

O USO DE REDES NEURAIS AUTO-
ORGANIZÁVEIS NA VISUALIZAÇÃO DE
FORMAÇÃO DE AGRUPAMENTOS A PARTIR DO
CONHECIMENTO ACENTUAL DE APRENDIZES
BRASILEIROS DE INGLÊS

Ana Cristina Cunha da Silva¹

Resumo: *This study aimed at investigating how Brazilian learners of English organize their knowledge about lexical stress of a specific word category at an early stage of L2 acquisition with the help of an unsupervised neural network, a self-organizing map (SOM), also called Kohonen network. The basic hypothesis tested was whether the parameterization of the speech signal from learner's utterances through processing techniques such as Linear Predictive Coding (LPC), which consisted of the input of the network, would be effective in the classification of learners and their utterances. The study consisted of an empirical part and a computational one. The participants were beginner students aged between 18 and 25. Preliminary results indicate that the combination of LPC+SOM allowed the creation of well-defined category clusters, which is an important step in data classification to aid language proficiency level determination, and computer-assisted pronunciation teaching.*

Palavras-chave: *Self-organizing map. Word stress. Linear Predictive Coding. U-matrix.*

¹ Professora Adjunta da Universidade Estadual do Piauí.

1. INTRODUÇÃO

Os modelos conexionistas têm desempenhado um importante papel no desenvolvimento da linguagem em várias áreas, tais como aquisição lexical e aquisição de pronomes (MCCLELLAND, 2009), sistematicidade sintática (MCCLELLAND et al., 2010), modelagem de distúrbios de linguagem (MIIKKULAINEN, 1997) e análise prosódica (GAUTHIER, 2009), apenas para citar alguns.

A maioria desses trabalhos se baseou em arquiteturas de redes neurais supervisionadas recorrentes ou *feedforward* (BLANC; DOMINEY, 2003; CUMMINS ET AL., 1999; ELMAN, 1990; KAZNATCHEEV, 2010), como as redes de Elman e o Perceptron Multicamadadas. Todavia, os modelos de redes neurais auto-organizáveis (*Self-organizing maps* - SOM), comumente chamados de rede de Kohonen, ao longo dos anos têm sido também usados como um modelo exploratório primário de fenômenos lingüísticos (Cf. FARKAS; CROCKER, 2008; GAUTHIER, 2009; LI ET AL., 2004; LI ET AL., 2007; MIIKKULAINEN, 1997).

Por exemplo, Li e seus colegas simularam a aquisição lexical em crianças utilizando um modelo de rede neural auto-organizável. O principal objetivo da pesquisa foi usar as propriedades da preservação topográfica do mapa auto-organizável (doravante SOM) para estudar a emergência de categorias lingüísticas e sua organização ao longo dos estágios de aprendizado lexical. O modelo conseguiu capturar uma série de fenômenos importantes que ocorriam na aquisição lexical inicial da criança e teve implicações significativas para os modelos de aquisição lingüística baseados em redes neurais auto-organizáveis.

Igualmente, ao se utilizar do SOM, Gauthier et al. (2009) investigou se e como as crianças poderiam aprender foco prosódico diretamente de um insumo (*input*) contínuo da fala. Os autores exploraram como o foco poderia ser aprendido por meio de sinais acústicos contínuos em Mandarim, que eram produzidos com tons lexicais co-ocorrentes e por vários falantes. Os resultados desse estudo mostraram que redes neurais podem desenvolver agrupamentos não supervisionados de foco específico a partir de sinal de fala dinâmico contínuo, produzido por vários falantes em várias condições de tom lexical, o que pode levar, no fim das contas, a aquisição do foco prosódico.

Após esta breve introdução de alguns trabalhos que se utilizaram direta ou indiretamente da rede SOM, passe-se ao objeto particular de

pesquisa neste artigo: o acento lexical, um dos elementos prosódicos mais importantes. O acento em língua inglesa tem múltiplas funções que variam de um papel enfático, através de uma intensidade contrastiva para indicar relações sintáticas entre palavras, como denota a oposição de pares de palavras de substantivos e verbos (e.g. **Object**, **objEct**; **dEsert**, **desErt**; **cOnflict**, **conflict**, etc.

Em português brasileiro (doravante PB) existe uma tendência de os substantivos trissílabos e polissílabos serem paroxítonos. Os verbos trissílabos e os polissílabos no português brasileiro possuem uma tendência a serem oxítonos. E ainda existe outra tendência: os adjetivos trissílabos e polissílabos em português são paroxítonos. Eis logo abaixo uma lista de alguns itens lexicais que exemplificam as tendências citadas acima, organizados por categoria lexical com seus respectivos padrões acentuais. O símbolo (´) representa acento primário:

Substantivo	Verbo	Adjetivo
conduta	conduzir	--
objeto	objetar	--
conflito	--	conflitante/ conflituoso
--	contestar	contestador / contestável
deserto	desertar	desertante
separação	separar	separado / separável
advogado	advogar	--
apropriação	apropriar (-se)	a.pro.priado
estimativa/estimação	estimar	estimado
elaboração	elaborar	elaborado

De acordo com algumas áreas da fonologia da interlíngua, o acento lexical (ou de palavra, como pode ser também chamado) é o elemento que mais se responsabiliza pelos casos de transferência lingüística, isto é, a influência de um sotaque predominante da L1 (primeira língua) na L2 (segunda língua) (cf. ARCHIBALD, 1994; BAPTISTA, 1981; MAIRS, 1989; SILVA, 2005).

A despeito de alguns trabalhos prévios sobre modelagem conexionista envolvendo traços prosódicos para o desenvolvimento e identificação da linguagem (BLANC; DOMINEY, 2003; CUMMINS et al., 1999; GAUTHIER et al., 2009), ao que parece, ainda não houve nenhuma investigação sistemática ou análise exploratória do fenômeno de

transferência negativa de padrão acentual de L1 para L2 por meio de um modelo conexionista utilizando a rede SOM.

Uma pesquisa que merece destaque é a de Zimmer (2004), que investigou a transferência do conhecimento fonético-fonológico do português brasileiro para o inglês (variedade norte-americana) a partir da incidência de nove processos de transferência sobre as elocuições dos participantes durante a recodificação leitora de palavras. Embora o foco da pesquisa não tenha sido o estudo dos fenômenos relacionados a tonicidade a autora incluiu entre os processos em suas simulações o elemento prosódico acento. A autora se utilizou de uma rede conexionista do tipo *feedforward* e para codificar a tonicidade, três componentes foram considerados: redução, duração e amplitude da vogal. Na pesquisa apresentada neste artigo utilizou-se os coeficientes LP e os coeficientes cepstrais (MFC) como codificação do *input* da rede já que condensam muita energia, frequência fundamental (F_0), banda de frequência, ou seja, informações genuinamente ligadas à prosódia da fala e principalmente ao acento, como será melhor descrito na metodologia a seguir.

À exceção dos estudos citados acima, não há estudos que aplicaram redes neurais artificiais para investigar a transferência do conhecimento fonológico relacionado ao acento em aprendizes brasileiros de língua inglesa, nem mesmo estudos que tenham levado em conta como o conhecimento sobre o acento lexical está organizado em um estágio de aquisição inicial de L2.

A partir do exposto, este artigo tem como objetivo principal investigar se e como a rede de Kohonen (SOM) é capaz de construir grupos bem definidos (*clusters*) a partir do sinal de fala das elocuições dos aprendizes. O intuito máximo é descobrir se padrões de transferência de conhecimento acentual de L2 possam vir a ser visualizados por meio da análise dos mapas-auto organizáveis, ou seja, do *output* oferecido pela rede SOM.

Partindo-se do pressuposto de que a rede neural organizará os aprendizes em função da pronúncia dos itens lexicais (ver Anexo1), hipotetiza-se que a rede neural será capaz de fazer generalizações e classificar os aprendizes em função não apenas do acento lexical mas também em relação às vogais, consoantes e outras categorias linguísticas. O estudo computacional constatou que a parametrização do sinal de fala proveniente da elocução dos aprendizes por meio de técnicas

de processamento de sinal (Codificação preditiva linear) conteve informações suficientes para a geração de mapas que possibilitassem a visualização de algumas categorias.

As próximas seções descrevem sucintamente a rede de Kohonen e sua arquitetura, a metodologia das simulações, o processamento dos dados de entrada da rede (*input*), o *corpus*, o perfil dos participantes, a simulação em si seguida de uma breve discussão e as considerações finais.

2. A REDE DE KOHONEN

Um mapa auto-organizável (*Self-Organizing Map* - SOM) 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).

Os mapas de Kohonen são utilizados usualmente como ferramenta de visualização de dados, permitindo que interrelações existentes em conjuntos complexos de dados possam ser percebidas e classificadas com mais exatidão. 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.

A seguir apresenta-se uma breve descrição do algoritmo SOM original, introduzido por Kohonen (2001). Denota-se $\mathbf{m}_i(t)$ $\mathbf{m}_i(t) \in \mathbb{R}^p$ como vetor de peso do i -ésimo neurônio no mapa. Depois de inicializar todos os pesos de vetores aleatoriamente ou de acordo com alguma heurística, cada iteração do algoritmo SOM envolve dois passos. Um primeiro, para um dado input de vetor $\mathbf{x}(t) \in \mathbb{R}^p$, encontra-se o neurônio vencedor atualizado $i^*(t)$, como se segue:

$$i^*(t) = \arg \min_{\forall i} \{ \|\mathbf{x}(t) - \mathbf{m}_i(t)\| \}. \quad \text{eq. 1}$$

Em que t denota as iterações do algoritmo. Então, é necessário ajustar os pesos de vetores do neurônio vencedor a aqueles que estão em sua vizinhança:

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \eta(t)h(i^*, i; t)[\mathbf{x}(t) - \mathbf{m}_i(t)], \quad \text{eq. 2}$$

Em que $0 < \eta(t) < 1$ é a taxa de aprendizagem e $h(i^*, i; t)$ é a ponderação de função gaussiana que limita a vizinhança do neurônio vencedor:

$$h(i^*, i; t) = \exp\left(-\frac{\|\mathbf{r}_i(t) - \mathbf{r}_{i^*}(t)\|^2}{2\sigma^2(t)}\right), \quad \text{eq. 3}$$

Em que $\mathbf{r}_i(t)$ e $\mathbf{r}_{i^*}(t)$ são respectivamente as posições dos neurônios i e i^* em uma matriz de saída pré-definida em que os neurônios estão organizados em nós, e $\sigma(t) > 0$ define o raio da função de vizinhança no tempo t . A fim de garantir a convergência do algoritmo, $\eta(t)$ e $\sigma(t)$ decaem exponencialmente no tempo de acordo com as seguintes expressões:

$$\eta(t) = \eta_0 \left(\frac{\eta_T}{\eta_0}\right)^{(t/T)} \quad \text{and} \quad \sigma(t) = \sigma_0 \left(\frac{\sigma_T}{\sigma_0}\right)^{(t/T)}, \quad \text{eq. 4}$$

Em que $\eta_0(\sigma_0)$ e $\eta_T(\sigma_T)$ são os valores iniciais e finais de $\eta(t)$ e $\sigma(t)$.

A matriz-U (matriz de distância unificada) é a ferramenta de visualização utilizada para analisar os resultados gerados pela rede SOM através da demonstração da distância de estruturas dos dados de entrada. 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 na organização espacial destes no SOM.

Além das habituais propriedades de quantização vetorial, o mapa resultante ordenado também preserva a topologia das amostras de entrada no sentido de que padrões de entrada adjacentes são mapeados em neurônios adjacentes no mapa. Devido a esta propriedade de preservação da topologia, o SOM é capaz de refletir as informações de entrada nas relações espaciais dos dados no mapa.

3. METODOLOGIA

A pesquisa consistiu de uma parte empírica e de uma parte computacional. O *corpus* desta pesquisa é composto das gravações das entre-

vistas com 30 alunos de uma instituição de ensino superior da cidade de Fortaleza-CE, 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.

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 níveis de desenvolvimento lingüístico distintos². Estabeleceu-se como critério circunstancial de classificação e organização dos indivíduos a quantidade de horas-aula acumulada na disciplina de língua inglesa obtida por meio de entrevistas e questionários respondidos pelos participantes. Esclarecemos essa tomada de decisão na próxima seção.

As elocuições provenientes das entrevistas com os participantes foram gravadas e digitalizadas no programa *Sound Forge* versão 5.0 em arquivos de áudio tipo WAV (extensão .wav), a uma taxa de 44.100 mil amostras por segundo (44.1 KHz) e resolução de 16 bits, mono.

Após essa fase, procedeu-se à segmentação de cada frase e consequentemente cada palavra que representava o item lexical a ser investigado (*e.g. object, separate, desert, etc*) (Ver anexo 1).

Como 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, 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 e/ou coeficientes mel-cepstrais, ao sinal de fala subdividido em vários frames. Assim, conjuntos de vetores de coeficientes passam a representar numericamente o sinal de fala dos aprendizes.

Os vetores de coeficientes LP ou MFC foram então organizados ao longo das linhas de uma matriz de coeficientes. Para o exemplo dado, foram gerados 100 vetores de coeficientes, um vetor para cada frame, que corresponderam 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 (no caso, 10, 15 ou 20). As matrizes de coeficientes foram extraídas e salvas em uma planilha do tipo Excel, sendo

2 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.

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 matriz de coeficientes está armazenada adicionou-se uma última coluna a esta matriz contendo um rótulo para identificar aquele conjunto de dados. Por fim, os arquivos de texto relativos à elocução de uma palavra específica por todos os falantes foram concatenados em um só arquivo com o auxílio do software Matlab®.

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.

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 do SOM e dos 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. 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 para este fim 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.

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).

Um conjunto de simulações foi rodado com parâmetros que variavam conforme a necessidade de ajuste ao fenômeno estudado. Apresentaremos a seguir os resultados da simulação que verificou se a rede conseguiria organizar os aprendizes em função do conhecimento acentual de L1 para L2. Com essa simulação, visa-se a contribuir na resolução de questões de determinação de nível de proficiência linguística.

4. A SIMULAÇÃO

A simulação descrita a seguir objetivou analisar a formação de grupos (clusters) nos mapas em função do conhecimento prosódico (acento) do aprendiz e do seu tempo de exposição ao idioma.

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 idioma.

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 (*Length of Residence*). 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 neurológica, como a criação de mapas mentais para a fala (KUHL et al., 1992).

Com o intuito 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. 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 abaixo:

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

Quadro 1: Parâmetros de treinamento da rede SOM. 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.

Logo 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.

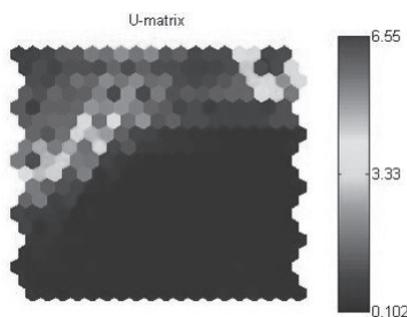


Figura 1: Matriz-U gerada para a rede 10 x 10 da palavra *object* (verbo) após treinamento de 250 épocas (ainda sem marcação dos grupos) Fonte: Elaboração da autora.

A figura 1 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-se 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. 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

2 (LPC 10), vê-se a sugestão de formação de três agrupamentos com erro de quantização final calculado a 0,427. Já na figura 3 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.

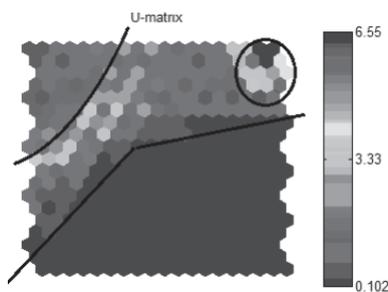


Figura 2: Matriz-U com entrada formada por 10 coeficientes LP. Índice DB indica 3 clusters e o erro de quantização final é igual a 0,427. Fonte: Elaboração da autora.

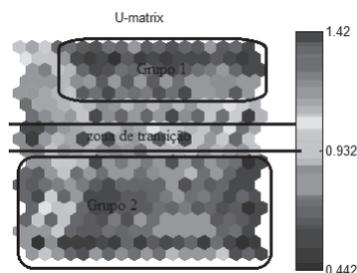


Figura 3: Matriz-U com entrada formada por 10 coeficientes MFC. Índice DB indica 2 clusters e o erro de quantização final é igual a 1,636. Fonte: Elaboração da autora.

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 fala (LPC 10).

5. DISCUSSÃO DOS RESULTADOS

A rotulação do conjunto de dados consiste numa etapa fundamental no processo de treinamento da rede. Rotular os mapas permite enten-

der de que forma a rede separou os falantes (nesse caso, entendem-se também como aprendizes ou locutores). 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 4), os vetores LPC relativos aos 34 participantes (30 aprendizes brasileiros + 4 falantes nativos de língua inglesa) 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.

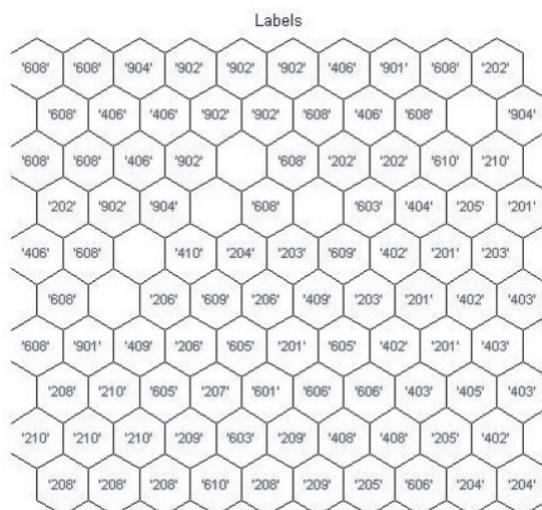


Figura 4: Mapa rotulado de uma rede SOM bidimensional 10 x 10 com estrutura de vizinhança hexagonal. Os rótulos correspondem a informações sobre falantes (aprendiz e nativos). Fonte: Elaboração da autora.

A Figura 5 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.

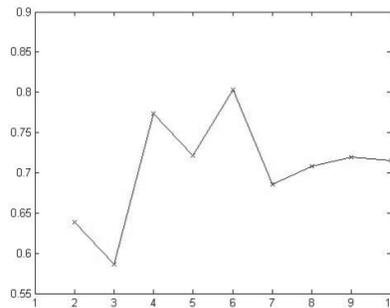


Figura 5: Gráfico para o índice Davies-Bouldin em função do número k de agrupamentos em uma rodada de treinamento típica. Fonte: Elaboração da autora.



Figura 6: Representação dos grupos no mapa treinado para k=3. Fonte: Elaboração da autora.

Conforme o número de agrupamentos identificados, a figura 6 exhibe 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.

Contrastando a figura 6 (mapa colorido sugerindo a formação de grupos) com a figura 7, que contém a representação dos grupos numerados na matriz-U, verifica-se a correspondência entre as áreas separadas no mapa colorido com as áreas circuladas e numeradas na matriz-U. A área azul corresponde ao grupo 4, a área de cor azul-piscina corresponde ao grupo 1 e a zona de cor amarela ao grupo 2 (ver figura 6).

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.

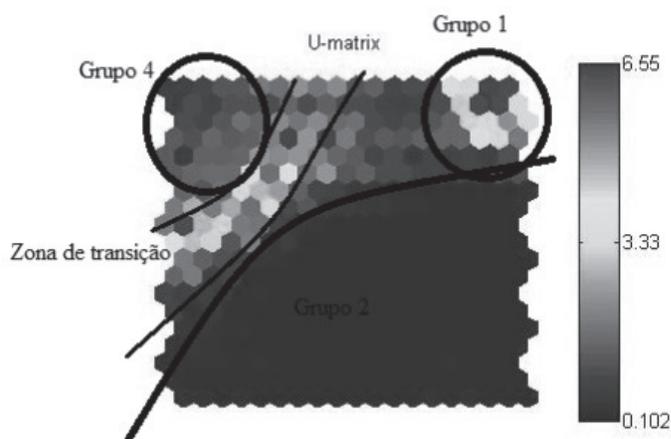


Figura 7: Matriz-U marcada com a representação dos grupos sugeridos pelo índice DB. Fonte: Elaboração da autora.

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 tentar organizar os alunos por nível de proficiência levando em consideração o tempo de exposição ao idioma, criaram-se seis níveis: nível 1, nível 2, nível 3, nível 4, nível 5, nível 6 para os falantes nativos (ver quadro 2). Como mostra a figura 9, 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.

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.	Aprendiz no.	Aprendiz no.	Aprendiz no. 210	Nativos
202	206	406	409		
203	207	407	608		
204	208	408	609		
205	209	605	610		
	401	606			
	402	607			
	403				
	404				
	405				
	601				
	602				
	603				
	604				

Quadro 2: Divisão dos participantes por horas-aula. Fonte: elaboração da autora

Foram feitas as devidas substituições a fim de treinar a rede com os novos rótulos de informações. Para tanto, 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 8 reforça o fato de que a rede foi capaz de segregar por características prosódicas semelhantes, ou seja, a rede neural 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 devidamente codificado e parametrizado é uma pista importante para a determinação do nível de proficiência lingüística em todas as fases de processamento de informações.

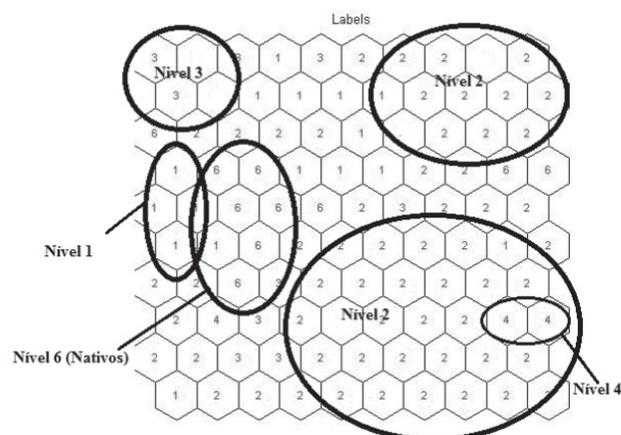


Figura 8: Mapa rotulado de acordo com o nível de proficiência (1, 2, 3, 4, 6). Fonte: Elaboração da autora.

Como se pode observar na figura 8, 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? Várias podem ser as justificativas. Todavia, 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 dito no início da seção, 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.

Para deixar novamente claro, os aprendizes só passaram por uma triagem com fins para o cálculo de tempo de exposição ao idioma. Se, por ventura, 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.

Além de a rede de Kohonen se mostrar como uma ferramenta de auxílio na determinação de nível de proficiência lingüística, 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 A1³, 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. É possível determinar o nível em que esse novo aprendiz se encontra sem o uso obrigatório de um teste padrão. Basta visualizar a topografia

3 Referência ao Marco Comum Europeu (Common European Framework of Reference - CERF) de Proficiência Lingüística, disponível em <http://www.cambridgeol.org/exams/exams-info/cefr.html>

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 foi capaz de reconhecer 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 (ou por que não dizer 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 em língua estrangeira.

CONSIDERAÇÕES FINAIS

A simulação aqui apresentada 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. Ao que parece, a rede conseguiu apreender informações sobre o histórico dos aprendizes e os classificou/agrupou a partir desse critério. Os padrões que emergiram da rede neural durante as simulações podem ser compreendidos como uma sugestão de nível de proficiência linguística que a rede oferece para esses aprendizes. Na realidade, a rede neural pode vir a determinar nível de proficiência linguística 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 quanto ao seu nível de proficiência linguística, desde que seja ajustada com o *input* adequado. Os resultados com base nessa

ferramenta de visualização também sugerem que o seu uso é bastante útil na sugestão de nível de proficiência linguística em LE/L2.

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.

BIBLIOGRAFIA

- ARCHIBALD, J. A formal model of learning L2 prosodic phonology. *Second Language Research* 10(3), 215–240, 1994.
- BAPTISTA, B. O. *An analysis of errors of Brazilians in the placement of English word stress*. Master's thesis, Postgraduate Program on Linguistics, Federal University of Santa Catarina, Brazil, 1981.
- BLANC, J. M., DOMINEY, P. F. Identification of prosodic attitudes by a temporal recurrent network. *Cognitive Brain Research* 17, 693–699, 2003.
- CUMMINS, F., GERS, F., SCHMIDHUBER, J. Language identification from prosody without explicit features. In: *Proceedings of the 6th European Conference on Speech Communication and Technology (EURO-SPEECH'99)*. pp. 305–308.
- ELMAN, J. L. Finding structure in time. *Cognitive Science* 14, 179–211, 1990.
- FARKAS, I.; CROCKER, M.W. Syntactic systematicity in sentence processing with a recurrent self-organizing network. *Neurocomputing* 71, 1172–1179, 2008.
- FLEGE, J., BIRDSONG, 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.
- GAUTHIER, B., SHI, R., XU, Y. Learning prosodic focus from continuous speech input: A neural network exploration. *Language Learning and Development* 5, 94– 114, 2009.

- KAZNATCHEEV, A. A connectionist study on the interplay of nouns and pronouns in personal pronoun acquisition. *Cognitive Computation* 2, 280–284, 2010.
- KOHONEN, T. *Self-Organizing Maps*. Springer, 3a. ed. 2001.
- 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.
- LI, P.; FARKAS, I.; MACWHINNEY, B. Early lexical development in a self organizing neural network. *Neural Networks* 17, 1345–1362, 2004.
- LI, P.; ZHAO, X.; MACWHINNEY, B. Dynamic self-organization and early lexical development in children. *Cognitive Science* 31, 581–612, 2007.
- MAIRS, J. L. Stress assignment in interlanguage phonology: an analysis of the stress system of Spanish speakers learning English. In: Gass, M & Schatcther, J. (orgs.) *Linguistic Perspectives on Second Language Acquisition*. Cambridge, USA: Cambridge University Press, 1989.
- MCCLELLAND, J. L. The place of modeling in cognitive science. *Topics in Cognitive Science* 1, 11–38, 2009.
- MCCLELLAND, J. L., BOTVINICK, M. M., NOELLE, D.C., PLAUT, D.C., ROGERS, T.T., SEIDENBERG, M.S., SMITH, L. B. Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences* 14, 348–356, 2010.
- MIKKULAINEN, R. Dyslexic and category-specific aphasic impairments in a self organizing feature map model of the lexicon. *Brain and Language* 59, 334–366, 1997.
- 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.
- 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, Ingrid; MATZENAUER, Carmen L. (Compiladoras). *TEP (Textos em Psicolingüística)* [CD-ROM]. 1. ed. Pelotas: Educat, 2006.

ANEXO 1 - LISTA DE ITENS LEXICAIS AMBÍGUOS (SUBSTANTIVOS, ADJETIVOS E VERBOS) USADOS NOS TESTES DE PRODUÇÃO

Itens Lexicais	Substantivo (Noun)	Verbo (Verb)	Adjetivo (Adjective)
Ab.stract [*]	/'æbstrækt/	/əb'strækt/	/'æbstræk, əb'strækt/
2. Ap.prox.i.mate		/ə'prɒksə'ment/	/'æprɒksəmt/
Ap.pro.pri.ate		/ə'prɒpəri,et/	/'æ'prɒpəriit/
A.dult	/ə'dʌlt, æ'dʌlt/		/'æ'dʌlt, æ'dʌlt/
Ad.vo.cate	/'ædvəkət/	/'ædvə,kæt/	
Ag.gre.gate	/'ægri:gt/	/'ægri,geit/	/'ægri:gt/
Com.pound*	/'kɑmpaʊnd/	/'kəm'paʊnd/	/'kɑmpaʊnd, kəm'paʊnd/
Con.duct	/'kɒndʌkt/	/kən'dʌkt/	
Con.flict	/'kɒn'flikt/	/kən'flikt/	
Con.test	/'kɒntest/	/kən'test/	
Con.trast	/'kɒntræst/	/kən'træst/	
Con.tract	/'kɒntrækt/	/kən'trækt/	
Choco.late	/'tʃɒklɪt/		
Con.tent	/'kɒntent/	/kən'tent/	/'kən'tent/
De.sert	/'dezɜ:t/	/di'zɜ:t/	
E.lab.o.rate		/'læbə'reit/	/'læbrɪt/
Es.ti.mate	/'estəmt/	/'estə,meit/	
Grad.u.ate	/'grædʒu:t/	/'grædʒu,et/	/'grædʒu:t/
In.crease	/'mkrɪs, 'ɪŋkris/	/'m'krɪs/	
In.sult	/'ɪnsʌlt/	/'ɪn'sʌlt/	
Mod.e.rate	/'mɒdərit/	/'mɒdə'reit/	/'mɒdərit/
Ob.ject	/'ɒbdʒɪkt, 'ɒbdʒekt/	/əb'dʒekt/	
Pro.duce	/'prɒdʌs, 'prɒdʌs/	/'prə'dʌs/	
Per.mit	/'pɜ:mit/	/'pɜr'mit/	
Pre.sent	/'prezənt/	/'pri'zənt/	/'prezənt/
Pro.gress	/'prɒgrəs, prɒgrɛs/		
Pro.ject	/'prɒdʒekt, 'prɒʒɪkt/	/'prə'dʒekt/	
Pro.test ^{2*}	/'prɒutest/	/'prə'test, 'prɒutest/	
Pre.cip.i.tate*	/'pri'sɪpə'teɪt, pri'sɪpə'tɪt/	/'pri'sɪpə'teɪt/	/'pri'sɪpə'tɪt/
Re.cord	/'rekɔ:d/	/'ri'kɔ:rd/	/'rekɔ:rd/
Sep.a.rate		/'sepə'reit/	/'sepri:t/
Sus.pect	/'sʌspekt/	/sə'spekt/	/'sʌspekt/
Sub.ject	/'sʌbdʒɪkt/	/səb'dʒekt/	/'sʌbdʒɪkt/

* itens lexicais que possuem duas pronúncias diferentes quanto ao acento numa dada categoria lexical.