

A LEITURA EM LÍNGUA ESTRANGEIRA E OS EFEITOS DA FREQUÊNCIA E DA CONSISTÊNCIA DO INSUMO LEXICAL EM L2

Márcia C. Zimmer*

Resumo: Este trabalho investiga o papel da consistência e da frequência do input lexical na leitura em L2 de forma a destacar duas importantes características do conexionismo dinamicista na leitura: a combinação não-linear de restrições probabilísticas e a divisão de trabalho. Essas características são complementares e, juntas, explicam a alta interação observada em investigações baseadas na premissa de que o input é uma combinação rica de diferentes tipos de informação cujas tendências emergentes guiam a aquisição. Para ilustrar isso, são apresentadas evidências de dois estudos, que tratam: 1) da interação entre os efeitos da frequência e da consistência do input lexical na prevalência da assimilação vocálica durante uma tarefa de leitura de palavras com estudantes brasileiros de inglês; 2) do papel dos fonemas na construção da compreensão textual em inglês (L2).

Palavras-chave: leitura em L2; frequência e consistência lexical; conexionismo.

1 INTRODUÇÃO

As diversas abordagens da aprendizagem¹ da linguagem – seja da língua materna ou da estrangeira – vêm apresentando um movimento pendular em que o foco concentra-se ou no contexto ou no cérebro do aprendiz, colocando maior ou menor ênfase no papel do insumo linguístico. Bloom (1974) foi um dos primeiros a rejeitar tendências pendulares, apontando três elementos fundamentais para a aquisição da linguagem: o insumo, o aprendiz e o contexto interacional.

* Professora do Programa de Pós-Graduação em Letras/ Linguística Aplicada da Universidade Católica de Pelotas. Doutora em Linguística Aplicada pela PUCRS. Email: marcia.zimmer@gmail.com.

¹ Neste trabalho não se adota a distinção, entre os termos ‘aquisição’ e ‘aprendizagem’, seja em língua materna ou estrangeira. Esses termos serão usados de modo intercambiável. Da mesma forma, os termos ‘língua estrangeira’ e ‘L2’ serão tratados aqui como sinônimos.

Essa intrincada relação entre fatores linguísticos, cognitivos e sócio-culturais pode ser examinada a partir do paradigma conexcionista, que destaca o insumo, colocando-o como um elemento de interface entre o contexto (ambiente) e o aprendiz (cognição).

O papel do insumo está estreitamente ligado ao tipo de processamento feito pelo aprendiz, e esse *continuum* entre insumo e processamento linguístico coloca em evidência a cognição. A aprendizagem da língua materna ou da língua estrangeira está, de fato, tão entranhada na cognição que relações entre sua produção e compreensão com o meio físico em que é processada às vezes se perdem. A aquisição da linguagem está estreitamente ligada e depende de vários mecanismos cognitivos fundamentais, como percepção, memória, formação de conceitos, resolução de problemas. Todos esses processos estão ligados à extrema capacidade dos aprendizes de fazer associações, que se deve à plasticidade cerebral e à transferência neuronal. Assim, estudar o processamento da linguagem implica fazer interface com as Ciências Cognitivas.

Partindo do pressuposto de que o conexionismo atribui a construção do conhecimento linguístico à aprendizagem e à transferência, e de que o insumo faz a interface entre ambiente e processos cognitivos, o presente trabalho destaca o efeito da frequência e da consistência do *input* na aprendizagem da leitura em língua estrangeira e está organizado da seguinte maneira: a) a primeira seção trata das principais características do conexionismo que estão em consonância com abordagens dinâmicas da aquisição da linguagem, ilustrando-as em modelos de leitura; b) a segunda parte focaliza os efeitos da frequência e da regularidade na leitura de palavras; c) a terceira e a quarta seções relatam dois estudos em leitura (L2) que trabalham com a noção de frequência e consistência. São apresentados, então, dados de dois estudos que envolvem a leitura em língua inglesa (L2) de palavras e de textos, respectivamente. O primeiro diz respeito à recodificação leitora de palavras em língua estrangeira e à assimilação vocálica (ZIMMER, 2004; ZIMMER, 2007), e o segundo conjunto de dados traz evidências de que a identificação de fonemas em palavras de baixa frequência auxilia a compreensão textual (MOTTA, 2007). Por fim, segue a conclusão do artigo.

2 CONEXIONISMO, DINAMICISMO E MODELOS DE LEITURA

A abordagem conexionista da aquisição da linguagem resgata uma faceta dinâmico-emergentista da linguagem, essencial para que nós, linguistas, não fiquemos restritos a abordagens simbólico-cognitivistas da linguagem que terminem por impedir o diálogo com as demais ciências cognitivas. A dinamicidade é uma das características da aquisição da linguagem que, felizmente, tem sido colocada em destaque por vários modelos conexionistas (ELMAN, 2004; MAIA; CLEERMANS, 2005). A abordagem dinâmica da cognição está intimamente ligada a idéias sobre a corporificação (*embodiment*) da mente e sobre a sensibilidade ao contexto da cognição humana, pois enfatiza aspectos em comum entre processos neurais e cognitivos por um lado e eventos fisiológicos e contextuais por outro (PORT, 2002). A hipótese dinâmica é inspirada em pesquisas da neurociência, tendo iniciado há mais de 50 anos, quando foi publicado um artigo sobre o comportamento individual de neurônios (HODGKINS; HUXLEY, 1952). Os modelos conexionistas deram novo fôlego à hipótese dinâmica através do desenvolvimento de algoritmos matemáticos de aprendizagem (RUMELHART; MCCLELLAND, 1986), que permitiram a implementação de modelos computacionais que simulam a aprendizagem linguística e a percepção, além de outros processos cognitivos.

De fato, as duas abordagens – a dinâmica e a conexionista – compartilham vários pressupostos sobre a cognição. Um deles é o de que fenômenos que tradicionalmente eram estudados de forma isolada, como se não fossem relacionados, passaram a ser vistos sob uma perspectiva unificadora. Nessa perspectiva, a atribuição de um conjunto de princípios teóricos comuns a uma variada gama de fenômenos cognitivos resultou na descoberta de inesperadas características comuns a eles. Se olharmos de maneira intuitiva para os sistemas morfológico e de escrita da linguagem, veremos muito pouco em comum. O sistema morfológico é um componente da linguagem natural que emergiu na espécie humana há centenas de milhares de anos, ao passo que o sistema de escrita é uma ferramenta inventada para representar a fala que começou a se desenvolver há apenas alguns milhares de anos. No

entanto, ambos são sistemas linguísticos caracterizados pela quase-regularidade (SEIDENBERG; ZEVIN, 2006). A caracterização do conhecimento e do processamento em domínios quase regulares é vista de maneiras distintas pelo paradigma simbólico e pelo conexionista. Como já se sabe, o primeiro propõe que os aspectos sistemáticos da linguagem são representados e processados como conjuntos de regras explícitas. Entretanto, como essa sistematicidade é parcial, faz-se necessário um mecanismo para lidar com as exceções. A distinção entre mecanismos baseados em regras e mecanismos que manipulam exceções, cada qual operando em consonância com princípios fundamentalmente distintos, forma o âmago de abordagens simbólicas da linguagem denominadas de “rota dupla”. A abordagem conexionista, por sua vez, postula a necessidade de um único mecanismo para lidar tanto com itens regulares como com os irregulares/inconsistentes. Um exemplo polêmico é o da aquisição do passado na língua inglesa, que vem gerando polêmica entre psicólogos e linguistas há mais de 20 anos. Outro, não menos controverso, é o da leitura de palavras.

Mas em que consiste a semi-sistematicidade que caracteriza esses dois sistemas linguísticos? Pode-se destacar que ambos são caracterizados por regularidades, mas admitem formas que se desviam de tendências centrais em diferentes vieses e em diferentes graus; os vieses que constituem as chamadas “exceções” do sistema linguístico se sobrepõem parcialmente às formas regulares e raramente são arbitrários (ALBANO, 1999, 2002; SEIDENBERG; ZEVIN, 2006). Esses vieses, as formas desviantes das tendências centrais, tendem a se agrupar entre as palavras de alta frequência da língua.

Quando se fala em leitura, entra em jogo o mapeamento probabilístico feito entre códigos de diferentes naturezas, como é o caso do sistema de escrita, que abarca o conhecimento fonético-fonológico, o conhecimento grafêmico e o conhecimento semântico. Na abordagem conexionista, a computação ou processamento desses conhecimentos de naturezas diversas toma a forma de interações cooperativas e competitivas entre um grande número de unidades simples de processamento semelhantes aos neurônios. As redes conexionistas aprendem através do ajuste de pesos em conexões entre unidades, de uma forma sensível a como a estrutura estatística do ambiente influencia o comportamento da rede. Como resultado, não existe uma dicotomia

abrupta entre elementos que obedecem às regras e elementos que não o fazem; pelo contrário, todos os elementos coexistem dentro de um sistema único – daí a denominação “rota única” dada aos modelos conexionistas.

Essa integração entre os sistemas fonético-fonológico, grafêmico e semântico ressalta a importância de outro princípio extremamente relevante na pesquisa sobre a leitura, tanto em L1 como em L2: a divisão de trabalho. Segundo esse princípio, a aprendizagem de uma habilidade complexa requer que se divida o problema entre subsistemas em interação (SEIDENBERG, 1992; PLAUT et al., 1996; HARM; SEIDENBERG, 2004; ZIMMER, 2004, 2007; SEIDENBERG; ZEVIN, 2006). A carga que cada um dos subsistemas processa depende do que está acontecendo em outros subsistemas que também estão contribuindo para o processamento de determinada tarefa. Outra propriedade importante da divisão de trabalho é a dimensão de *tempo* compartilhada por todos esses domínios, que permite sintonia em tempo real entre os domínios, onde a dinâmica de um sistema influencia o *timing* do outro sistema (ELMAN, 2005).

Outro aspecto fundamental na abordagem conexionista de cunho dinâmico, que complementa a divisão de trabalho, é a combinação não-linear de restrições. Um aprendiz adulto de L2, ou uma criança aprendendo a falar, por exemplo, identifica as fronteiras entre as palavras no fluxo da fala a partir da combinação de vários tipos de informação que, tomadas isoladamente, não seriam confiáveis ou robustas o suficiente para sustentar a aquisição da linguagem. Como afirma Albano, ao discutir os resultados de suas pesquisas sobre a natureza dinâmica e distribuída da representação fônica no PB,

os modelos fônicos dinâmicos são teorias extremamente abstratas e em contínua construção. Ainda que rejeitados pelo *main stream* da área [...] seguem seu destino científico de desvendar fatos que só uma ótica atenta à temporalidade da fala pode enxergar. Uma das suas vantagens foi aqui demonstrada mais uma vez: revelar regularidades sutis, encobertas, ou insuspeitas, do conhecimento fônico [...]. (ALBANO, 2007)

Essas regularidades encobertas são às vezes tão sutis que passam despercebidas aos olhos e ouvidos acostumados a contrastes categóricos

binários. Como o olhar aqui tenta desvendar relações distribucionais menos óbvias presentes no input, a opção é pela gradiência.

Cabe ressaltar, ainda, que as características aqui destacadas, como a divisão do trabalho e a combinação não-linear de restrições, apontam para uma propriedade essencial do input: a composicionalidade funcional da informação. Observa-se, aqui, que a composicionalidade, em termos de conexionismo, difere da visão clássica, em que é caracterizada pela habilidade de combinar constituintes recursivamente, distribuindo-se de forma previzível e linear nas unidades linguísticas (Van GELDER, 1990). Assim, embora tida como fundamental para a produtividade linguística, cognitiva e perceptual, a composicionalidade é vista como existente numa distribuição não tão obviamente concatenativa, mas sim detectável em vieses funcionais, às vezes menos frequentes em relação a tendências centrais. Portanto, a composicionalidade, na pesquisa conexionista, não deve ser restrita à componencialidade concatenativa, que está mais próxima de um conceito combinatorial sintagmático clássico.

Uma vez definidos os principais conceitos e delimitadas as propriedades-chave com as quais se pretende analisar os dados aqui apresentados, pode-se resumir brevemente o que nos diz a literatura conexionista sobre os efeitos da frequência e da consistência das palavras na leitura.

3 OS EFEITOS DA FREQUÊNCIA E DA CONSISTÊNCIA LEXICAL NA LEITURA DE PALAVRAS

Já no início desta seção, é preciso colocar com clareza que os estudos conexionistas sobre os efeitos da frequência e da consistência na leitura partem das palavras como insumo principal. Essa opção nada tem de redutor, pelo contrário; essa concepção sinaliza uma atribuição maior de importância ao léxico na aprendizagem da linguagem em geral e da leitura em particular. Dessa forma, neste artigo, toma-se a definição conexionista de que

as palavras são entendidas como estímulos que operam diretamente sobre estados mentais. As propriedades fonológicas,

sintáticas e semânticas de uma palavra são reveladas pelos efeitos que ela exerce sobre esses estados. (ELMAN, 2004, p.301)

A definição de Elman compreende uma noção já defendida por Rumelhart (1979) há quase três décadas: a de que as palavras não têm significado, mas constituem dicas (*cues*) para a construção do significado. A concepção de que diferentes tipos de conhecimento estão embutidos na linguagem é contemplada com muita felicidade por emergentistas como MacWhinney (2005, 2007), que também parte da idéia de que as diferentes pistas distribucionais contidas nas palavras cooperam e competem entre si para a construção do significado. Essa nova concepção sobre as palavras “confere um potencial ainda maior às palavras para interagir com flexibilidade com outras pistas” (ELMAN, 2004, p. 306).

Como essas pistas interagem durante a leitura, formando os vieses mencionados na seção anterior? A resposta a essa pergunta constitui um desafio que está sendo pesquisado com muito entusiasmo pelos conexionistas. Nos modelos conexionistas de leitura, o conhecimento grafo-fônico-fonológico está intimamente relacionado ao processamento que o leitor ou a rede faz. Na leitura, esse processamento está intimamente relacionado às variáveis da consistência e da frequência lexical. A consistência de uma palavra é uma variável contínua (GLUSHKO, 1979) que expressa o grau de semelhança – grafêmica, fonético-fonológica – de uma palavra com outras do léxico. A semelhança grafêmica é postulada em termos do corpo grafêmico das palavras, o que corresponderia fonologicamente à rima, e consistiria, basicamente, de vogal e consoante(s). Dessa forma, conflitos entre possíveis pronúncias de uma determinada palavra são resolvidos por interações cooperativas e competitivas baseadas em como os corpos grafêmicos dessas palavras se assemelham aos de todas as outras palavras conhecidas em termos de mapeamento grafo-fonêmico. Isso significa que o grau de consistência está relacionado aos vizinhos ortográficos e fonológicos das palavras. Palavras regulares consistentes são as que têm muitas palavras “amigas”, ou seja, palavras cujos corpos grafêmicos e suas respectivas pronúncias seguem a conversão grafo-fônica regular da língua. Palavras regulares inconsistentes (i.e., *gave*, *paid*, *foe*) são aquelas com um grande número de palavras amigas no léxico (i.e., *save*, *maid*, *Poe*) e um pequeno número de palavras-exceção

(i.e., have, said, shoe); palavras ambíguas, por sua vez, contêm corpos grafêmicos ocorrendo em um número equilibrado de palavras de palavras amigas e “inimigas” (palavras contendo corpos grafêmicos semelhantes às palavras-exceção, com um mapeamento grafo-fonêmico inconsistente), como em brown, blow; known, gown. As palavras-exceção apresentam um número imenso de palavras inimigas e pouquíssimas palavras amigas (SEIDENBERG; McCLELLAND, 1989, p. 535-539; PLAUT et al., 1996, p. 59-62).

A variável frequência está relacionada à frequência de uso das palavras de uma língua. As palavras de alta frequência, por exemplo, são reconhecidas mais rápida e acuradamente do que as palavras com que os leitores se deparam com menos frequência. Nota-se, também, que o efeito da experiência – que se traduz na alta frequência de algumas palavras na língua – parece ter mais impacto sobre as palavras-exceção do que sobre as palavras regulares, que seguem a conversão grafo-fônica (CGF) da língua (SEIDENBERG, 1992; SEIDENBERG; HARM, 2004).

Essa interação entre frequência e consistência emerge porque o desempenho na leitura de uma palavra consistente não depende tanto da experiência prévia com aquela palavra, mas é fortemente auxiliado pelo que há em comum entre a palavra em questão e outras palavras que têm correspondências grafo-fonêmicas semelhantes. O contrário ocorre com a pronúncia de palavras-exceção, que é muito mais fortemente determinada pela experiência com essas palavras (MACDONALD; CHRISTIANSEN, 2002). Isso nos leva a concluir que, enquanto a leitura de palavras regulares segue o modelo prototípico de aprendizagem, a leitura de palavras-exceção é mais influenciada pelo encontro com cada exemplo, cada *token* da palavra. Essa fusão prototípico-exemplar de modelos de aprendizagem e formação de conceitos é uma interessante propriedade da aprendizagem em redes conexionistas, que captam não somente as similaridades mais características do insumo, mas também as singularidades mais marcantes, as idiosincrasias presentes no mesmo².

Em suma, é importante destacar que a frequência do insumo lexical reflete o grau de experiência do aprendiz com determinada

² Isso se deve à capacidade que a rede tem de processar tanto *insumos* ortogonais como insumos linearmente independentes (RUMELHART; MCCLELLAND, 1986; MCCLELLAND; RUMELHART 1988).

palavra, ao passo que a consistência diz respeito ao fato de a pronúncia de uma dada palavra seguir ou não a CGF regular da língua. Conforme foi colocado anteriormente, a frequência das palavras lidas espelha a probabilidade do quanto o leitor pode ter-se deparado com essas palavras anteriormente, o que expressa a medida da experiência do leitor com determinado repertório lexical e fonético-fonológico. Essa experiência, por conseguinte, pode ser expressa em termos de velocidade na leitura, auxiliando a compreensão textual. Desse modo, as palavras de alta frequência são reconhecidas mais rápida e acuradamente do que as palavras com que os leitores se deparam com menos frequência. Para ilustrar isso, nas seções 4 e 5 são apresentados dados de dois estudos que trazem evidência sobre a composicionalidade entre frequência e consistência na leitura oral de palavras e leitura de textos em inglês (L2), respectivamente.

4 A RECODIFICAÇÃO LEITORA DE PALAVRAS EM LÍNGUA ESTRANGEIRA E A ASSIMILAÇÃO VOCÁLICA

Zimmer (2004, 2007) investigou, dentre vários outros processos de transferência do conhecimento fonético-fonológico do PB (L1) durante a leitura oral de palavras em inglês (L2)³, a incidência do processo de assimilação vocálica em função da frequência e da consistência do *input* durante a leitura oral de palavras da língua inglesa (palavras regulares de alta frequência, palavras regulares de baixa frequência, palavras-exceção de alta frequência, palavras-exceção de baixa frequência). Para isso, foi realizada uma pesquisa de campo transversal entre 156 estudantes em diferentes estágios de aprendizagem de língua inglesa – 50 no nível básico (1), 57 no nível intermediário (2), 34 no intermediário-avançado (3) e 15 no avançado (4) – alunos de universidades de Porto Alegre e região metropolitana.

As produções dos sujeitos foram gravadas em arquivos *.wav*, transcritas foneticamente e posteriormente analisadas com o intuito de observar a incidência de processos de transferência. A taxa de uso do

³ Para uma visão detalhada do trabalho empírico, ver Zimmer (2004); para uma visão esmiuçada da simulação computacional, ver Zimmer e Plaut (2007); para uma visão da tese de doutorado inteira, que engloba a pesquisa empírica e a computacional, ver Zimmer (2007).

processo de assimilação vocálica foi analisada a partir do efeito conjunto do nível de proficiência e do tipo de palavra - regular de alta frequência, regular de baixa frequência, exceção de alta frequência e exceção de baixa frequência. Para isso, foram realizadas Análises de Variância com fator duplo e medidas repetidas – uma vez que um mesmo sujeito teve de pronunciar palavras de quatro diferentes grupos de palavras –, considerando nível de proficiência e tipo de palavra como fatores independentes. Uma análise de medidas repetidas de variação, com nível de proficiência como fator inter-sujeitos e tipo de palavra como fator intra-sujeitos, confirmou que a diminuição de incidência desse processo é fortemente influenciada pelo aumento no nível de proficiência ($F(3,152)= 44,406$, $p = 0,000$). Além disso, o tipo de palavra teve um efeito estatisticamente significativo na utilização do processo ($F(1,152)= 146,664$, $p = 0,000$). Houve, também, um efeito de interação entre nível de proficiência e tipo de palavra ($F(3,152)= 3,601$, $p = 0,015$), o que sugeriu que, dependendo do tipo de palavra, a prevalência dos processos por nível se altera, como se observa na figura 1.

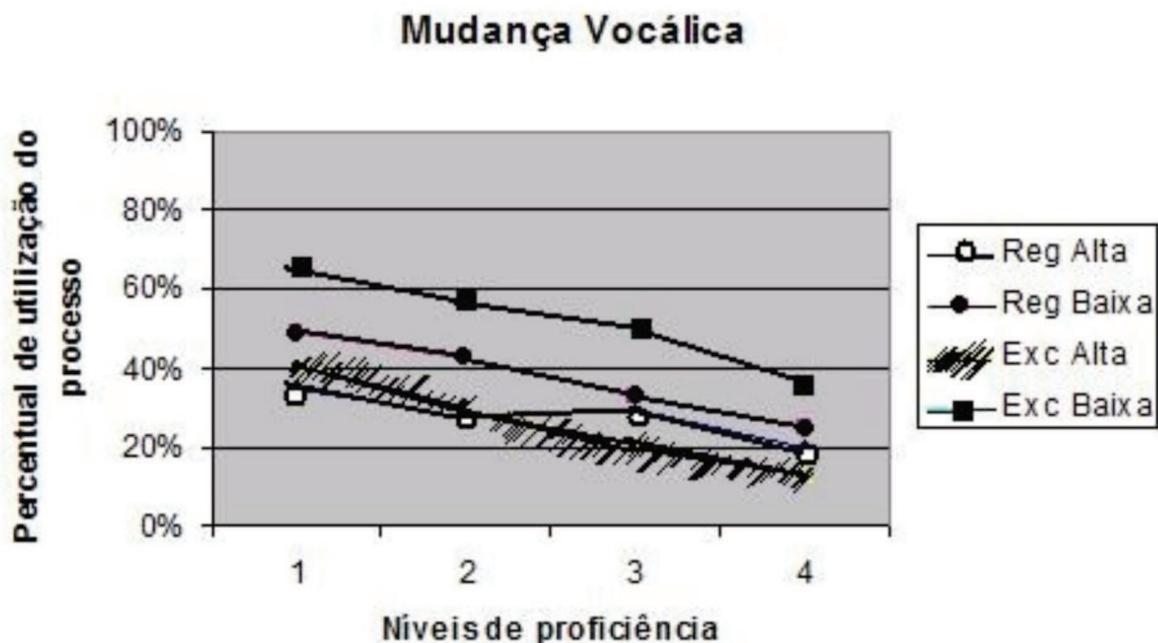


Figura 1 – Médias de incidência do uso da assimilação vocálica pelos grupos de proficiência nas quatro categorias de palavras.

A interação observada parece ter sido causada pelo desempenho dos sujeitos dos níveis 3 e 4 na leitura de palavras regulares de alta frequência, uma vez que o esperado era que a incidência do processo continuasse decrescendo no nível 3, ao invés de aumentar, como de fato ocorreu. Se não houvesse interação, a incidência do processo sobre as palavras regulares de alta frequência teria sido menor do que sobre as palavras-exceção de alta frequência nos níveis 3 e 4. Uma conclusão que se pode tirar dessa análise é que o processo de assimilação vocálica incidiu mais sobre o grupo de palavras de baixa frequência do que sobre o grupo de palavras de alta frequência.

Conforme mencionado na segunda seção deste artigo, a interação entre frequência e regularidade/consistência emerge porque o desempenho na recodificação de uma palavra regular não é auxiliado somente pela experiência prévia com aquela palavra, mas também pelo conhecimento prévio de palavras que têm CGFs semelhantes. O sucesso na pronúncia de palavras-exceção, por sua vez, é muito mais fortemente determinado pela experiência com essas palavras (MACDONALD; CHRISTIANSEN, 2002).

O processo de assimilação vocálica incidiu mais sobre o grupo de palavras de baixa frequência do que sobre o grupo de palavras de alta frequência. No que tange à influência maior da frequência sobre a performance dos sujeitos, há que se levar em conta que, quando o aprendiz/leitor da L2 chega a um patamar de proficiência mais elevado, o que provavelmente implica uma maior exposição à língua, parece haver uma maior interação entre regularidade e frequência. Isso pode ser observado, por exemplo, na performance dos sujeitos dos níveis 3 e 4 na recodificação de palavras regulares e palavras-exceção de alta frequência (fig. 1), em que a regularidade e/ou inconsistência dos itens lexicais parecem ser diluídas pela frequência. Em outras palavras, parece que o fato de os indivíduos mais proficientes na L2 terem sido expostos com muita frequência a essas palavras neutraliza as diferenças relativas à consistência nelas presente.

De fato, a própria natureza da interação entre frequência e regularidade pode variar ao longo da exposição do leitor à língua. Seidenberg (1992), ao investigar essa interação na recodificação de palavras regulares e irregulares na língua inglesa, descobriu que os leitores proficientes não tinham mais dificuldade de ler palavras

irregulares em comparação com as palavras regulares, exceto quando eram expostos a essas palavras irregulares com frequência muito baixa. Os leitores pouco proficientes, por sua vez, demoravam mais tempo para ler – e, por conseguinte, processar – as palavras irregulares. Seidenberg (1992) interpretou esses resultados como indicadores de que os leitores proficientes dispunham de um conjunto maior de palavras irregulares de alta frequência do que os leitores pouco proficientes, que liam menos e tinham menos prática na pronúncia dessas palavras. Isso indica que a variação na quantidade de exposição à língua escrita exerce efeitos diferentes sobre o processamento de informações linguísticas mais inconsistentes, gerando diferenças individuais que parecem ser de natureza qualitativa.

Essas diferenças individuais se fazem notar ainda com maior intensidade na leitura em L2, exatamente devido à maior variação na exposição à língua escrita, que faz com que se cogite a hipótese de que na aquisição da língua estrangeira a frequência de exposição às palavras da L2 não é igual à da L1, uma vez que o aprendiz da L2 é exposto às palavras da língua estrangeira por meio de materiais impressos em grande parte diversos daqueles aos quais o falante da L1 é exposto.

Outro aspecto importante a ser apontado é que os sujeitos ainda produziam a assimilação vocálica nos níveis 3 e 4. Por que, então, os alunos de níveis mais avançados não obtiveram um desempenho ótimo neste tipo de insumo regular e frequente?

A resposta a essa questão parece residir em dois focos: 1) o input que os alunos recebem; e 2) a percepção do tipo de input produzido na linguagem ambiente. Esses dois focos parecem estar relacionados à divisão do trabalho e à combinação não-linear de restrições presentes no input. O primeiro aspecto diz respeito à qualidade do estímulo ao qual o aluno é exposto. Isso não se refere somente ao insumo produzido pelo professor tradicional, presente na instrução formal, mas também à fala produzida por outros interlocutores, como colegas de aprendizagem da língua estrangeira, na sua esmagadora maioria falantes do PB, bem como a estímulos auditivos advindos de fontes alternativas, como canções, diálogos autênticos apresentados em fitas de áudio e vídeo, etc. Isso pode enviesar a frequência com que se ouvem determinadas palavras e expressões, bem como a maneira de produzi-las e percebê-las. O segundo foco mencionado acima diz respeito a como e com que

intensidade esse estímulo é percebido, e traz ao palco o papel de filtragem (KUHL, 2000) ou assimilação perceptual (BEST et al., 2001) exercido pela língua materna na percepção de contrastes da L2. Uma percepção enviesada do que é produzido pode afetar a combinação das dicas presentes no insumo, uma vez que, dependendo do contexto em que se insere o triângulo da divisão de trabalho explanada anteriormente, algumas informações, como a fonético-fonológica, podem ser combinadas com outras, como a grafêmica, de maneira diferente da de um nativo, criando novos vieses na produção oral, que pode contribuir para caracterizar o sotaque.

5 A RELAÇÃO ENTRE A COMPREENSÃO DE FONASTEMAS E A COMPREENSÃO TEXTUAL

Nesta seção, discute-se se itens lexicais de baixa frequência, como as palavras contendo fonastemas⁴, podem conter dicas ou pistas que auxiliem na construção do sentido textual em L2. O recorte dos dados referentes aos fonastemas deve-se ao fato de serem eles ótimos exemplos da composicionalidade não-linear presente na linguagem, já que os fonastemas não parecem ser categoricamente produtivos como morfemas e fonemas, mas, mesmo assim, funcionam como um pareamento semi-sistemático – embora arbitrário à primeira vista – entre o domínio fonético-fonológico, o domínio da escrita e o semântico⁵. Além disso, esse pareamento semi-sistemático também constitui um exemplo de interação entre frequência e consistência, sendo essa última medida pela vizinhança grafêmica, fonológica e também semântica entre os fonastemas.

De fato, Hinton, McClelland e Rumelhart (1986), ao discorrerem sobre mapeamentos arbitrários entre domínios, mencionam o fato de que muitas palavras monomorfêmicas podem conter “fragmentos peculiares

⁴ Fonastemas são sequências recorrentes na língua e constituem unidades maiores que um fonema e menores que um morfema, apresentando pareamento entre som e significado, mas não são morfemas contrastivos. Um exemplo é o onset *gl-* em inglês, como em *glimmer*, *glisten*, *glean*, etc., que ocorrem num grande número de palavras relacionadas à luz ou visão.

⁵ Ver Bergen, 2004, para uma maior discussão sobre a visão clássica de composicionalidade nos estudos sobre a linguagem e o que representam os fonastemas.

que têm um significado associado” (p. 97). Esses autores citam como exemplos, advindos de uma comunicação pessoal com Lakoff, palavras iniciando com *sn-*, “que geralmente significam algo desagradável relacionado ao lábio ou nariz (*sneer, snarl, snigger*)” (HINTON et al., 1986, p. 97). O interesse de Lakoff em fonastemas parece ter continuado por muito tempo, até que um de seus orientandos, Bergen (2004), fez deles seu objeto de estudo em tese de doutorado.

Os fonastemas aparecem, em geral, em palavras de conteúdo, como substantivos e verbos, e são pouco frequentes (BERGEN, 2004). Apesar de haver uma grande semelhança entre morfemas e fonastemas, há algumas diferenças que comprovam que os fonastemas não podem ser considerados um tipo especial de morfema (SHISLER, 1997). Em primeiro lugar, os morfemas são considerados segmentáveis, diferentemente dos fonastemas. Se retirarmos o fonastema *gl-* da palavra *glimmer*, o que sobra (*-immer*) não é um morfema, nem um pareamento de forma e significado. Em segundo lugar, os morfemas desempenham uma função sintática; os fonastemas, não. Por exemplo, o morfema ‘a’ (artigo definido, feminino, singular) tem a função sintática de determinante. Por fim, os morfemas só podem ser inseridos em determinados lugares dentro das palavras, o que não é o caso dos fonastemas, que podem aparecer no início, no meio, ou no fim de uma determinada palavra.

Vários estudos, em diferentes línguas, como o inglês (FIRTH, 1930) e o sueco (ABELIN, 1999), sugerem a existência de fonastemas, e sua difusão pode ser percebida na criação e identificação de neologismos, pois quando instados a inventarem ou criarem novas palavras, participantes de pesquisas tendem a seguir padrões previsíveis, considerando os fonastemas da sua língua (BERGEN, 2004).

Em seu estudo experimental sobre fonastemas, Bergen mostrou que, apesar de serem não-composicionais, os fonastemas não somente desempenham um papel no processamento da linguagem, mas também mostram efeitos de *priming*⁶, que são muito parecidos com os suscitados por morfemas composicionais. O autor concluiu que, apesar de os fonastemas não serem considerados morfológicos, o pareamento de

6 Processo pelo qual estímulos iniciais específicos, como uma palavra, ativam representações mentais correlatas que aumentam a capacidade para processar os estímulos subsequentes.

forma-significado dentro de palavras veiculado por eles é representado como parte do sistema linguístico.

Motta (2007) investigou se leitores brasileiros proficientes em inglês (L2) utilizam algumas estratégias *bottom-up*, consideradas cognitivas, como estratégias metacognitivas diante de palavras de baixa frequência. Para isso, conduziu um estudo empírico com 10 estudantes brasileiros adultos com nível elevado de proficiência leitora em inglês (L2), comprovada pelos escores na seção de leitura do TOEIC (Test of English for International Communication), instrumento já validado. O estudo empírico foi dividido em duas fases distintas, realizado em sessões alternadas. A fase piloto constou de três tarefas distintas: leitura de um texto, perguntas de compreensão e protocolos verbais, e incluiu um treinamento dos sujeitos, uma vez que o estudo foi realizado no computador com utilização do *software Translog*. Na segunda fase ou estudo propriamente dito, foram trabalhadas as mesmas tarefas que no estudo piloto, porém num texto de opinião sobre moda, de 380 palavras, contendo cinco fonastemas⁷. Nessa fase, os sujeitos responderam a 25 perguntas de compreensão textual. Após a compreensão do texto, uma entrevista retrospectiva foi conduzida em sessão à parte, objetivando esclarecer como os sujeitos pensaram ou agiram para chegar a cada resposta dada. Para tanto, foi utilizada a função *replay* do *Translog*, onde os sujeitos podiam visualizar a pergunta e sua respectiva resposta, discutindo-a com a pesquisadora.

Os resultados relativos à relação entre o número de fonastemas cujos significados os sujeitos conseguiram acertar (NFC) e o escore final no teste de compreensão em leitura (NTRC) aparecem na figura 2, abaixo.

Observa-se que os participantes que obtiveram os melhores escores em compreensão leitora, como os sujeitos 1, 7 e 8, foram precisamente aqueles que obtiveram os números de acertos máximos (5) em relação ao significado dos fonastemas. Essa, no entanto, é uma análise meramente descritiva dos dados. Na análise estatística de diversas variáveis estudadas por Motta através da correlação de Spearman (Rho)⁸ entre variáveis quantitativas, observou-se uma

⁷ Os fonastemas que se encontravam no texto foram: glitter, glimmer, glisten, glow, glossy.

⁸ O teste não-paramétrico de Spearman foi usado em virtude do baixo número de participantes.

importante correlação entre o número de palavras contendo fonastemas corretamente compreendidas e o desempenho dos sujeitos no teste de compreensão leitora: $Rho = 0,819$ ($p < 0,01$). Fica claro que, respeitado o pequeno tamanho da amostra em questão, é necessário ter cautela na interpretação desses resultados. Contudo, é importante destacar que, dentre oito correlações feitas entre as mais diversas variáveis investigadas, essa é a mais forte de todas. É importante que esse estudo seja replicado com um número maior de sujeitos, pois se as tendências apontadas por Motta (2007) forem confirmadas, haverá importantes implicações teóricas sobre o caráter não linear da composicionalidade encontrada na relação grafo-fônico-semântica distribuída nos fonastemas.

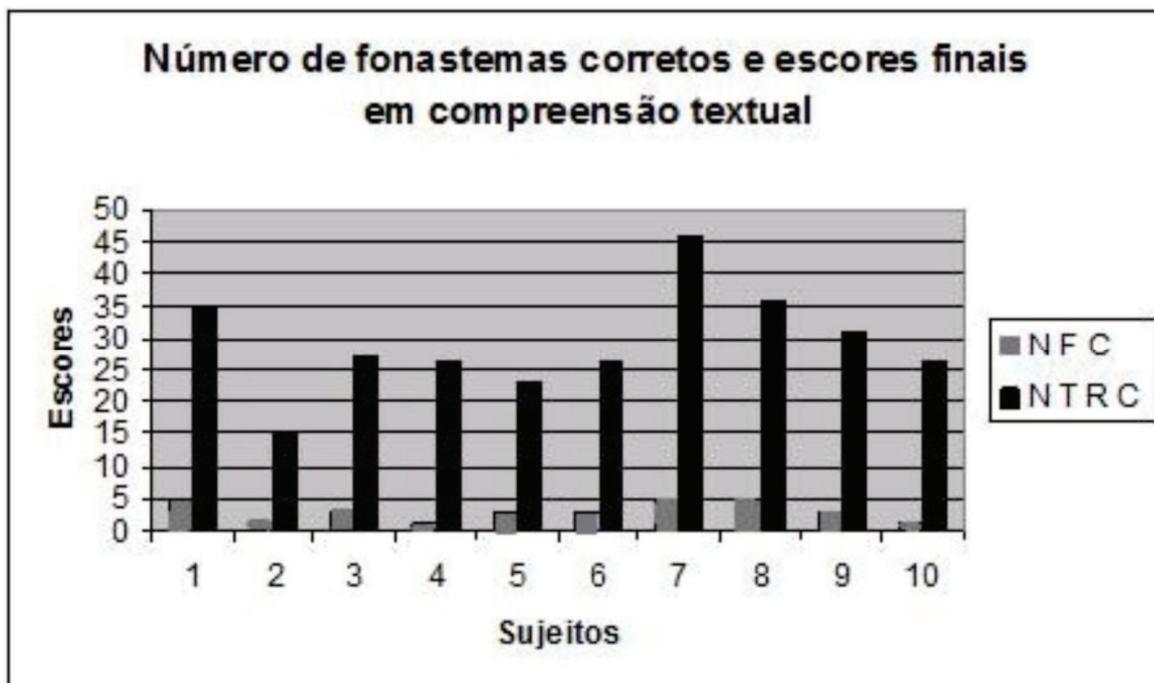


Figura 2 – Gráfico descritivo da relação entre número de fonastemas corretos e escores em compreensão textual dos participantes da pesquisa.

Outro resultado interessante apurado por Motta é que a maioria dos participantes, ao serem entrevistados durante o protocolo verbal retrospectivo - em que tiveram acesso às respostas que haviam escrito para o teste de leitura - usaram a estratégia de recodificação leitora das palavras contendo fonastemas, repetindo-as em voz alta muitas vezes e associando-as com outras que continham os mesmos fonastemas. Dessa

forma, a maioria dos sujeitos aumentou seus escores de compreensão das palavras contendo fonastemas durante os protocolos verbais.

Pode-se afirmar, portanto, que a recodificação leitora de fonastemas funciona não apenas como uma estratégia ascendente (*bottom up*) – que é apresentada como cognitiva na literatura *mainstream* –, mas também como uma estratégia metacognitiva. Assim, durante a recodificação leitora de palavras contendo fonastemas, fonar significa não apenas produzir sons, mas também refletir sobre eles, associando-os a um significado guiado pelo viés da frequência de ocorrência de determinado fonastema no léxico da língua.

Esses resultados indicam, assim, que a composicionalidade não-linear que os fonastemas imprimem às palavras pode auxiliar não apenas na construção de sentido das palavras, mas também pode funcionar como desencadeadora de estratégias metacognitivas motivadoras de atividades fônicas – como a recodificação leitora – que antes eram vistas como meramente ascendentes.

6 CONCLUSÃO

Neste trabalho, os efeitos da consistência e da frequência do insumo lexical na leitura em L2 foram abordados com o objetivo de destacar duas importantes propriedades do conexionismo dinamicista: a combinação não-linear de restrições probabilísticas e a divisão do trabalho. Para esse fim, foram mobilizados dados, recortados de dois estudos empíricos sobre a leitura em L2, em relação: 1) à interatividade entre os efeitos da frequência e da consistência na prevalência do processo de assimilação vocálica durante a leitura de palavras em L2; 2) ao papel dos fonastemas na construção da compreensão leitora em textos. O insumo, cuja consistência e frequência são destacadas em estudos sobre leitura, é o lexical, e a dobradinha frequência-consistência presente nas palavras parece ser complementar, assim como são complementares a combinação de restrições probabilísticas e a divisão de trabalho entre os sistemas que interagem durante o processamento da leitura. Juntas, essas características são responsáveis pelo alto nível de interação entre diferentes tipos de informação linguística, cujos vieses

distribucionais guiam a aquisição da linguagem e a compreensão da leitura.

REFERÊNCIAS

ABELIN, Asa. **Studies in sound symbolism**. Tese de doutorado inédita. Goëteborg: Goëteborg University, 1999.

ALBANO, Eleonora Cavalcante. Representações dinâmicas e distribuídas: indícios do português brasileiro adulto e infantil. **Letras de Hoje**, v. 42, p. 131-150, 2007.

_____. A pulsação da fala sob a letra: pela quebra de um silêncio histórico no estudo do som de fala. **Cadernos de Estudos Linguísticos**, v.1, n.42, p.7-20, 2002.

_____. Criatividade e gradiência num léxico sem derivações. In: L. GRIMM-CABRAL; J. MORAIS. (Org.). **Investigando a linguagem**: ensaios em homenagem a Leonor Scliar-Cabral. Florianópolis-SC: Mulheres, 1999. p. 35-54.

BERGEN, Benjamin K. The psychological reality of phonaesthemes. **Language**, v. 80, n. 2, p. 290-311, 2004.

BEST, Catherine; McROBERTS, Gerald; GODELL, Elizabeth. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. **Journal of the Acoustical Society of America**, 109, p. 775-994, 2001.

BLOOM, L. Talking, understanding, and thinking. In: SCHIEFELBUSCH, R; LLOYD, L. (Eds). **Language perspectives**: acquisition, retardation, and intervention. Baltimore: University Park Press, 1974.

ELMAN, Jeffrey. An alternative view of the mental lexicon. **Trends In Cognitive Sciences**, v. 7, p. 301-306, 2004.

ELMAN, Jeffrey. Connectionist models of cognitive development: Where next? **Trends In Cognitive Sciences**, v. 9, p. 111-117, 2005.

FIRTH, John. **Speech**. London: Oxford University Press, 1930.

HARM, Michael; SEIDENBERG, Mark. Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. **Psychological Review**, n. 111, p. 662-720, 2004.

ZIMMER – A leitura em língua estrangeira e...

- HERNANDEZ, Arturo; LI, Ping; MacWHINNEY, Brian. The emergence of competing modules in bilingualism. **Trends in Cognitive Sciences**, v. 9, n. 8, p. 220-225, 2005.
- HINTON, Gilmort E.; MCCLELLAND, James L.; RUMELHART, David E. Distributed representations. In: RUMELHART, David E.; MCCLELLAND, James L. **Parallel distributed processing: explorations in the microstructure of cognition**. Cambridge, MA: MIT, v. 1, 1986. p. 77-109.
- HODGKIN, A.; HUXLEY, A. A quantitative description of membrane current and its application to conduction and excitation in nerve. **Journal of Physiology**, v. 17, p.500–544, 1952.
- KUHL, Patricia. A new view of language acquisition. **Proceedings of the National Academy of Science**, n. 97, p. 11850-11857, 2000.
- MacDONALD, Maryellen; CHRISTIANSEN, Morten. Reassessing working memory: a reply to Just & Carpenter and Waters & Caplan. **Psychological Review**, v. 109, n. 1, p. 35-54, 2002.
- MacWHINNEY, Brian. Emergent fossilization. In: HAN, Z; ODLIN, Terence (Eds.) **Perspectives on fossilization**. Clevedon, UK: Multilingual Matters, 2007.
- MAIA, Tiago; CLEEREMANS, Axel. Consciousness: converging insights from connectionist modeling and neuroscience. **Trends in Cognitive Sciences**, v. 9, n. 8, p. 271-279, 2005.
- McCLELLAND, James; RUMELHART, David. Learning in PDP models: the pattern associator. In: _____. **Explorations in parallel distributed processing: a handbook of models, programs, and exercises**. Cambridge, MA: MIT Press, p. 83-120. 1988.
- MOTTA, Alayde Guimarães. **O continuum cognição-metacognição nas estratégias bottom-up empregadas na compreensão textual em L2**. 2007. 113 f. Dissertação (Mestrado) - Curso de Letras, Programa de Pós-graduação em Leitura e Cognição, UNISC, Santa Cruz do Sul, 2007.
- PORT, Robert F. Dynamical systems hypothesis in cognitive science. In: NADEL, L. (Ed). **Encyclopedia of cognitive science**, v. 1, 2002. p. 1027-1032.
- PLAUT, David; McCLELLAND, James; SEIDENBERG, Mark; PATTERSON, Karalyn. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. **Psychological Review**, n.103, p. 56-115, 1996.

RUMELHART, David. Some problems with the notion that words have literal meanings. In: ORTOGONY, A (Ed.), **Metaphor and thought**. Cambridge: Cambridge University Press, 1979. p. 71–82.

RUMELHART, David; McCLELLAND, James. On learning the past tenses of English verbs. In: McCLELLAND, James; RUMELHART, David. **Parallel distributed processing: Explorations in the microstructure of cognition**. Cambridge, MA: MIT, v. 2, 1986. p. 216-271.

SEIDENBERG, Mark. Beyond orthographic depth: equitable division of labor. In: FROST, R; KATZ, L (Eds.), **Orthography, phonology, morphology, and meaning**. Amsterdam: North-Holland, 1992.

SEIDENBERG, Mark; ZEVIN, Jason. Connectionist Models in Developmental Cognitive Neuroscience: critical Periods and the Paradox of Success. In: MUNAKATA, Y; JOHNSON, M. (Eds.). **Attention & performance XXI: processes of change in brain and cognitive development**. Oxford: Oxford University Press, 2006.

SHISLER, Benjamin K. **The influence of phonesthesia on the English language**. 1997. Disponível em:
<www.geocities.com/SOHO/studios/9783/index2.html> Acesso em: 17 jan. 2006.

Van GELDER, Timothy. Compositionality: a connectionist variation on a classical theme. **Cognitive Science**, v.14, 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 Psicolinguística)* [CD-ROM]. 1. ed. Pelotas: Educat, 2006.

ZIMMER, Márcia Cristina. Um estudo conexionista da transferência do conhecimento fonético-fonológico do PB (L1) para o inglês (L2) na leitura oral. In: POERSCH, Marcelino; ROSSA, Adriana. **Processamento da linguagem e conexionismo**. Santa Cruz: EDUNISC, 2007, p. 105-154.

ZIMMER, Márcia Cristina; PLAUT, David. Simulações conexionistas da transferência do conhecimento fonético-fonológico do PB (L1) para o inglês (L2). In: POERSCH, Marcelino; ROSSA, Adriana. **Processamento da linguagem e conexionismo**. Santa Cruz: EDUNISC, 2007, p. 264-299.

Recebido em 06/10/08. Aprovado em 15/09/09.

Title: L2 reading and the effects of frequency and consistency of the l2 lexical input

Abstract: This paper looks into the role of lexical input consistency and frequency in L2 reading to highlight two important features of dynamic connectionism in reading: nonlinear combination of probabilistic constraints and division of labour. Such features are complementary and together they account for the interaction observed in investigations grounded on the postulate that the input is a rich combination of different types of information whose emerging biases guide acquisition. For that purpose, evidence is presented from two studies, which deal with: 1) the interaction between the effects of English lexical frequency and consistency in the prevalence of vowel assimilation during a word naming task among Brazilian students of English; 2) the role of phonastemes in building textual comprehension in English (L2).

Keywords: L2 reading; lexical frequency and consistency; connectionism.