem sido criadas como algoritmos de aprendizado não-supervisionado, essa rede é flexível o suficiente a ponto de superar um número considerável de arquiteturas neurais supervisionadas devotadas a tarefas de TSP, por exemplo, as redes recorrentes, argumento mais do que suficiente para considerar a rede de Kohonen como ideal para a simulação dos fenômenos linguísticos aqui pesquisados.

4.4 Aplicações dos Mapas Auto-Organizáveis ao Processamento de Fala Natural

A máquina de escrever fonética (*Phonetic Typewriter*) (KOHONEN, 1988) foi um exemplo de aplicação dos mapas auto-organizáveis à produção de saídas fonéticas desejadas. O objetivo desta rede era informar os prováveis fonemas constituintes daquela palavra quando uma palavra era pronunciada, no caso do idioma Finlandês. Para tanto, a rede foi treinada com os fonemas do idioma, em que estes, dependendo de suas características, eram agrupados com outros que possuíam características similares. Assim, a cada fonema pronunciado um determinado neurônio de saída correspondente era ativado.

Outra aplicação foi o mapa semântico de palavras de Ritter e Kohonen (1990). Os autores apresentaram um dos pontos fortes da rede SOM, isto é, a sua capacidade de formar um mapa de itens de entrada que diferem uns dos outros em uma forma multifacetada. Eles testaram que tipo de mapa seria formado por itens lexicais, o que diferiria um do outro em várias dimensões léxico-semânticas e sintáticas.

A parte mais difícil do *design* do modelo da rede foi determinar a representação de entrada para cada palavra. A solução foi representar cada palavra pelo contexto em que elas apareciam nas frases. A representação de entrada consistia em duas partes: uma que servia como um identificador da palavra individual, e outro que representava o contexto em que a palavra aparecia. A parte

anterior foi representada por 7 nós e um padrão de atividade gerada aleatoriamente foi associado a cada palavra. A última parte do *design* do modelo consistia em concatenar as representações da palavra anterior com a palavra que se seguia, representando, assim, um total de 14 nós. Em termos práticos, essa modelagem representou um custo computacional elevado em comparação ao comportamento de uma rede neural que dispensa o contexto, no caso, a rede de Kohonen.

O mapa, ou seja, a camada de saída foi uma matriz bidimensional de 10x15 nós. Depois de 2000 apresentações de 10.000 sentenças geradas aleatoriamente, o mapa formado foi examinado. Ao fim do treinamento da rede, foi percebido que não apenas classes de palavras (substantivo, verbo e advérbio) foram mapeadas como também subclasses de palavras léxico-semânticas foram capturadas. Os verbos transitivos e verbos intransitivos formaram sub-regiões dentro da região (ou também chamado de espaço de formação) “verbo”. Além disso, nomes de pessoas, animais e substantivos inanimados formaram sub-regiões dentro da região “substantivo”. Vale ressaltar que o contexto foi definido apenas por palavras antecessoras e sucessoras. Negishi (1998) afirma que o fato de o modelo ter capturado subclasses de palavras indica a força da abordagem conexionista em capturar transições graduais entre regras e comportamento idiossincrático.

Gauthier et al. (2009) usaram modelos conexionistas para explorar se e como as crianças poderiam aprender foco prosódico⁵⁰ diretamente de *input* de fala contínua. Em três simulações utilizando redes neurais auto-organizáveis, os autores exploraram como o foco poderia ser aprendido a partir de sinais acústicos contínuos em Mandarim, que foram produzidos com tons lexicais co-ocorrentes e por vários falantes. Os resultados desse estudo mostraram que redes neurais não-supervisionadas podem desenvolver agrupamentos específicos de foco a partir de sinais de fala dinâmicos contínuos, produzidos por vários falantes em várias condições de tom lexical, o que pode eventualmente conduzir à aquisição do foco.

Li et al. (2004) simularam a aquisição lexical em crianças utilizando um modelo de rede neural auto-organizável. O objetivo principal da pesquisa era usar as propriedades de preservação topográficas do SOM para estudar a emergência e a organização de categorias linguísticas ao longo dos estágios de aprendizagem lexical. A combinação de aprendizagem Hebbiana e auto-organização pôde dar

⁵⁰ O foco é uma função comunicativa, que serve para colocar ênfase em uma determinada parte de um enunciado.

conta do processo de como o aprendiz consolida relações entre forma de palavras e representações de significado lexical com base na frequência que essas representações podem ocorrer e quão fortemente elas são co-ativadas na representação. Os resultados das simulações demonstraram como a organização de mapas locais pode dar surgimento a categorias emergentes ao longo de estágios de aprendizado com uma plasticidade inicial substancial e competição inicial para a organização e reorganização de membros de categorias.

Com relação à confusão lexical em aprendizagem de vocabulário inicial, as simulações permitiram identificar um número de fatores cruciais que levavam às confusões observadas na linguagem infantil inicial. De forma curiosa, a confusão lexical ocorre mais frequentemente com substantivos do que em qualquer outra categoria de palavras, já que os substantivos se encontram mais densamente populados no mapa.

O modelo proposto por Li et al. (2004) apresentou uma alternativa emergentista à hipótese nativista de que as categorias lexicais são conectadas no cérebro. As simulações mostraram que a representação de categorias linguísticas pode emergir na topologia da rede como um produto natural do processo de auto-organização no aprendizado lexical.

Os estudos sobre aquisição de L2 que se utilizaram de modelos de mapas auto-organizáveis, logo acima resenhados, oferecem resultados promissores por advogarem a favor de uma comparação razoável entre as topologias e mecanismos de aprendizagem. Tal plausibilidade de analogia só se torna possível em virtude de os mapas auto-organizáveis seguirem padrões de organização muito semelhantes aos padrões que emergem na rede neural biológica durante o processo de estruturação do léxico mental. As estruturas neurais artificiais são formuladas sob o conceito de auto-organização de unidades básicas (neurônios) e regidas por princípios de aprendizagem competitiva.

Já que há a possibilidade de simulação de aquisição lexical bem como a simulação de fenômenos de transferência do conhecimento lexical, vislumbra-se também ser possível a simulação da transferência de padrões prosódicos de aprendizes de LE. No próximo capítulo, serão apresentados os procedimentos metodológicos seguidos para a realização das simulações na rede neural aqui apresentada.

4.5 Resumo do capítulo

Neste capítulo foram apresentados os princípios básicos que regem a rede neural artificial auto-organizável. Foi apresentada também uma revisão bibliográfica dos estudos de aquisição de L2 que se utilizaram desses modelos neurais. Os mapas auto-organizáveis (*Self-Organizing Maps* – SOM) são estruturas neurais artificiais formuladas sobre o conceito de auto-organização de unidades básicas (neurônios) segundo algum processo de aprendizagem competitiva.

A escolha da topologia e do mecanismo de aprendizagem depende de fatores como tipo do dado a ser analisado, grau de generalização desejado, tipo de aplicação etc. São propriedades do SOM a ordenação topológica, o casamento de densidade e a seleção de características. A rede apresenta enormes vantagens para a análise de fenômenos lingüísticos, entre elas a mais preponderante sendo a categorização do *input* de uma maneira não-supervisionada, o que é muito útil em vários níveis de categorização linguística.

O próximo capítulo apresenta o aporte metodológico adotado nesta pesquisa bem como os critérios para a seleção do *corpus*, o perfil dos participantes, os procedimentos envolvidos na coleta de dados, a aplicação dos testes e as variáveis investigadas.

5 METODOLOGIA DE PESQUISA

No presente capítulo, estão descritos os critérios para a seleção do *corpus*, o perfil dos participantes, os procedimentos envolvidos na coleta de dados, a aplicação dos testes e as variáveis dos fenômenos investigados, atrelados aos respectivos problemas. São apresentados ainda a metodologia do processamento dos dados de entrada da rede, o projeto e a simulação da rede, bem como a ferramenta de visualização dos mapas auto-organizáveis.

5.1 Caracterização dos participantes da pesquisa, o *corpus* e o método de coleta de dados

A escolha dos participantes que constituem o *corpus* da pesquisa se restringiu a um curso de extensão em língua inglesa, oferecido pela Casa de Cultura Britânica (CCB), na Universidade Federal do Ceará (UFC), à comunidade acadêmica. A escolha do local onde o *corpus* foi coletado se explica pelo fato de a pesquisadora ter sido professora temporária do curso e ter tido a oportunidade de observar na fala dos alunos uma quantidade considerável de fenômenos fonológicos típicos do processo de aquisição de inglês como L2. Vale ressaltar que os alunos que compuseram o *corpus* desta pesquisa não foram alunos da pesquisadora no semestre em que os dados foram coletados.

Após a aplicação de um questionário em sala de aula (ver apêndice B), 30 participantes foram selecionados para a entrevista sobre seu histórico em língua inglesa e, em seguida, foram aplicadas as tarefas de leitura de frases. Para fins de controle, o *corpus* contou também com um grupo de 4 falantes nativos.

O *corpus* desta pesquisa é composto das gravações das entrevistas com alunos que, à época em que o *corpus* foi gravado, cursavam os semestres II, IV e VI da CCB, com idade entre 18 e 25 anos, todos brasileiros, de ambos os sexos e que não tinham feito nenhuma viagem para um país de língua inglesa até o momento da entrevista.

Ressalta-se como informação importante: há duas possibilidades de admissão no curso básico de língua inglesa onde o *corpus* foi coletado: por teste de

admissão, em que o aluno começa no primeiro semestre ou através de teste de nível (*placement test*). Geralmente oferecem-se vagas a partir do semestre II até o semestre VI. Uma parte dos alunos que opta por ingressar no curso por teste de nível, muitas vezes, é proveniente de outros cursos de inglês da cidade, tendo, assim, uma experiência anterior ao idioma.

A aplicação de questionários em sala de aula, as entrevistas com os aprendizes escolhidos e a aplicação das tarefas aconteceram no período de outubro de 2003 a junho de 2004. As entrevistas, bem como a aplicação das tarefas de produção, foram realizadas num ambiente silencioso e tinha como instrumentos de pesquisa folhas de papel com frases que continham, pelo menos, um item lexical investigado por frase. Para tanto, foi utilizados um gravador digital Sony® e um microfone unidirecional da marca multilaser.

No total, foram colhidas 10 (dez) gravações por semestre, ou seja, 10 indivíduos entrevistados por semestre, contabilizando trinta entrevistas no total com tempo de gravação com cada entrevistado variando entre 30 (trinta) minutos e 1 (uma) hora, o que contabilizou um total de aproximadamente 30 horas. O material sonoro foi digitalizado e segmentado no programa Sound Forge 5.0 e posteriormente transcrito foneticamente (ver Apêndice F). Durante a fase de transcrição, contou-se com a colaboração de dois foneticistas: José Álber Campos Uchoa e Otaciano Noronha Filho.

O método de coleta utilizado foi o transversal (*cross-sectional studies*), em que se observa a língua de um indivíduo em somente um intervalo de tempo. Esse tipo de estudo geralmente compara diferentes grupos de indivíduos em um período e se opõe aos estudos chamados longitudinais (*longitudinal studies*) nos quais se compara a língua do mesmo indivíduo (ou indivíduos) em, pelo menos, dois intervalos de tempo.

5.2 As tarefas e os testes

Os alunos realizaram duas tarefas. A primeira tarefa consistia na leitura de 30 frases (ver Apêndice A), repetindo-as no mínimo duas vezes e tendo um intervalo de 2 segundos entre cada uma delas. A segunda tarefa se referia à escolha lexical (substantivo, verbo ou adjetivo) por parte do aluno de cada palavra contida

nas frases em que este respondia em folha separada. O objetivo da tarefa era documentar as estratégias de tomada de decisão do aluno em relação à sua escolha. Não foi realizado nenhum teste formal para a determinação do nível de proficiência dos aprendizes. Essa decisão metodológica será explicada na próxima seção.

5.3 A organização dos participantes da pesquisa

Antes do processo de simulação na rede neural e com o objetivo de ajudar na interpretação do mapa, decidiu-se alocar os 30 participantes em 5 níveis de desenvolvimento distintos, com base no critério de tempo de exposição ao idioma⁵¹.

De acordo com Thomas (1994), existem quatro formas de avaliar a proficiência linguística em L2, a saber: (a) julgamentos impressionistas; (b) *status* institucional (por exemplo, 2º semestre do 3º ano de inglês); (c) testes criados especificamente para a pesquisa; e (d) testes padronizados (TOEFL, *Michigan Proficiency Test* etc).

Pelo fato de existirem muitos testes padronizados para aferir a proficiência de aprendizes de LE, o campo da aquisição de L2 se depara constantemente com dificuldades ao tentar comparar estudos já realizados, já que não há um ponto de corte exato entre os níveis e que seja aceito em absoluto por todos os pesquisadores. Muitas vezes, a categoria entendida como avançada de um pesquisador pode corresponder a uma categoria intermediária de outro (GASS; SELINKER, 1993, p. 71).

Com base neste fato, estabeleceu-se como critério circunstancial de classificação e organização dos indivíduos a quantidade de horas-aula acumulada na disciplina de língua inglesa (ver apêndice D – Sumário dos Participantes) obtida por meio de entrevista e questionário respondido pelos participantes (ver apêndice B - Ficha de Informações do Participante). Após consultar outras obras que lidaram com classificação de falantes de L2 em níveis de proficiência, descobriu-se que tal

⁵¹ Para maiores informações e critérios mais detalhados sobre a determinação de nível de proficiência, consultar o ACTFL (*American Council on the Teaching of Foreign Languages*) Proficiency Guidelines em <<http://www.actfl.org/i4a/pages/index.cfm?pageid=1>> bem como The Common European Framework of Reference em <http://www.coe.int/t/dg4/linguistic/cadre_en.asp>.

critério já foi utilizado anteriormente por Ellis (1982) e Praxedes Filho (2007). Posteriormente, saber-se-á se os dados, assim organizados, serão corroborados ou não pelos resultados oferecidos pela rede neural. As simulações na rede neural contidas no próximo capítulo investigam a formação de grupos com características fônicas semelhantes, uma possível indicação de classificação em níveis de proficiência.

Vale ressaltar que o tempo de exposição ao *input* é um fator tão relevante quanto à frequência (regularidade) e à consistência (qualidade) do *input* no processo de construção do conhecimento de L2. Todavia, tempo de exposição e frequência não são fatores intercambiáveis. Pode-se explicar isso contrastando os exemplos de dois alunos nomeados de A e B: “A” tem aula de inglês de 50 minutos de duração duas vezes por semana há 5 anos e “B” tem aulas de inglês (50 minutos também) 5 vezes por semana e há um ano. Se somarmos o tempo de exposição ao idioma a partir da frequência de A e de B, verifica-se que B tem um tempo de exposição maior. No entanto, não se pode avaliar a sua experiência com o idioma porque, como se sabe, tempo de exposição a uma determinada língua não determina diretamente a fluência e a acurácia.

Por outro lado, é coerente afirmar que os fatores tempo de exposição ao idioma e frequência tomados isoladamente não são critérios suficientes para a determinação de nível de proficiência linguística. Alguns trabalhos que se propuseram a determinar a proficiência linguística de aprendizes estrangeiros (FLEGE et al., 2006) chamam atenção para as sérias restrições desses critérios como únicos determinantes de nível de proficiência. Há outros fatores que contribuem diretamente para o acúmulo de conhecimento de L2, como a qualidade do ensino e do aprendizado, a motivação e a concentração do aprendiz, a experiência anterior do indivíduo, a idade e até mesmo outros componentes de base neurobiológica, como a criação de mapas mentais para a fala (KUHL et al., 1992).

A tabela 5.1 contém a estrutura curricular das instituições de ensino regular, no tocante à disciplina de língua inglesa, com as respectivas horas-aula acumuladas e seus níveis de proficiência resultantes. A tabela foi adaptada com base no estudo de Praxedes Filho (2007) e devidamente complementada pela autora da presente pesquisa ao inserir a última linha (cursinho ou pré-vestibular),

com vistas a descrever fidedignamente a realidade dos dados fornecidos pelos participantes entrevistados:

Tabela 5.1: Classificação de nível de proficiência de língua inglesa de acordo com horas-aula acumuladas (distribuição da estrutura curricular da disciplina de língua inglesa em 2004.1 nas escolas particulares de Fortaleza)

Série	Nível	Horas-aula por nível	Horas-aula acumuladas	Nível de proficiência
5ª série do 1º grau	1	2 semestres letivos = 2 1 sem. letivo = 45 h-a	90 h-a	Iniciante
6ª série do 1º grau	2	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a	
7ª série do 1º grau	3	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a	Pré-intermediário
8ª série do 1º grau	4	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a	
1º ano do 2º grau	5	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a	Intermediário
2º ano do 2º grau	6	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a	
3º ano do 2º grau	7	2 semestres letivos = 2 1 sem. let. = 45 h-a	90 h-a ⁵²	Avançado
Cursinho pré-vestibular ⁵³	8	semestres letivos = a depende (1 ou 2) 1 sem. let. = 45 h-a	45 ou 90 h-a ^{*54}	
Total de horas-aula			720 horas-aula	

Fonte: Elaboração da autora.

Os dados sobre a quantidade de horas-aula correspondente a cada série/ano da escola regular foram fornecidos pelo MEC, Brasil, no ano de 2005, antes da mudança de currículo no ensino médio e fundamental. Os dados estão organizados de acordo com a estrutura curricular das escolas particulares na cidade de Fortaleza.

Na tabela 5.2, tem-se a sugestão de nível de proficiência do curso de extensão em língua inglesa, CCB. A grade curricular está organizada por semestre e toma as horas-aula de cada semestre cursado como base para determinação de nível de proficiência. Os dados sobre a quantidade de horas-aula correspondente à

52 Durante o 3º ano do 2º grau, havia a possibilidade de o aluno pré-vestibulando escolher a língua espanhola como opção de língua para o vestibular.

53 Há cursinho pré-vestibular que possui uma duração de seis meses (intensivo).

54 Nem todos os alunos se submeteram ao cursinho.

cada semestre foram fornecidos pela Coordenação das Casas de Cultura – UFC no ano de 2004.

Tabela 5.2: Estrutura curricular do curso de extensão em língua inglesa no semestre 2004.1.

Semestre	Horas-aula por nível	Nível de proficiência
1	60 h-a	Iniciante
2	60 h-a	
3	60 h-a	
4	60 h-a	Pré-Intermediário
5	60 h-a	
6	60 h-a	Intermediário
7	60 h-a	
Total de horas-aula	420 horas	-
Cursos preparatórios (TOEFL, FCE, CAE ⁵⁵)	180 h-a (divididas em três semestres)	Avançado

Fonte: Elaboração da autora.

Com base no histórico escolar na disciplina de língua inglesa e a carga horária do curso de inglês que os aprendizes estavam cursando ao serem entrevistados, apresenta-se a tabela 5.3 de divisão dos participantes por horas-aula acumuladas. Para informações mais detalhadas sobre o histórico escolar de cada indivíduo, ver sumário dos participantes no apêndice D:

⁵⁵ TOEFL: Test of English as a Foreign language.
 FCE: First Certificate of English.
 CAE: Certificate of Advanced English.

Tabela 5.3: Divisão dos participantes por horas-aula

Nível 1 (abaixo de 800 h-a)	Nível 2 (de 800 a 1000 h-a)	Nível 3 (de 1000 a 1200 h-a)	Nível 4 (de 1200 a 1400 h-a)	Nível 5 (de 1400 a 1600 h-a)	Nível 6 (sem considerar)
Aprendiz no. 201	Aprendiz no. 206	Aprendiz no. 406	Aprendiz no. 409	Aprendiz no. 210	Nativos
202	207	407	608		
203	208	408	609		
204	209	605	610		
205	401	606			
	402	607			
	403				
	404				
	405				
	601				
	602				
	603				
	604				

Fonte: Elaboração da autora.

5.4 Determinação das variáveis operacionais na simulação

As variáveis se constituem nos processos de transferência fonológica observados em nível microprosódico (isto é, em nível de palavra). Todas as análises que serão futuramente apresentadas aconteceram em função da divisão dos participantes em grupos distintos com base no tempo de exposição à língua inglesa, como foi elencado acima, ou seja, cinco níveis (nível 1, nível 2, nível 3, nível 4 e nível 5).

Eis a lista de fenômenos investigados com seus rótulos específicos:

1. Redução vocálica
2. Inserção vocálica
3. Apagamento vocálico
4. Transferência de padrão acentual

5.5 A metodologia das simulações

Todas as simulações foram realizadas no sistema operacional da Microsoft Windows 7 Home Basic, em um *notebook* com processador Intel® Core™ 2 Duo CPU, 2,20 GHz, 2201 Mhz, 3 Gb de memória RAM e 250 Gb de Hard Drive.

A seguir serão apresentadas as cinco etapas necessárias para as simulações, que vai desde o tratamento computacional dos dados de entrada até a geração dos mapas, ou seja, a saída da rede.

5.3.1 Etapa de Aquisição dos dados e segmentação das palavras

As elocuições dos participantes foram gravadas e digitalizadas no programa *Sound Forge* versão 5.0 em arquivos de áudio tipo WAV (extensão .wav), a uma taxa de amostragem de 44.100 mil amostras por segundo (44.1 KHz) e resolução de 16 bits, mono. Arquivos WAV (ou WAVE) são o formato-padrão de arquivo de áudio da Microsoft e IBM para armazenamento de áudio em computadores pessoais. Todas as frases foram ditas na forma declarativa.

Após essa fase, procedeu-se à segmentação de cada frase e conseqüentemente cada palavra que representava o item lexical a ser investigado (*e.g. object, separate, desert, etc*). A segmentação de palavras conectadas na cadeia da fala representou um desafio devido à dificuldade de detecção de extremos (*endpoints detection*). Estudos sobre Reconhecimento Automático de Voz (RAV) alertam para essa dificuldade na fase do Bloco de Pré-Processamento e explicam que a área de segmentação de palavras, por si só, abarca um campo vasto de pesquisa pelo nível de complexidade notado. No caso do presente estudo, não foi necessário utilizar a técnica de detecção de extremos porque se optou pela segmentação manual, na qual se contou com o auxílio de um foneticista.

A fim de diminuir a complexidade de preparação dos dados, decidiu-se que as elocuições de fala não seriam normalizadas para a extração de ruídos, decisão esta que não prejudicou nem alterou em nenhum momento a qualidade dos dados de análise.

5.3.2 Etapa de Processamento ou Parametrização do Sinal de Fala

A entrada da rede neural, também chamada de vetores de atributos ou de características, será processada pelos neurônios de modo a gerar uma representação compacta da informação presente nos dados originais. No caso da rede SOM, esta produzirá uma representação que preserva, entre neurônios

vizinhos, relações de proximidade (vizinhança) presentes nos dados originais. A representação gerada pela rede SOM é codificada nos vetores de pesos dos neurônios e pode ser visualizada através da matriz-U.

O sinal de fala não pode ser usado diretamente para alimentar a rede por conter milhares de amostras, o que tornaria seu processamento muito lento e também por ser muito ruidoso, o que dificulta sobremaneira a extração de conhecimento (SOUZA JR., 2009, p. 10). Além disso, o sinal de fala não é estacionário no tempo. Para a análise do sinal de fala, deve-se recorrer à segmentação do sinal e, nesse caso, considerá-lo *quasi* estacionário por partes.

A solução é representá-lo numericamente em um conjunto de coeficientes obtidos a partir da aplicação de técnicas matemáticas, tais como coeficientes de predição linear (LPC) e/ou coeficientes mel-cepstrais (MFCC), ao sinal de fala subdividido em vários *frames*. Assim, conjuntos de vetores de coeficientes passam a representar numericamente o sinal de fala dos aprendizes. Diz-se, então, que o sinal de fala é parametrizado pelos coeficientes LP/MFC. Essa fase de tratamento dos dados em que algoritmos atuam sobre o sinal de fala a fim de representá-lo de forma mais compacta é chamada de Extração de Características e foi realizada por meio da ferramenta PRAAT, versão 5.1.25.

Tanto a análise por predição linear, para obtenção dos coeficientes LP, como a análise do mel-cepstrum, para obtenção dos coeficientes MFC, necessitam da especificação de alguns parâmetros. Por exemplo, numa extração dos coeficientes LP, é preciso definir: (i) a ordem do filtro autorregressivo (AR), que na presente tese foram testados os valores 10, 15 ou 20; (ii) o tamanho ou comprimento do quadro (*frame*), que é a duração do segmento em que o sinal de fala é subdividido para ser considerado estacionário e ser possível a aplicação de técnicas lineares, foi definido em 25 milissegundos; (iii) o tamanho do passo, definido em 5ms; e (iv) a frequência do filtro de pré-ênfase do sinal, definido em 50Hz. A figura 5.1 mostra, de forma detalhada, o *input* da rede:

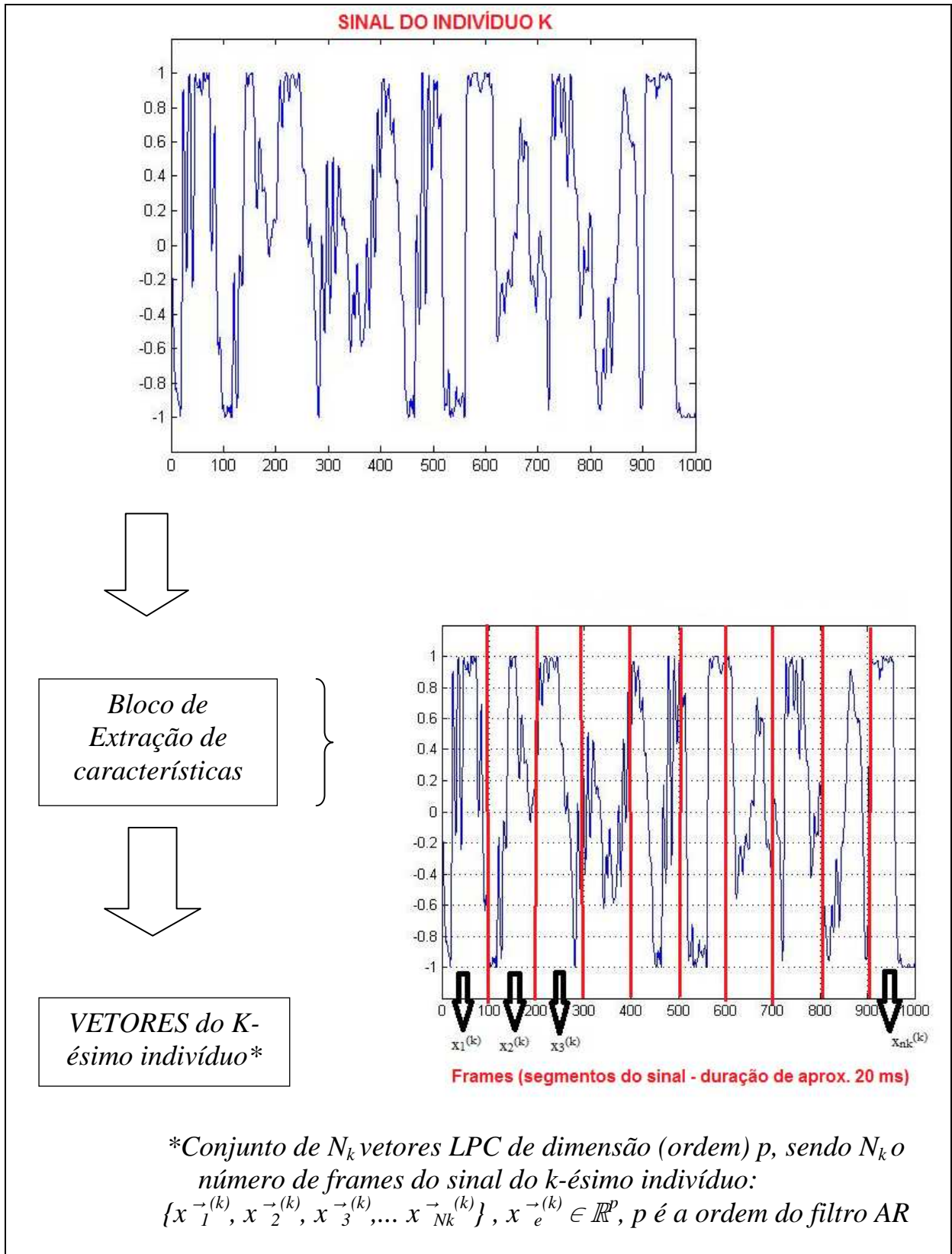


Figura 5.1: Input da rede
 Fonte: Elaboração da autora.

O número de *frames* depende da duração do sinal de fala original e do tamanho do frame propriamente dito. Por exemplo, para uma palavra cuja elocução resulta em sinal com duração total de 2s e com duração do frame de 20ms, ter-se-á um total de 100 *frames*. Cada frame contribui com um vetor de coeficientes de tamanho fixo (10, 15 ou 20, na presente análise).

Os vetores de coeficientes LP ou MFC são então organizados ao longo das linhas de uma matriz de coeficientes. Para o exemplo dado, são gerados 100 vetores de coeficientes, um vetor para cada frame, que correspondem a 100 linhas da matriz de coeficientes. O número de colunas dessa matriz é igual ao número de coeficientes extraídos da análise LPC ou MFCC. As matrizes de coeficientes⁵⁶ são extraídas e salvas em uma planilha do tipo Excel, sendo posteriormente convertidas em arquivos de texto (ASCII) para serem usadas no treinamento da rede neural.

É importante enfatizar que cada palavra pronunciada por um falante gera uma matriz de coeficientes. Logo, para identificar qual falante pronunciou qual palavra, faz-se necessário uma etapa de rotulação dos dados. Partindo do arquivo-texto em que uma certa matriz de coeficientes está armazenada adicionou-se uma última coluna a esta matriz contendo um rótulo para identificar aquele conjunto de dados.

Os rótulos podem identificar tanto o falante quanto a categoria linguística na qual se insere a palavra pronunciada. A rotulação é feita a fim de relacionar os conjuntos de vetores a seus locutores. Por fim, os arquivos de texto relativos à elocução de uma palavra específica por todos os falantes são concatenados em um só arquivo com o auxílio do Matlab⁵⁷. A tabela abaixo apresenta os rótulos com seus respectivos significados como também os exemplos de rotulação para os casos de categorias linguísticas avaliadas. Na primeira coluna estão os rótulos utilizados e a segunda coluna explica o significado de cada um deles:

⁵⁶ Na fase de teste piloto, durante a geração das matrizes no programa Praat de versão 30918.0.0.0 encontrado um erro, um problema (bug) no programa que consistia na criação equivocada de uma coluna adicional à matriz de dados, não permitindo assim a rotulação apropriada dos dados para sua utilização posterior. O *bug* foi reportado aos desenvolvedores do software e os códigos fontes do programa foram atualizados e o problema detectado corrigido.

⁵⁷ Os códigos do Matlab para os exemplos de simulação reportados aqui estão todos incluídos na parte de anexo desta tese (ver Anexo A).

SIGNIFICADO DOS RÓTULOS			
Identificação dos aprendizes		Processos	
201	O primeiro dígito indica um participante que cursava o segundo semestre e os dois últimos dígitos indicam sua ordem nas entrevistas (01 = 1º)	0	No caso de transferência de padrão acentual do PB. A ausência da transferência foi marcada com 0 (zero).
406	indica um participante que cursava o quarto semestre e foi o sexto entrevistado	6.2901 p. ex.	No caso de representação de categoria lexical, o primeiro dígito representa a frase (1 - 20), o segundo dígito representa a categoria (1 para substantivo, 2 para verbo e 3 para adjetivo) e os três últimos dígitos representam a identificação do aprendiz.
608	para indicar um participante matriculado no semestre, oitavo na ordem de entrevista	1, 2, 3, 4, 5, 6	Nível de proficiência (ver tabela de Divisão dos participantes por horas-aula)
903	para identificar o terceiro nativo entrevistado	1	Fenômeno de inserção de vogal em posição de coda. A ausência do fenômeno foi marcado com 0 (zero).

Quadro 5.1: Rótulos dos vetores de características com seus respectivos significados.
Fonte: Elaboração da autora.

De posse do quadro acima, vale ressaltar que são utilizados os mesmos vetores de características, porém com rótulos diferentes. A análise denominada Multi-Rótulo (*Multi-Label Analysis*) é considerada uma abordagem nova na análise de agrupamentos e tem como objetivo último verificar qual rotulação é mais adequada ao tipo de parametrização usada (LPC/MFC). A estratégia de análise Multi-Rótulo permite, em princípio, verificar que propriedades linguísticas do sinal de fala são melhor codificadas nos coeficientes LP/MFC. A apresentação desta técnica de análise de dados consiste na principal contribuição desta pesquisa.

A abordagem usando modelos conexionistas supervisionados, tais como a rede Perceptron multicamadas (MLP) ou a rede de Elman, usam como entrada da rede neural características linguísticas codificadas e extraídas por um especialista (eg.: foneticista) (Cf. ZIMMER, 2006). A abordagem aqui apresentada, por outro lado, é totalmente não-supervisionada, pois os coeficientes LP ou MFC carregam implicitamente as características temporais e prosódicas de uma elocução. A abordagem usando LPC-SOM ou MFCC-SOM é uma proposta inédita que esta tese traz. Ao usar a rede SOM, está-se testando a hipótese de que a parametrização do

sinal de fala por coeficientes LP ou MFC é eficiente na categorização dos falantes por características prosódicas.

5.3.3 Etapa de configuração da rede neural

A rede neural foi implementada no ambiente Matlab (MATHWORKS, 2002), VERSÃO 7. Foi utilizado o pacote SOM Toolbox, um conjunto de rotinas de programação desenvolvidas por um grupo de pesquisa na Finlândia liderado por Teuvo Kohonen. O programa é responsável por gerar a Matriz-U, os mapas de agrupamentos e o índice Davies-Bouldin⁵⁸, medida para determinar a quantidade de agrupamentos existente no conjunto de dados formado pelos coeficientes LPC e cepstrais.

Vê-se na figura 5.2 a interface do programa Matlab. A janela do meio da tela mostra o programa que está sendo executado, a janela logo abaixo é a janela de comando e a janela ao lado é o espaço de trabalho, em que informações dos dados de *input* da rede podem ser facilmente acessadas.

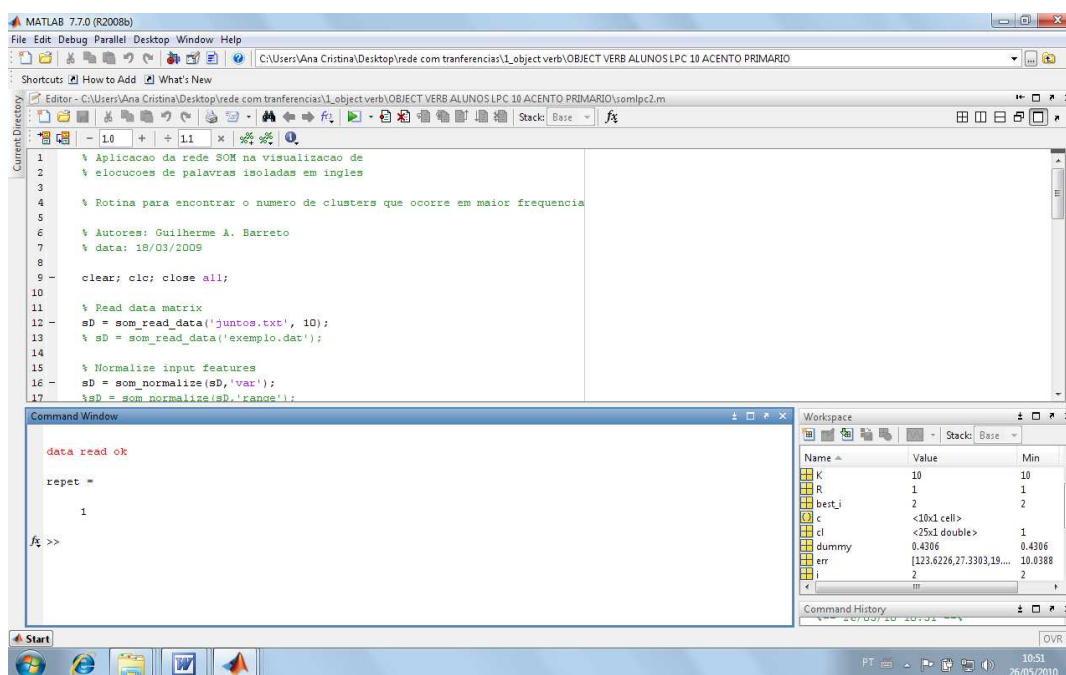


Figura 5.2: Representação da interface gráfica do ambiente Matlab durante o processo de inicialização e treinamento da rede.

Fonte: Figura gerada no programa Matlab.

⁵⁸ Para saber mais sobre índices relativos internos, consultar Jain e Dubes (1988).

A rede SOM, assim como a maioria dos modelos conexionistas, necessita que alguns parâmetros sejam pré-especificados de modo a garantir um correto treinamento. Kohonen (2001) ressalta que a determinação do tamanho da rede SOM e dos seus parâmetros de aprendizagem é um processo empírico, baseado na experiência do usuário e em métodos de tentativa e erro. Após alguns testes preliminares, foi escolhido um mapa bidimensional, com arranjo hexagonal (geralmente utilizado por ter uma boa capacidade de projeção de dados), função de vizinhança gaussiana, iniciação linear dos pesos e aprendizagem em lote.

Outro item de fundamental importância no projeto da rede SOM é a definição do número de neurônios. Outros parâmetros livres a serem ajustados para o correto treinamento da rede SOM são o número total de épocas de treinamento, o número de repetições da simulação e o raio inicial da função de vizinhança. A seguir, discutir-se-ão as heurísticas para especificação de tais parâmetros livres.

5.3.4 Determinação dos parâmetros da rede

A determinação dos parâmetros da rede SOM que ofereçam os melhores resultados para a visualização dos agrupamentos e que satisfaçam as propriedades inerentes da rede (dimensionalidade do mapa – 3 x 3, 5 x 5, 10 x 10; número de épocas de treinamento, número de repetições do treinamento etc.), pode vir a influenciar o resultado produzido pelo algoritmo, o que torna a aplicabilidade da rede SOM a um problema prático um grande desafio (BOSCARIOLI, 2008).

5.3.5 A simulação e suas fases

Quando aplicado ao problema de interesse, tem-se que o processo de simulação da rede SOM e a análise do resultado do treinamento envolvem os seguintes passos:

1. Inicialização e treinamento (aprendizagem) da rede;
2. Avaliação da qualidade do mapa gerado usando os índices - erro de quantização (E_Q) e erro topológico (E_T);

3. Geração e armazenamento da matriz-U e mapas rotulados para cada resultado de treinamento;
4. Validação dos agrupamentos através do índice Davies-Bouldin (DB);
5. Tabulação no programa Excel de todos os resultados das medidas de desempenho da rede (erro de quantização e erro topológico).

Cada um destes passos está melhor detalhado a seguir:

✓ Passo 1: Treinamento da rede

Para treinar um mapa SOM 10x10, é necessário construir um conjunto de dados que podem ser constituídos de coeficientes LP ou coeficientes MFC para cada item lexical (ver Apêndice C) e pronunciados por todos os falantes. A cada nova palavra é necessário construir um novo conjunto de dados e referenciar esse conjunto de dados no código fonte do programa Matlab. Uma vez escrito o programa em Matlab usando os comandos da SOMtoolbox, toda a análise (treinamento do mapa, geração das figuras, etc.) é automática.

✓ Passo 2: Avaliação do mapa

Existe um conjunto razoável de medidas quantitativas para avaliação da qualidade do mapa gerado após o processo de aprendizagem (KOHONEN, 2001). Escolheram-se duas destas métricas, o erro da quantização vetorial e o erro topológico. O erro de quantização, $E_Q(t)$, ajuda a avaliar a convergência da rede e pode fornecer uma estimativa do número necessário de neurônios a serem dispostos no mapa antes de se treinar a rede (CRUZ, 2007). O vetor de erro de quantização é definido como a diferença entre o vetor de entrada atual e o vetor de pesos do neurônio vencedor correspondente, ou seja:

$$e_q(t) = x(t) - w_{j^*}(t),$$

sobre o qual opera uma medida de distância, em geral, euclidiana, dando origem a uma grandeza escalar denominada **erro de quantização** associado ao vetor $x(t)$:

$$e_q(x(t)) = \| x(t) - w_{j^*}(t) \| = \| e_q(t) \| = \sqrt{\sum_{j=i}^p [x_j(t) - w_{j^*}(t)]^2} ,$$

na qual N é a dimensão de $x(t)$. Já o erro topológico é uma medida que busca por pontos no espaço de entrada em que o segundo vizinho mais próximo, ou segundo vencedor, não pertence à vizinhança do vencedor (SOUZA JR., 2009) e é definido por:

$$E_T = \frac{1}{N} \sum_{i=1}^N u_i,$$

em que

$$u_i = \begin{cases} 1, & i^{**} \notin V_{i^*}, \\ 0, & \text{caso contrário,} \end{cases}$$

em que i^{**} é o segundo neurônio vencedor para o padrão \mathbf{x}_i , V_{i^*} é o conjunto de neurônios vizinhos imediatos do vencedor i^* , e N é o número de vetores de entrada. De acordo com Souza Jr. (2009, p. 42), “para cada vetor de entrada deve-se encontrar o segundo neurônio mais próximo. (...) Ao término da apresentação de todos os padrões, a medida é normalizada dividindo-se o número de erros pelo número de vetores de entrada”.

✓ Passo 3: Interpretação do mapa

Avaliar os mapas poderia, a princípio, se basear no cálculo da quantização de erro do mapa. No entanto, em muitas aplicações, uma análise visual da localização relativa dos aprendizes ou palavras no mapa é mais útil na análise do problema. Dentre todas as formas de representação visual do resultado da ordenação topológica do mapa destaca-se a matriz-U (SILVA, 2004). A matriz-U apresenta, de forma inteligível, a estrutura de vizinhança entre os protótipos vizinhos e compara as distâncias entre eles.

Essa estrutura de vizinhança, que é codificada nos protótipos de uma rede SOM treinada, reflete a estrutura de vizinhança dos dados no espaço original. É importante ressaltar que a informação provida da matriz-U é de natureza qualitativa e difusa, ou seja, pode até ser possível visualizar a formação de grupos, mas a fronteira entre eles é altamente difusa. Há que se desenvolver a habilidade de

visualizar fronteiras e tal habilidade é desenvolvida com bastante treino e tempo de observação. Para a visualização da Matriz-U, foi utilizada a SOM toolbox.

✓ Passo 4: Análise e validação de agrupamentos

Em reconhecimento de padrões, agrupar dados é fazer uso de uma ampla escala de métodos numéricos. As redes de Kohonen são altamente eficazes em tarefas de formação de agrupamento. O objetivo é encontrar um esquema que os agrupe em um número finito de 'g' grupos. "Com este agrupamento, pode-se identificar relacionamentos úteis entre os dados, com similaridade e diferenças antes não encontradas." (BOSCARIOLI, 2008, p. 29).

Conforme mencionado no Passo 3, a Matriz-U fornece um meio qualitativo para avaliar como relações de vizinhança presentes nos dados originais estão sendo preservadas nas relações de vizinhança entre os protótipos da rede. No quesito validação quantitativa dos agrupamentos sugeridos pela rede SOM, faz-se necessário o uso de certos índices consagrados na literatura, como o índice DB.

O índice DB é um índice de validação de agrupamentos e indica quantos grupos podem existir em um mapa. Como proposto por Davies & Bouldin (1979), o índice é uma função da razão da soma da dispersão dentro dos agrupamentos pela dispersão entre os agrupamentos (PAKHIRA et al., 2004 *apud* CRUZ, 2007, p. 39). O autor continua dizendo que:

O melhor agrupamento é dado pelo menor valor do índice DB para uma seqüência de valores calculados para diferentes números de agrupamentos K . É importante notar que o valor do índice DB é nulo para o agrupamento trivial, em que cada padrão pertence a um agrupamento individual, devendo ser computado apenas quando cada agrupamento contiver um número razoável (ou seja, maior do que 1) de padrões.

A figura 5.3 ilustra de modo simples como a arquitetura da rede SOM é utilizada na análise do problema de interesse para esta tese:

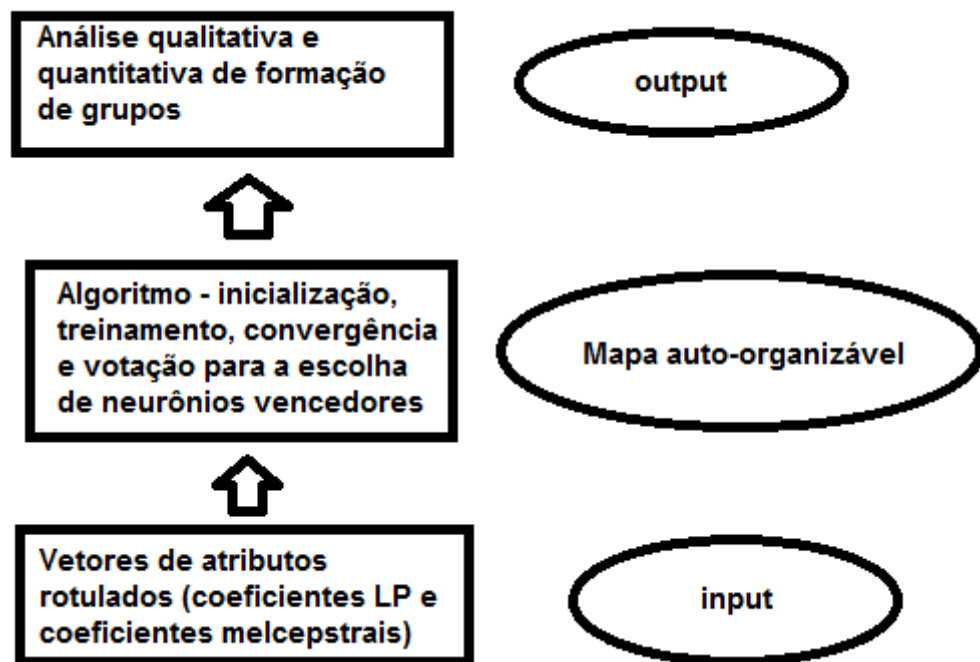


Figura 5.3: Arquitetura da rede neural.
 Fonte: Elaboração da autora.

5.6 A ferramenta computacional: a SOM toolbox no Matlab

A SOM Toolbox é uma biblioteca de funções para o Matlab versão 7.7.0 (R2008b) com a implementação da rede SOM, que é um algoritmo de rede neural baseado em aprendizado não-supervisionado.

A rede SOM é uma ferramenta para análise da projeção de relações não-lineares e complexas presentes em dados de alta dimensão em estruturas de baixa dimensão. A rede tenta comprimir a informação presente nos dados enquanto simultaneamente tenta preservar as relações métricas e topológicas mais importantes dos dados. A análise visual da projeção levada a cabo pela rede SOM pode produzir abstrações úteis no entendimento das relações complexas presentes nos dados. Esses dois aspectos, visualização e abstração, podem ser utilizados de várias formas em tarefas complexas, tais como análises de processos industriais, percepção de máquinas, controle e comunicação (KOHONEN, 1988, 1998 e 2001).

Uma rede SOM consiste de neurônios organizados em uma grade regular de baixa dimensão. Cada neurônio é um vetor de peso d -dimensional (também

chamado de vetor-protótipo, vetor de código ou vetor-referência) em que d é igual à dimensão dos vetores de entrada. Os neurônios são conectados a neurônios adjacentes por uma relação de vizinhança que dita a topologia ou a estrutura do mapa.

Na SOM toolbox, a topologia é dividida em dois fatores: estrutura de entrelaçamento local (hexagonal ou retangular) e forma de mapa global (folha, cilindro ou toroide). A rede SOM pode ser pensado como uma rede que se espalha sobre uma nuvem de dados. O algoritmo de treinamento do SOM movimenta os vetores de peso de forma a abraçar um grande número de dados e assim organizar o mapa. Neurônios avizinados na grade obtêm vetores de peso similares. Duas variantes do algoritmo de treinamento do SOM⁵⁹ têm sido implementadas na SOM Toolbox: treinamento sequencial e treinamento em lote.

5.4 O uso da SOM Toolbox

Os tipos de dados que podem ser processados com a SOM Toolbox devem ter o formato de planilha do Excel ou tabela de dados. Cada linha (row) da tabela é um vetor de atributos, contendo as variáveis de entrada. Conseqüentemente, as colunas representam as variáveis do conjunto de dados. Estas variáveis podem ser as propriedades de um objeto, ou um conjunto de medidas mensuradas em um tempo específico (VESANTO et al., 2000). No problema tratado nesta tese, as variáveis consistem nos coeficientes LP ou MFC, e o número de coeficientes define a dimensão do vetor de entrada da rede SOM.

A SOM toolbox pode manipular tanto dados numéricos quanto dados categóricos, mas somente a primeira opção pode ser utilizada no algoritmo SOM. Na SOM toolbox, dados categóricos podem ser inseridos na forma de rótulos (*labels*) associados com cada amostra de análise. Eles podem ser considerados como “bilhetinhos” de *post-it's* atachados a cada vetor de atributos. O usuário pode consultá-las mais tarde para verificar a qual classe pertence um vetor de entrada específico.

⁵⁹ Para uma descrição mais completa do SOM e sua implementação no MatLab conferir Kohonen (1995) e Vesanto et al. (2000).

Antes de ser trazido para o Matlab, o conjunto de dados deve ser construído de forma a agrupar informações relacionadas de alguma forma. O processamento dos dados, em geral, envolve transformações simples ou normalizações das variáveis existentes. Na SOM toolbox, especificamente, a função 'som_normalize' pode ser usada para implementar mudanças lineares e logarítmicas de escala (*scalings*) nas variáveis de interesse, além de normalização dos histogramas destas variáveis de modo a mantê-las com média zero e variância unitária. O escalonamento de variáveis é de fundamental importância na SOM toolbox, já que a rede SOM usa a métrica euclidiana para medir a distância entre vetores⁶⁰.

Existem dois algoritmos de inicialização dos pesos (aleatório e linear) e dois de treinamento (treinamento sequencial e treinamento em lote). Por padrão, a inicialização linear e o algoritmo de treinamento em lote são usados no treinamento da rede. A maneira mais simples de iniciar e treinar um SOM é usar a função 'som_make', que implementará ambos os processos utilizando os parâmetros pré-selecionados automaticamente. O treinamento é feito em duas fases: treinamento inicial (*rough training*) com o raio de vizinhança maior e alta taxa de aprendizagem. Se um controle mais rígido sobre os parâmetros de treinamento é desejado, o usuário pode optar pela especificação manual do tipo de inicialização dos pesos, tipo de treinamento e todos os outros parâmetros livres da rede SOM.

Há uma variedade de métodos de visualização do SOM. Na SOM Toolbox, a ferramenta básica é a função 'som_show'. Essa função pode ser usada para mostrar a Matriz-U e os planos de componentes do SOM. A Matriz-U visualiza as distâncias entre unidades vizinhas no mapa e desta forma a estrutura de agrupamento (*cluster*) no mapa. Valores altos da Matriz-U indicam uma fronteira de agrupamento e áreas uniformes de valores baixos indicam os agrupamentos propriamente ditos. Cada plano de componente mostra os valores de uma variável em cada unidade do mapa. Apesar dessas visualizações, informações adicionais podem ser ainda mostradas, tais como rótulos, histograma de dados e trajetórias (VESANTO et al., 2000).

Com a função 'som_grid', muitas outras visualizações avançadas são possíveis. A função é baseada na ideia de que a visualização de um conjunto de

⁶⁰ Para mais informações e detalhes sobre normalização e escalonamento linear de variáveis conferir Vesanto et al. (2000).

dados simplesmente consiste de um conjunto de objetos, cada um com uma posição, cor e forma única. Além disso, conexões entre objetos e relações de vizinhança podem ser implementadas usando linhas. Com o comando 'som_grid', o usuário é capaz de determinar valores arbitrários a cada uma dessas propriedades.

5.5 Os experimentos

Um conjunto de quatro simulações foi rodado com parâmetros que variavam conforme a necessidade de ajuste dos dados a serem analisados. Todas as simulações foram conduzidas em uma rede neural auto-organizável bidimensional, hexagonal, com função de vizinhança gaussiana e aprendizagem em lote. Eis logo abaixo a lista de simulações conduzidas:

- Simulação 1 – Simulação da representação de categorias lexicais;
- Simulação 2 – Simulação de nível de proficiência em função do conhecimento prosódico;
- Simulação 3 – Simulação da transferência de padrões acentuais em substantivos, verbos e adjetivos;
- Simulação 4 – Simulação da transferência de padrões vocálicos (substituição, apagamento, redução, inserção).

5.6 Resumo do Capítulo

A rede SOM é um modelo conexionista altamente recomendado para a fase de compreensão dos dados no processo de descoberta do conhecimento (*knowledge discovery*), embora também possa ser usado para a preparação, modelagem e classificação de dados. Com as propriedades apresentadas neste capítulo, a SOM toolbox apresenta-se como uma ferramenta útil na visualização de dados de alta dimensão.

A análise Multi-Rótulo (*Multi-Label Analysis*) é uma abordagem nova na análise de agrupamentos e objetiva verificar os padrões de rotulação de dados mais adequados. A estratégia de análise Multi-Rótulo permite, em princípio, verificar que propriedades linguísticas do sinal de fala são melhor codificadas nos coeficientes

LP/MFC.

A abordagem de natureza não-supervisionada aqui apresentada, LPC-SOM ou MFCC-SOM, é uma proposta inédita que esta tese traz. O objetivo de se utilizar essa nova abordagem é testar a hipótese de que a parametrização do sinal de fala por meio desta abordagem de coeficientes LP ou MFC seja eficiente na categorização dos falantes ao se levar em conta características prosódicas.

O próximo capítulo descreve passo a passo os procedimentos relativos às simulações mencionadas acima. Serão apresentados também os resultados, seguidos de uma discussão geral da avaliação da rede.

6 RESULTADOS DAS SIMULAÇÕES

Neste capítulo, serão apresentados os resultados das simulações, seguidos de suas análises e discussão geral da avaliação da rede. Todos os resultados apresentados foram simulados com os parâmetros descritos na seção 5.5 do capítulo da metodologia de pesquisa.

As quatro simulações apresentadas a seguir utilizaram-se do mesmo conjunto de dados, no entanto, sendo estes rotulados de maneira diferenciada e de acordo com as variáveis a serem testadas.

6.1 Simulação da representação de categorias lexicais

6.1.1 Metodologia da simulação

O presente experimento simula a representação de categorias lexicais e investiga a capacidade de separação de palavras em categorias lexicais diferentes (verbo, substantivo e adjetivo) com base em traços suprasegmentais (acento primário e redução vocálica). Alguns autores já utilizaram arquiteturas de rede neural do tipo recorrente simples (SRN – *simple recurrent network*), tal como a rede de Elman (Cf. ELMAN, 1990 e 1991), para representar o tempo explicitamente e estabelecer algum tipo de memória numa rede neural artificial, utilizando para isso um grupo de unidades de contexto entre a camada de *input* e a camada oculta. Apesar de as redes de Kohonen não possuírem uma camada oculta que fornece informações semânticas no processo de decisão e convergência da saída da rede, elas são competentes para executar a tarefa de organização das categorias lexicais com base numa característica prosódica que está intimamente ligada ao tempo e ao contexto: o acento.

Com o objetivo de constatar a competência da rede em segregar grupos com características semelhantes, os verbos e substantivos pronunciados pelos falantes nativos que compuseram o banco de dados foram codificados na entrada da rede em coeficientes LP e coeficientes MFCC com número de coeficientes variando

a 10, 15 e 20. Os parâmetros necessários para treinamento da rede SOM estão mostrados no quadro 6.1:

Parâmetro	Opção Escolhida
Tipo de treinamento	<i>batch</i> (em lote)
Topologia do mapa	Bidimensional (planar)
<i>Dimensão do mapa</i>	10 x 10
Estrutura da vizinhança	hexagonal
Número máximo de clusters	$k = 10$
número de repetições	1
épocas	250 (50 – <i>rough training</i> , 200 – <i>fine tuning</i>)
raio vizinhança inicial	2
raio vizinhança final	1
taxa de aprendizagem inicial	<i>Não é utilizado no modo de treinamento batch</i>
taxa de aprendizagem final	<i>Não é utilizado no modo de treinamento batch</i>

Quadro 6.1 – Parâmetros de treinamento da rede SOM para o experimento 1.

Fonte: Elaboração da autora.

6.1.2 Resultados da simulação

A fim de encontrar a quantidade neurônios que satisfariam as propriedades do mapa antes de treinar a rede, determinaram-se duas medidas de avaliação: o erro de quantização e o erro topológico. Observe-se a Tabela 6.1.

Tabela 6.1: Resultados para algoritmo SOM – teste de desempenho da rede.

Algoritmo	Erro de quantização	Erro topológico
Lpc 10	0,383	0,025
Lpc 15	0,390	0,042
Lpc 20	0,452	0,043
Cepstral 10	0,399	0,029
Cepstral 15	0,451	0,034
Cepstral 20	0,426	0,031

Fonte: Elaboração da autora.

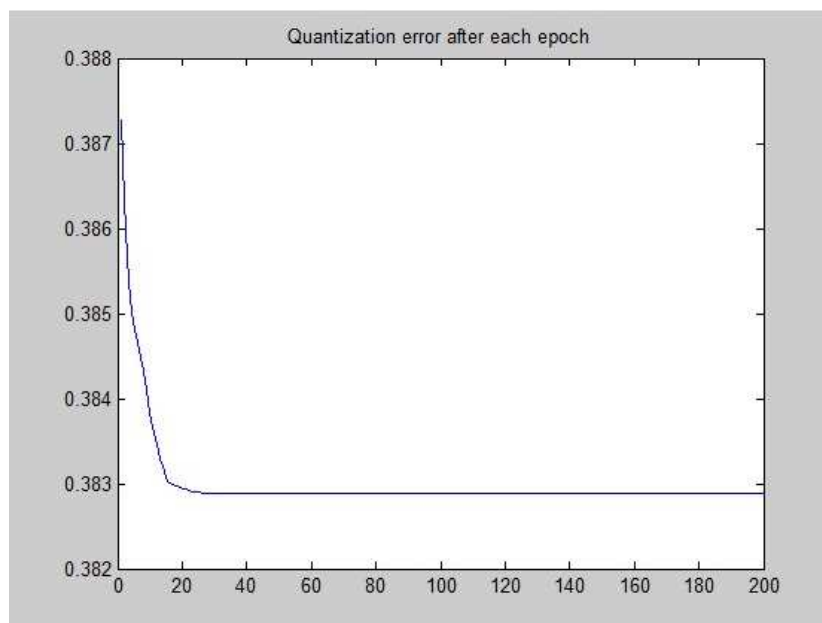


Figura 6.1: Gráfico típico do erro de quantização por épocas
 Fonte: Elaboração da autora.

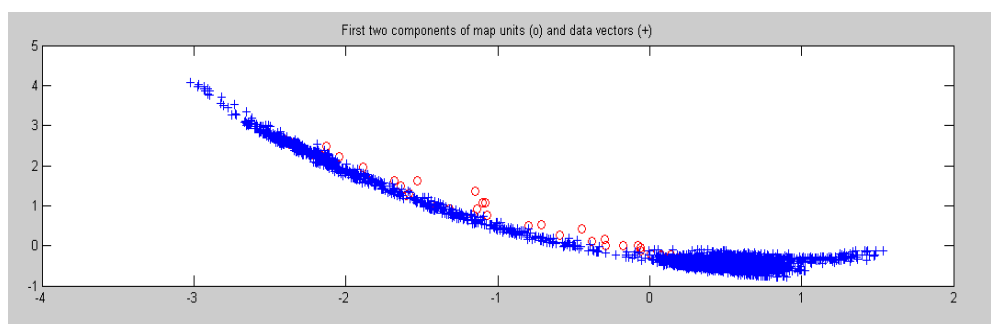


Figura 6.2: Gráfico típico do erro topológico
 Fonte: Elaboração da autora.

Como explicado na seção 4.5 do capítulo sobre redes de Kohonen, vetores de entrada que estão próximos no espaço de entrada tendem a ser agrupados em neurônios próximos no mapa, de modo que os padrões de elocução que definem as diversas categorias lexicais de interesse deverão tender a formar grupos em certas zonas do mapa. A matriz-U é a ferramenta de visualização utilizada nesta tese para analisar os resultados gerados pela rede SOM.

Os resultados da Matriz-U apresentaram a rede sempre segregando e fazendo a distinção clara entre os três grupos (substantivos, verbos e adjetivos), com zonas de fronteira bem definidas. Cada neurônio é rotulado pelo método do voto

majoritário, ou seja, ele recebe o rótulo da classe com maior número de ocorrências dentro do conjunto de vetores de entrada para os quais aquele neurônio foi considerado vencedor. O vetor de pesos de cada neurônio rotulado passa a ser um protótipo da categoria lexical indicada no rótulo daquele neurônio.

A figura 6.3 mostra a matriz-U (matriz de distância unificada) resultante do treinamento de um mapa auto-organizável bidimensional com 100 neurônios (10x10). A região circulada de preto representa os neurônios vencedores – protótipos – associados ao vetor de entrada da categoria lexical “substantivo” enquanto que a região azul fora do círculo que está acima da região de fronteira em cores claras representa os neurônios associados aos vetores de entrada relacionados à categoria lexical “verbo”. Cabe ressaltar que, nessa região, os vetores de peso são muito similares entre si, ou seja, quanto mais denso (cor azul escura), mais o conjunto está unido e o grau de similaridade entre os pontos é maior, conforme demonstrado na figura 6.3.

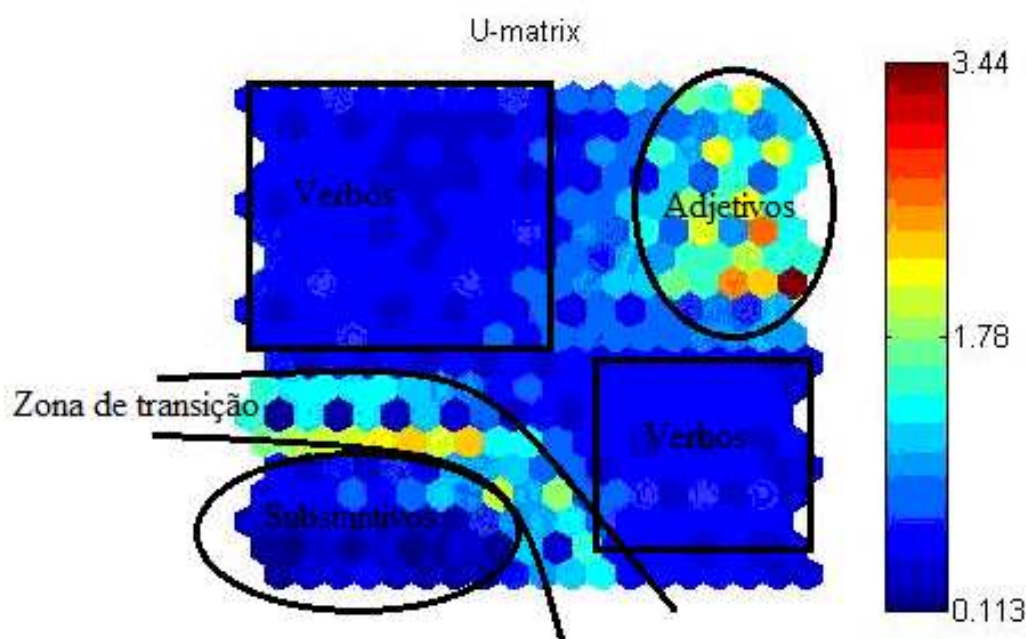


Figura 6.3: Matriz-U da simulação da representação de categorias lexicais.
Fonte: Elaboração da autora.

No mapa rotulado mostrado na figura 6.4 podemos ver a disposição dos neurônios com seus respectivos rótulos. A rotulação segue a seguinte ordem: a categoria lexical é codificada no segundo dígito – 1 para substantivo e 2 para verbo.

Na região circulada em preto há a presença de dois neurônios representantes da categoria de verbo (6.2901). No entanto, se usarmos a função 'freq', observamos que um destes dois neurônios teve 54 vetores de entrada para os quais o neurônio foi vencedor rotulado com verbo. Ao observar o 2º neurônio vencedor, observa-se que há 38 vetores de entrada para a categoria verbo, o que não chega a ser uma diferença significativa. Note-se a região de fronteira que se fez a partir de alguns neurônios sem rótulos, um sinal claro de indecisão, ou seja, quando a rede não consegue escolher o rótulo para o neurônio.

Ao se observar a matriz-U novamente e contrastá-la com o mapa rotulado, constatar-se-á que a região azul escura logo acima da zona de transição tem dominância de neurônios que representam a categoria dos verbos (segundo dígito após o ponto - número 2).

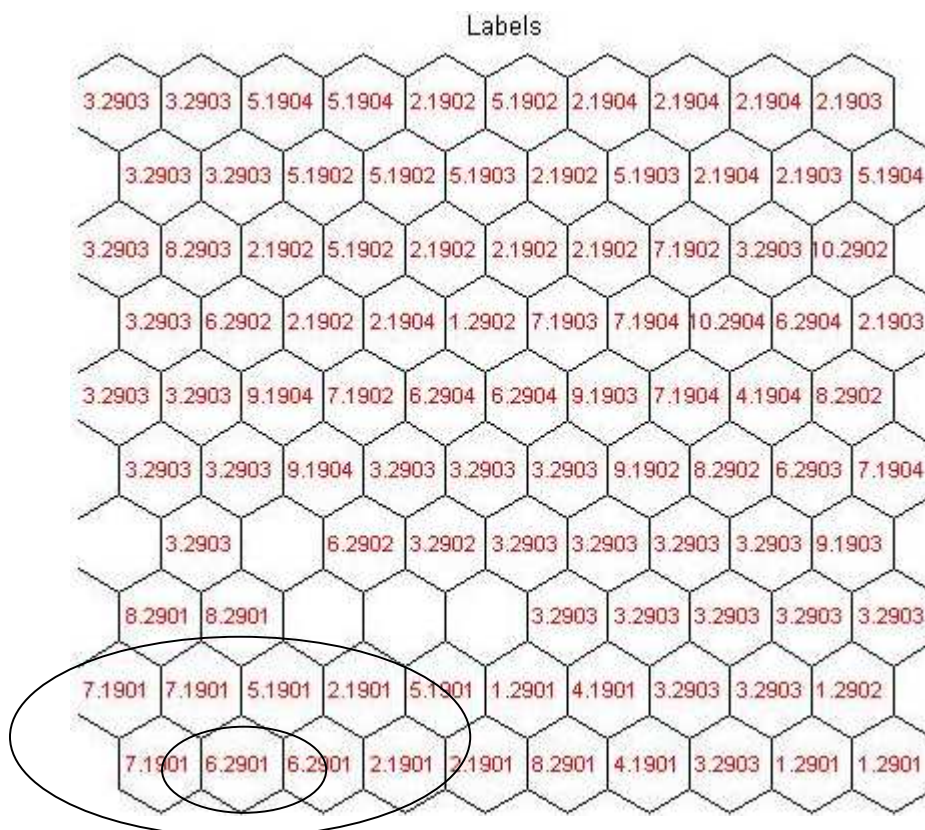


Figura 6.4: Mapa rotulado com as características lexicais e de identificação dos aprendizes. Fonte: Elaboração da autora.

6.1.3 Discussão geral

Com base na análise dos resultados dessa simulação, pode-se afirmar que categorias linguísticas podem emergir na topologia da rede como um produto natural do processo de auto-organização dos dados. Note-se que o processo é totalmente não-supervisionado, baseando-se apenas na similaridade entre os vetores de entrada. Os rótulos são usados apenas para facilitar a interpretação da matriz-U e do mapa. O experimento comprova a habilidade da rede SOM em segregar vetores por similaridade com base nos vetores de pesos dos neurônios vencedores, ou seja, os padrões que emergiram durante a simulação determinaram que a rede foi capaz de segregar os dados de entrada por categoria lexical e corrobora pesquisas prévias sobre a natureza emergente e auto-organizável de categorias linguísticas na representação lexical inicial (LI et al., 2004; LI, 2009).

6.2 Análise da formação de grupos em função do conhecimento prosódico do aprendiz

6.2.1 Metodologia da simulação

O objetivo do experimento foi analisar se a rede organizava os aprendizes quanto às suas características fonéticas e prosódicas em grupos relativamente homogêneos e se esta organização condizia com a classificação por tempo de exposição ao idioma, proposto na seção 5.4 da metodologia. Destaca-se aqui o uso da matriz-U como uma ferramenta de visualização útil no auxílio do julgamento do conhecimento linguístico dos aprendizes, o que ajudará a alocá-los em grupos relativamente homogêneos e conseqüentemente na determinação do nível de proficiência.

Partiu-se da hipótese inicial de que a rede seria capaz de agrupar os falantes em grupos formados de acordo com suas características prosódicas. Tais características estariam “implicitamente representadas” nos vetores de características dos falantes, que consistem nos coeficientes LP ou MFC. Em princípio, a rede SOM é adequada para este tipo de problema, pois os mapas são altamente eficazes em tarefas de formação de agrupamento, desde que os vetores de características sejam eficientes na extração das propriedades prosódicas

presentes na palavra pronunciada. Outra hipótese testada foi a adequação dos coeficientes LP ou MFC na extração implícita de características prosódicas da fala. Em última instância, está-se tentando categorizar a "forma" como uma certa palavra foi pronunciada por vários aprendizes.

Com o objetivo de determinar qual a parametrização (LPC e MFCC) e o número de coeficientes que mais se adequavam ao problema, recorreu-se à criação de um conjunto de dados para cada combinação do tipo de parametrização e do número de coeficientes. Em seguida, especificaram-se os parâmetros de treinamento da rede, mostrados no quadro 6.2. Note-se que a única diferença em relação aos parâmetros especificados no quadro 6.1 está no número de repetições do treinamento, que agora passa a ser igual a 20:

Parâmetro	Opção Escolhida
Tipo de treinamento	<i>batch</i> (em lote)
Topologia do mapa	Bidimensional (planar)
<i>Dimensão do mapa</i>	10 x 10
Estrutura da vizinhança	hexagonal
Número máximo de clusters	$k = 10$
<i>número de repetições</i>	20
<i>épocas</i>	250 (50 – <i>rough training</i> , 200 – <i>fine tuning</i>)
<i>raio vizinhança inicial</i>	2
<i>raio vizinhança final</i>	1
<i>taxa de aprendizagem inicial</i>	Não é utilizado no modo de treinamento <i>batch</i>
<i>taxa de aprendizagem final</i>	Não é utilizado no modo de treinamento <i>batch</i>

Quadro 6.2 – Parâmetros de treinamento da rede SOM para o experimento 2.
Fonte: Elaboração da autora.

Os mapas foram inicializados com 100 neurônios, um número escolhido heurísticamente por ser grande o suficiente para dar suporte à formação de grupos, mas pequeno o suficiente para evitar *overfitting* (superespecialização do aprendizado). Optou-se por este tamanho de rede porque, de acordo com Silva (2004), “para redes pequenas (e.g. 5x5) a estrutura da matriz-U apresenta-se complexa e não fornece subsídios para a análise dos dados; já para redes muito grandes (e.g. 30x30) percebe-se uma super-especialização do mapa”, o que também não é aconselhado.

Na Tabela 6.2, estão apresentados os resultados quanto ao número de agrupamentos sugeridos pelo índice Davis-Bouldin a cada rodada de treinamento, para a palavra 'object' (verbo) usando LPC como técnica de extração de características do sinal de fala, com ordem de predição 10. Esses resultados ajudam na análise do desempenho da rede:

Tabela 6.2 – Número de clusters sugerido pelo índice DB.

Repetição	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Número de grupos	2	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3	3	3	3

Fonte: Elaboração da autora.

Como se pode observar, o número 3 aparece mais vezes e, portanto, se configura como o melhor representante para a formação de grupos nas simulações feitas. Na tabela abaixo, são apresentados os resultados quanto ao número de agrupamentos para cada palavra em cada algoritmo de extração de parâmetros da fala e ordem, com base na consulta da função 'best i*', com k variando entre 2 e 10:

Tabela 6.3: Número de agrupamentos depois da aplicação do índice DB para cada palavra em cada algoritmo de extração de parâmetros da fala e ordem de predição.

Coeficientes Itens lexicais	LPC 10	LPC 15	LPC 20	MFCC 10	MFCC 15	MFCC 20
1. Object (verbo)	3	3	4	4	4	10
2. object (subst.)	3	3	4	4	4	3
3. conduct (v.)	3	3	4	4	4	2
4. conduct (s.)	2	4	5	2	3	4
5. conflict (subst)	2	4	3	2	3	4
6. conflict (verb)	3	3	9	4	3	3
7. contest (subst.)	3	3	4	4	4	4
8. contest (v.)	3	3	4	4	4	3
9. desert (subs)	4	4	4	3	4	4
10. desert (verb)	3	3	7	3	3	4
11. separate (verb)	3	3	4	3	3	4
12. separate (adj.)	5	5	4	5	5	4
13. advocate (subs.)	6	5	6	3	4	5

14. advocate (v.)	4	4	5	3	4	5
15. appropriate (adj.)	4	4	4	4	4	4
16. appropriate (v.)	4	4	4	2	4	4
17. estimate (subst.)	3	3	4	4	3	3
18. estimate (v.)	3	4	3	4	3	4
19. elaborate (v.)	4	5	3	3	5	5
20. elaborate (adj.)	7	6	9	4	5	10

Fonte: Elaboração da autora.

Abaixo está a matriz-U resultante do conjunto de dados formado pelos vetores de características dos 30 aprendizes e mais 4 falantes nativos pronunciando a palavra *object* (verbo), usando 10 coeficientes LP.

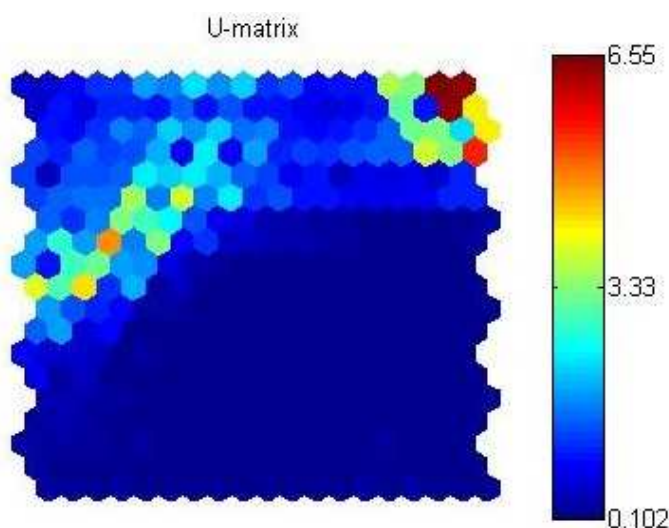


Figura 6.5: Matriz-U gerada para a rede 10 x 10 da palavra *object* (verbo) após treinamento de 250 épocas (ainda sem marcação dos grupos)

Fonte: Elaboração da autora.

A figura 6.5 mostra a matriz-U resultante do treinamento de um mapa auto-organizável bidimensional com 100 neurônios (10x10). As regiões escuras denotam a proximidade dos vetores de peso dos neurônios vizinhos no mapa. Observe que a parte inferior do centro à direita possui neurônios próximos, formando o maior e principal agrupamento. Além disso, existe outro grupo no canto superior direito do mapa e outro sendo separado por uma zona de transição no canto superior esquerdo em que os dois grupos (*clusters*) estão separados por regiões

claras (verde e amarelo). As zonas de fronteira são sempre representadas na matriz-U por neurônios de cores mais claras (ver seção 4.7). Na teoria, as zonas de transição também podem ser consideradas como grupo.

Quanto à investigação do parâmetro que melhor facilitasse a visualização da segregação da rede, foram testados os coeficientes LP e os coeficientes cepstrais (MFC). A validação de agrupamentos pôde ser feita tanto via “Erro de Quantização” quanto por Índice DB. Na figura 6.6 (LPC 10), vê-se a sugestão de formação de três agrupamentos com erro de quantização final calculado a 0,427. Já na figura 6.7, tem-se o coeficiente Mel 10 sugerindo somente dois agrupamentos com erro de quantização superior (1,636), o que já descarta a futura utilização dessa matriz resultante de coeficientes MFC.

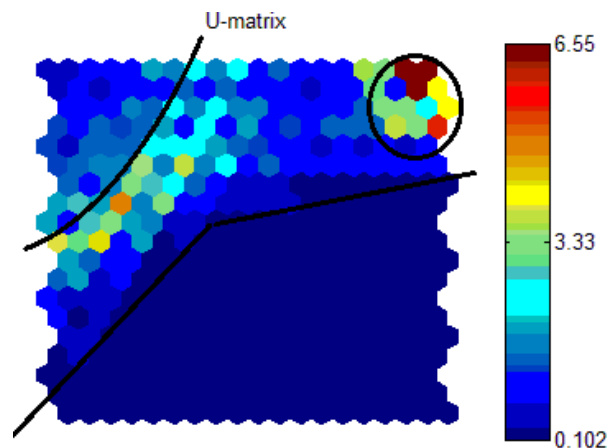


Figura 6.6: Matriz-U com entrada formada por 10 coeficientes LP. Índice DB indica 3 clusters e o erro de quantização final é igual a 0,427
Fonte: Elaboração da autora.

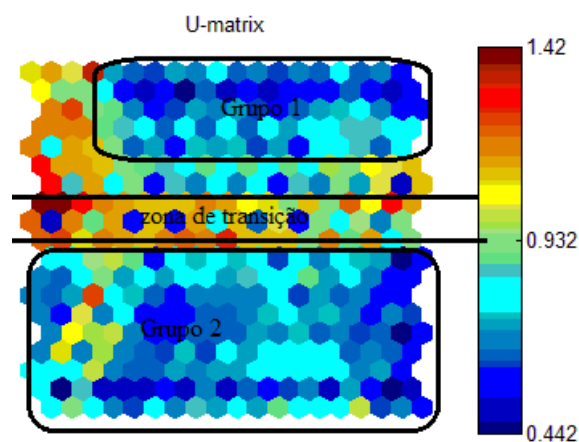


Figura 6.7: Matriz-U com entrada formada por 10 coeficientes MFC. Índice DB indica 2 clusters e o erro de quantização final é igual a 1,636.
Fonte: Elaboração da autora.

Já que os coeficientes LP 10 revelaram-se mais adequados para segregar os grupos por características prosódicas, deu-se prosseguimento aos testes da rede para cada item lexical contido na lista do apêndice C usando a mesma parametrização dos sinais de voz (LPC 10).

6.2.2 O Mapa rotulado

A rotulação do conjunto de dados consiste numa etapa fundamental no processo de treinamento da rede (ver seção 5.5.2). Rotular os mapas permite entender de que forma a rede separou os falantes. Ao contrário do que se poderia pensar, não se deve interpretar o rótulo de cada neurônio como um indicativo de que este neurônio representa exatamente um determinado aprendiz, mas sim que aquele neurônio é mais representativo de uma elocução com as mesmas características daquele determinado aprendiz. Isto ocorre porque o processo de rotulação é baseado no voto majoritário, de modo que elocuições de outros indivíduos podem também ter sido mapeadas naquele neurônio específico, embora em menor número.

À guisa de exemplificação, em um mapa com estrutura de vizinhança hexagonal (observe a marcação com a seta na figura 6.8), os vetores LPC relativos aos 34 indivíduos foram mapeados no neurônio indicado, mas somente 1 deles ganhou por voto majoritário, pois vetores de características do indivíduo 608 (aprendiz que pertence ao semestre 6 e foi o oitavo no ranqueamento em relação ao tempo de exposição ao idioma) ocorreram em maior número.

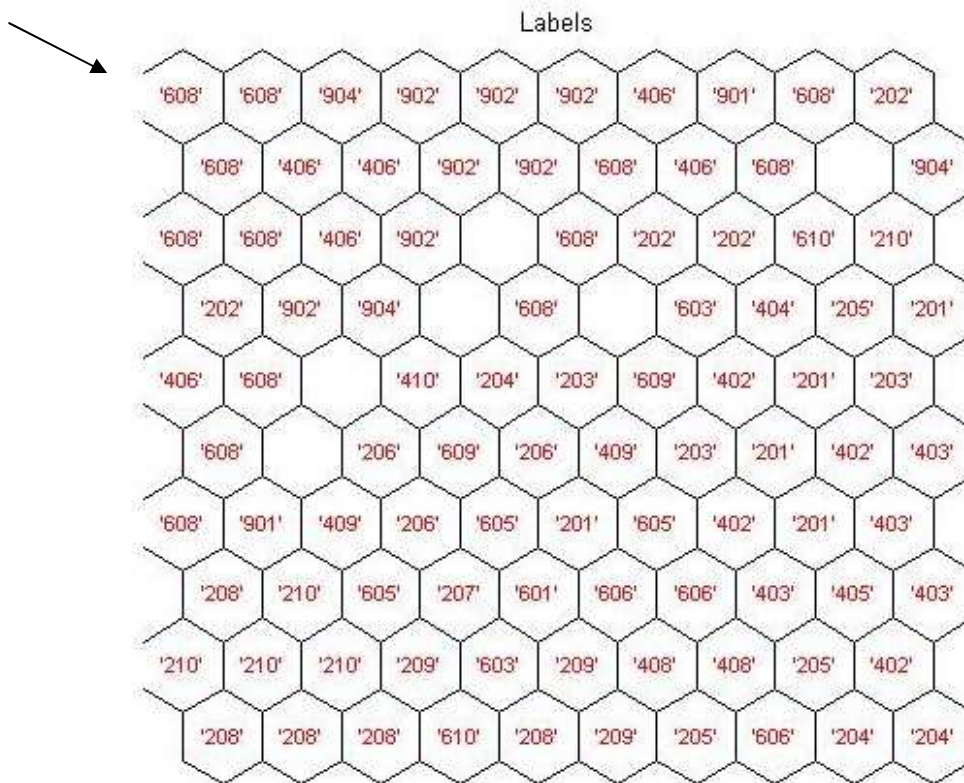


Figura 6.8: Mapa rotulado de uma rede SOM bidimensional 10 x 10 com estrutura de vizinhança hexagonal. Os rótulos correspondem a informações sobre falantes (aprendiz e nativos).
 Fonte: Elaboração da autora.

A Figura 6.9 mostra um gráfico do índice Davies-Bouldin (DB) para o mapa gerado. A métrica de DB é representada em função do número de agrupamentos e é utilizada com fins de validação. O menor valor para o índice DB sugere o melhor número de agrupamento que, no caso, é igual a 3 para ambos os exemplos abaixo. O número de grupos variou de 2 (valor mínimo estipulado) a 10.

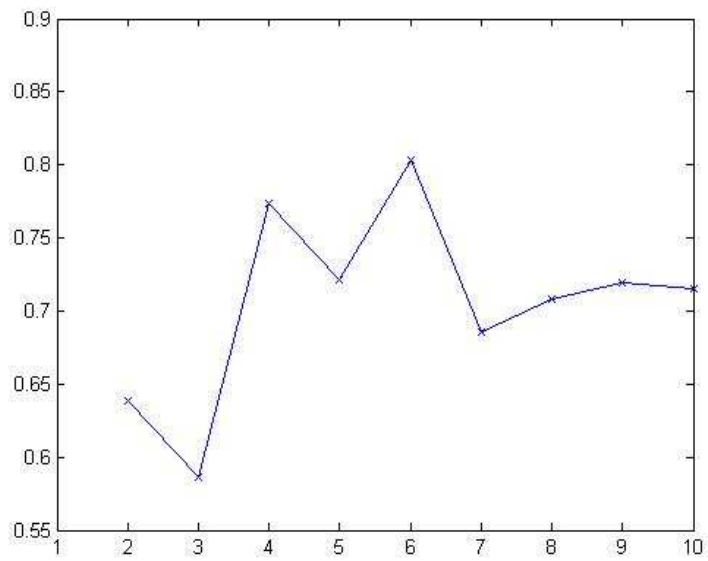


Figura 6.9: Gráfico para o índice Davies-Bouldin em função do número k de agrupamentos em uma rodada de treinamento típica.
 Fonte: Elaboração da autora.

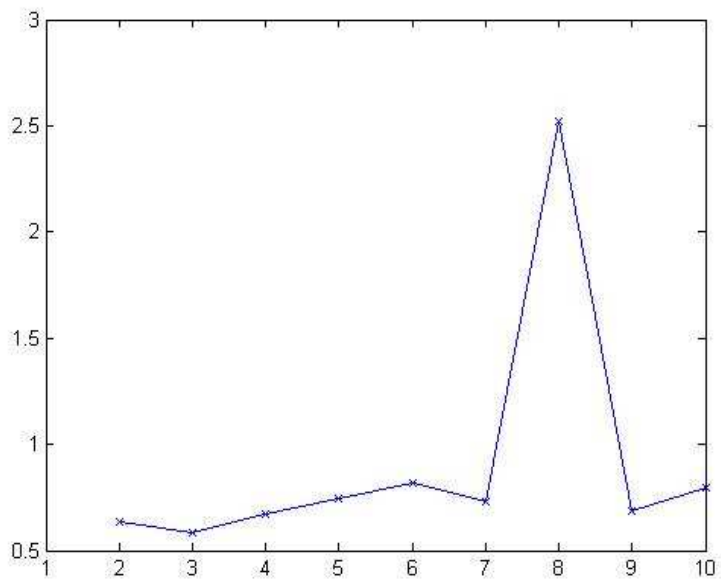


Figura 6.10: Representação dos grupos no mapa treinado para k=3.
 Fonte: Elaboração da autora.

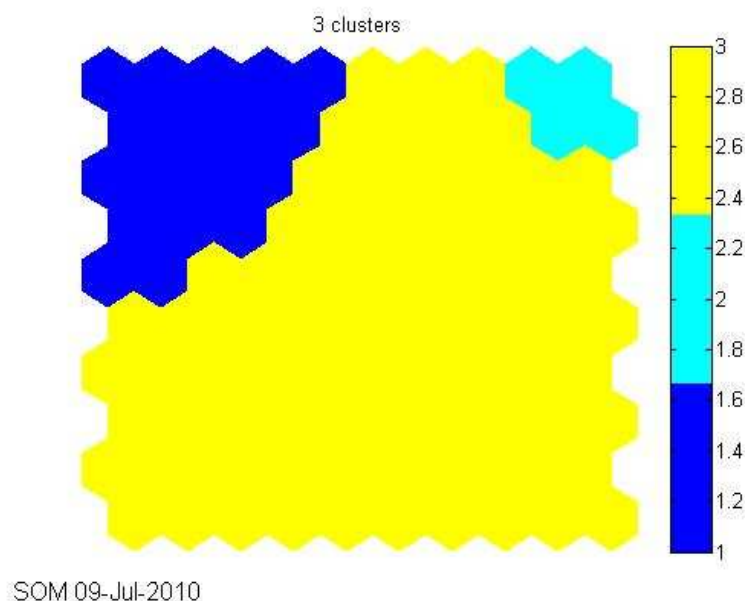


Figura 6.11: Representação dos grupos no mapa treinado para $k=3$.
 Fonte: Elaboração da autora.

Conforme o número de agrupamentos identificados, a figura 6.11 exibe a organização dos grupos no mapa. Ao se interpretar o mapa, chega-se à conclusão de que três grupos representam melhor a distribuição dos dados. Assim, pode-se dizer que a rede sugere que os aprendizes estão separados em três grupos, cada um contendo propriedades linguísticas relevantes que os diferenciam.

De posse da tabela 5.3 (Divisão dos participantes por tempo de exposição ao idioma) contida na seção 5.5 do capítulo 5 - e contrastando a figura 6.12 com a representação dos grupos no mapa (figura 6.11), verifica-se a correspondência entre as áreas circuladas no mapa. A área azul corresponde ao grupo 4, a área azul-piscina corresponde ao grupo 1 e a zona de cor amarela ao grupo 2 (ver figura 6.11).

Como se pode observar, o que garante um julgamento completo e confiável da saída da rede é a combinação da análise da matriz-U colorida, o mapa rotulado com as informações acerca do nível de proficiência linguística do aprendiz e o mapa colorido. A análise individual de um só exemplo de representação da saída da rede não garante o sucesso na interpretação dos dados.

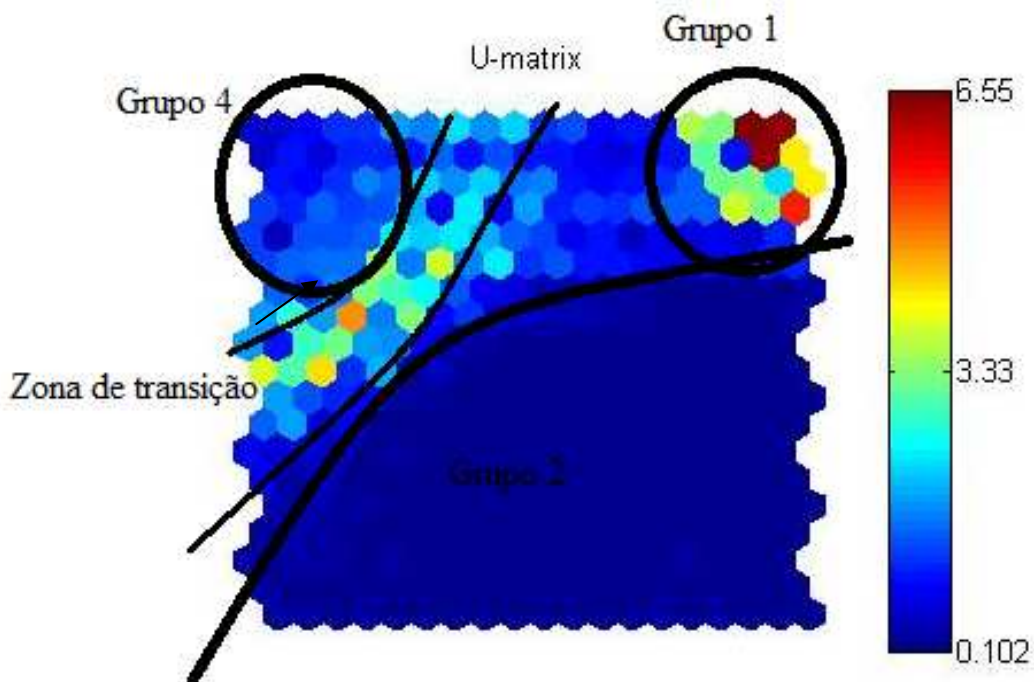


Figura 6.12: Matriz-U marcada com a representação dos grupos sugeridos pelo índice DB.
 Fonte: Elaboração da autora.

Na matriz-U da figura acima, os vetores atípicos são identificados por regiões pequenas e separadas das demais por áreas de transição (SILVA, 2004). Isso significa dizer que a região correspondente ao grupo 4 possui rótulos vencedores para neurônios representantes de dados atípicos.

Ao fazermos referência ao critério de organização adotado, na seção 5.3 da metodologia, ou seja, o de organizar os alunos por nível de proficiência levando em consideração o tempo de exposição ao idioma (nível 1, nível 2, nível 3, nível 4, nível 5, nível 6 para os falantes nativos) substituiu-se a rotulação referente à identificação do falante (por exemplo, 201) pelo nível de proficiência (tempo de horas-aula acumuladas) em que esse aluno se encontrava, nível 1. Melhor explicando, o indivíduo 201, de acordo com os critérios de organização escolhido aqui se insere no nível de proficiência 1.

Foram feitas as devidas substituições a fim de treinar a rede com os novos rótulos de informações. Com efeito, adotou-se o coeficiente cepstral 10, para uma

única palavra (object – verb). A rede aparentemente seguiu um padrão na forma de segregação dos dados. O mapa rotulado na figura 6.13 reforça o fato de que a rede foi capaz de segregar por características prosódicas semelhantes, ou seja, ela aprendeu por similaridade, o que facilitou o processo de agrupamento de indivíduos por tempo de exposição ao idioma. Relacionando esses dois fatos, depreendeu-se que o conhecimento prosódico é uma pista importante para a determinação do nível de proficiência linguística.

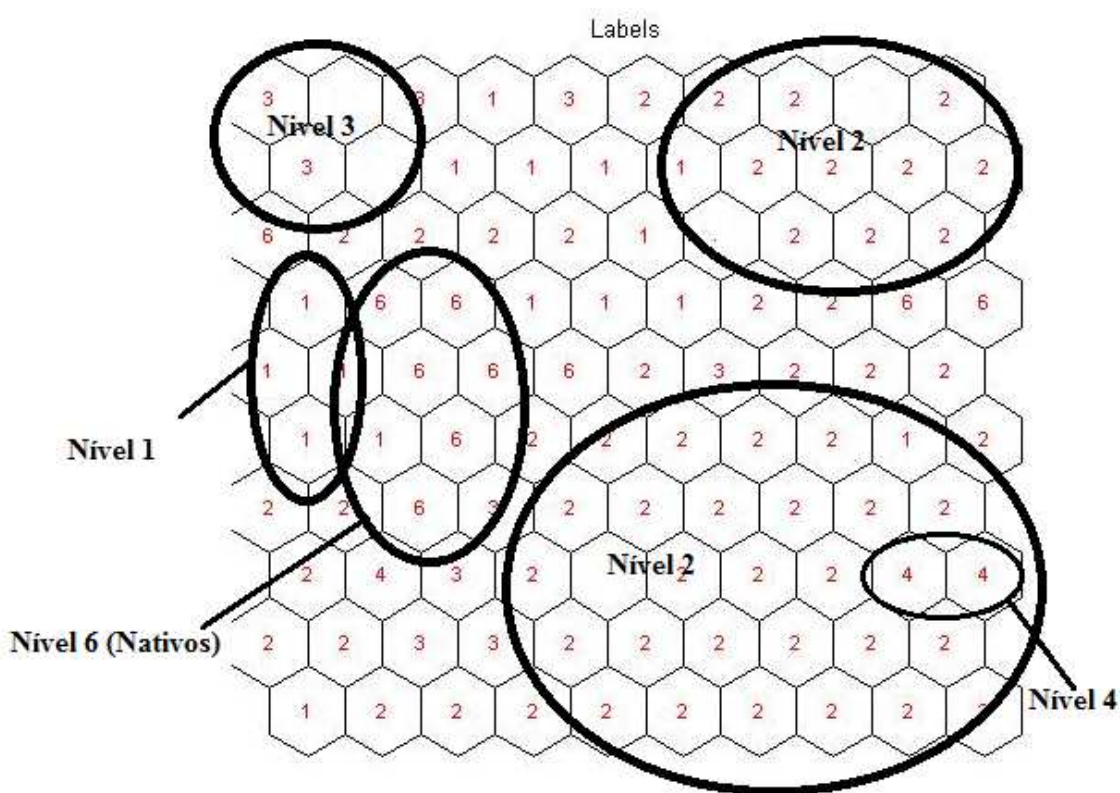


Figura 6.13: Mapa rotulado de acordo com o nível de proficiência (1, 2, 3, 4, 6).
Fonte: Elaboração da autora.

Como se pode observar na figura 6.13, nenhum neurônio foi rotulado com o nível de proficiência número 5. Isto pode ter acontecido devido ao tipo de rotulação utilizada (voto majoritário), de modo que os rótulos de indivíduos no nível 5 (menos frequentes) ficaram “escondidos” pelos rótulos dos indivíduos mais frequentes. Deve ser lembrado que, estatisticamente falando, o rótulo que aparece na figura acima é a medida de maior ocorrência da distribuição de rótulos que são mapeados em cada neurônio.

Deve-se chamar a atenção para a situação que se apresenta na figura acima. O grupo formado por aprendizes do nível de proficiência 1, que teoricamente é o grupo que possui aprendizes com menor tempo de exposição à língua inglesa, está figurando ao lado do grupo 6 (nativos). Coincidência? Confusão da rede? Insuficiência nos dados de entrada? Os aprendizes que figuravam nesses neurônios foram inspecionados e comparados aos outros colegas do mesmo nível em todas as pronúncias das palavras em questão e observou-se que estes não transferiram o padrão acentual de L1 para L2.

Foram realizados testes com as outras palavras com o método de extração de características LPC e os resultados foram condizentes com os resultados acima. Observou-se que houve a formação de 5 agrupamentos nos outros conjuntos de dados testados na rede, um sinal claro de que a rede seguiu o mesmo padrão de organização. A formação da topologia dos mapas relativos aos demais itens lexicais corrobora a proposta de existência do mesmo número de agrupamentos básicos organizados por tempo de exposição ao idioma (horas-aula acumuladas na disciplina de língua inglesa).

Como já falado na seção 5.4, o processo de determinação de nível de proficiência lingüística não é tarefa fácil para o especialista. O uso da ferramenta de visualização Matriz-U, portanto, é inovador no auxílio de julgamento da proficiência quanto aos aspectos acentuais de língua inglesa.

É claro que se os aprendizes tivessem sido submetidos previamente a um teste padrão de proficiência lingüística, por exemplo, o *Michigan Proficiency Test* ou o TOEFL, poder-se-ia contrastar esses resultados com os resultados oferecidos pela rede com mais acuidade. Uma providência a ser tomada daqui para frente é submeter todos os futuros participantes da pesquisa a algum desses testes padrão, alocá-los em nível de proficiência com critério básico definido e, numa fase posterior, treinar os conjuntos de dados e armazená-los para serem inseridos novamente na rede. Com os indivíduos alocados de maneira mais rígida e organizados sob critérios bem mais definidos, acredita-se que a análise da topologia do mapa far-se-á de forma mais segura e confiável.

Uma outra contribuição da rede SOM é uso da norma euclidiana para calcular a distância que um falante, cujo conjunto de dados seja inserido posteriormente na rede, esteja de outro indivíduo com características já mapeadas.

Por exemplo, pode-se treinar a rede com vetores de atributos de um número igual de falantes por nível de proficiência (e.g.: 5 falantes do nível A1⁶¹, 5 do nível B2, 5 do nível C1) juntamente com os vetores de características de um aprendiz que ainda não se submeteu a uma classificação rígida antes (ver Anexo C). É possível determinar o nível em que esse novo aprendiz se encontra sem o uso obrigatório de um teste padrão. Basta visualizar a topografia dos dados na matriz-U para depreender em que agrupamento ele se insere após o treinamento da rede, se nos agrupamentos A1, B2 ou C3 ou em outros.

Em suma, a rede reconheceu padrões recorrentes nas formas de agrupamento dos neurônios vencedores (protótipos), refletindo o grau de aproximação e baseando-se em medidas de similaridade nos dados de entrada da rede. Tal tarefa desempenhada pela rede é um indicativo/sugestão de nível de proficiência que deve ser retomado em trabalhos futuros que se dediquem à análise dos critérios de determinação, de classificação e avaliação linguística.

6.3 Simulação da transferência de padrões acentuais (acento primário) para substantivos, verbos e adjetivos

6.3.1 Metodologia da simulação

O presente experimento objetivou investigar se a rede seria capaz de evidenciar, através da formação de agrupamentos, os processos de transferência do padrão acentual de PB (português brasileiro) para o inglês, mais especificamente o acento primário nas categorias lexicais substantivo, verbo e adjetivo. Dessa vez, especificou-se a rede em uma topologia bidimensional com 25 neurônios (5x5) em vizinhança hexagonal, já que se teria que vislumbrar o aparecimento de dois grupos somente – o grupo que transferiria o padrão de PB para o Inglês e o grupo que não transferiria. Os parâmetros de treinamento da rede são mostrados no quadro 6.3.

⁶¹ Referência ao Marco Comum Europeu (Common European Framework of Reference - CERF) de Proficiência Linguística.

Parâmetro	Opção Escolhida
Tipo de treinamento	<i>batch</i> (em lote)
Topologia do mapa	Bidimensional (planar)
<i>Dimensão do mapa</i>	5 x 5
Estrutura da vizinhança	hexagonal
Número máximo de clusters	$k = 10$
<i>número de repetições</i>	1
<i>épocas</i>	250 (50 – <i>rough training</i> , 200 – <i>fine tuning</i>)
<i>raio vizinhança inicial</i>	2
<i>raio vizinhança final</i>	1
<i>taxa de aprendizagem inicial</i>	Não é utilizado no modo de treinamento <i>batch</i>
<i>taxa de aprendizagem final</i>	Não é utilizado no modo de treinamento <i>batch</i>

Quadro 6.3: Parâmetros de treinamento da rede SOM para o experimento 3.
 Fonte: Elaboração da autora.

As figuras 6.14 e 6.15 são relativas à formação do mapa que teve como entrada uma matriz de dados de falantes rotulados por um número (semestre e ranqueamento) acrescido do rótulo “er” quando um certo participante erra a pronúncia ao transferir o padrão de L1 para L2. Vê-se claramente a sinalização de formação de dois grupos: um que transfere o padrão acentual (grupo maior e mais coeso) e um que não transfere (grupo menor e mais disperso).

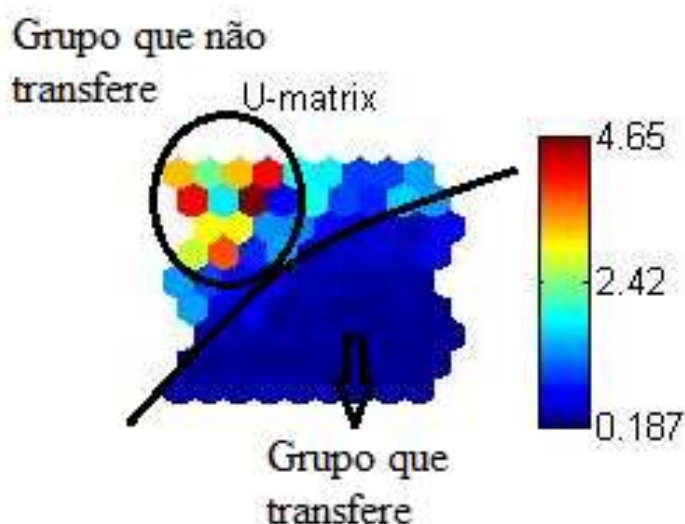


Figura 6.14: Matriz-U evidenciando a formação de 2 grupos: um que contém indivíduos que transferem o padrão de colocação do acento primário de L1 para L2 e outro que não transfere.
 Fonte: Elaboração da autora.

No mapa rotulado com a identificação dos aprendizes quanto ao semestre adicionado do padrão de transferência para a palavra object (verb) (ver tabela de falsos cognatos no Apêndice E) usando coeficientes LP 10, há uma correspondência entre as topologias dos respectivos mapas. O mapa rotulado corrobora a formação do mesmo número de agrupamentos nas mesmas áreas de formação na matriz-U. Vale ressaltar que a interpretação do mapa se vale das duas formas de visualização para garantir uma análise fidedigna do conjunto de dados.

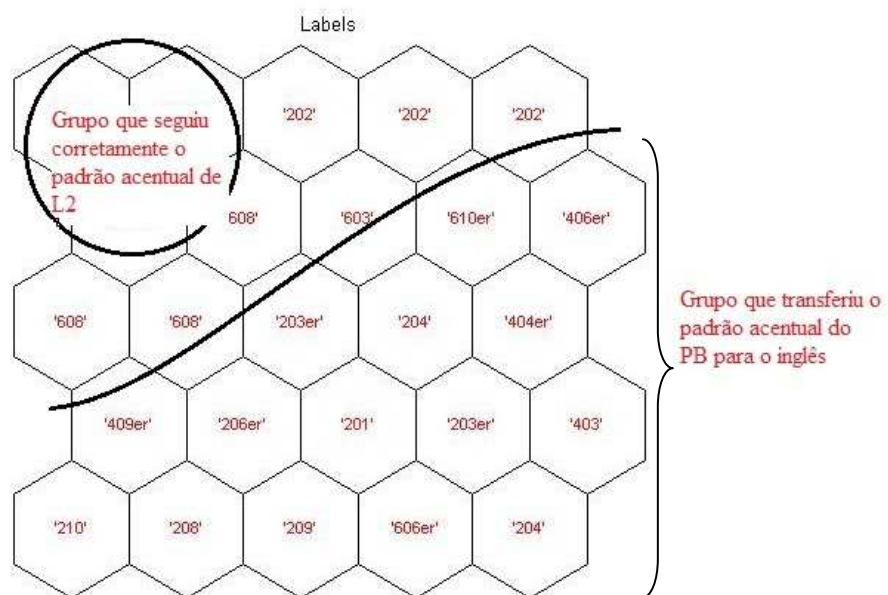


Figura 6.15: Mapa rotulado com “erro” para os conjuntos de dados referentes à palavra object (verbo), usando LPC 10, treinados com dados dos 30 aprendizes.
 Fonte: Elaboração da autora.

A fim de obter uma compreensão maior da organização dos dados, foram seguidos dois critérios de rotulação da rede. No primeiro, mantiveram-se os números de identificação dos falantes sem nenhuma informação sobre seus erros na computação do acento de L2, ou seja, a transferência do padrão de L1 para L2. No segundo critério, adicionou-se o rótulo “er” para indicar o erro na computação do acento primário de palavra daquele falante. Cabe ressaltar que os vetores de atributos carregam informação que pode ter uma interpretação ampla. Todavia, as categorias linguísticas que emergiram na rede puderam dar margem de interpretação sobre o fenômeno de transferência acentual. Tal fato não parece ser

fruto do acaso, pois a rede organiza por similaridade e os vetores referentes aos indivíduos que cometeram erros de substituição vocálica e consonantal, bem como os que realizaram o processo de inserção e apagamento de segmentos vocálicos, também emergiram durante o processo de treinamento e atualização dos pesos sinápticos da rede.

Os neurônios que aparecem sem a notação do rótulo “er” representam os vetores de peso dos neurônios vencedores dos grupos de aprendizes que não transferiram o padrão acentual do PB para o inglês. A análise dos rótulos de neurônios vencedores que ficaram em segundo lugar pôde nos dar garantir a segregação correta da rede, como mostra a figura 6.16:

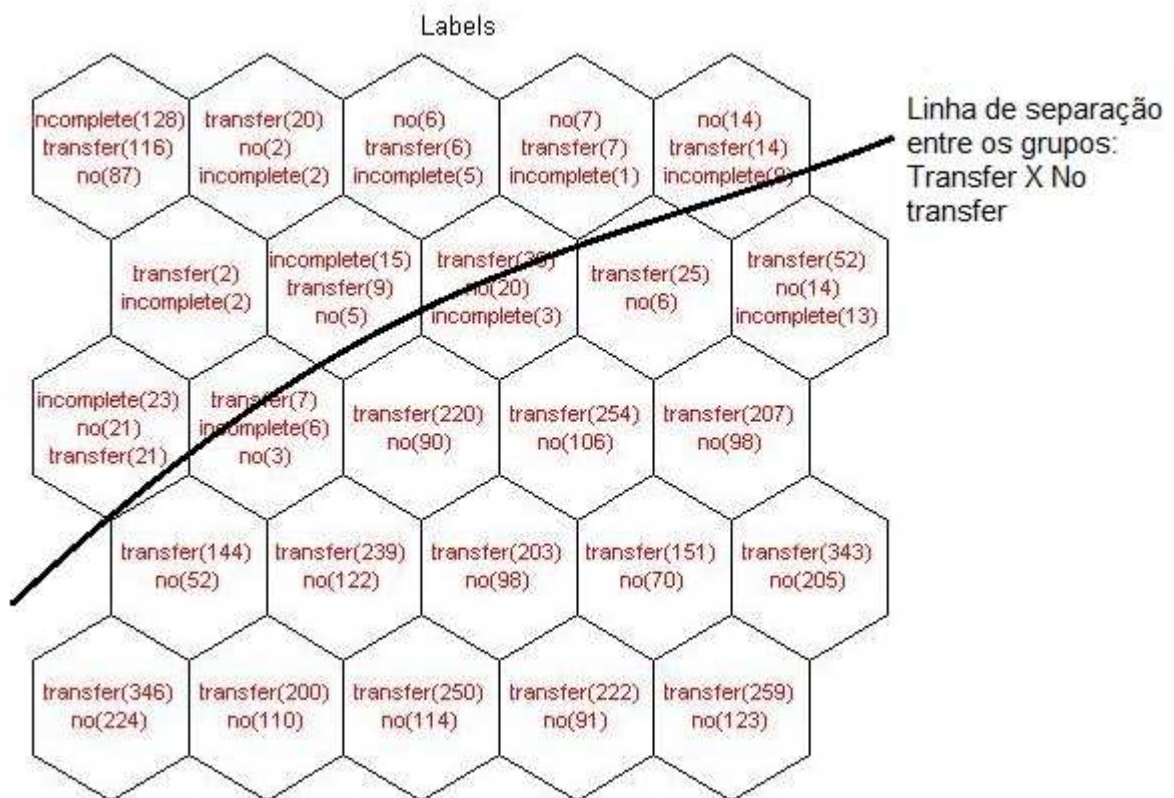


Figura 6.16: Mapa rotulado com informações sobre a freqüência das classes (rótulos) para cada neurônio vencedor.

Fonte: Elaboração da autora.

Durante o preparo do *input* da rede, rotulou-se o conjunto de dados com diferentes codificações - rótulo alfanumérico (e.g.: 201er), rótulo “transfer” e “no transfer”, e rótulo ‘er’. Em todos os resultados para a palavra *object* (LPC 10), a matriz-U seguiu o mesmo padrão de segregação não importando o rótulo dado ao

conjunto de dados. Tal fato faz chegar a uma conclusão importante: as linhas de exemplos de vetores de atributos são exclusivamente processados sem levar em consideração a variação do título do rótulo durante a convergência da rede.

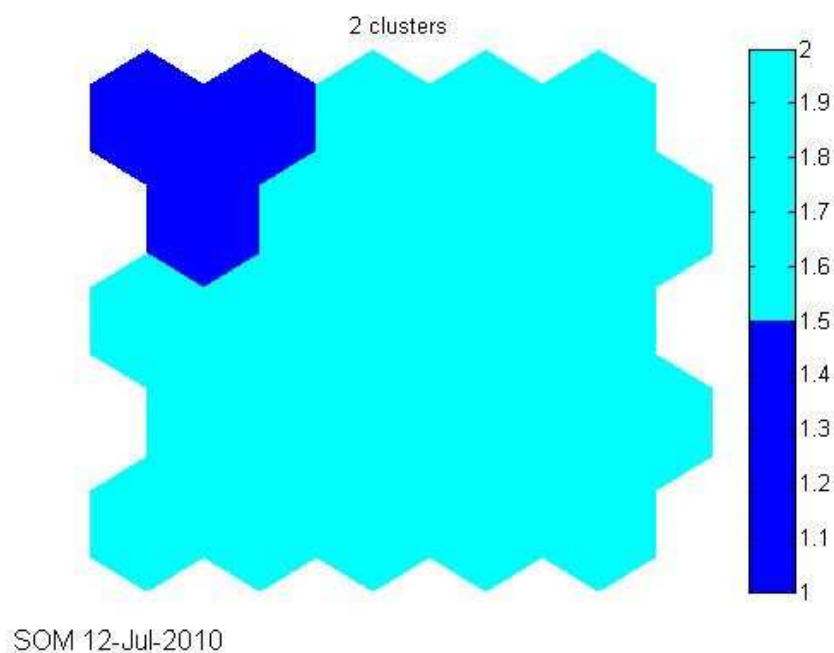


Figura 6.17: Mapa colorido sugerindo a formação de dois grupos, conforme indicação do índice Davies-Bouldin.

Fonte: Elaboração da autora.

O teste da rede aconteceu para os 19 itens lexicais restantes com conjunto de dados extraídos usando coeficientes MFC de diferentes tamanhos. Em todos os outros resultados, observou-se a emergência do mesmo padrão.

6.4 Simulação da Transferência de Padrões Vocálicos (substituição, apagamento, inserção e redução vocálica)

Ao contrário do que se esperava, não houve resultados promissores nesta simulação porque a rede não produziu um padrão de organização dos protótipos em agrupamentos em relação aos cinco níveis de proficiência determinados por tempo de exposição ao idioma. Um mapa rotulado com transcrições fonéticas referentes a cada neurônio vencedor foi criado no PaintBrush® para facilitar a visualização das características escolhidas para a decisão de agrupamento.

Quase sempre a rede sugeriu a organização de protótipos em grupos que não faziam sentido na análise dos dados. O que se esperava era que a rede segregasse os agrupamentos que tivessem realizações vocálicas semelhantes. Para tentar resolver o problema, decidiu-se rotular os arquivos com extensão .txt (matrizes) que compuseram o padrão de entrada da rede com um número que denotasse a ausência ou a presença do processo de transferência de padrão vocálico.

Tomemos o exemplo de inserção de vogal em posição de coda de palavra (e.g. “It’s a small metal object”: [ɔbi'dʒektʃi]). Adotou-se a codificação 201.0 e 201.1; o primeiro dígito após a vírgula representa a ausência ou presença do fenômeno de transferência linguística na elocução da palavra. Assim, o dígito 1 (um) representou a presença do processo fonológico (inserção de vogal em posição de coda) e o 0 (zero) representou a ausência do processo. Todavia, tal codificação não gerou os resultados esperados já que não foi visualizada no mapa rotulado a presença de pelo menos duas categorias, ou seja, o agrupamento dos vetores de atributos que representavam a inserção e o agrupamento dos vetores que não realizavam a inserção de vogal em posição de coda.

A razão para não ter aparecido 1's no mapa é a frequência (quantidade) de indivíduos em que inserção de vogal em posição de coda é observada. No jargão da área de reconhecimento de padrões, as classes estão desbalanceadas ou desequilibradas. i.e., com um número diferente de indivíduos em cada uma delas. Com o objetivo de observar a emergência de neurônios rotulados com 1's, seria aconselhável repetir o experimento com a mesma quantidade de indivíduos em cada categoria.

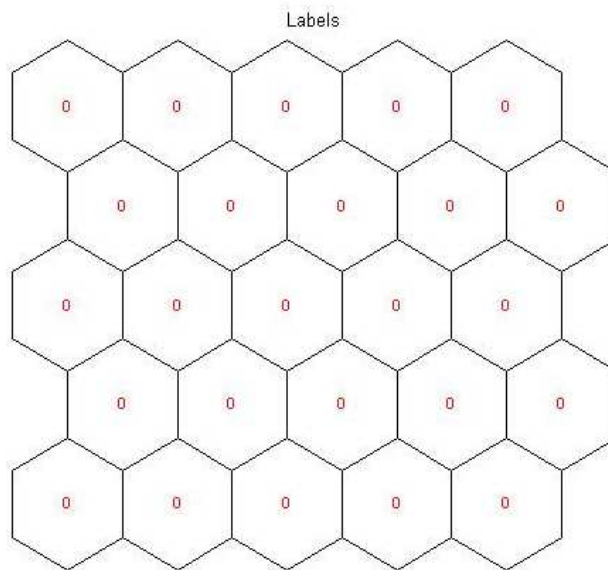


Figura 6.18: Mapa rotulado com codificação relativa aos fenômenos de inserção de vogal em posição de coda.
 Fonte: Elaboração da autora.

Outro fator que invalidou a análise foi o fato de a rede sugerir mais de dois agrupamentos, já que a expectativa era de dois grupos somente. Após vinte repetições do treinamento, as sugestões de agrupamento ficaram entre 4 e 9, números que não correspondem à expectativa da simulação.

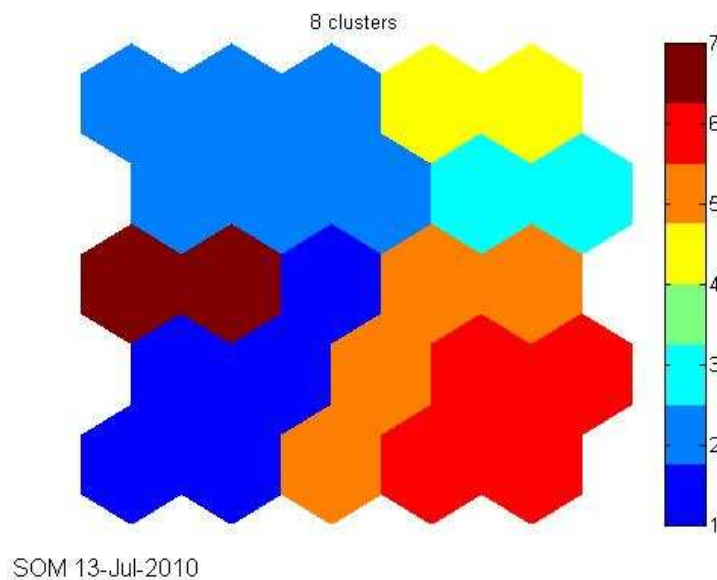


Figura 6.19: Mapa colorido sugerindo a formação de clusters.
 Fonte: Elaboração da autora.

A matriz-U também não permitiu nenhuma visualização de dados que contrariasse o alto número de agrupamentos sugeridos. Ao final da observação dos resultados, não houve concordância entre os três mapas, cada qual treinado com o mesmo conjunto de dados.

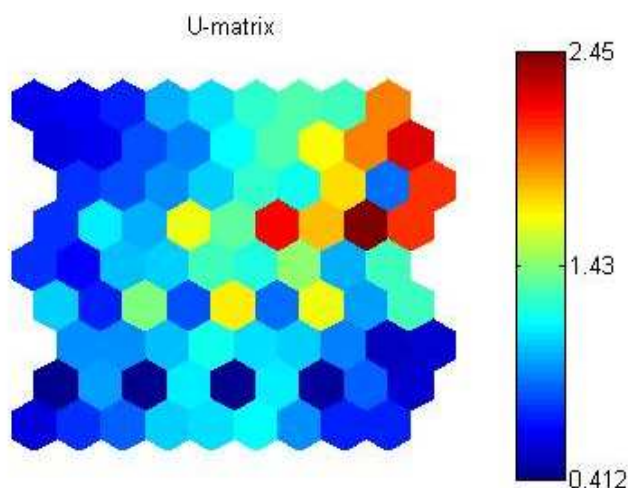


Figura 6.20: Matriz-U sem nenhuma representação significativa de clusters.
Fonte: Elaboração da autora.

Chegou-se a essa conclusão após várias tentativas de codificação do conjunto de dados que alimentava a rede serem consideradas como inválidas. Talvez se outro algoritmo de extração de características tivesse sido utilizado, diferente dos dois aqui utilizados - LPC e MFCC, poder-se-ia visualizar a formação de agrupamentos, ou até mesmo a emergência de categorias que representassem processos de transferência vocálica e consonantal durante o treinamento da rede.

Já se sabe que a rede recorrente simples pode simular a transferência desses processos, como proposto por Zimmer (2006). Note-se, porém, que uma comparação direta não pode ser feita entre as duas propostas porque a abordagem supervisionada não usa os mesmos *inputs* da abordagem supervisionada. Além disso, o problema do ponto de vista não-supervisionado é mais complexo, pois o algoritmo tenta extrair sozinho as propriedades presentes nos dados, sem o auxílio de um “professor”, como se diz no jargão da área de redes neurais supervisionadas.

A despeito de as tentativas terem sido frustradas em relação à transferência de padrões vocálicos, tal fato despertou a curiosidade da pesquisadora para simular a formação de instâncias prototípicas, ou magnetos perceptuais (KUHL, 1991) em um novo experimento que ocorrerá futuramente.

Com respeito à realização do fenômeno de redução vocálica, a simulação não ofereceu resultados motivadores. Na realidade, houve o aparecimento de questões novas e não uma resposta a questões anteriores. O conjunto de dados foi extraído utilizando-se coeficientes LP de ordem 10 da palavra *appropriate* (adj), item lexical em que se observa claramente a realização da vogal reduzida (*schwua*, [ə]). Os dados referentes aos vetores de atributos de 30 aprendizes foram primeiramente inseridos em uma rede SOM bidimensional de dimensão 10 x 10.

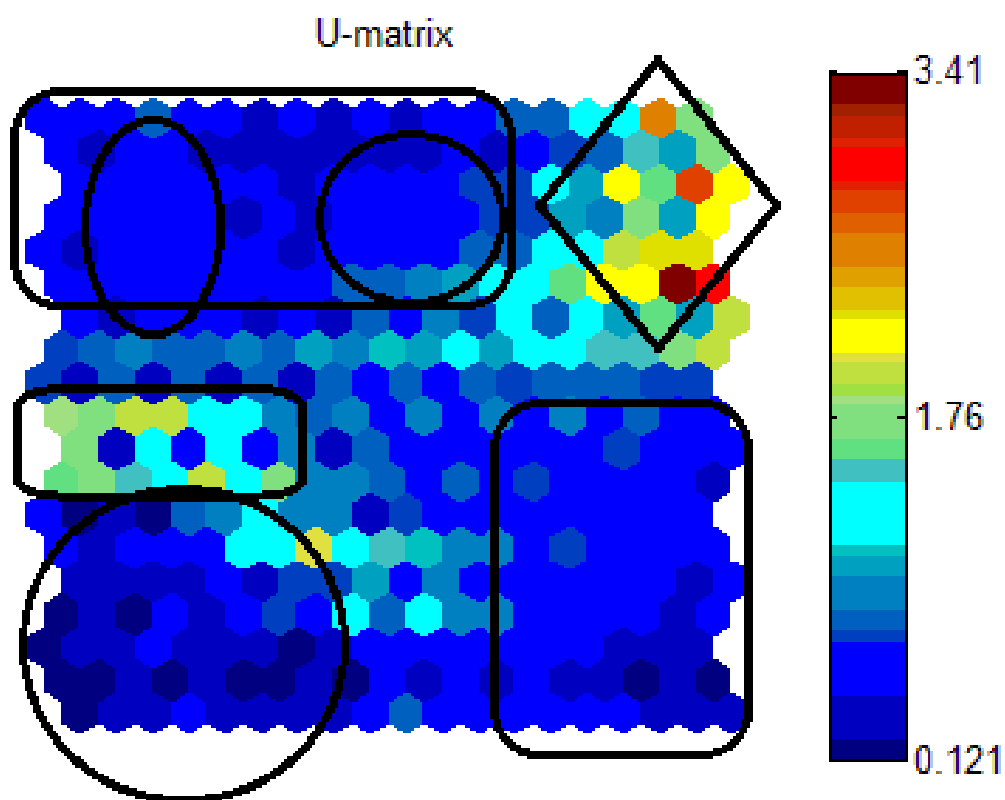


Figura 6.21: Marcação da formação de grupos na Matriz-U.
Fonte: Elaboração da autora.

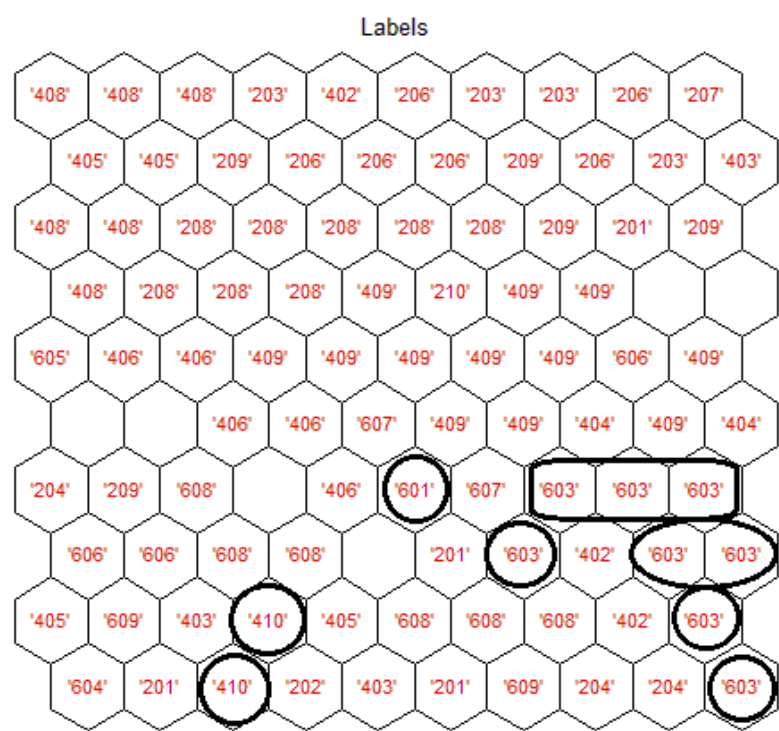


Figura 6.22: Mapa rotulado com identificação de indivíduos sem boa formação de agrupamento. Fonte: Elaboração da autora.

Não se obteve uma visualização satisfatória da organização do mapa, então se decidiu mudar para uma grade formada por 16 neurônios, dispostos da forma 4 x 4, como se vê na figura 6.23. A rede sugeriu a formação de 4 grupos, assim dispostos:

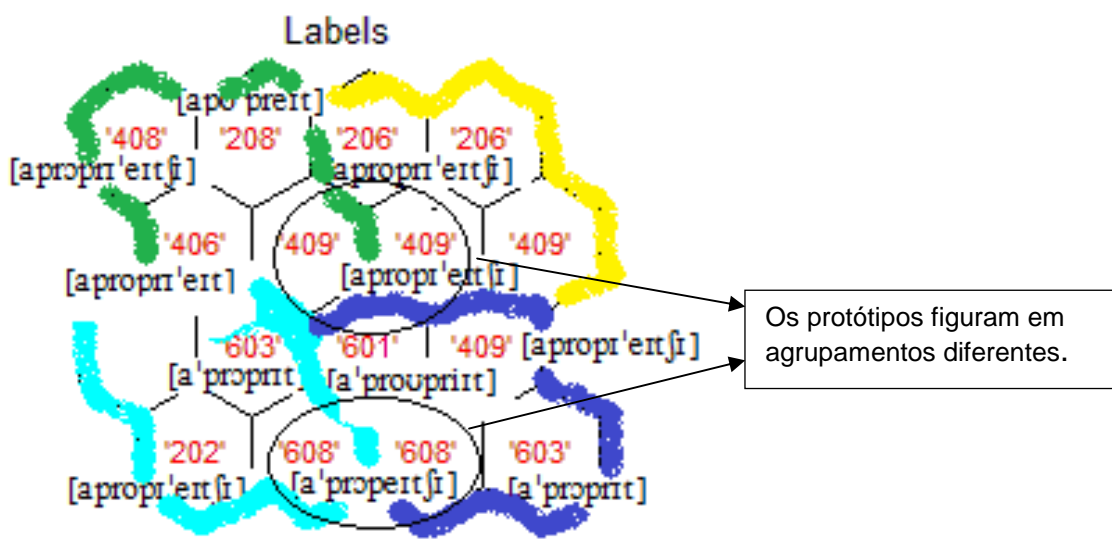


Figura 6.23: Divisão do mapa rotulado sugerida pela rede em quatro grupos. Fonte: Elaboração da autora.

A região circulada por azul escuro possui dois neurônios que representam casos de redução vocálica (neurônio rotulado com os vetores de atributos do aprendiz 601 e 603) enquanto os outros neurônios desses agrupamentos não possuem nem o neurônio vencedor, muito menos as segundas unidades vencedoras sendo representadas pelos vetores de alguma elocução em que haja a realização de vogal reduzida.

A primeira razão para tal inabilidade da rede reside no fato de haver uma falta de proporção entre a quantidade de valores relativos à redução vocálica e entre as outras elocuições. Entre o número de 30 falantes que pronunciou essa palavra, somente 4 realizaram a redução da vogal. Como a rede trabalha com padrões estatísticos, seria o caso de testar novamente outro conjunto de dados com informações balanceadas para garantir a confiabilidade na representação da amostra.

Outro problema percebido no mapa foi a presença de um mesmo rótulo em dois agrupamentos diferentes. O rótulo 608 aparece nos agrupamentos de cor azul claro e azul escuro, e o rótulo 409 no agrupamento verde, amarelo e azul escuro. O problema detectado pode ser explicado sob o prisma de algumas teorias sobre categorização prototípica. Como a meta de algumas perspectivas teóricas (LANGACKER, 1986) sobre categorização é capturar generalizações ao se descrever qualquer traço comum entre os membros de uma determinada categoria, o analista não irá escapar da tarefa árdua de generalizar e especificar traços até chegar à conclusão de que categorias complexas partem de protótipos. Ao se considerar os fonemas como categorias complexas, mais simulações podem ser realizadas a fim de explicar o cenário para a evolução dessas categorias complexas.

Rosch (1977) aponta que as categorias não são definidas somente com base em atributos igualmente divididos por todos os membros de classes conceituais. “O fato de pertencer a uma categoria não se constitui como um fenômeno decisivo, sendo, de fato, uma questão de gradação que vai de um protótipo a casos limítrofes” (MACEDO, 2006, p. 7).

Além da área de categorização lexical, a teoria dos protótipos tem exercido grande influência na área de reconhecimento de padrões, análise de agrupamentos e aquisição fonológica. A teoria prototípica tem sugerido ao longo de sua evolução que a solução para o problema da representação fonológica esteja no

fato de que nós guardamos em nossa mente um protótipo para cada fonema, uma forma idealizada que pode ser comparada com outras formas variantes encontradas no discurso. Outras simulações poderiam ser realizadas tendo como objetivos comprovar a teoria de formação de protótipos na representação fonológica de L1, que servirá de base para a proposição de uma nova teoria de aquisição de L2, já que a formação de categorias fonético-fonológicas de L2 se apoia diretamente no inventário de L1, como mesmo dizem os estudos sobre transferência do conhecimento fonológico.

A Teoria dos Exemplos (BYBEE, 2001) propõe uma alternativa a essa explicação que provavelmente explique o problema da rede mapear os mesmos neurônios em grupos distintos. O objetivo principal desta teoria é entender como as representações fonológicas se organizam. Em vez de classificar entidades do mundo real com relação a um bom exemplo, são armazenados traços múltiplos de todos os exemplos já encontrados, o que se opõe à teoria dos protótipos.

Além de a teoria sugerir que informações fonéticas detalhadas bem como informações sociolinguísticas estejam presentes nas representações mentais, a teoria dos exemplos ainda propõe que generalizações ocorrem em vários níveis, desde o segmental silábico até o fonológico. Além disso, os itens lexicais que constituem as representações mentais, sejam elas lexicais ou fonológicas, definem o mapeamento dessas mesmas representações.

Dessa forma, os casos de redução de segmentos passam a ser tratados como casos de aquisição fonológica, como se acredita ter ocorrido aqui na simulação da transferência de redução vocálica. Conclui-se, portanto, que não há como simular um processo de transferência fonético-fonológica cuja variação sofreu a interferência direta de variáveis idiossincráticas ou cujo padrão aquisicional seja inexistente na língua do aprendiz, como é o caso da redução vocálica no PB.

Já que o objetivo maior da teoria dos exemplos é entender como se dá aquisição de padrões sonoros variáveis, motiva-se esta pesquisadora a continuar a discussão em outro momento. Para isso, propõe-se fazer uso de um conjunto de dados colhidos da língua nativa e simular a aquisição de mais padrões com algoritmos de extração de parâmetros que levem em conta informações mais detalhadas do espectro vocálico.

Outra explicação para o problema detectado nesta simulação poderia ser cedida pela teoria de Sistemas Fuzzy de Zadeh (1965). A Teoria de Conjuntos Fuzzy foi concebida com o objetivo de fornecer um ferramental matemático para o tratamento de informações de caráter impreciso ou vago. Dada a capacidade dos seres humanos de lidarem com informações imprecisas na resolução de problemas bastante complexos, é possível usar estratégias de natureza lingüística geralmente imprecisas. A lógica Fuzzy procura modelar os modos imprecisos do raciocínio que têm um papel fundamental na habilidade humana de tomar decisões.

Os conceitos de Lógica Fuzzy nasceram inspirados na lógica proposicional (tradicional). Na teoria dos Conjuntos Fuzzy, que se opõe aos Conjuntos “Crispy”, a noção de pertinência é bem definida: elementos pertencem ou não pertencem a um dado conjunto A (em um universo X). Existem conjuntos cujo limite entre pertinência e não-pertinência é vago. A regra é disparada desde que exista um grau de similaridade diferente de zero entre a premissa 1 e o antecedente da regra, sendo que o resultado é um conseqüente que tem um grau de similaridade diferente de zero com o conseqüente da regra. Ao que parece, essa teoria ofereceria uma alternativa razoável para a resolução do problema acima exposto enfrentado pela rede neural.

Diante das opções cima apresentadas, cabe à pesquisadora escolher o *background* teórico mais adequado antes de realizar novas simulações com um novo conjunto de dados.

6.5 Sumarização dos resultados

O primeiro experimento simulou a representação de categorias lexicais tendo por base o elemento prosódico acento. O objetivo da simulação foi testar se as redes de Kohonen são competentes o suficiente para executar a tarefa de organização das categorias lexicais com base nessa característica prosódica sem levar em conta o contexto. Os resultados da matriz-U, que se basearam no erro de quantização e no erro topológico, apresentaram um padrão claro de segregação entre os três grupos apresentados à rede (substantivos, adjetivos e verbos). A partir do experimento rodado, pode-se depreender que categorias linguísticas podem

emergir na topologia da rede como um produto natural do processo de auto-organização dos dados.

A segunda simulação analisou a formação de grupos em função do conhecimento prosódico do aprendiz e do seu tempo de exposição ao idioma. Determinou-se a parametrização do sinal de fala em coeficientes que se mostraram mais adequados para a codificação e condensação dos elementos prosódicos (LPC e MFCC) em nível lexical (acento primário e acento secundário). Destacou-se a matriz-U como uma ferramenta de visualização útil no auxílio do julgamento do conhecimento linguístico dos aprendizes. Os resultados com base nessa poderosa ferramenta de visualização sugerem que o seu uso é bastante útil na sugestão de nível de proficiência linguística em LE/L2.

A terceira simulação esboçou a transferência de padrões acentuais (acento primário) para substantivos, verbos e adjetivos. O objetivo foi investigar se a rede seria capaz de evidenciar, através da formação de agrupamentos, os processos de transferência do padrão acentual de PB (português brasileiro) para o inglês, mais especificamente o acento primário nas categorias lexicais substantivo, verbos e adjetivos. Pôde-se perceber claramente a emergência de padrões similares em todos os testes. Em resumo, a rede foi capaz de agrupar os falantes que transferiram os padrões acentuais do PB para o inglês e de forma análoga também foi capaz de segregar os grupos no mapa.

O quarto experimento simulou a transferência de padrões vocálicos (substituição, apagamento, inserção e redução vocálica). Infelizmente, não houve resultados promissores nesta simulação, porque a rede não produziu um padrão de organização dos protótipos em agrupamentos em relação aos cinco níveis de proficiência determinados por tempo de exposição idioma.

CONCLUSÃO

A presente pesquisa se justifica pela escassez de estudos que se utilizem de redes neurais artificiais auto-organizáveis para explicar como o conhecimento acentual se organiza na fase inicial de aquisição de L2 e ainda pelo interesse de se visualizar como essa rede neural organizaria os aprendizes em função desse conhecimento de natureza prosódica.

O objetivo principal desta tese foi analisar como a rede de Kohonen organiza os aprendizes brasileiros de inglês em função do conhecimento acentual de língua inglesa, quando da pronúncia de itens lexicais com categorias determinadas pela posição do acento primário, acento secundário e redução vocálica e quais informações subjacentes foram relevantes no processo de formação de agrupamentos. Ao final das simulações constatou-se que a ferramenta SOM conseguiu capturar fenômenos de transferência do conhecimento acentual.

A pesquisa contou com a modelagem da aprendizagem de padrões linguísticos de L2, utilizando o sinal de fala numericamente codificado como *input* de uma rede neural não-supervisionada. Partiu-se da hipótese de que a parametrização do sinal de fala por meio da abordagem de coeficientes LP ou MFC seria eficiente na categorização dos falantes por características prosódicas.

A nova proposta teórico-metodológica apresentada aqui por meio da Análise Multi-rótulo (*Multi-label Analysis*) é inédita e a ela estão atreladas as vantagens do uso da rede de Kohonen (rede neural não-supervisionada), que também tem utilização inédita para esta área investigada, assim como o uso da técnica de LPC e MFCC na codificação do *input* da rede.

O trabalho também contribuiu para a resolução da questão de processamento de sinal acústico da fala já que a formulação do *input* e arquitetura da rede neural deu um grande passo em direção à tarefa de simulação de captação, decodificação e classificação do sinal linguístico para finalmente organizá-lo em classes ou agrupamentos (saída da rede). Vale ressaltar que o método de extração de características do sinal da fala aqui utilizado merece grande destaque, já que conseguiu capturar as principais características atreladas ao fenômeno aqui

analisado: o acento lexical, energia, frequência fundamental (F_0) e banda de frequência.

A segregação do mapa em regiões de agrupamentos densos sugeriu que os aprendizes foram agrupados por características fonético-acústicas semelhantes. De acordo com as rodadas de experimentos, confirmou-se que a rede discriminou os falantes por características prosódicas e os organizou de acordo com as similaridades relativas a essas características.

Uma possibilidade de organização da rede descartada a partir da visualização dos dados foi a de organização por padrão vocálico e consonantal. Pelo que se pôde observar nos experimentos, uma única unidade fonológica pinçada de um conjunto imenso de possibilidades não se constitui como traço relevante e suficiente para a emergência de padrões.

A formação de agrupamentos na matriz-U sugeriu pontos críticos no sistema lingüístico do aprendiz, o que fornece informações valiosas sobre a construção do sistema lingüístico do aprendiz (interlíngua). Com isso, se comprova a capacidade da rede neural em captar mudanças sutis no desenvolvimento lingüístico à aprendizagem de acento de língua inglesa.

Ao que parece, a rede conseguiu apreender informações sobre o histórico dos aprendizes e os classificou/agrupou a partir desse critério. Para confirmar essa hipótese, poder-se-ia submeter os participantes da pesquisa a um teste de nível padrão antes de extrair suas características do sinal de fala. Os padrões que emergiram da rede neural durante as simulações podem ser compreendidos como uma sugestão de nível de proficiência lingüística que a rede oferece para esses aprendizes. Na realidade, a rede neural pode vir a determinar nível de proficiência lingüística desde que se tenha uma amostra considerável dos dados e que a ferramenta de extração de características seja adequada (filtros autorregressivos).

Em suma, os resultados das simulações realizadas demonstraram a ferramenta de visualização, Matriz-U, como uma poderosa aliada no processo classificação de aprendizes quanto à competência/*performance* e, se vier a ser ajustada com o *input* adequado, quanto ao seu nível de proficiência lingüística.

A rede de Kohonen também pode vir a ser usada para avaliar o grau de distância que um grupo de aprendizes está do grupo de falantes nativos. De posse de medidas (distância euclidiana), podem-se comparar os segmentos realizados em

cada neurônio vencedor ou nos agrupamentos formados com os dados de falantes nativos e se ter uma idéia clara do percurso desenvolvimental feito durante a aquisição do padrão acentual do inglês pelos aprendizes.

Quanto ao dimensionamento ideal da rede, verificou-se que as zonas de fronteiras aparecem somente nos mapas de dimensão 10 x 10. Essas mesmas zonas na matriz-U correspondem a neurônios interpolados no mapa rotulado chegando a representar fenômenos atípicos no conjunto de dados. Melhor explicando, os neurônios da região de fronteira da matriz-U representam protótipos de categorias fuzzy.

A hipótese de que os coeficientes LP e os coeficientes cepstrais (MFC) condensam muita energia, frequência fundamental (F_0), banda de frequência foi confirmada. Essas informações estão intimamente ligadas à prosódia da fala e principalmente ao acento, fato este de fundamental importância para a codificação do *input* da rede. Com isso, a parametrização do sinal de fala por coeficientes LP e MFC foi eficiente na categorização dos falantes por características.

A rede neural analisou as propriedades (regularidades) estatísticas dos atributos relativos ao conjunto de dados dos falantes como um todo e organizou as informações por grau de similaridade em grupos bem definidos. Como os dados foram rotulados de forma diferenciada (*Multi-label analysis*), a emergência de certas categorias no mapa auto-organizável ficou bem mais fácil.

A hipótese que levou em consideração que a rede neural seria capaz de discriminar os coeficientes LP e MFC dos aprendizes e segregar os grupos de aprendizes usando como base os processos de transferência foi confirmada, tornando-se um sinal claro de indicação de que a propriedade auto-organizável da rede é um indicativo nível de proficiência linguística ou, no mínimo, de descrição de aquisição de L2.

Outra hipótese confirmada é a de que a codificação da rede com os coeficientes LP/MFC podem descartar o uso de memória associativa (ELMAN, 1990; 1991) e do contexto anterior e posterior (HONKELA, 1998) das palavras segmentadas. A escolha desta ferramenta de extração de características do sinal da fala reduziu consideravelmente o custo computacional do experimento. Em pesquisas anteriores a esta, hipotetizou-se que a extração das características do acento somente seria possível numa rede recorrente simples, pelo fato deste

elemento prosódico se desenrolar em uma escala temporal. Todavia, os filtros autorregressivos são dotados de memória embutida o que facilita a análise do sinal e conseqüentemente a quantização vetorial no treinamento das redes.

Até então, somente as vantagens da rede de Kohonen foram apresentadas. Sobre as desvantagens da rede, a primeira a ser revelada é que a rede de Kohonen é dotada de um grau elevado de subjetividade nos critérios de formação de categorias. Espera-se que mais trabalhos em lógica *fuzzy* e sistemas fuzzy venham a resolver esse problema. A segunda desvantagem é que a rede de Kohonen oferece resultados parciais e necessita da análise e interpretação de especialistas para dirimir questões levantadas pela formação de agrupamentos. Todavia, se esforços em conjunto forem empreendidos, todos os pontos considerados acima poderão vir a ser melhorados.

Espera-se que os resultados aqui expressos, aliados a outros já realizados dessa mesma natureza, possam contribuir de forma positiva com teorias de aquisição fonológica que se propõem a estudar esses fenômenos. Uma delas é a Teoria Neural da Linguagem (*Neural Theory of Language - NTL*) (Cf. FELDMAN, 2006).

Embora muitos cientistas acreditem que ainda seja precipitado formular teorias explícitas que associem aquisição, desenvolvimento e processamento de línguas à computação neural, a procura por uma Teoria Neural da Linguagem se materializa como uma alternativa para resolver problemas importantes não tratados anteriormente por outra metodologia ou teoria linguística mais tradicional, além de representar uma tentativa ambiciosa de aproximar teorias já consolidadas, com vistas a sugerir modelos mais evoluídos de processamento da linguagem.

Implicações e direcionamentos futuros da pesquisa

Quanto às implicações metodológicas desta pesquisa para a linha de pesquisa de linguística aplicada ao ensino de línguas, ressalta-se que os achados da presente pesquisa não possuem uma aplicação direta no ensino de língua estrangeira, porém sugerem a utilização do procedimento metodológico aqui apresentado a novas pesquisas que tenham como objetivo esclarecer as teorias sobre processamento de sinais como também avaliar o desempenho de aprendizes

em LE. Os resultados aqui apresentados não almejam orientar a composição de material didático para o ensino de prosódia de língua inglesa, todavia servem de consulta válida aos que buscam uma nova fonte de análise da aquisição fonológica do acento de língua inglesa sob uma perspectiva conexionista.

A despeito de se achar que as teorias conexionistas não podem ter uma implicação no tratamento de pronúncia de língua estrangeira, há uma possibilidade de retomada das teorias de base conexionista para a elaboração de novas metodologias do ensino de pronúncia de LE e para o tratamento do erro relacionado à computação do acento de palavra.

A possível razão para esses tipos de erro (computação do acento primário e secundário) se deve ao fato de que os aprendizes utilizam-se da mesma rede já marcada para o português quando tentam falar inglês. Sabe-se que utilizar a mesma rede é mais fácil a produção dos sons da LE como sendo sons da língua materna. Da mesma forma, a entonação, o acento e o tempo ficam prejudicados. A consequência é drástica: a palavra semelhante nas duas línguas é pronunciada de forma quase igual à L1, tornando-a uma palavra híbrida, mal pronunciada e irreconhecível na LE.

Por ser a marcação da rede feita através do reforço das sinapses, acredita-se que o método de ensino que se propõe a ensinar LE deve ter como base um mesmo modelo. Ora, se este modelo de marcação é o modelo pelo qual os seres humanos aprendem a falar sua primeira língua este deve ser o modelo mais adequado para o aprendizado de línguas, independentemente da idade de seus aprendizes. O ensino de LE, portanto, deve centrar-se na criação de novas conexões, na marcação e no reforço de novas redes neurais. Pelo exposto, o modelo de ensino de LE baseado em uma teoria conexionista deve valorizar a repetição como um processo natural de marcação da rede neural e criação de novos circuitos neurais.

Assim, lança-se aqui a proposta de elaboração de treinamentos para a superação das dificuldades encontradas ao expor os aprendizes a um *input* sintetizado artificialmente com características acústicas exageradas.

A despeito de a rede neural aqui utilizada ter-se revelado como uma ferramenta de grande valia na determinação e formação de grupos que possuem características similares, aconselha-se o seu aperfeiçoamento para uma possível

aplicação futura no auxílio de determinação de nível de proficiência lingüística em língua estrangeira em testes formais. Após novas rodadas e simulações com um conjunto de dados maior e com parâmetros mais estáveis, sinaliza-se a possibilidade de criação de um aplicativo para geração de mapas que façam um monitoramento do desenvolvimento do conhecimento dos aprendizes ao longo do tempo.

Posteriormente, pode-se proceder a testagem de outras abordagens estatísticas e de classificação, a saber, Árvores de Classificação e Regressão (CARTs - *Classification and Regression Trees*), Modelos de Markov e Redes MLP para a análise de um grupo maior de dados e com o objetivo de analisar outros elementos prosódicos (entonação, por exemplo). Caberia até se pensar no uso de redes recorrentes ou redes seqüenciais para analisar a transferência do conhecimento prosódico além da palavra em aprendizes brasileiros de língua inglesa em níveis distintos de proficiência, bem como simular a questão da dinamicidade da fala.

Outra providência seria proceder à extração de características prosódicas através de transformadas *wavelets* discretas (DTW) em lugar das aqui utilizadas. Seria bem vindo o teste de outros algoritmos de extração de parâmetros do sinal de fala e combinados de forma diferente (e.g. Mel+Cepstral).

Análises futuras deverão incluir outras fontes de características, tais como: energia, duração, taxa de cruzamento por zero, entre outros, de forma a possibilitar uma descrição mais precisa de certas propriedades lingüísticas que se pretende avaliar.

REFERÊNCIAS

American Council for the Teaching of Foreign Languages. 1983. **ACTFL Proficiency Guidelines**. Revised 1985. Hastings-on-Hudson, NY: ACTFL Materials Center. Disponível em: <<http://www.sil.org/lingualinks/LANGUAGELEARNING/OtherResources/ACTFLProficiencyGuidelines/contents.htm>>. Acesso em: 21 de ago. 2010.

AAMODT, S.; WANG, S. **Bem-Vindo ao Seu Cérebro**. São Paulo: Cultrix, 2009.

ADÁMEK, J. **Neural Networks Controlling Prosody of Czech language**. Magister thesis. Department of software engineering. Univerzita Karlova v Praze. Prague, 2002.

ARCHIBALD, J. A formal model of prosodic learning. **Second Language Research** v. 10, n. 3, p. 215-240, 1994.

_____. **Second language phonology**. Amsterdam/Philadelphia: John Benjamins, 1998.

ABAURRE, B.; GALVES, C. As diferenças rítmicas entre o português europeu e o português brasileiro: uma abordagem otimalista e minimalista. In: **D.E.L.TA.**, v. 14, n. 2, 1998, p. 377- 403.

ÁGOSTON, TÓTH. **Perspectives on the lexicon**. Doctorate Thesis. Hungria. 2005.

ALBANO, E., BARBOSA, P., GAMA-ROSSI, A., MADUREIRA e SILVA, A. A interface fonética-fonologia e a interação prosódia-segmentos. Estudos Lingüísticos XXVII, **Anais... XLV Seminário do Grupo de Estudos Lingüísticos de Estado de São Paulo - GEL'97**. Unicamp, Campinas - SP, p.135-143, 1998.

ALBANO, E. C. **O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro**. Campinas: Mercado de Letras, ALB, Fapesp, 2001.

BARRETO, G. A. **Redes Neurais Não-Supervisionadas para Processamento de Seqüências Temporais**. Dissertação de Mestrado, Depto. Engenharia Elétrica, Escola de Engenharia de São Carlos (EESC), Universidade de São Paulo, 1998.

_____. **Redes Neurais Não-Supervisionadas Temporais para Identificação e Controle de Sistemas Dinâmicos**. Tese de Doutorado, Depto. Engenharia Elétrica, Escola de Engenharia de São Carlos (EESC), Universidade de São Paulo, 2002.

_____. Time series prediction with the self-organizing map: A review. In: **Perspectives on Neural-Symbolic Integration**, Pascal Hitzler and Barbara Hammer (eds.), Spriger-Verlag, 2007.

BARBOSA, P. A. Syllable-Timing in Brazilian Portuguese: uma crítica a Roy Major. **D.E.L.TA.**, v. 16, n. 2, p.369-402, 2000.

_____. **Incursões em torno do ritmo da fala.** Tese de livre docência. Campinas: Pontes Editores, São Paulo, 2006.

BISOL, L. O acento e o pé binário. **Letras de Hoje.** Porto Alegre, v. 29, n. 4, p. 25-36, dez. 1994.

BOLINGER, D. **Intonation and its parts.** Melody in spoken English. London: Edward Arnold, 1985.

BITTENCOURT, G. **XII Ciclo de Cursos e Palestras da Engenharia Elétrica - XI CIPEEL – Conexionismo.** 2006

BOSCARIOLI, C. **Análise de agrupamentos baseada na topologia dos dados e em mapas auto-organizáveis.** Tese de doutorado. Escola Politécnica da Universidade de São Paulo, 2008.

BRAGA, D.; FREITAS, D.; FERREIRA, H. **Processamento Linguístico Aplicado à Síntese da Fala.** CLME, 2003. Disponível em: <<http://hfilipe.no.sapo.pt/publicacoes/CLME2003.pdf>>. Acesso em: 24 ago 2009.

BROWN, H. D. **Principles of language learning and teaching.** 4. ed. New York: Longman, 2000.

BYBEE, J. A view of phonology from a cognitive and functional perspective. **Cognitive Linguistics.** v. 4, p. 285-306. 1994.

_____. **Phonology and language use.** Cambridge: Cambridge, 2001.

_____. Cognitive process in gramaticalization. In THOMASELLO, M., editor, **The New Psychology of Language**, volume II. New Jersey: Lawrence Erlbaum Associates Inc.. 2002.

CASTRO, L. N. de. **Fundamentals on Neurocomputing. Basic concepts, algorithms, and applications.** Taylor e Francis Group, 2006.

CARICATTI, A. M.; WEIGANG, L. **Reconhecimento de Locutores em Língua Portuguesa com Modelos de Redes Neurais e Gaussianos.** Proceedings of V Brazilian Conference on Neural Networks - V Congresso Brasileiro de Redes Neurais p. 25-30, April 2-5, 2001

CIELO, C. A flexibilidade do paradigma conexcionista. **Letras de Hoje.** Porto Alegre. v. 33, n.2, p. 43-49, jun. 1998.

CAGLIARI, L. C. Da importância da prosódia na descrição de fatos gramaticais. In ILARI, R. (org.) **Gramática do Português Falado.** v. II: Níveis de Análise Lingüística. Campinas, SP: Editora da UNICAMP, 1993.

_____. O acento: o problema do problema. In: Estudos Lingüísticos. XXV **Anais...** Seminários do GEL. Rio Preto: São Paulo, 1995.

CALLOU, Dinah; LEITE, Yonne. **Iniciação à fonética e à fonologia**. 9. ed. Rio de Janeiro: Zahar, 2003.

CHOMSKY, N. **Aspectos da teoria da sintaxe**. Lançado como Saussure Jakobson Hjelmslev Chomsky. Textos Seleccionados, edited by V. Civita. Sao Paulo: Abril S.A. Cultural, 1985.

CHUN, Dorothy M. Signal analysis software for teaching discourse intonation. **Language Learning & Technology**. volume 2, number 1, pp. 74-93, july 1998. Disponível em <<http://llt.msu.edu/vol2num1/article4/>> Acesso em: 23 mar. 2010.

COLLISCHONN, G. Acento secundário em português. **Letras de Hoje**. Porto Alegre, v. 29, n. 4, p. 43–53, dez. 1994.

_____. O Acento em Português. In BISOL, L. (org.) **Introdução a estudos de fonologia do português brasileiro**. 2 ed. Porto Alegre: EDIPUCRS, 1999. p. 125-158.

CORDER, S.P. The significance of learners' errors. In: RICHARDS, J. C. (org.) **Error analysis: perspectives on second language acquisition**. London: Longman, 1974.

CRYSTAL, David. **Dicionário de fonética e fonologia**. 2 ed. Rio de Janeiro: Zahar, 2000.

CRISTOFARO-SILVA, T. **Fonética e fonologia do português**. São Paulo: Contexto, 2001.

CRUZ, Magnus Alencar da. **Avaliação de Redes Neurais Competitivas em tarefas de Quantização Vetorial: Um Estudo Comparativo**. Programa de Pós-Graduação em Engenharia de Teleinformática (PPGETI), Universidade Federal do Ceará (UFC), 2007.

CUNHA, Celso; CINTRA, Lindley. **Nova gramática do português contemporâneo**. 2. ed. Rio de Janeiro: Nova Fronteira, 1985.

DAVIES, D. L.; BOULDIN, D.W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 1, n. 2, p. 95–104, 1979.

DELLER, Jr. R.; PROAKIS, J. G.; HANSEN, J. H. L. **Discrete-time Processing of Speech Signals**. Englewood Cliffs, NJ: Prentice-Hall, 2000.

DOMINEY, P. F.; BOUCHE, J-D. Developmental Stages of Perception and Language Acquisition in a Perceptually Grounded Robot. In Berthouze, L., Kozima, H., Prince, C. G., Sandini, G., Stojanov, G., Metta, G., and Balkenius, C. (Eds.) **Proceedings of the Fourth International Workshop on Epigenetic Robotics Lund University**

Cognitive Studies, 117. Disponível em: <<http://www.lucs.lu.se/LUCS/117/dominey.pdf>>. Acesso em: 05 ago. 2008.

DOUGHTY, CATHERINE J. and MICHAEL H. LONG. The Scope of Inquiry and Goals of SLA. **The Handbook of Second Language Acquisition**. Doughty, Catherine J. and Michael H. Long (eds). Blackwell Publishing, 2005. Blackwell Reference Online. Disponível em: <http://www.blackwellreference.com/subscriber/tocnode?id=g9781405132817_chunk_g97814051328172>. Acesso em: 14 Nov. 2007

ELLIS, R. The origins of interlanguage. **Applied Linguistics** 3. v. 3, p. 207 -223, 1982.

ELLIS, N. Constructions, Chunking, and Connectionism: The emergence of second language structure. In: DOUGHTY, C. J.; LONG, M. H. **The handbook of second language acquisition**. Blackwell Publishing, 2003.

ELMAN, J. Distributed representations, simple recurrent networks, and grammatical structure. **Machine Learning**. v.7, p. 195–225, 1991.

FELDMAN, J. **From Molecule to Metaphor**. A Neural Theory of Language. MIT Press, 2006.

FERREIRO, Emília; TEBEROSKY, Ana. **Psicogênese da Língua escrita**. Porto Alegre: Artes Médicas, 1985.

FIGUEIREDO, F. J. Q. **Aprendendo com os erros**: uma perspectiva comunicativa do ensino de línguas. Goiânia: Editora da UFG, 1997.

FLORES, R. D.; NUNES, R. C.; MARTINS, J. B. S. **Estimação e Cálculo de Coeficientes de Predição Linear em Sinais de Voz**. Jundiaí: CRICTE, 2006.

FLEGE, J., BIRDSOING, D., BIALYSTOK, E., MACK, M., SUNG, H., & TSUKADA, K. Degree of foreign accent in English sentences produced by Korean children and adults. **Journal of Phonetics**. v. 34, p. 153-175, 2006.

FODOR, J. e PYLYSHYN, Z. **Connectionism and cognitive architecture**: a critical analysis. *Cognition*. Cambridge, Mass., M.I.T., v. 28, p. 3-71, 1988.

FODOR, J. D. Psycholinguistics cannot escape prosody. **Speech Prosody 2002 electronic proceedings**. 2002. Disponível em: < <http://www.lpl.univ-aix.fr/sp2002/pdf/fodor.pdf>. Acesso em: 23 set. 2007.

FURUI, S. **Digital speech processing, synthesis and recognition**. 2. Ed. New York. Marcel Dekker, 2001.

GASS, S.; SELINKER, L. **Language transfer in language learning**. John Benjamins: Philadelphia, 1993, p. 1-17.

GAUTHIER, B; SHI, R; XU, YI. Simulating the acquisition of lexical tones from continuous dynamic input. **Journal of Acoustical Society of America**. n.121, 5, May 2007.

_____. Learning Prosodic Focus from Continuous Speech Input: A Neural Network Exploration. **Language Learning and Development**. v. 5, p. 94–114, 2009.
GOVENDER, N.; KUUN,C.; ZIMU, V.; BARNARD, E.; DAVEL, M. **Computational models of prosody in the Nguni languages**. Disponível em: <<http://www.meraka.org.za/pubs/govender06computational.pdf>>. Acesso em: 27 jul. 2009.

HALLIDAY, M. A. K. Categories of the theory of grammar. In: HALLIDAY: **System and function in language**, edited by G. R. Kress. p. 52-72, Oxford: OUP, 1961.

HAYES, B. **A metrical theory of stress rules**. Tese de Doutorado. Universidade de Yale, 1981.

HAYES, B. **Metrical stress theory: principles and case studies**. Texto não publicado, UCLA, 1992.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**. 2nd edition, Prentice-Hall, 1999.

HIRST, D.J.; DI CRISTO, A. (orgs.). **Intonation Systems: A survey of Twenty Languages**. Cambridge, Cambridge University Press, 1998.

HOGG, R.; McCULLY, C. B. **Metrical phonology – a coursebook**. Cambridge: Cambridge University Press, 1987.

HONKELA, T. **Self-organizing maps in natural language processing**. Doctorate thesis. University of Helsinki.1997.

HULST, H. VAN DER. **Cognitive Phonology**. University of Connecticut. s/d. Disponível em: <<http://odur.let.rug.nl/~koster/DenBesten/VanderHulst.pdf>>. Acesso em: 14 fev 2007.

JAIN, A K.; DUBES, R. C. **Algorithms for Clustering**. Cap. IV - Cluster Validity, pp. 143-222. Prentice Hall, 1988.

JAMES, Carl. **Errors in language learning and use: exploring error analysis**. Harlow, UK: Longman, 1998.

JENKINS, J. Research in teaching pronunciation and intonation. **Annual Review of Applied Linguistics** (2004), 24, 2004.

KATAMBA, F. **Morphology**. England: Palgrave, 1993.

KOCH, I. G. V.; CUNHA-LIMA, M. L. Do cognitivismo ao sociocognitivismo. In: MUSSALIM, F.; BENTES A. C. (orgs.). **Introdução à Lingüística: fundamentos epistemológicos**. São Paulo: Cortez, 2004.

KELSO, J. A. Scott. **Dynamic Patterns: The Self-organization of Brain and Behavior.** MIT Press, 1995.

KOHONEN, T. The 'Neural' Phonetic Typewriter. **Computer**, v. 21, n. 3, p. 11-22, Mar. 1988.

KOHONEN, T. The self-organizing map. **Neurocomputing**. v. 21, p. 1–6, 1998.

KOHONEN, T. **Self-organizing Maps**. 3. ed. Berlin: Springer, 2001.

KREIDLER, C. W. **The pronunciation of English: A course book in phonology.** Oxford: Blackwell, 1989.

KRASHEN, S. D. **Second language acquisition and second language learning.** Oxford, Pergamon, 1981.

KUHL, P. K.; WILLIAMS, K. A.; LACERDA, F.; STEVENS, K. N.; LINDBLOM, B. Linguistic experience alters phonetic perception in infants by 6 months of age. **Science**, 31 January, v. 255, p. 606-608, 1992.

KUHL, P. K. A new view of language acquisition. **Proceedings of the National Academy of Science**, n. 97, p. 11850-11857, 2000.

LADEFOGED, Peter. **A course in phonetics**. 3. ed. USA: Hartcourt Brace, 1993.

LEE, Seung Hwa. A regra do acento do português. **Letras de Hoje**, Porto Alegre. v. 29, n. 4, p. 37-42, dez. 1994.

LEVIS, J. COMPUTER TECHNOLOGY IN TEACHING AND RESEARCHING PRONUNCIATION. **Annual Review of Applied Linguistics** (2007) 27, 184–202. CUP: U.S.A. 2008

LIBERMAN, M.; PRINCE, A. On stress and linguistic rhythm. **Linguistic Inquiry**. v. 8 n. 2, p. 249-336. 1977.

LANGACKER. An introduction to cognitive grammar. **Cognitive science**, n. 10, p. 1-40, 1986.

LILJENCRANTS, J. Speech signal processing. In: HARDCASTLE, William J.; LAVER, John. **The Handbook of Phonetic Sciences**. Cambridge: Blackwell Handbooks in Linguistics, 1997.

LI, P.; FARKAS, I.; MacWHINNEY, B. Early lexical development in a self organizing neural network. **Neural networks**. v. 17, p. 1345 – 1362, 2004.

LIMA, Amaro de A. **Análises comparativas em sistemas de reconhecimento de voz.** Tese de doutorado. Universidade Federal do Rio de Janeiro. 2000.

LISKIN-GASPARRO, Judith E. The ACTFL Proficiency Guidelines: A Historical Perspective. In: Theodore V. Higgs, ed., **Teaching for Proficiency: The Organizing Principle**. Lincolnwood, IL: National Textbook Company, p. 11-42, 1984.

LYONS, John. **Introdução à lingüística teórica**. São Paulo: Companhia Editora Nacional, 1987.

MAJOR, R. C. Stress and rhythm in Brazilian Portuguese. In: **Language**, v. 61, n. 2, p. 259-282, 1985.

MACIEL, F. J.; ROTHE-NEVES, R. **Investigações experimentais da entonação no português brasileiro: uma revisão de literatura**. In: Laboratório de Fonética da UFMG. (Org.). IX Congresso Nacional de Fonética e Fonologia e III Congresso Internacional de Fonética e Fonologia. Belo Horizonte, 2006.

MACWHINNEY, Brian. Models of emergence of the language. **Annual Review of Psychology**. n. 49, p. 199--227, 1998.

MACEDO, A. C. P. S. de. Categorization and Metaphor: An overview. **Revista do Gelne**. v. 6, n. 2, p. 7-16. 2004.

_____. Cognição e lingüística. In: Ana Cristina Pelosi Silva de Macedo; Heloísa Pedroso de Moraes Feltes; Emília Maria Peixoto Farias. (Org.). **Cognição e lingüística: territórios, mapeamentos e percursos**. Porto Alegre e Caxias do Sul: EDPU CRS e EDUCS, 2008, v. 01, p. 09-37.

MAFRA, A. **Reconhecimento automático de locutor em modo independente de texto por Self-Organizing Maps**. Dissertação de Mestrado – Escola Politécnica da Universidade de São Paulo, 2002.

MAIRS, J. L. Stress assignment in interlanguage phonology: an analysis of the stress system of Spanish speakers learning English. In: Gass, M & Schatther, J. (orgs.) **Linguistic Perspectives on Second Language Acquisition**. Cambridge, USA: Cambridge University Press, 1989.

MASSINI-CAGLIARI, Gladis. **Acento e ritmo**. São Paulo: Contexto, 1992.

MATEUS, Maria Helena Mira; FALÉ, Isabel e FREITAS, Maria João. **Fonética e Fonologia do Português**. Lisboa: Universidade Aberta. Parte III, 2005.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophysics*. v. 5, n. 115-133. 1943.

MEDLER, D. A. **A brief history of connectionism**, 1998. Disponível em: <<http://www.blutner.de/NeuralNets/Texts/Medler.pdf>.> Acesso em: 24 jun. 2009.

MOLHOLT, G. Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, 22(1), 91-111, 1988(1988).

MORAES, João Antônio de. Em torno da entonação: alguns problemas teóricos. In: **Cultura lingüística**. v. 1, p. 63-78, 1982.

_____. Acentuação lexical e acentuação frasal em português: um estudo acústico-perceptivo. In: **Estudos Lingüísticos e Literários** 17, Universidade Federal da Bahia, p. 39-57, 1995.

_____. F0 declination in Brazilian Portuguese in read and spontaneous speech. In: 14th International Congress of Phonetic Sciences. 1999, San Francisco. **Proceedings of the 14th International Congress of Phonetic Sciences**, p. 2323-2326, 1999.

MÜLLER, DANIEL NEHME. **COMFALA - Modelo Computacional do Processo de Compreensão da Fala**. Tese de doutorado. Porto Alegre: PPGC da UFRGS, 2006.

NESPOR, Marina; VOGEL, Irene. **Prosodic phonology**. Dordrecht; Foris, 1986.

NEGISHI, M. **Everything linguists wanted to know about connectionism but were afraid to ask**. 1998. Disponível em: <citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.49.73&rep>. Acesso em: 13 mai. 2008.

ODLIN, Terence. **Language transfer: cross-linguistic influence in language learning**. 3 ed. Cambridge: Cambridge University Press, 1989.

ORSINI, M. T. Análise entonacional das construções de tópico. In: IX Congresso Nacional de Lingüística e Filologia, IX, 2005, Rio de Janeiro. **Anais... IX Congresso Nacional de Lingüística e Filologia**, p. 132-139. 2005.

PAKHIRA, M. K.; BANDYOPADHYAY, S.; MAULIK, U. Validity index for crisp and fuzzy clusters. **Pattern Recognition**, v. 37, n. 3, p. 487-501, 2004.

PINKER, S. Language Acquisition. Capítulo a ser publicado no livro organizado por L. R. Gleitman, M. Liberman, and D. N. Osherson (Eds.), **An Invitation to Cognitive Science**. 2 ed. v. 1: Language. Cambridge, MA: MIT Press. Disponível em: <<http://users.ecs.soton.ac.uk/harnad/Papers/Py104/pinker.langacq.html>>. Acesso em: 10 set. 2006.

PLUNKETT, K. **Language Acquisition and Connectionism**. London: Psychology Press, 1998.

POERSCH, J. M. O paradigma simbólico é demasiadamente rígido para explicar determinados problemas de aquisição lingüística. **Letras de Hoje**. Porto Alegre. v. 33, n. 2, p 37-42, jun. 1998.

_____. A leitura como fonte de saber lingüístico: processos cognitivos. **Letras de Hoje**. Porto Alegre, v. 36, n. 3, p. 401- 407, setembro, 2001.

_____. A new paradigm for learning language: connectionism artificial intelligence. **Linguagem & Ensino**, v. 8, n. 1, p. 165-183, 2005.

_____. Simulações conexionistas: a inteligência artificial moderna. **Linguagem em (Dis)curso**. Tubarão, v. 4, n. 2, p. 441-458, jan./jun. 2004.

POERSCH, J. M; ROSSAS, A. A (Orgs.). **Processamento da linguagem e conexionismo**. EDUNISC: Rio Grande do Sul, 2007.

PRATOR, Jr., Clifford H.; ROBINETT, Betty Wallace. **Manual of American English pronunciation**. 4. ed, Orlando: Harcourt Brace & Company, 1985.

PRAXEDES FILHO, P. H. L. **A corpora based study of the development of EFL Brazilian learners' interlanguage from simplification to complexification in the light of systemic-functional grammar**. Tese de doutorado. UFSC, 2007.

ROSENBLATT, F. **Principles of neural dynamics**. New York: Spartan, 1962.

ROACH, P. **English phonetics and phonology**. 2 ed. Cambridge: Cambridge University Press, 1991.

RUMELHART, D. E. e MCCLELLAND, J. On learning the past tenses of English verbs. In: MCCLELLAND, J. e RUMELHART, D., and the PDP group. **Parallel distributed processing: explorations in the microstructure of cognition**. v. 2, 1986, p. 216-271. Cambridge, MA: MIT.

RUMELHART, D.E., HINTON, G.E. e WILLIAMS, R.J. Learning internal representations by error propagation. In: RUMELHART, D. e MCCLELLAND, J. **Parallel distributed processing: explorations in the microstructure of cognition**. Cambridge, MA: MIT, v. 1, p. 318-362, 1986.

RITTER, H; KOHONEN, T. Learning semantotopic maps from context. **Proceedings of IJCNN 90**, Washington D.C., I., 1990.

SCARPA, E. M. **Estudos de Prosódia**. Campinas: Editora da Unicamp, 1999.

SCLIAR-CABRAL, Leonor. Representação mental das unidades de processamento? In: III Congresso Internacional da ABRALIN, Rio de Janeiro, 13-15/03/03, 2003, Rio de Janeiro. **Anais...** III Congresso Internacional da ABRALIN. Rio de Janeiro: www.abralin.org. p. 39-43. 2003.

SELINKER, L. Interlanguage. In: RICHARDS, J. C. (org.) **Error analysis: perspectives on second language acquisition**. London: Longman, 1974.

SHI, R., WERKER, J.F., MORGAN, J.L. Newborn infants' sensitivity to perceptual cues to lexical and grammatical words, **Cognition**, v. 72, Issue 2, 1999.

SHAFER, R. W. Homomorphic systems and cepstrum analysis of speech. In: BENESTY, J.; SONDHI, M. M.; HUANG, Y. **Springer Handbook of speech processing**. [S.1.]: Springer, p. 161-180, 2008.

SOUZA JR. A. H. de. **Avaliação de rede neurais auto-organizáveis para reconhecimento de voz em sistemas embarcados**. Dissertação de mestrado.

Fortaleza. Programa de Pós-Graduação em Engenharia de Teleinformática, UFC. 2009.

SILVA, M. A. S. da. **Mapas auto-organizáveis na análise exploratória de dados geoespaciais multivariados**. Dissertação de mestrado em computação aplicada. INPE, 2004.

SILVA, Ana Cristina Cunha da. **A produção e a percepção do acento em pares mínimos de língua inglesa por aprendizes brasileiros**. Dissertação (Mestrado em Lingüística) – UFC, Fortaleza, 2005.

_____. Um estudo sobre a redução vocálica em pares mínimos de língua inglesa por aprendizes brasileiros. In: ARAGÃO, M. do S. S. de (org.). **CD ROM - GT de Fonética e Fonologia**. Anpoll 2008. 2009.

STEINBERG, Martha. **Pronúncia do inglês norte-americano**. São Paulo: Ática, 1985.

TAO, J-H; CAI, L-H. **The Context-Based Method Of Creating Chinese Prosodic Model**. ISSPR, 1998.

TAYLOR, L.; JONES, N. **Cambridge ESOL exams and the Common European Framework of Reference (CEFR)**. Research Notes 24/1. 2006.

TEIXEIRA, J. de F. **Mentes e máquinas: uma introdução à ciência cognitiva**. Porto Alegre: Artes Médicas, 1998.

THOMAS, M. Assessment of L2 proficiency in second language acquisition research. **Language learning**. n. 44, p. 307-336, 1994.

ULTSCH, Alfred. **U*-Matrix: a Tool to visualize Clusters in high dimensional Data**. University of Marburg, Department of Computer Science, Technical Report, Nr. 36, December 2003.

VARELA, F. J. Conocer. **Las ciencias cognitivas: tendencias y perspectivas**. Cartografía de las ideas actuales. 2ed. Barcelona: Gedisa, 1998.

VAINIO. M. **Artificial Neural Network Based Prosody Models for Finnish Text-to-Speech Synthesis**. Doctorate thesis, 2001.

VAISSIÈRE, Jacqueline. **Language-independent prosodic features**. In: A. Cutler e D. R. Ladd (eds.), p. 53-66, 1983.

VESANTO, Juha; HIMBERG, Johan; ALHONIEMI, Esa; PARHANKANGAS, Juha. **Self-organizing map in Matlab: The SOM Toolbox**. SOM Toolbox Team. Helsinki University of Technology, 2000.

WATKINS, M. A. **Variability in vowel reduction by Brazilian speakers of English**. Tese de Doutorado Não Publicada, Universidade Federal de Santa Catarina,

Florianópolis, 2001.

XAVIER, M. F.; MATEUS, M. H. (orgs.) **Dicionário de termos lingüísticos**. v. I. Lisboa: Cosmos, 1990.

ZIMMER, Márcia Cristina. **A transferência do conhecimento fonético-fonológico do português brasileiro (L1) para o inglês (L2) na recodificação leitora: uma abordagem conexionista**. Tese de doutorado. PUCRS, Porto Alegre, 2004. In: FINGER, I.; MATZENAUER, C. (Compiladoras). TEP (Textos em Psicolingüística) [CD-ROM]. 1. ed. Pelotas: Educat, 2006.

ANEXOS

ANEXO A – PROGRAMAS DA REDE NEURAL

Código do Matlab nº 1

```
% Aplicacao da rede SOM na visualizacao de
% elocucoes de palavras isoladas em ingles

% Autores: Guilherme A. Barreto & Amauri Holanda
% data: 12/12/2008

clear; clc; close all;

% Read data matrix
sD = som_read_data('lpc10object.dat');

% Normalize input features
sD = som_normalize(sD,'var');
%sD = som_normalize(sD,'range');

% Create (specify and train) a map
sMap = som_make(sD,'msize',[10 10],'training',[50 200]);

% The best matching unit (BMU or winner) of each sample is found from the map, and the
% species label is given to the map unit.
sMap = som_autolabel(sMap,sD,'vote');

% Create a U-matrix with no labels
som_show(sMap,'umat','all','empty','Labels')

% Add labels to the U-matrix
som_show_add('label',sMap,'Textsize',8,'TextColor','r','Subplot',2)

% Clustering of the Map
[c,p,err,ind] = kmeans_clusters(sMap, 10); % find at most 10 clusters
figure; plot(1:length(ind),ind,'x-')
[dummy,i] = min(ind)
cl = p{i};

figure;
som_show(sMap,'color',{p{i},sprintf('%d clusters',i)}); % visualize
colormap(jet(i)), som_recolorbar % change colormap
```

Código do Matlab nº 2

% Rotina para concatenacao de todos os arquivos de um falante

```
% load separate201r.txt
% X=separate201r;
% load separate202r.txt
% X=[X; separate202r];
% load separate203r.txt
% X=[X; separate203r];
% load separate204r.txt
% X=[X; separate204r];
% load separate205r.txt
% X=[X; separate205r];
% load separate206r.txt
% X=[X; separate206r];
% load separate207r.txt
% X=[X; separate207r];
% load separate208r.txt
% X=[X; separate208r];
% load separate209r.txt
% X=[X; separate209r];
% load separate210r.txt
% X=[X; separate210r];
% load separate401r.txt
% X=[X; separate401r];
% load separate402r.txt
% X=[X; separate402r];
% load separate403r.txt
% X=[X; separate403r];
% load separate404r.txt
% X=[X; separate404r];
% load separate405r.txt
% X=[X; separate405r];
% load separate406r.txt
% X=[X; separate406r];
% load separate407r.txt
% X=[X; separate407r];
% load separate408r.txt
% X=[X; separate408r];
% load separate409r.txt
% X=[X; separate409r];
% load separate410r.txt
% X=[X; separate410r];
% load separate601r.txt
% X=[X; separate601r];
% load separate602r.txt
% X=[X; separate602r];
% load separate603r.txt
% X=[X; separate603r];
% load separate604r.txt
% X=[X; separate604r];
% load separate605r.txt
% X=[X; separate605r];
% load separate606r.txt
% X=[X; separate606r];
% load separate607r.txt
% X=[X; separate607r];
```

```
% load separate608r.txt
% X=[X; separate608r];
load objectnative2r.txt
X= objectnative2r;
load separate1.txt
X=[X; separate1];

save juntos.txt X -ascii
```

Código do Matlab nº3

```
% Rotina para adicionar coluna com rotulo do falante
clear; clc;

load objectnative2.txt % Carrega arquivo do falante
x=objectnative2; % Troca de nome para facilitar
r=903; % Rotulo numerico do falante (trocar sempre)

[nL nC]=size(x); % Determina tamanho da matriz de dados
rot=repmat(r,nL,1); % Cria coluna de rotulos iguais

x=[x rot]; % Acrescenta coluna de rotulos

save objectnative2r.txt x -ascii
```

ANEXO B - JANELA DE COMANDO – SOM TOOLBOX NO MATLAB

Comand Window for C:\Users\Ana Cristina\Documents\FRASES\15_we will take appropriate action once the investigation is over\appropriate_adj\lpc 10

data read ok

repet =

1

Initialization...

Training using batch algorithm...

Rough training phase...

Training: 0/ 1 s

Training: 0/ 2 s

Training: 0/ 2 s

Training: 0/ 2 s

.

.

.

Finetuning phase...

Training: 0/ 3 s

Training: 0/ 3 s

Training: 0/ 4 s

Training: 0/ 5 s

Training: 0/ 5 s

.

.

.

Training: 5/ 5 s

Training: 5/ 5 s

Training: 5/ 5 s

Training: 5/ 5 s

Training: 5/ 5 s

Final quantization error: 0.417

Final topographic error: 0.024

ANEXO C – COMMON EUROPEAN FRAMEWORK OF REFERENCE⁶²

Quadro de habilidades por nível

ALTE (Association of Language Testers in Europe) Can Do Statements: overall general ability			
CEFR LEVELS	Listening/Speaking	Reading	Writing
C2	CAN advise on or talk about complex or sensitive issues, understanding colloquial references and dealing confidently with hostile questions.	CAN understand documents, correspondence and reports, including the finer points of complex texts.	CAN write letters on any subject and full notes of meetings or seminars with good expression and accuracy.
C1	CAN contribute effectively to meetings and seminars within own area of work or keep up a casual conversation with a good degree of fluency, coping with abstract expressions.	CAN read quickly enough to cope with an academic course, to read the media for information or to understand non-standard correspondence.	CAN prepare/draft professional correspondence, take reasonably accurate notes in meetings or write an essay which shows an ability to communicate.
B2	CAN follow or give a talk on a familiar topic or keep up a conversation on a fairly wide range of topics.	CAN scan texts for relevant information, and understand detailed instructions or advice.	CAN make notes while someone is talking or write a letter including non-standard requests.
B1	CAN express opinions on abstract/cultural matters in a limited way or offer advice within a known area, and understand instructions or public announcements.	CAN understand routine information and articles, and the general meaning of non-routine information within a familiar area.	CAN write letters or make notes on familiar or predictable matters.
A2	CAN express simple opinions or requirements in a familiar context.	CAN understand straightforward information within a known area, such as on products and signs and simple textbooks or reports on familiar matters.	CAN complete forms and write short simple letters or postcards related to personal information.
A1	CAN understand basic instructions or take part in a basic factual conversation on a predictable topic.	CAN understand basic notices, instructions or information.	CAN complete basic forms, and write notes including times, dates and places.

⁶² Quadro extraído de <http://www.cambridgeesol.org/exams/exams-info/cefr.html>

APÊNDICES

APÊNDICE A - CORPUS EMPREGADO NA PESQUISA

Part I – Noun/Verb

1. I **object** to going to a bar.
2. It's a small metal **object**.
3. I want you to **conduct** this research.
4. The senator's **conduct** is being investigated.
5. She needs to be in **conflict** with her parents.
6. Different software often **conflict** in my computer.
7. They have a kissing **contest** every year.
8. Politicians will **contest** the election.
9. Camels were meant to live in the **desert**.
10. Never let your confidence **desert** you.

Part II – Adjective/Verb⁶³

11. We need to **separate** the good apples from the bad ones.
12. Let's put them in **separate** baskets.
13. She's a passionate **advocate** of natural birth.
14. The extremists openly **advocate** violence.
15. We will take **appropriate** action once the investigation is over.
16. Politicians **appropriate** \$1 billion for anti-drug campaigns.
17. The garage gave me an **estimate** of the cost of repairing my car.
18. They **estimate** it will cost at least R\$ 20,00.
19. I didn't **elaborate** the plan myself.
20. He's got an **elaborate** tattoo of an eagle.

Part III - Adjective/Noun⁶⁴

Chestnuts have a high water content.

I'm content with my job.

The company spent an aggregate of \$2 million on the product.

He has a huge aggregate income and investment.

Marines protect the U.S. Embassy compound.

"Ice cream" is a compound noun.

My childhood memories are the subject of my first book.

When you are in a foreign country, you are subject to its laws.

Some children find it difficult to talk to adult people.

You need to deal with your problems in an adult way.

63 As frases 13 e 14 foram extraídas do CD-ROM do LONGMAN ADVANCED AMERICAN DICTIONARY (2001) enquanto que as frases 17 e 18 foram extraídas do PRATOR, Jr., Clifford H. & ROBINETT, Betty Wallace. Manual of American English Pronunciation. 4 ed, Orlando: Harcourt Brace, 1985. O restante das frases foram criadas pela autora desta tese.

64 As frases 21, 25, 28 e 30 foram extraídas do CD-ROM do LONGMAN ADVANCED AMERICAN DICTIONARY (2001). O restante das frases foi criado pela autora desta pesquisa.

APÊNDICE B - FICHA DE INFORMAÇÕES DO PARTICIPANTE

Nome: _____ Sexo: _____ Idade: _____
Semestre: _____ Horário: _____ Sala: _____ Professor(a): _____
Tel de contato (outro que não seja celular): _____
Melhor horário para um possível contato: _____

Por favor, responda às perguntas abaixo com o máximo de veracidade a fim de contribuir com uma pesquisa acadêmica: (Marque com um X os campos que forem convenientes)

1. É Brasileiro (a)?

Sim _____ Não _____

2. Qual região passou maior parte da Infância e Adolescência?

Nordeste _____, Norte _____, Sudeste _____, Sul _____, Centro-Oeste _____
Que cidade e estado mais exatamente?

3. Qual o seu nível de escolaridade?

2º grau cursando _____ 2º grau completo _____ Aluno da graduação _____
Graduado _____ Aluno de Pós-Graduação _____ Pós-Graduado _____

4. Caso esteja no 1º semestre ou foi aprovado por teste de nível, esta é a primeira vez que estuda a Língua Inglesa? (Não leve em consideração a disciplina escolar ou Inglês Instrumental)

Sim _____ Não _____

5. Se já estudou, por quanto tempo?

_____ anos, _____ meses, _____ semanas

6. Era uma instituição pública ou privada?

Pública _____ Privada _____

7. Já morou ou passou algum período em algum país de Língua Inglesa?

Sim _____ Não _____

8. Se a resposta anterior foi positiva, quanto tempo?

_____ anos, _____ meses, _____ semanas Qual país?

9. Qual foi o propósito da viagem?

Turismo _____ Trabalho _____ Estudo _____ Outro _____

10. Quanto tempo dedica por dia ao estudo da Língua Inglesa? (Não contar com o tempo de aula)

_____ horas, _____ minutos

11. Fala outra língua além do Português? (Não incluir Inglês)

Não _____ Sim _____ Qual? _____

APÊNDICE C - LISTA DE ITENS LEXICAIS AMBÍGUOS (SUBSTANTIVOS, ADJETIVOS E VERBOS) USADOS NOS TESTES DE PRODUÇÃO

Itens Lexicais	Substantivo (Noun)	Verbo (Verb)	Adjetivo (Adjective)
1. Ab.stract*	/ˈæbstrækt/	/əb'strækt/	/ˈæbstræk, əb'strækt/
2. Ap.prox.i.mate	..	/ə'praksə,meɪt/	/ə'praksəmit/
3. Ap.pro.pri.ate	..	/ə'proupri,eɪt/	/ə'prouprɪt/
4. A.dult	/ə'dʌlt, æ'dʌlt/	..	/ə'dʌlt, æ'dʌlt/
5. Ad.vo.cate	/ˈædvəkət/	/ˈædvə,kəɪt/	..
6. Ag.gre.gate	/ˈægriɡɪt/	/ˈægri,geɪt/	/ˈægriɡɪt/
7. Com.pound*	/ˈkampaʊnd/	/ˈkəm'paʊnd/	/ˈkampaʊnd, kəm'paʊnd/
8. Con.duct	/ˈkɒndʌkt/	/kən'dʌkt/	..
9. Con.flict	/ˈkɒnflɪkt/	/kən'flɪkt/	..
10. Con.test	/ˈkɒntest/	/kən'test/	..
11. Con.trast	/ˈkɒntræst/	/kən,træst/	..
12. Con.tract	/ˈkɒntrækt/	/kən'trækt/	..
13. Choco.late	/ˈtʃɒklɪt/
14. Con.tent	/ˈkɒntent/	/kən'tent/	/kən'tent/
15. De.sert	/ˈdezɜrt/	/dɪ'zɜrt/	..
16. E.lab.o.rate	..	/ɪ'læbə,reɪt/	/ɪ'læbrɪt/
17. Es.ti.mate	/ˈestəmit/	/ˈestə,meɪt/	..
18. Grad.u.ate	/ˈgrædʒuɪt/	/ˈgrædʒu,eɪt/	/ˈgrædʒuɪt/
19. In.crease	/ˈɪnkris, ˈɪŋkris/	/ɪn'kris/	..
20. In.sult	/ˈɪnsʌlt/	/ɪn'sʌlt/	..
21. Mod.e.rate	/ˈmɒdəreɪt/	/ˈmɒdə,reɪt/	/ˈmɒdəreɪt/
22. Ob.ject	/ˈɒbdʒɪkt, ˈɒbdʒekt/	/əb'dʒekt/	..
23. Pro.duce	/ˈprɒdʌs, ˈpraʊdʌs/	/prə'dʌs/	..
24. Per.mit	/ˈpɜrmit/	/pɜr'mɪt/	..
25. Pre.sent	/ˈprezənt/	/pri'zent/	/ˈprezənt/
26. Pro.gress	/ˈprɒgrəs, praʊgrəs/
27. Pro.ject	/ˈprɒdʒekt, ˈpraʒɪkt/	/prə'dʒekt/	..
28. Pro.test*	/ˈpraʊtest/	/prə'test, ˈpraʊtest/	..
29. Pre.cip.i.tate*	/prɪ'sɪpə,teɪt, prɪ'sɪpətɪt/	/prɪ'sɪpə,teɪt/	/prɪ'sɪpətɪt/
30. Re.cord	/ˈrekɜrd/	/rɪ'kɜrd/	/ˈrekɜrd/
31. Sep.a.rate	..	/ˈsepə,reɪt/	/ˈsepɪt/
32. Sus.pect	/ˈsʌspekt/	/sə'spekt/	/ˈsʌspekt/
33. Sub.ject	/ˈsʌbdʒɪkt/	/səb'dʒekt/	/ˈsʌbdʒɪkt/

* itens lexicais que possuem duas pronúncias diferentes quanto ao acento numa dada categoria lexical.

APÊNDICE D - SUMÁRIO DOS PARTICIPANTES

Identificação do Falante	Idade	Grau de escolaridade	Quant. de horas-aula na escola (média ponderada)	Quantidade de horas-aula acumuladas no curso extensão em língua inglesa atual	Tempo somado de estudo de língua inglesa em outros cursos livres	Total de horas-aula de estudo (exposição) de língua inglesa (tempo aproximado em horas-aula)
201	18	Aluno da graduação	630 h-a	60 h-a	-	690
202	19	2º grau completo	630 h-a	60 h-a	-	690
203	16	Cursando 2º grau	630 h-a	60 h-a	-	690
204	20	Aluno da graduação	630 h-a	60 h-a	-	690
205	32	Graduado	630 h-a	60 h-a	-	690
206	19	Aluno da graduação	630 h-a	60 h-a	6 meses (ou 60 h-a)	750
207	27	graduado	630 h-a	60 h-a	6 meses (ou 60 h-a)	750
208	20	Aluno da graduação	630 h-a	60 h-a	1 ano (ou 120 h-a)	810
209	25	graduado	630 h-a	60 h-a	1 ano (ou 120 h-a)	810
210	20	Aluno da graduação	630 h-a	60 h-a	6 anos (ou 720 h-a)	1410
401	19	Aluno da graduação	630 h-a	180 h-a	-	810
402	15	Cursando 2º grau	540 h-a	180 h-a	-	810
403	21	Aluno da graduação	630 h-a	180 h-a	-	810
404	34	graduado	540 h-a	180 h-a	-	810
405	31	Aluno de pós-graduação	630 h-a	180 h-a	-	810
406	20	graduado	630 h-a	180 h-a	2 anos (240 h-a)	1050
407	22	Aluno da graduação	630 h-a	180 h-a	2 anos (240 h-a)	1050
408	20	Aluno da graduação	630 h-a	180 h-a	2 anos (240 h-a)	1050
409	16	Cursando 2º grau	630 h-a	180 h-a	4 anos (480 h-a)	1290
410	31	2º grau completo	630 h-a	180 h-a	11 anos (1.320 h-a)	2130
601	20	Aluno da graduação	630 h-a	300 h-a	-	930
602	22	Aluno da graduação	630 h-a	300 h-a	-	930
603	21	Aluno da graduação	630 h-a	300 h-a	-	930

604	19	2º grau completo	630 h-a	300 h-a	-	930
05	20	Aluno da graduação	630 h-a	300 h-a	2 anos (ou 240 h-a)	1170
606	21	Aluno da graduação (Letras)	630 h-a	300 h-a	2 anos (ou 240 h-a)	1170
607	23	Aluno da graduação	630 h-a	300 h-a	2 anos (ou 240 h-a)	1170
608	20	Aluno da graduação	630 h-a	300 h-a	3 anos (ou 360 h-a)	1290
609	20	Aluno da graduação	630 h-a	300 h-a	3 anos (ou 360 h-a)	1290
610	21	Aluno da graduação	630 h-a	300 h-a	3 anos e 6 meses (ou 420 h-a)	1350

APÊNDICE E – PADRÃO ACENTUAL DOS ITENS LEXICAIS AMBÍGUOS DO PORTUGUÊS BRASILEIRO E DO INGLÊS COM SEUS RESPECTIVOS COGNATOS E FALSOS COGNATOS

PB (PORTUGUÊS BRASILEIRO)			INGLÊS		
Substantivo	Verbo	Adjetivo	Substantivo	Verbo	Adjetivo
Con'duta	Condu'zir	?	'conduct	To con'duct	?
Ob'jeto	Obje'tar	?	'object	To ob'ject	?
Con'flito	?	Conflitante/ conflituoso	'conflict	To con'flict	Con'flicting
Existe ?	Contes'tar	contestador	'constest	To com'test	?
De'serto	Deser'tar	Desertante?	'desert	To de'sert	?
separação	Sepa'rar	Sepa'rado	Sepa'ration	To 'sepa,rate	'separate
Advo'gado (4 ou 5 sílabas)	Advo'gar (3 ou 4 sílabas)	?	'advocate	To 'advo,cate	?
Apropria'ção	Apropri'ar (se)	A.pro.pri.'a.do (5 sílabas)	Appropri'ation	To ap'propriate	Ap'propriate
Estima'tiva/ estimação	Esti'mar	estimado	Estim'ation	To 'esti,mate	'estimate
Elabora'ção	Elabo'rar	Elabo'rado	Elabor'ation	To E'labo,rate	E'laborate

Legenda:
' = acento primário
, = acento

A tabela acima resulta nas seguintes conclusões:

Tendência acentual dos substantivos trissílabos e polissílabos no PB: paroxítono.

Tendência acentual dos verbos trissílabos e polissílabos no PB: oxítonos.

Tendência acentual dos adjetivos trissílabos e polissílabos no PB paroxítona (Atentar para a Restrição da Janela de Três Sílabas: o acentual só pode recair nas três últimas sílabas).

APÊNDICE F - TRANSCRIÇÕES FONÉTICAS

1. Palavra: “Object” [əb'dʒɛkt] (Categoria lexical: Verbo)

Identificação do Aprendiz + padrão acentual	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ob'dʒɛktʃ]	
202 (correto)	[ɔbi'dʒɛktʃi]	
203 (errado)	['ɔbidʒɛ]	
204 (correto)	[obi'ʒɛktʃi]	
205 (errado)	['ɔbidʒɛktʃ]	
206 (errado)	['ɔbdʒɛkt]	
207 (correto)	[obi'dʒɛk]	
208 (correto)	[obi'dʒɛkt]	
209 (correto)	[obi'dʒɛ]	
210 (correto)	[ɔubi'dʒɛkti]	
401 (errado)	['ɔbdʒɛkt]	
402 (correto)	[ɔb'dʒɛkt]	
403 (correto)	[ɔb'ʒɛkt]	
404 (errado)	['ɔbdʒɛk]	
405 (errado)	['ɔbiʒɛt]	
406 (errado)	['əubʒɛkt]	
407 (errado)	['əubdʒɛkt]	
408 (correto)	[ɔb'ʒɛkt]	
409 (errado)	['ɔbdʒɛkti]	
410 (errado)	['ɔbidʒɛk]	
601 (errado)	['ɔbdʒɛ]	
602 (correto)	[ɔb'dʒɛkt]	
603 (correto)	[ɔb'dʒɛkt]	
604 (errado)	['ɔbdʒɛ]	
605 (correto)	[ɔb'dʒɛkt]	
606 (errado)	['ɔbdʒɛkt]	
607 (correto)	[ɔb'dʒɛk]	
608 incompleto	[?'dʒɛktʃ]	
609 (errado)	['ɔbdʒɛkt]	
610 (errado)	['ɔbʒɛk]	

2. Palavra: “Object” [ˈabdʒɪkt, ˈabdʒɛkt] - (Categoria lexical: Substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ˈɔbdʒɛktʃ]	
202 (errado)	[ɔbiˈʒɛktʃi]	
203 (errado)	[ɔbˈʒɛkt]	
204 (errado)	[ɔbiˈdʒɛktʃi]	
205 (errado)	[ɔbˈdʒɛkt]	
206 (correto)	[ˈɔbdʒɛkt]	
207 (errado)	[ɔbˈdʒɛktʃ]	
208 (errado)	[ɔbiˈʒɛkt]	
209 (correto)	[ˈɔbiʒɛt]	
210 (correto)	[ˈɔbʒɛkt]	
401 (correto)	[ˈɔbʒɛkt]	
402 (correto)	[ˈɔbʒɛkt]	
403 (correto)	[ˈɔbiʒɛkt]	
404 (correto)	[ˈɔbdʒɛkt]	
405 (errado)	[ɔbˈʒɛkt]	
406 (correto)	[ˈɔbʒɛt]	
407 (correto)	[ˈɔbdʒɛkt]	
408 (errado)	[ɔbˈʒɛkt]	
409 (correto)	[ˈɔbdʒɛt]	
410 (correto)	[ˈɔbdʒɛkt]	
601 (correto)	[ˈɔbdʒɛkt]	
602 (errado)	[ɔbˈdʒɛktʃ]	
603 (correto)	[ˈɔbdʒɛkt]	
604 (errado)	[ɔbˈdʒɛkt]	
605 (errado)	[ɔbˈdʒɛkt]	
606 (errado)	[ɔbiˈdʒɛktʃ]	
607 (errado)	[ɔbˈʒɛkt]	
608 (correto)	[ˈɔbdʒɛktʃ]	
609 (correto)	[ˈabdʒɛktʃ]	
610 (correto)	[ˈɔbdʒɛkt]	

3. Palavra: “Conduct” [kən'dʌkt] - (Categoria lexical: Verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[kõ'duktʃ]	
202 (correto)	[kõ'duktʃi]	
203 (correto)	[kən'dʌkt]	
204 (errado)	['kõduktʃ]	
205 (correto)	[kən'dukt]	
206 (correto)	[kən'dəktʃi]	
207 (correto)	[kən'dʌkt]	
208 (correto)	[kõ'duktʃ]	
209 (correto)	[kõ'duk]	
210 insuficiente	[kən'dəkt]	
401 (correto)	[kən'dək]	
402 (correto)	[kõ'duktʃi]	
403 (correto)	[kõ'dʌk]	
404 (correto)	[kõ'duktʃi]	
405 (correto)	[kõ'dʒuktʃi]	
406 (correto)	[kõ'dəkt]	
407 (correto)	[kən'dəktʃ]	
408 (correto)	[kõ'duktʃi]	
409 (correto)	[kõ'dʒok]	
410 (correto)	[kən'dʌk]	
601 (correto)	[kõ'dukt]	
602 (correto)	[kən'dəkt]	
603 (correto)	[kən'duk]	
604 (pausa)	[kõ:'dukt]	
605 (correto)	[kõ'duktʃi]	
606 (correto)	[kən'dʌktʃ]	
607 (correto)	[kən'duktʃ]	
608 (correto)	[kən'dʒʌktʃ]	
609 (correto)	[kən'dəkt]	
610 (correto)	[kən'dʌk]	

4. Palavra: “Conduct” [ˈkɒndʌkt] - (Categoria lexical: Substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (errado)	[kõ'duktʃ]	
202 (errado)	[kõ'duktʃi]	
203 (errado)	[kən'dəkt]	
204 (errado)	[kõ'duktʃ]	
205 (errado)	[kon'dukt]	
206 (errado)	[kõ'dukt]	
207 (errado)	[kən'dʌktʃi]	
208 (errado)	[kõ'dukt]	
209 (errado)	[kõ'dukt]	
210 (hesitação)	[kõ'duk]	
401 (errado)	[kən'dəkt]	
402 (errado)	[kõ'dʒuktʃi]	
403 (hesitação)	[kõ'dʌkt]	
404 (errado)	[kən'dukt]	
405 (errado)	[kən'dukt]	
406 (errado)	[kən'dəkt]	
407 (errado)	[kən'dəkt]	
408 (errado)	[kõ'duktʃi]	
409 (hesitação)	[kõ'dʒukt]	
410 (errado)	[kõ'dʌkt]	
601 (pausa)	[kõ'dukt]	
602 (errado)	[kõ'dʌkt]	
603 (errado)	[kən'dukt]	
604 (errado)	[kõ'dukt]	
605 (errado)	[kõ'dukt]	
606 (correto)	[ˈkɒndʌkt]	
607 (errado)	[kõ'duktʃi]	
608 (correto)	[ˈkɒ:dʌkt]	
609 (errado)	[kən'dəkt]	
610 (errado)	[kõ'dʌkt]	

5. Palavra: “Conflict” [ˈkənˌflɪkt] - (Categoria lexical: Substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ˈkɔ̃flɪtʃ]	
202 (pausa)	[kɔ̃ˈflɪktʃi:]	
203 (pausa)	[ˈkɔ̃flɪt]	
204 (errado)	[kɔ̃ˈflɪtʃɪ]	
205 (correto)	[ˈkənflɪt]	
206 (correto)	[ˈkɔ̃flɪtʃ]	
207 (correto)	[ˈkənflɪt]	
208 (errado)	[kɔ̃ˈflɪt]	
209 (errado)	[kɔ̃ˈflɪtʃɪ]	
210 (correto)	[ˈkɔ̃flɪt]	
401 (correto)	[ˈkɔ̃flɪt]	
402 (errado)	[kɔ̃ˈflɪt]	
403 (errado)	[kɔ̃ˈflɪktʃ]	
404 (correto)	[ˈkɔ̃flɪktʃ]	
405 (hesitação)	[ˈkɔ̃flɪ]	
406 (correto)	[ˈkɔ̃flɪkt]	
407 (errado)	[kɔ̃ˈflɪktʃɪ]	
408 (errado)	[kɔ̃ˈflɪktʃi:]	
409 (correto)	[ˈkɔ̃flɪt]	
410 (correto)	[ˈkɔ̃flɪkt]	
601 (errado)	[kɔ̃ˈflɪt]	
602 (correto)	[ˈkɔ̃flɪt]	
603 (errado)	[kɔ̃ˈflɪt]	
604 (errado)	[kɔ̃ˈflɪt]	
605 (errado)	[kɔ̃ˈflɪkt]	
606 (correto)	[ˈkənflɪt]	
607 (correto)	[ˈkɔ̃flɪkt]	
608 (hesitação)	[ˈkɔ̃flɪt]	
609 (correto)	[ˈkɔ̃flɪt]	
610 (correto)	[ˈkɔ̃flɪkt]	

6. Palavra: “Conflict” [kən'flikt] - (Categoria lexical: Verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (errado)	[¹ kõflitʃɪ]	
202 (pausa)	[kõ'flitʃɪ:]	
203 (correto)	[kõ'flit]	
204 (correto)	[kõ'fliktʃɪ]	
205 (errado)	[¹ kõflitʃɪ]	
206 (correto)	[kõ'fliktʃɪ]	
207 (correto)	[kõ'fliktʃɪ]	
208 (correto)	[kõ'flikt]	
209 (errado)	[¹ kõflit]	
210 (correto)	[kõ'flitʃ]	
401 (correto)	[kõ'flit]	
402 (hesitação)	[kõn'fliktʃɪ]	
403 (correto)	[kõ'fliktʃɪ]	
404 (correto)	[kõ'fliktʃ]	
405 (correto)	[kõ'flikt]	
406 (errado)	[¹ kõflit]	
407 (correto)	[kõ'flitʃ]	
408 (correto)	[kõ'flitʃɪ]	
409 (correto)	[kõ'flitʃ]	
410 (correto)	[kõ'flikt]	
601 (correto)	[kõ'flitʃ]	
602 (correto)	[kõ'flikt]	
603 (correto)	[kõ'flit]	
604 (correto)	[kõ'flit]	
605 (correto)	[kõ'fliktʃɪ]	
606 (errado)	[¹ kõflikt]	
607 (correto)	[kõ'flikt]	
608 (correto)	[kõ'fliktʃ]	
609 (errado)	[¹ kønflit]	
610 (pausa)	[¹ konflit]	

7. Palavra: “Contest” [ˈkɒntɛst] - (Categoria lexical: Substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ˈkɒtɛʃtʃ]	
202 (hesitação)	[kɔ̃ˈtɛʃtʃi]	
203 (errado)	[kɔ̃ˈtɛst]	
204 (errado)	[kɔ̃ˈtɛst]	
205 (errado)	[kɔ̃ˈtɛst]	
206 (correto)	[ˈkɒntɛʃ]	
207 (correto)	[ˈkɒtɛs]	
208 (errado)	[kɔ̃ˈtɛst]	
209 (correto)	[ˈkɒtɛst]	
210 (correto)	[ˈkɒtɛst]	
401 (correto)	[ˈkɒtɛs]	
402 (errado)	[kɔ̃ˈtɛst]	
403 (errado)	[kɔ̃ˈtɛʃt]	
404 (errado)	[kɔ̃ˈtɛst]	
405 (errado)	[kɔ̃ˈtɛʃtʃ]	
406 (errado)	[kɔ̃ˈtɛst]	
407 (correto)	[ˈkɒtɛʃtʃi]	
408 (errado)	[kɔ̃ˈtɛst]	
409 (errado)	[kɔ̃ˈtɛkstʃ]	
410 (errado)	[kɔ̃ˈtɛst]	
601 (errado)	[kɔ̃ˈtɛs]	
602 (correto)	[ˈkɒtɛst]	
603 (correto)	[ˈkɒntɛst]	
604 (errado)	[kɔ̃ˈtɛʃtʃi]	
605 (errado)	[kɔ̃ˈtɛs]	
606 (correto)	[ˈkɑ̃tɛst]	
607 (errado)	[kɔ̃ˈtɛst]	
608 (errado)	[kɔ̃ˈtʃɛst]	
609 (correto)	[ˈkɒntɛst]	
610 (errado)	[kɔ̃ˈtɛst]	

8. Palavra: “Contest” [kən'tɛst] - (Categoria lexical: Verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[kõ'tɛʃ]	
202 (correto)	[kõ'tɛst]	
203 (correto)	[kõ'tɛst]	
204 (correto)	[kõ'tɛʃt]	
205 (correto)	[kon'tɛs]	
206 (hesitação)	['kõtɛs]	
207 (errado)	['kãtɛst]	
208 (correto)	[kõ'tɛkst]	
209 (correto)	[kõ'tɛst]	
210 (errado)	['kõtɛs]	
401 (correto)	[kõ'tɛst]	
402 (correto)	[kõ'tɛks]	
403 (correto)	[kõ'tɛst]	
404 (correto)	[kõ'tɛst]	
405 (pausa)	[kõ'tɛʃɪs]	
406 (errado)	['kõtɛst]	
407 (errado)	['kõtɪs]	
408 (correto)	[kõ'tɛstɪ]	
409 (correto)	[kõ'tɛks]	
410 (correto)	[kõ'tɛks]	
601 (correto)	[kõ'tɛs]	
602 (correto)	[kõ'tɛs]	
603 (errado)	['kantɛst]	
604 (hesitação)	[kõ'tɛs]	
605 (correto)	[kõ'tɛs]	
606 (errado)	['kontɛst]	
607 (correto)	[kõ'tɛst]	
608 (correto)	[kõ'tjɛʃtʃɪ]	
609 (correto)	[kən'tɛst]	
610 (correto)	[kõ'tɛst]	

9. Palavra: “desert” [ˈdɛzɜrt] - (Categoria lexical: substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (errado)	[dɪˈzɜrtʃɪ]	
202 (errado)	[dɪˈzɜrtʃ]	
203 (errado)	[dɪˈzɜrtʃ]	
204 (errado)	[deˈzart]	
205 (errado)	[deˈzɜrt]	
206 (errado)	[deˈzart]	
207 (correto)	[ˈdʒɪzə:]	
208 (errado)	[deˈzɜrt]	
209 (errado)	[dɪˈzɜrt]	
210 (correto)	[ˈdɜzɜr]	
401 (correto)	[ˈdɜzɜrt]	
402 (errado)	[dʒɪˈzɜrt]	
403 (errado)	[dɪˈzɜrt]	
404 (correto)	[ˈdezart]	
405 (errado)	[deˈzɜrt]	
406 (correto)	[ˈdezɜrt]	
407 (errado)	[dɪˈsɜrt]	
408 (errado)	[dʒɪˈzɜrt]	
409 (correto)	[ˈdɜzɜrt]	
410 (correto)	[ˈdɜzɜrt]	
601 (correto)	[ˈdʒɪzɜrt]	
602 (errado)	[dɪˈzart]	
603 (errado)	[deˈzart]	
604 (correto)	[ˈdezɜrt]	
605 (errado)	[dɪˈsɜrtʃ]	
606 (correto)	[ˈdɜzɜrtʃ]	
607 (errado)	[dʒɪˈzɜrtʃ]	
608 (errado)	[dɪˈzɜrtʃ]	
609 (errado)	[dɪˈzɜrt]	
610 (errado)	[deˈsɜrt]	

10. Palavra: “Desert” [dɪ'zɜrt] - (Categoria lexical: Verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[dʒɪ'sɛ:rtʃɪ]	
202 (correto)	[de'zɜrtʃ]	
203 (correto)	[dʒɪ'sɜrt]	
204 (correto)	[de'zɜhtʃu] link com a palavra seguinte	
205 (correto)	[de'zɛht]	
206 (correto)	[de'zɛrtʃ]	
207 (errado)	['dʒɪzɛrt]	
208 (correto)	[de'zɛh]	
209 (correto)	[dɪ'zɜrt]	
210 (errado)	['dɛzɜrt]	
401 (correto)	[dɪ'zɜrt]	
402 (correto)	[dɪ'zɜrt]	
403 (correto)	[dɪ'zɛht]	
404 (correto)	[de'zɜrt]	
405 (correto)	[dɪ'zɜrt]	
406 (correto)	[de'zɜrtʃ]	
407 (correto)	[dɪ'sɜrt]	
408 (correto)	[dɪ'zɜrt]	
409 (correto)	[de'zɜrtʃɪ]	
410 (correto)	[dɪ'sɜrt]	
601 (correto)	[dɪ'sɜrt]	
602 (correto)	[dɪ'zɜrt]	
603 (correto)	[dɪ'sɜrt]	
604 (errado)	['dezɜrt]	
605 (correto)	[dɪ'sɜht]	
606 (correto)	[dɪ'zɜrt]	
607 (correto)	[dɪ'zɜrtʃɪ]	
608 ?	['sɜrt]	
609 (correto)	[dɪ'zɜrtʃ]	
610 (correto)	[de'satʃju]	

11. Palavra: “separate” [ˈsepəˌreɪt] - (Categoria lexical: verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[sepəˈreɪtʃɪ]	
202 (errado)	[sepəˈreɪtʃɪ]	
203 (errado)	[sepəˈreɪtɪ]	
204 (errado)	[sepəˈreɪtʃɪ]	
205 (errado)	[sepəˈreɪt]	
206 (correto)	[sepəˈreɪtʃ]	
207 (errado)	[ˈsepəreɪt]	
208 (errado)	[sepəˈreɪt]	
209 (correto)	[ˈsepəˌreɪt]	
210 (correto)	[ˈsepəˌreɪt]	
401 (correto)	[ˈsepəreɪt]	
402 (correto)	[sepəˈreɪt]	
403 (correto)	[sepəˈreɪtʃɪ]	
404 (correto)	[sepəˈreɪt]	
405 (errado)	[sepəˈreɪt]	
406 (correto)	[ˈsepəreɪ]	
407 (correto)	[ˈsepəˌreɪt]	
408 (errado)	[sepəˈreɪt]	
409 (correto)	[sepəˈreɪt]	
410 (correto)	[ˈsepəˌreɪt]	
601 (correto)	[ˈseprɪt]	
602 (errado)	[sepəˈreɪ]	
603 (correto)	[ˈsepəreɪt]	
604 (correto)	[ˈsepəreɪt]	
605 (errado)	[sepəˈreɪtʃɪ]	
606 (correto)	[ˈsepreɪtʃɪ]	
607 (correto)	[ˈsepəˌreɪt]	
608 (correto)	[sepəˈreɪt]	
609 (correto)	[ˈsepəreɪ]	
610 (correto)	[ˈsepəreɪt]	

12. Palavra: “separate” [ˈsepəɹɪt] - (Categoria lexical: adjetivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ˈsepəɹeɪtʃɪ]	
202 (errado)	[sepəˈreɪt]	
203 (errado)	[sepəˈreɪt]	
204 (errado)	[ˈsepəɹeɪtʃ]	
205 (errado)	[sepəˈreɪt]	
206 (correto)	[sepəˈreɪtɪ]	
207 (errado)	[ˈsepɹɪt]	
208 (errado)	[sepəˈreɪt]	
209 (correto)	[sepəˈreɪtɪ]	
210 (correto)	[ˈsepəɹeɪt]	
401 (correto)	[ˈsepɹɪt]	
402 (correto)	[sepəˈreɪt]	
403 (correto)	[sepəˈreɪtʃɪ]	
404 (correto)	[ˈsepɹɪt]	
405 (errado)	[sepəˈreɪ]	
406 (correto)	[ˈsepəɹeɪt]	
407 (correto)	[ˈsepəɹeɪt]	
408 (errado)	[sepəˈreɪtʃɪ]	
409 (correto)	[ˈsepəɹeɪtʃɪ]	
410 (correto)	[ˈsepəɹeɪtəd]	
601 (correto)	[ˈsepɹɪt]	
602 (errado)	[ˈsepəɹeɪtʃɪ]	
603 (correto)	[ˈsepɹɪt]	
604 (correto)	[ˈsepəɹeɪt]	
605 (errado)	[ˈsepəɹeɪtʃɪ]	
606 (correto)	[ˈsepɹɪt]	
607 (correto)	[sepəˈreɪtʃɪ]	
608 (correto)	[ˈsepɹeɪtʃɪ]	
609 (correto)	[ˈsepɹɪt]	
610 (correto)	[ˈsepəɹeɪt]	

13. Palavra: “advocate” [ˈædvəkət] (Categoria lexical: substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ədʒɪˈvəkətʃɪ]	
202 (errado)	[ædʒɪvoˈkeɪtʃɪ]	
203 (errado)	[ɛdʒɪvoˈkeɪt]	
204 (errado)	[ədʒɪˈvoketʃɪ]	
205 (errado)	[ædʒɪvoˈkeɪtʃɪ]	
206 (correto)	[ɛdʒɪvoˈkeɪtʃɪ]	
207 (errado)	[ɛvdʒɪˈvouketʃɪ]	
208 (errado)	[ədʒɪvoˈkeɪtʃɪ]	
209 (correto)	[ædʒɪvoˈkeɪt]	
210 (correto)	[ədʒɪvoˈkeɪt]	
401 (correto)	[ədʒɪvoˈkeɪt]	
402 (correto)	[ævoˈkeɪt]	
403 (correto)	[ədʒɪvoˈkeɪtʃɪ]	
404 (correto)	[ɛdʒɪvoˈkeɪtʃɪ]	
405 (errado)	[ɛ:dʒɪˈvokt]	
406 (correto)	[adˈvoget] [ədʒɪˈvoket] pronunciou a mesma palavra duas vezes	
407 (correto)	[ædʒɪvoˈkeɪt]	
408 (errado)	[ədʒɪvoˈgeɪt]	
409 (correto)	[ədʒɪvoˈkeɪtʃɪ]	
410 (correto)	[ɛdʒɪvoˈkeɪt]	
601 (correto)	[ædvoˈkeɪt]	
602 (errado)	[ɛdʒɪˈvəkət]	
603 (correto)	[ɛdvoˈkeɪt]	
604 (correto)	[ədʒɪvoˈkeɪtʃɪ]	
605 (errado)	[ɛdʒɪˈvəkətʃɪ]	
606 (correto)	[ˈædvoket]	
607 (correto)	[ədʒɪvəʊˈkeɪtʃɪ]	
608 (correto)	[ɛdˈvɔkətʃɪ]	
609 (correto)	[ɛd#ˈvɔkət] (há uma pequena (pausa) de tantos ms)	
610 (correto)	[advəˈkætʃ]	

14. Palavra: “advocate” [ˈædvəˌkæt] - (Categoria lexical: verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[adʒɪvəˈkeɪt]	
202 (errado)	[adʒɪvəˈgeɪt]	
203 (errado)	[adʒɪvəˈkeɪt]	
204 (errado)	[adʒɪvəˈkeɪtʃɪ]	
205 (errado)	[ɛdvəˈkeɪt]	
206 (correto)	[ɛdʒɪvəˈkeɪt]	
207 (errado)	[dɛdʒɪvəˈkeɪtʃɪ]	
208 (errado)	[advəˈkeɪ]	
209 (correto)	[ædʒɪvəˈkeɪt]	
210 (correto)	[adʒɪvəˈkeɪt]	
401 (correto)	[ɛdʒɪvəˈkeɪtʃɪ]	
402 (correto)	[ævəˈkeɪtʃɪ]	
403 (correto)	[adʒɪvəˈkeɪtʃɪ]	
404 (correto)	[ædʒɪvəˈkeɪtʃɪ]	
405 (errado)	[ˈædʒɪvəkɛɪtʃɪ]	
406 (correto)	[adʒɪvəˈkeɪtʃɪ]	
407 (correto)	[adʒɪvəˈkeɪtʃɪ]	
408 (errado)	[adʒɪvəˈkeɪtʃɪ]	
409 (correto)	[ɛdʒɪvəʊˈkeɪtʃɪ]	
410 (correto)	[adʒɪvəˈkeɪt]	
601 (correto)	[ɛdʒɪvəˈkeɪtʃɪ]	
602 (errado)	[ɛdvəˈkeɪtʃɪ]	
603 (correto)	[adʒɪvəˈkeɪt]	
604 (correto)	[advəˈkeɪtʃɪ]	
605 (errado)	[ɛdʒɪvəˈkeɪtʃɪ]	
606 (correto)	[ˈædvəkɛɪt]	
607 (correto)	[ɛdʒɪvəˈkeɪtʃɪ]	
608 (correto)	[ɛdʒɪvəˈkeɪtʃɪ]	
609 (correto)	[adʒɪˈvəʊkɛɪt]	
610 (correto)	[avəˈkætʃɪ]	

15. Palavra: “appropriate” [ə'prouprɪt] - (Categoria lexical: adjetivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
202 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
203 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
204 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
205 (errado)	[apɹo'pɹeɪt]	
206 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
207 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
208 (errado)	[apɹo'pɹeɪt]	
209 (correto)	[apɹo'pɹeɪtʃɪ]	
210 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
401 (correto)	[apɹɔpɹɪ'eɪt]	
402 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
403 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
404 (correto)	[apɹo'pɹeɪtʃɪ]	
405 (errado)	[ɔpɹɔpɹɪ'eɪtʃɪ]	
406 (correto)	[apɹɔpɹɪ'eɪt]	
407 (correto)	[a'pɹɔpɹɪt]	
408 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
409 (correto)	[apɹɔpɹɪ'eɪtʃɪ]	
410 (correto)	[a'pɹouprɪt]	
601 (correto)	[a'pɹouprɪt]	
602 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
603 (correto)	[a'pɹɔpɹɪt]	
604 (correto)	[apɹɔpɹɪ'eɪt]	
605 (errado)	[apɹɔpɹɪ'eɪtʃɪ]	
606 (correto)	[ə'pɹɔpɹeɪt]	
607 (correto)	[a'pɹɔpɹeɪtʃɪ]	
608 (correto)	[a'pɹɔpɹeɪtʃɪ]	
609 (correto)	[a'pɹɔpɹeɪt]	
610 (correto)	[a'pɹouprɪt]	

16. Palavra: “appropriate” [ə'proupri,eɪt] - (Categoria lexical: verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[apɹɔpɹi'eɪtʃɪ]	
202 (errado)	[apɹɔpɹi'eɪʃən]	
203 (errado)	[apɹɔpɹi'eɪt]	
204 (errado)	[apɹɔpɹi'eɪtʃɪ]	
205 (errado)	[apɹo'preɪtɹ:]	
206 (correto)	[apɹɔ'peɪtʃɪ]	
207 (errado)	[apɹɔpɹi'eɪtʃɪ]	
208 (errado)	[apɹɔ'preɪtʃɪ]	
209 (correto)	[apɹɔ'pɹi'eɪtɪ]	
210 (correto)	[apɹɔpɹi'eɪtʃɪ]	
401 (correto)	[a'pɹɔpɹeɪt]	
402 (correto)	[apɹo'preɪtɪ]	
403 (correto)	[apɹɔpɹi'eɪtʃɪ]	
404 (correto)	[apɹɔpɹi'eɪtʃɪ]	
405 (errado)	[apɹɔpɹi'eɪʃɪs]	
406 (correto)	[a'pɹɔpɹeɪt]	
407 (correto)	[a'pɹɔpɹi'eɪt]	
408 (errado)	[apɹɔ'preɪtʃɪ]	
409 (correto)	[a'pɹɔpɹi'eɪtʃɪ]	
410 (correto)	[a'pɹɔpɹi'eɪt]	
601 (correto)	[a'pɹɔpɹi'eɪtʃɪ]	
602 (errado)	[apɹɔpɹi'eɪʃɪ]	
603 (correto)	[a'pɹɔpɹi'eɪt]	
604 (correto)	[apɹɔpɹi'eɪtʃɪ]	
605 (errado)	[apɹɔpɹi'eɪtʃən]	
606 (correto)	[ə'pɹɔpɹeɪt]	
607 (correto)	[apɹɔpɹi'eɪtʃɪ]	
608 (correto)	[a'pɹɔpɹi'eɪtʃɪ]	
609 (correto)	[a'pɹɔpɹi'eɪt]	
610 (correto)	[apɹɔpɹi'eɪtʃɪ]	

17. Palavra: “estimate” [ˈestəməɪt] - (Categoria lexical: substantivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ɪtʃɪ'meɪtʃɪ]	
202 (errado)	[esti'meɪtʃɪ]	
203 (errado)	[esti'meɪtʃɪ]	
204 (errado)	[estʃɪ'meɪtʃɪ]	
205 (errado)	[esti'meɪt]	
206 (correto)	[ˈestimeɪtʃɪ]	
207 (errado)	[esti'meɪtʃɪ]	
208 (errado)	[esti'meɪt]	
209 (correto)	[esti'meɪt]	
210 (correto)	[esti'meɪtʃɪ]	
401 (correto)	[estʃɪ'meɪtʃɪ]	
402 (correto)	[esti'meɪtʃɪ]	
403 (correto)	[estʃɪ'meɪtʃɪ]	
404 (correto)	[esti'meɪt]	
405 (errado)	[esti'meɪ]	
406 (correto)	[esti'meɪtʃɪ]	
407 (correto)	[esti'meɪt]	
408 (errado)	[eʃtʃɪ'meɪtʃɪ]	
409 (correto)	[esti'meɪt]	
410 (correto)	[ˈestmeɪt]	
601 (correto)	[esti'meɪt]	
602 (errado)	[esti'meɪtʃɪ]	
603 (correto)	[ˈestimeɪt]	
604 (correto)	[esti'meɪt]	
605 (errado)	[eʃtʃɪ'meɪtʃɪ]	
606 (correto)	[ˈestimeɪt]	
607 (correto)	[esti'meɪtʃɪ]	
608 (correto)	[desti'meɪtʃ]	
609 (correto)	[esti'meɪt]	
610 (correto)	[estʃu'meɪtʃɪ]	

18. Palavra: “estimate” [ˈestəˌmeɪt] - (Categoria lexical: verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[eʃtʃɪ'meɪtʃɪ]	
202 (errado)	[esti'meɪtʃɪ]	
203 (errado)	[ɪstrɪ'meɪt]	
204 (errado)	[estʃɪ'meɪtʃɪ]	
205 (errado)	[eɪstrɪ'meɪt]	
206 (correto)	[eʃtʃɪ'meɪtʃɪ]	
207 (errado)	[estʃɪ'meɪtʃɪ]	
208 (errado)	[esti'meɪt]	
209 (correto)	[ɪstrɪ'meɪt]	
210 (correto)	[esɪ'meɪt]	
401 (correto)	[esti'meɪ]	
402 (correto)	[ɪstrɪ'meɪt]	
403 (correto)	[estʃɪ'meɪtʃɪ]	
404 (correto)	[esti'meɪt]	
405 (errado)	[ˈeʃtʃɪmeɪt]	
406 (correto)	[ɪʃtʃɪ'meɪtʃɪ]	
407 (correto)	[esti'meɪt]	
408 (errado)	[eʃtʃɪ'meɪtʃɪ]	
409 (correto)	[ˈestɪmeɪt]	
410 (correto)	[ˈestɪmeɪt]	
601 (correto)	[esti'meɪt]	
602 (errado)	[ɪstrɪ'meɪt]	
603 (correto)	[ˈestɪmeɪ]	
604 (correto)	[ɪʃtʃɪ'meɪtʃɪ]	
605 (errado)	[estʃɪ'meɪtʃɪ]	
606 (correto)	[ˈestmeɪt]	
607 (correto)	[esti'meɪt]	
608 (correto)	[esti'meɪtʃɪ]	
609 (correto)	[ɪstrɪ'meɪt]	
610 (correto)	[ɪʃtʃɪ'meɪtʃɪ]	

19. Palavra: “elaborate” [ɪˈlæbəˌreɪt] - (Categoria lexical: verbo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ɛlabo'reɪtʃɪ]	
202 (errado)	[ɛlabɔ'reɪtʃɪ]	
203 (errado)	[ɛlabɔ'reɪtʃɪ]	
204 (errado)	[ɛləbo'reɪtʃɪ]	
205 (errado)	[ɛləbɔ'reɪtʃɪ]	
206 (correto)	[ɛlabɔ'reɪtʃɪ]	
207 (errado)	[ɛləbo'reɪtʃɪ]	
208 (errado)	[ɛlabo'reɪ]	
209 (correto)	[ɛlabo'reɪt]	
210 (correto)	[ælabo'reɪt]	
401 (correto)	[ɛlabo'reɪt]	
402 (correto)	[ɛla'bɔreɪt]	
403 (correto)	[ɛlɛbo'reɪtʃɪ]	
404 (correto)	[ɛlɛbo'reɪ]	
405 (errado)	[ɛlabɔ'reɪt]	
406 (correto)	[ɛlabo'reɪt]	
407 (correto)	[ɛlabo'reɪt]	
408 (errado)	[ɛlabɔ'reɪtʃɪ]	
409 (correto)	[ɛlɛbo'reɪt]	
410 (correto)	[ɛ'laboreɪt]	
601 (correto)	[ˈɔbreɪt]	
602 (errado)	[ɛlabo'reɪt]	
603 (correto)	[ɛ'laboreɪt]	
604 (correto)	[ɛlabo'reɪt]	
605 (errado)	[ɛlabo'reɪtʃɪ]	
606 (correto)	[ɛ'labreɪt]	
607 (correto)	[ɛ'laboˌreɪt]	
608 (correto)	[ɛlɛbo'reɪtʃɪ]	
609 (correto)	[e'laboreɪt]	
610 (correto)	[ɛ'lɛboreɪtʃ]	

20. Palavra: “elaborate” [ɪˈlæbrɪt] - (Categoria lexical: adjetivo)

Identificação do Falante	Transcrição fonética	Processos fonéticos e fonológicos observados (para uso do foneticista)
201 (correto)	[ɛlabo'reɪtʃɪ]	
202 (errado)	[ɛlabo'reɪtʃɪ]	
203 (errado)	[ɛlabo'reɪtʃɪ]	
204 (errado)	[ɛɛɪbo'reɪtʃɪ]	
205 (errado)	[ɛ̃nabo'reɪtʃɪ]	
206 (correto)	[ˈɛboreɪtʃɪ]	
207 (errado)	[e'lebreɪtʃɪ]	
208 (errado)	[ɛlabo'reɪt]	
209 (correto)	[ɛlabo'reɪted]	
210 (correto)	[ɛlabo'reɪs]	
401 (correto)	[e'lebrɪt]	
402 (correto)	[ɛlabo'reɪtʃɪ]	
403 (correto)	[alɛbo'reɪtʃɪ]	
404 (correto)	[ɛ'leboɪ]	
405 (errado)	[ɛlabo'reɪtʃ]	
406 (correto)	[ɛɛbo'reɪt]	
407 (correto)	[ɛ'laboreɪtʃ]	
408 (errado)	[ɛbo'reɪtʃɪ]	
409 (correto)	[ɛɛbo'reɪtʃɪ]	
410 (correto)	[ɛɛbo'reɪt]	
601 (correto)	[e'laboreɪ]	
602 (errado)	[ɛlabo'reɪtʃɪ]	
603 (correto)	[e'leboɪ]	
604 (correto)	[ɛlabo'reɪ]	
605 (errado)	[labo'reɪtʃɪ]	
606 (correto)	[ɛ'labreɪtʃɪ]	
607 (correto)	[ɛ'laboreɪt]	
608 (correto)	[ɛ'leboɪtʃɪ]	
609 (correto)	[ɛ'leboɪ]	
610 (correto)	[ɛ'leboɪt]	