

Reconhecimento de palavras reais e de não-palavras em crianças de 1^a a 4^a série: uma tarefa de decisão lexical

Ângela Maria V. Pinheiro
FAFICH/UFMG

Anna Edith B. da Costa
Faculdade Metropolitana de Belo Horizonte

Francis Ricardo R. Justi
FAFICH/UFMG

Abstract

This paper describes the results of a lexical decision task which was applied to groups of children in 1st – 4th grades. The main objective was to obtain preliminary measurements on which to base the development of a final version of the same type of task, to be included in the computerized evaluation tool AVACLE

(cognitive assessment of reading and writing abilities), already in the final stage of development by Pinheiro and collaborators. 68 children were presented individually with a mixed list of words and non-words and instructed to press the Yes-key for words and the No-key for non-words. Decision reaction time (DRT) and error percentage were recorded. As expected, words were recognised more rapidly than non-words for all grades. In terms of error, this lexical effect was only significant for 3rd graders. The frequency effect was significant for both DRT and accuracy only for 2nd graders and on DRT for 1st and 3rd graders. The results show how the orthographic lexicon develops, and are discussed according to the information processing approach. The test proved to be efficient and, with minimal modifications, will be used as part of the final version of the AVACLE evaluation tool.

INTRODUÇÃO

Com base no “Procedimento de Avaliação Cognitiva da Leitura e da Escrita” (SEYMOUR, 1986, 1987, 1992), Pinheiro e colaboradores estão correntemente desenvolvendo um instrumento computadorizado para a avaliação cognitiva das capacidades de leitura e de escrita denominado AVACLE. Ele será aplicado em uma amostra representativa de crianças mineiras cursando o ensino fundamental nas redes pública e particular de ensino de Belo Horizonte, com o objetivo de criar tabelas normativas para essa população. O trabalho, que toma como referencial a teoria de “reconhecimento de palavras” – a investigação dos fatores que exercem influência no processamento de palavras isoladas –, baseia-se na suposição de que a competência na leitura e na escrita, definida respectivamente como a habilidade de ler palavras reais e não-palavras em voz alta e em silêncio e de mostrar compreensão do material lido, e a habilidade de produzir uma escrita convencional para palavras, é o produto de um conjunto de processos psicológicos subjacentes.

O método de Seymour – que atualmente constitui referência para a construção de conjuntos semelhantes de procedimentos em diferentes línguas alfabéticas – consiste no uso de *tarefas de leitura* e de *tomada de ditado* de palavras reais e de não-palavras. As tarefas de leitura subdividem-se em tarefas de leitura em voz alta e tarefas de decisão lexical e semântica – o julgamento sobre as propriedades lexical ou semântica de um estímulo. A seleção dessas tarefas é feita tomando-se como referência as rotas (e sistemas) de um modelo de processamento da leitura e da escrita. Os estímulos de ambas as tarefas são apresentados na tela de um computador e as medidas de *tempo de reação vocal* (TRV – tempo entre reconhecer o estímulo e pronunciá-lo) e de *tempo de reação de decisão* (TRD

– tempo entre reconhecer o estímulo e apertar a tecla “sim” ou “não”) são, respectivamente, registradas para as tarefas de leitura em voz alta e de decisão, além das medidas de taxas de precisão.

As tarefas de *leitura em voz alta* e de *tomada de ditado* do procedimento de Seymour foram testadas com crianças mineiras por Pinheiro (1994 e 1999). Com base nesses estudos e na introdução de tarefas de decisão, o AVACLE consiste nas seguintes tarefas que testam as rotas lexical e fonológica, bem como estruturas e processos específicos dos modelos derivados da teoria de Morton (1980), tais como o modelo de Ellis e Young (1988) e o de Coltheart, Rastle, Perry, Langdon e Ziegler (2001):

- 1) Leitura em voz alta e ditado de palavras reais (teste da rota lexical) e de não-palavras (teste da rota fonológica): palavras reais (variando em frequência de ocorrência, regularidade ortográfica e em número de sílabas) e não-palavras (construídas com a mesma estrutura ortográfica e silábica das palavras reais) são testadas em blocos separados.
- 2) Decisão lexical (teste do sistema de reconhecimento de palavras – léxico ortográfico): uma lista contendo uma mistura de palavras reais (variando em frequência de ocorrência e número de letras) e de não-palavras (construídas com a mesma estrutura ortográfica e silábica das palavras reais) são apresentadas aleatoriamente na tela do computador. Os sujeitos são instruídos a apertar a tecla “sim”, se o estímulo for uma palavra real, e a tecla “não”, se for uma não-palavra.
- 3) Decisão semântica (teste do acesso e da recuperação do significado das palavras da memória semântica): o nome de uma categoria aparece na tela do computador acompanhada de outra palavra. O sujeito deve apertar a tecla “sim”, se a segunda palavra pertencer à mesma categoria da primeira palavra, ou a tecla “não”, caso contrário.

O presente trabalho lida apenas com o item 2. Relata a aplicação de uma tarefa de **decisão lexical** em crianças da 1^a à 4^a série do ensino fundamental com o **objetivo geral** de obter medidas

preliminares que irão fornecer os subsídios para o desenvolvimento de uma versão final dessa tarefa que, por sua vez, irá compor o conjunto de tarefas mencionadas.

A tarefa de decisão lexical testa o *léxico ortográfico*, o sistema do modelo de processamento da leitura que reconhece unidades lexicais (palavras reais) sem acesso a seu significado e sua pronúncia. Envolve o julgamento lexical de uma seqüência de letras ser ou não ser uma palavra real, o que é feito por meio de um toque a teclas específicas de um mecanismo acoplado ao computador. Assim, apresenta-se ao sujeito uma lista contendo uma mistura de palavras reais e de não-palavras (um estímulo de cada vez) e espera-se que ele aperte a tecla “sim”, quando o estímulo apresentado for uma palavra real, e a tecla “não”, quando o estímulo for uma não-palavra. Admite-se que essa tarefa ofereça informações sobre o código ortográfico formado no léxico ortográfico e sobre os processos envolvidos no reconhecimento de palavras (SEYMOUR, 1986; ANDREWS, 1997; COLTHEART *et al.*, 2001).

O léxico ortográfico indica que uma palavra já foi vista antes, mas, para ser entendida, a sua representação semântica deve ser ativada no *sistema semântico* e, para ser pronunciada corretamente, a sua forma falada deve ser ativada no *léxico de produção da fala*, o que constitui uma das formas de leitura lexical. Diante de um estímulo não-familiar ou de uma não-palavra (que, por definição, não tem entrada nos léxicos e no sistema semântico), o sujeito pode responder de três formas, segundo Ellis e Young (1988):

1. Declarando que não é uma palavra real e, portanto, pressionando a tecla “não”.
2. Permitindo que a representação ortográfica de uma palavra familiar visualmente semelhante à não-palavra apresentada seja ativada. Nesse caso, o sujeito comete um erro: pode, por exemplo, tomar a não-palavra “cosa” pela palavra “casa”.
3. Tentando uma pronúncia com base em uma analogia com uma palavra real semelhante à não-palavra, no caso de leitura em voz alta.

É interessante notar que as respostas de número 2 e 3 ocorrem, na maioria dos casos, em decorrência da influência da semelhança ortográfica do estímulo com uma outra palavra, o que é denominado efeito de vizinhança ortográfica. Dessa forma, a não-palavra-alvo é tomada por uma palavra vizinha, resultando em erro, seja na tarefa de decisão lexical, seja na tarefa de leitura em voz alta. A definição mais tradicional de uma palavra vizinha (ortograficamente similar) é a proposta por Coltheart, Davelaar, Jonasson e Besner (1977), que consideram uma palavra vizinha qualquer palavra que pode ser criada ao se mudar uma letra da palavra ou da não-palavra-estímulo, enquanto se preservam as posições das demais letras. Assim, podemos considerar ‘laço’ e ‘lago’ como vizinhos ortográficos da não-palavra ‘labo’. O número de palavras que pode ser criado utilizando o processo proposto por Coltheart *et al.* (1977) é chamado de medida “N” e refere-se ao total de vizinhos da palavra-alvo. Além dessa medida, nos estudos sobre o efeito de vizinhança, também se tem utilizado a medida “NF” que se refere à frequência de ocorrência dos vizinhos da palavra-alvo, ou seja, indica se ela possui vizinhos de alta frequência ou não (ANDREWS, 1997; ZIEGLER e PERRY, 1998). Embora, na tarefa de decisão lexical, ainda existam controvérsias sobre a influência exata do efeito de vizinhança ortográfica no reconhecimento de palavras reais, parece existir um certo consenso de que, para as não-palavras, esse efeito é inibitório, ou seja, em comparação às não-palavras com poucos vizinhos, aquelas com um grande número de vizinhos ortográficos tendem a gerar tempos de reação maiores e menor precisão ao serem reconhecidas (ANDREWS, 1997; ZIEGLER e PERRY, 1998; CARREIRAS, PEREA e GRAINGER, 1997; PEREA e POLLATSEK, 1998; HUNTSMAN e LIMA, 2002).

Em relação às palavras reais, além das medidas N e NF, elas podem variar (nas tarefas de decisão lexical e de leitura em voz alta) em frequência de ocorrência – sendo classificadas em palavras de alta e de baixa frequência – e em número de letras. Normalmente, na tarefa de decisão lexical, escolhem-se palavras irregulares para a leitura por serem esses estímulos processados preferencialmente pela

rota lexical. As não-palavras, que também podem variar nas medidas N e NF, são formadas pela substituição de uma ou mais letras das palavras reais de referência (ou pela inversão de sílabas), mantendo-se, no novo estímulo, a mesma estrutura ortográfica e silábica, assim como o mesmo número de letras das palavras reais de origem (ver Anexo 1). O contraste entre palavras de alta e baixa frequência e o de palavras reais e não-palavras permitem-nos averiguar os efeitos de frequência e de lexicalidade, respectivamente. O resultado do controle da variável número de letras de palavras reais e de não-palavras – efeito de comprimento – não será apresentado aqui.

De acordo com os modelos adotados, e tomando as palavras reais, o reconhecimento mais rápido e correto de palavras de alta frequência em comparação às de baixa frequência – **efeito de frequência** – é interpretado como uma indicação do envolvimento do processo lexical na leitura, enquanto um efeito de frequência exagerado (maior do que normalmente encontrado) indica que as palavras de baixa frequência estão sendo processadas pela rota fonológica.

Considerando o contraste entre palavras reais e não-palavras, as reais, por terem suas identidades ortográficas arquivadas no léxico ortográfico, apresentam vantagem em termos de processamento e/ou de precisão em comparação às não-palavras. Essa vantagem é conhecida como **efeito de lexicalidade** que, como o efeito de frequência, indica a utilização preferencial da via lexical. Embora as palavras reais possam ser lidas pela rota fonológica (por meio do uso das regras de correspondência letra-som), na leitura eficiente são lidas pela rota lexical por meio da ativação de suas representações ortográficas (no léxico ortográfico), semânticas (no sistema semântico) e fonológicas (no léxico de produção da fala) pré-estocadas. Esse processo é muito mais rápido e eficiente do que o fonológico, usado para a leitura de palavras desconhecidas e não-palavras. Por outro lado, a ausência de vantagem no processamento de palavras reais em comparação às não-palavras indica leitura fonológica e, dependendo do nível de leitura do sujeito, grave disfuncionamento do processo lexical.

A investigação desses efeitos, utilizando estímulos referentes à língua portuguesa falada no Brasil, reveste-se de especial relevância tanto teórica quanto prática. Do lado teórico, permite verificar a validade de achados empíricos e de teorias de pesquisadores que lidaram com estímulos relacionados a outros idiomas, sobre o processamento e dificuldades de leitura, no que diz respeito ao nosso idioma. Do ponto de vista prático, fornece subsídios para a construção de um instrumento padronizado de avaliação cognitiva de leitura, em português utilizado no Brasil, o que não só possibilitará o diagnóstico eficaz e eficiente das dificuldades de leitura, como também facilitará a proposição de ações preventivas e corretivas no tratamento dessas dificuldades.

O trabalho tem os seguintes objetivos: (1) verificar a ocorrência do efeito de frequência (leitura mais rápida e/ou mais correta de palavras de alta frequência em relação às palavras de baixa frequência) no nível do léxico ortográfico; (2) comparar a precisão e o TRD no reconhecimento de palavras reais e de não-palavras; (3) comparar o desempenho no julgamento de palavras reais e de não-palavras, em termos de TRD e precisão, de crianças das séries iniciais às finais do ensino fundamental; (4) obter subsídios para construção da versão final da tarefa de decisão lexical que comporá a bateria AVACLE (Avaliação cognitiva de leitura e escrita).

HIPÓTESES

1. Se o efeito frequência ocorre no nível do léxico ortográfico, espera-se que as palavras de alta frequência apresentem tempos menores de decisão do que os apresentados pelas palavras de baixa frequência.
2. Espera-se que as palavras reais apresentem vantagem em termos de processamento e/ou de precisão em comparação às não-palavras.
3. No que se refere ao desenvolvimento do léxico ortográfico, espera-se uma diminuição no tempo de processamento e/ou

aumento de precisão no reconhecimento de palavras reais das séries iniciais às finais. O mesmo resultado é esperado com relação às não-palavras. À medida que a identidade das palavras se torna mais bem estabelecida no léxico ortográfico, as não-palavras são rejeitadas mais rapidamente.

MÉTODO

Trata-se de um estudo transversal do desenvolvimento que compara o desempenho em uma tarefa de decisão lexical de sujeitos de várias séries escolares, não-portadores de necessidades especiais, portanto com grande probabilidade de serem leitores competentes.

Antes de prosseguirmos com a descrição dos procedimentos, cabem aqui algumas considerações sobre a análise estatística dos estudos que usam a medida de TR como variável dependente – especialmente as investigações que usam materiais lingüísticos. Nesses estudos, o tratamento dos dados consiste normalmente em tirar a média aritmética dos estímulos testados em cada condição para cada sujeito (por exemplo, no presente estudo, os estímulos em cada condição são palavras reais, palavras de alta e baixa freqüência e não-palavras) e fazer a análise estatística nas médias das condições para o número de sujeitos testados. Uma análise significativa desse tipo permite a generalização dos resultados para um grupo particular de sujeitos selecionados randomicamente da população de onde foram derivados. Esse procedimento é chamado de “análise de sujeitos”. No entanto, Clark (1973) chama a atenção para o problema em fazer-se apenas esse tipo de análise: é que os estímulos, os itens testados em cada condição são tratados como um fator fixo. Geralmente, um conjunto de estímulos é escolhido na crença de que os itens do conjunto são representativos de todos os itens que podem constituir uma certa condição e pressupõe-se que, se eles mostrarem um efeito particular, o mesmo efeito será encontrado na população de itens da condição considerada. Uma análise de sujeitos não oferece nenhuma informação sobre isso. Por exemplo, se 10 itens

fossem testados em cada uma de duas condições, teoricamente, uma diferença significativa entre as condições poderia ser demonstrada pela análise de sujeitos se apenas um dos itens diferisse do restante. O efeito desse item poderia ser reduzido pelos nove itens restantes na condição, mas, se todos ou a maioria dos sujeitos mostrassem um efeito nesse item, ele poderia ainda causar uma pequena diferença consistente entre as condições e então resultar em uma diferença significativa na análise de sujeitos entre as condições. Se outro conjunto de itens da condição em que se obtiveram diferenças significativas for testado, poderá não ser encontrado o mesmo efeito. Um resultado que é causado por somente um número limitado de itens em uma condição pode não ser devido a nenhuma característica desses itens que seja relevante para o experimento.

Segundo Clark (1973), para generalizar a partir de um conjunto particular de itens experimentais para a população de itens que forma uma condição, é necessário, pelo menos, fazer uma “análise de itens” em que os itens em cada condição sejam tratados como um efeito randômico e efetuar uma análise sobre as médias obtidas para cada item pelo conjunto de sujeitos, isto é, em cada condição, calcula-se a média de TR do grupo de sujeitos para cada item, e a análise estatística é feita nas médias resultantes. Uma análise de itens requer, para obter-se uma diferença significativa entre as condições, que se consiga, em primeiro lugar, uma diferença consistente sobre os itens. Conseqüentemente, uma análise desse tipo permite generalizações dos estímulos específicos do material usado em um experimento para a população de itens da qual o conjunto de itens experimentais foi derivado.

Com base no exposto, Pinheiro (1994, 1995), seguindo as recomendações de Clark (1973), conduziu em seu estudo sobre o desenvolvimento de reconhecimento e produção de palavras por crianças tanto a análise de sujeitos (análise das respostas produzidas pelos sujeitos) como a análise de itens (análise das respostas dadas a cada item). Aqui, adotamos para os resultados de tempo de reação o mesmo procedimento de Pinheiro. Assim, a análise de sujeito será

referida como AS e a análise de itens como AI. Para os erros, efetuaremos apenas a análise de sujeito.

Outro procedimento estatístico usado foi a identificação de sujeitos que se distanciaram do grupo de referência por apresentarem respostas discrepantes, conhecidas tecnicamente pelo termo “*outliers*”, que são respostas que se distanciam da média obtida pelo grupo por volta de 2,5 desvios-padrão em qualquer direção. Feita essa análise, optamos por excluir da análise estatística um sujeito de 1ª série que, em termos de tempo de processamento, foi considerado um *outlier*. Com isso, o número de sujeitos da primeira série decresceu de 21 para 20.

AMOSTRA

A amostra constituiu-se de 68 crianças da 1ª à 4ª série do ensino fundamental, matriculadas no Centro Pedagógico da UFMG, em Belo Horizonte. A composição da amostra por série foi a seguinte: 21 crianças da primeira série, 14 da segunda, 17 da terceira e 16 da quarta.

INSTRUMENTO

A tarefa experimental – uma lista contendo igual número de palavras reais e de não-palavras – foi administrada por meio de um microcomputador portátil, dirigido pelo programa *Mel Professional* (comercializado pela *Psychology Software Tools, Inc.*, uma empresa radicada em Pittsburgh, Estados Unidos) que possui recursos para medir o TRD (tempo entre o aparecimento do estímulo e o toque nas teclas “sim” ou “não”), em milésimos de segundo, e controlar intervalos entre estímulos. Os itens da lista foram apresentados em letras minúsculas, no centro da tela do monitor, e foram vistos a uma distância de 50 a 75 centímetros, dependendo dos movimentos dos sujeitos. A cada tentativa, um estímulo (uma palavra real ou uma não-palavra) era apresentado no centro da tela do micro, precedido de

um sinal de alerta. Assim que o sujeito apertava a tecla “sim” ou “não”, o TRD era marcado.

PROCEDIMENTO

O estudo foi realizado em um período de dois meses. Os testes foram aplicados por bolsistas dos projetos PAD e PIBIC e por alunas voluntárias. O procedimento consistiu na aplicação da tarefa de decisão lexical juntamente com mais duas outras tarefas: um teste de leitura em voz alta de palavras reais e um teste de decisão semântica. Descreveremos aqui apenas o procedimento da tarefa de decisão lexical.

Na tarefa de decisão lexical foi apresentada a cada criança uma lista composta de 24 itens: 12 palavras reais e 12 não-palavras. As palavras reais variavam em frequência de ocorrência (seis palavras de alta frequência (AF) e seis de baixa frequência (BF), retiradas da Contagem de Frequência de Ocorrência de Palavras de Pinheiro (1996), e em comprimento (quatro a seis letras). Todas as palavras reais eram palavras irregulares para a leitura: palavras contendo um <x> intervocálico e os vocálicos <o> ou <e> em posição paroxítona (PARENTE, SILVEIRA e LECOURS, 1997). As não-palavras (NP) foram construídas com as mesmas estruturas ortográficas e comprimentos das palavras reais. As listas de palavras reais e de não-palavras são apresentadas no Anexo 1. O efeito de vizinhança ortográfica não foi controlado nesse experimento.

Os sujeitos foram testados individualmente em uma sessão. Inicialmente, explicou-se à criança como executar as tarefas a serem desempenhadas, o que foi demonstrado em três treinamentos, um para cada tarefa aplicada (leitura em voz alta, decisão semântica e decisão lexical). O treinamento para a tarefa de decisão lexical foi introduzido por meio da seguinte instrução: “Você vai ver dois tipos de palavras escritas que serão apresentadas uma de cada vez. Um tipo de palavra existe no português, sendo, portanto, uma ‘palavra real’; o outro tipo é uma palavra que não existe no português, sendo,

portanto, uma ‘palavra inventada’. Aperte a tecla SIM quando for uma palavra real e a tecla NÃO quando for uma palavra inventada”. A seguir, introduziu-se o treinamento que consistiu de uma lista com 20 itens (10 palavras reais e 10 não-palavras). O treinamento para as três tarefas durava por volta de 15 minutos. Após essa sessão, os testes eram introduzidos e a seqüência de apresentação deles variava de sujeito a sujeito. Para cada sujeito, os itens de cada teste eram embaralhados pelo computador. O teste foi realizado no laboratório de Psicologia do DECAE da FAE/UFMG.

As respostas foram automaticamente classificadas como corretas ou incorretas pelo programa.

RESULTADOS

Os resultados consistiram em medidas de TRD para as respostas corretas e em porcentagem de erros para a leitura de palavras reais e de não-palavras. As médias de TRD e de proporção de erros do grupo de sujeitos de cada série foram comparadas por análise de variância geral (*overall*) com medidas repetidas. Os fatores intra-sujeitos foram freqüência e lexicalidade. Os resultados da análise de sujeito – apresentados na Tabela 1 – indicam que, no que se refere ao TRD, a variabilidade dos resultados em cada grupo sofre influência das variáveis controladas. Isso significa que a leitura de todos os grupos de sujeitos é afetada pela variável freqüência de palavras e pela variável lexicalidade. Esse resultado foi confirmado pela análise de itens: amplitude de $F = (2,21) 14,57 - 22,98$, $p < 0,001$. As variáveis freqüência e lexicalidade também influenciaram os erros cometidos, mas, ao contrário dos resultados de tempo de decisão, os efeitos não foram observados em todos os grupos: o efeito de freqüência foi significativo apenas para a 2ª série e o lexical para a 3ª e 2ª séries. Na 3ª série, o efeito foi maior e, na 2ª, limitou-se à comparação entre palavras de alta freqüência e não-palavras.

TABELA 1

Resumo dos resultados da análise de variância (análise de sujeito) feita nas médias de tempo de reação de decisão (TRD) e de porcentagens de erros dos fatores intra-sujeito (frequência e lexicalidade) para os quatro grupos de crianças da 1ª à 4ª série.

Nº de Sujeitos	Série	TR F	Nível Sig. F	% de Erros	Nível Sig.
21	1ª	(3,60) 20,022	0,000***	0,608	0,612 ns.
14	2ª	(3,39) 23,512	0,000***	4,046	0,013**
17	3ª	(3,48) 42,262	0,000***	5,021	0,004*
16	4ª	(3,45) 21,370	0,000***	0,490	0,691 ns.

* = 0,05; ** = 0,01; *** = 0,001

As análises subseqüentes consistiram de *t*-testes feitos nos resultados de cada grupo entre as médias de TRD e de porcentagem de erros para avaliar os efeitos de *frequência* e *lexical*. Na primeira situação, os itens de alta frequência (AF) e de baixa frequência (BF) foram comparados entre si e, na segunda, as palavras reais (com os dois níveis de frequência juntos) foram comparadas com as não-palavras (NPs) para avaliar o *efeito lexical geral*. A seguir, as médias de cada nível de frequência foram comparadas com as médias das não-palavras (palavras AF x NPs e palavras BF x NPs) – *efeito lexical específico*. Essa última comparação teve como finalidade a investigação do efeito lexical em um nível mais avançado, isto é, averiguar se a leitura de palavras de baixa frequência (que, em princípio, é processada mais lentamente e menos corretamente do que as palavras reais) apresenta vantagem em termos de tempo de processamento e/ou de precisão sobre a leitura de não-palavras. Esses resultados são apresentados nas Tabelas 2 e 3. Finalmente, o progresso dos quatro grupos foi comparado por *t*-teste a fim de avaliar o efeito de experiência na rapidez e precisão de reconhecimento de palavras reais e não-palavras.

TABELA 2

(1) Médias de TRD para as palavras AF, BF, PRs (palavras AF e BF juntas) e NPs;
 (2) efeitos de frequência e lexical (geral e específico) e resultados de t-teste (análise de sujeito) para os quatro grupos de crianças da 1ª à 4ª série.

Nº Suj.	Série	AF		BF		EF FREQ	T	SIG
		M	DP	M	DP			
20	1ª	2654	781	3040	877	387	(19)2,466	0,023*
14	2ª	2363	884	2860	1233	497	(13)2,696	0,018*
17	3ª	1653	635	1972	829	319	(16)3,288	0,005**
16	4ª	1586	383	1817	610	231	(15)1,945	0,071(*)

Nº Suj.	Série	PR		NP		EF LEXIC		
		M	DP	M	DP	GERAL	T	SIG
20	1ª	2829	744	3837	929	1008	(19)6,113	0,000***
14	2ª	2598	994	4089	1233	1491	(13)5,390	0,000***
17	3ª	1807	710	2672	945	865	(16)7,467	0,000***
16	4ª	1703	450	2756	898	1057	(15)5,008	0,000***

Nº Suj.	Série	AF X NP			BF X NP		
		EF LEXIC	T	SIG	EF LEXIC	T	SIG
20	1ª	1183	(19)6,448	0,000***	796	(19)4,360	0,000***
14	2ª	1726	(13)5,904	0,000***	1229	(13)4,143	0,001***
17	3ª	1019	(16)8,315	0,000***	700	(16)5,482	0,000***
16	4ª	1170	(15)5,758	0,000***	939	(15)4,03	0,001***

*=0,05; **=0,01; ***0,001

TABELA 3

(1) Médias de porcentagens de erros para as palavras AF, BF, PRs (palavras AF e BF juntas) e NPs; (2) efeitos de frequência e lexical (geral e específico) e resultados de t-teste (análise de sujeito) para os quatro grupos de crianças da 1^a à 4^a série.

Nº Suj.	Série	AF		BF		EF FREQ	T	SIG
		M	DP	M	DP			
20	1 ^o	5,83	9,79	10,00	11,34	4,17	(19)1,228	0,234
14	2 ^o	2,38	6,05	14,28	18,32	11,9	(13)2,925	0,012**
17	3 ^o	1,96	5,54	5,88	11,7	3,92	(16)1,289	0,216
16	4 ^o	2,08	5,7	3,13	6,71	1,05	(15)0,565	0,580

Nº Suj.	Série	PR		NP		EF LEXIC		
		M	DP	M	DP	GERAL	T	SIG
20	1 ^o	7,91	7,39	10,83	12,71	2,92	(19)0,098	0,445
14	2 ^o	8,33	11,32	12,5	10,71	4,17	(13)1,023	0,325
17	3 ^o	3,92	6,66	12,25	12,54	8,33	(16)2,675	0,017*
16	4 ^o	2,6	5,02	4,17	6,804	1,57	(15)0,764	0,457

Nº Suj.	Série	AF X NP			BF X NP		
		EF LEXIC	T	SIG	EF LEXIC	T	SIG
21	1 ^o	5,0	(19)1,283	0,214	0,083	(19)0,193	0,849
14	2 ^o	10,12	(13)3,076	0,009**	1,78	(13)0,323	0,752
17	3 ^o	10,29	(16)3,104	0,007**	6,37	(16)1,765	0,097(*)
16	4 ^o	2,09	(15)0,939	0,363	1,04	(15)0,459	0,653

* = 0,05; ** = 0,01; *** = 0,001

Como pode ser visto na Tabela 2, e de acordo com as expectativas, em cada série, as palavras reais foram reconhecidas mais rapidamente do que as não-palavras. A análise de itens confirmou os resultados apresentados na análise de sujeito para todas as séries nas três comparações efetuadas entre palavras reais e não-palavras: (1) PR x NP (amplitude de $t^1 = (1, 22) 4,715 - 5,851$, $p < 0,001$); (2) AF x NP

(amplitude de $t = (1, 16) 4,526 - 8,317$, $p < 0,001$); (3) BF x NP (amplitude de $t = (1, 16) 2,650 - 3,635$, $p < 0,001$). Na análise de erros (Tabela 3), o efeito lexical geral foi significativo apenas para as crianças da 3ª série. Esse efeito deveu-se ao contraste entre as palavras de alta frequência e não-palavras, já que o contraste entre palavras de baixa frequência e não-palavras não foi significativo (assim como não o foi para nenhuma das outras séries) indicando, portanto, que algumas palavras de baixa frequência foram consideradas como não-palavras por toda a amostra de crianças.

O efeito de frequência foi significativo em TRD e precisão na 2ª série e em TRD na 1ª e 3ª séries. Na 4ª série, apresentou-se bem próximo à significância em termos de TRD (Tabelas 2 e 3). A análise de itens confirmou o efeito de frequência em TRD para a 1ª série: $t = (1,10) 2,34$, $p < 0,05$ e para a 2ª série: $t = (1, 10) 2,66$, $p < 0,05$, indicando que o efeito encontrado na análise de sujeitos para a 3ª série deveu-se a apenas alguns itens e que pode ser generalizado somente para essa amostra de sujeitos, mas não para os itens testados.

A Tabela 2 também mostra que o desenvolvimento da leitura de palavras reais se caracteriza por uma diminuição de TRD da 1ª à 4ª série. A mesma tendência não foi observada para a leitura de não-palavras: o TRD para esses estímulos tendeu a ser levemente maior para a 2ª série em relação à 1ª série e para a 4ª série em relação à 3ª série. As comparações entre pares de séries *consecutivas* (*t*-testes) indicaram que a diminuição de tempo de processamento foi significativa para a diferença entre a 2ª e 3ª séries tanto para palavras reais (AS (1,29) = 2,580, $p < 0,05$; AI (1,22) = 8,496, $p < 0,001$) quanto para não-palavras (AS (1,29) = 3,623, $p < 0,01$; AI (1,22) = 6,732, $p < 0,001$), mostrando que é a essa altura que ocorrem ganhos significativos em termos de rapidez de reconhecimento de palavras e de rejeição de não-palavras. A mesma tendência foi observada para o desenvolvimento de palavras de alta e baixa frequência. Diferenças significativas entre pares de séries consecutivas foram encontradas apenas entre a 2ª e 3ª séries tanto para palavras de alta frequência (AS (1,29) = 2,597, $p < 0,5$; AI (1,10) = 9,338, $p < 0,001$), como para de baixa frequência (AS (1,29) = 2,390 $p \leq 0,05$; AI (1,10) = 7,582, $p < 0,001$).

Já em termos de erros, podemos ver na Tabela 3 que o desenvolvimento da precisão na leitura mostra um padrão de resultados semelhante ao apresentado em tempo de decisão: a porcentagem de erros tendeu a declinar no reconhecimento de palavras reais e, nas não-palavras, caracterizou-se por descontinuidade de decréscimo de erro nas séries intermediárias. As comparações entre pares de séries *consecutivas* para as porcentagens de erros nas palavras reais não mostraram resultados significativos e, nas não-palavras, a mesma comparação revelou decréscimo significativo nas porcentagens de erros somente entre a 3ª e 4ª séries ($AS(1,31) = 2,282, p < 0,05$; $AI(1,22) = 3,52, p < 0,05$).

Considerando o desenvolvimento da rapidez e da precisão na tarefa de decisão lexical ao longo das séries – efeito de experiência escolar –, os resultados evidenciam que, enquanto, em termos de tempo de processamento, é na passagem da 2ª para a 3ª série que se observam mudanças significativas em desempenho de palavras reais e não-palavras, o desenvolvimento da precisão não segue a mesma curva, pois não houve mudanças significativas nas palavras reais. O efeito de experiência escolar influenciou também a precisão, mas a melhora de desempenho nesta ocorreu mais tarde, entre as séries finais, e se limitou às não-palavras.

Esse padrão de resultados se assemelha aos resultados obtidos por Pinheiro (1994, 1995), nas tarefas de leitura em voz alta. A autora observou que, no desenvolvimento da leitura, só depois que as crianças obtêm um certo automatismo em tempo de processamento é que começam a mostrar ganhos em precisão. Em outras palavras, nas fases iniciais do processo de aquisição da leitura, primeiro adquire-se o automatismo e depois a precisão.

DISCUSSÃO

Como vimos, tanto o efeito de frequência quanto o lexical sinalizam para uma leitura preferencialmente pela via lexical. No caso desta pesquisa, esses efeitos mostram a formação do léxico

ortográfico ao longo do desenvolvimento. A presença do efeito de frequência e lexical já na leitura das crianças da 1ª série indica que, no português, como em outras línguas, o léxico ortográfico, mesmo em fases iniciais de formação e em termos de tempo de processamento, distingue as palavras de alta frequência das de baixa e as palavras reais das não-palavras. Esse dado revela que as palavras de alta e baixa frequência são processadas de forma diferente no léxico e que as palavras reais – mesmo as de baixa frequência –, apresentam vantagens em termos de reconhecimento em relação às não-palavras, denotando, portanto, o uso da via lexical para a leitura de palavras reais e da via fonológica para a leitura de não-palavras.

O efeito lexical observado na 1ª série se manteve forte na leitura das crianças das séries restantes. As palavras reais apresentam vantagem em comparação às não-palavras porque possuem representações ortográficas, semânticas e fonológicas arquivadas nos diferentes sistemas da rota lexical, o que permite reconhecimento instantâneo, ao contrário da leitura mediada pelo processo de conversão grafema-fonema (rota fonológica) efetuada no caso das não-palavras e de algumas palavras pouco familiares. Esses resultados confirmam a hipótese número 2, em que foi antecipado que as palavras reais apresentariam vantagem em termos de processamento e/ou de precisão em comparação às não-palavras.

O efeito de frequência observado em tempo de processamento na 1ª série estendeu-se, na 2ª série, aos erros, indicando que alguns dos itens de baixa frequência foram tratados como não-palavras. Na 3ª série, o efeito permanece apenas em tempo de processamento e se limitou à análise de sujeitos. Na 4ª série, a despeito do maior tempo de processamento das palavras de baixa frequência em relação às palavras de alta frequência, a diferença obtida não atinge significância na análise de sujeito e menos ainda na análise de itens. A diminuição da força do efeito na 3ª série, combinada com a sua ausência na 4ª série, pode indicar processamento automático e igualmente eficaz para os dois tipos de estímulos. Em decorrência de sucessivas exposições a palavras de baixa frequência, os leitores

experientes passam a reconhecer esses estímulos cada vez mais rapidamente, o que afeta o efeito de frequência, que passa a diminuir. Pinheiro (2001), analisando o efeito de frequência em uma tarefa de leitura em voz alta, observou que esse efeito tende a aparecer na leitura de alguns sujeitos, mas não de todos, e que o número de sujeitos que não mostram tal efeito aumenta como resultado da idade e de experiência acadêmica.

Uma explicação alternativa para a falta de diferença significativa entre as palavras de alta e baixa frequência na análise de itens na 3ª e 4ª séries seria que esse resultado pode ter ocorrido devido a um efeito de vizinhança ortográfica, que não foi controlado neste estudo. Em línguas com um nível de regularidade relativamente próximo ao do português, como o espanhol, tem-se encontrado um efeito inibitório da vizinhança ortográfica no reconhecimento de palavras, utilizando-se a tarefa de decisão lexical (ANDREWS, 1997; CARREIRAS *et al.*, 1997). Desse modo, é de se esperar que palavras e não-palavras que tenham um grande número de vizinhos ortográficos tendam a gerar tempos de reação mais longos e a produzir mais erros. Um efeito inibitório de vizinhança baseia-se no pressuposto de que existe uma certa competição no acesso lexical. Assim, quanto maior o léxico do leitor, maior é a probabilidade de haver palavras vizinhas que possam competir com a palavra-alvo armazenada em seu léxico.

Caso o efeito de vizinhança possa ser uma variável importante no reconhecimento de palavras no português brasileiro – investigação em fase de desenvolvimento por Justi e Pinheiro –, o que pode ter acontecido no presente estudo é que, à medida que os sujeitos da 3ª e 4ª séries se tornavam mais experientes na habilidade de ler, apresentavam uma maior probabilidade de serem afetados pelo efeito de vizinhança ortográfica, posto ser o seu léxico, provavelmente, maior do que o léxico dos sujeitos de 1ª e 2ª séries. Portanto, existe a possibilidade de que alguns itens possam ter sido afetados por seus vizinhos nos sujeitos mais experientes, sendo, portanto, processados mais lentamente do que os outros itens da mesma categoria, o que explicaria a falta de significância encontrada na análise de itens para

o efeito de freqüência. Por exemplo, na análise de itens das respostas das crianças da 4ª série, percebemos que a palavra de baixa freqüência “moço” foi processada muito mais lentamente do que as outras palavras de baixa freqüência (aproximadamente dois desvios-padrão acima da média dessas). Uma análise mais detalhada desse estímulo mostra que ele possui nove vizinhos ortográficos – “coço”, “poço”, “maço”, “meço”, “modo”, “mofo”, “moro”, “moto” e “moça” –, sendo que um deles, “modo”, é uma palavra de alta freqüência, o que a coloca como um competidor mais forte que a palavra-alvo “moço”, tornando alta a probabilidade de que essa palavra possa ter sofrido um efeito inibitório de seus vizinhos ortográficos. O maior tempo de processamento para a palavra “moço” em relação às outras palavras da mesma categoria pode, portanto, ter sido a causa da ausência de resultados significativos na análise de itens já que, como dissemos anteriormente, a ausência de confirmação de um efeito encontrado na análise de sujeito pela análise de itens – caso da 3ª série – indica que o efeito observado (ou quase observado, no caso da 4ª série) na primeira análise pode ter sido devido a apenas um ou alguns itens, e não aos itens como um todo.

Como a variável vizinhança ortográfica (tanto nas palavras reais como nas não-palavras) não foi manipulada nesse estudo, não encontramos outras palavras de baixa freqüência com muitos vizinhos para que pudéssemos averiguar se essas palavras, como “moço”, mostrariam altos tempos de processamento, o que inevitavelmente validaria a nossa suposição. Assim, apontamos não só para a relevância de estudos sobre o efeito de vizinhança ortográfica no português brasileiro, como também sugerimos que esse efeito venha a ser controlado em listas de palavras subseqüentes.

A despeito da ausência do efeito de freqüência na leitura das crianças mais velhas de nosso estudo, os resultados aqui apresentados corroboram a hipótese número 1. A ocorrência do efeito freqüência na leitura das crianças da 1ª e 2ª séries mostra que o *locus* (ou pelo menos um dos *locus*) desse efeito pode estar no léxico ortográfico, o que está de acordo com Morton (1969) e Ellis (1984).

No que se refere ao desenvolvimento do léxico ortográfico, espera-se uma diminuição no tempo de processamento e/ou um aumento de precisão no reconhecimento de palavras reais das séries iniciais às finais. O mesmo resultado é esperado com relação às não-palavras. À medida que a identidade das palavras se torna mais bem estabelecida no léxico ortográfico, as não-palavras são rejeitadas mais rapidamente.

Considerando o desenvolvimento do léxico ortográfico, a sensível diminuição no tempo de decisão tanto das palavras reais como das não-palavras nas séries intermediárias mostra que é a essa altura que ocorrem ganhos significativos em termos de rapidez de reconhecimento de palavras e de rejeição de não-palavras. A melhora de desempenho entre séries consecutivas ocorre mais tarde em termos de acurácia do que para tempo de processamento e é restrita à facilidade de rejeição de não-palavras. Esses resultados confirmam a hipótese 3, apesar de a diminuição de tempo de processamento e o aumento de precisão não terem ocorrido de forma contínua.

Finalmente, a tarefa de decisão lexical desenvolvida revelou-se eficaz e poderá, com poucas modificações (p. ex., a troca do item “moço” por um item equivalente de baixa frequência, mas com baixo “N”), ser usada na versão final da bateria AVACLE. Temos, pois, os primeiros parâmetros para definir o que se espera, em termos de reconhecimento de palavras, ao nível do léxico visual de entrada, de crianças nas séries iniciais do ensino fundamental. Qualquer distanciamento deste parâmetro pode denotar indícios de dificuldade de leitura.

NOTA

¹ Refere-se à amplitude dos *ts* obtidos da 1ª à 4ª série.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDREWS, S. The effect of orthographic similarity on lexical retrieval: resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, v. 4, n. 4, p. 439-461, 1997.
- CARREIRAS, M.; PEREA, M.; GRAINGER, J. Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, v. 23, n. 4, p. 857-871, 1997.
- CLARK, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behaviour*, v. 12, p. 334-359, 1973.
- COLTHEART, M.; DAVELAAR, E.; JONASSON, J.; BESNER, D. Access to the internal lexicon. In: DORNIC, S. (Ed.). *Attention and performance VI*. Hillsdale, NJ: Erlbaum, 1977.
- COLTHEART, M.; RASTLE, K.; PERRY, C.; LANGDON, R.; ZIEGLER, T. *DRC: A dual-route cascaded model of visual word recognition and reading aloud*. *Psychological Review*, v. 108, n. 1, p. 204-256, 2001.
- ELLIS, A. W. *Reading writing and dyslexia: a cognitive analysis*. London: Lawrence Erlbaum Associates, 1984.
- ELLIS, A., YOUNG, A. W. *Human cognitive neuropsychology*. London: Lawrence Erlbaum, 1988.
- HUNTSMAN, L.; LIMA, S. Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research*, v. 31, n. 3, p. 289-306, 2002.
- JUSTI, R. R.; PINHEIRO, A. M. V. *O efeito de vizinhança ortográfica nas palavras e não-palavras do português brasileiro: um estudo transversal*. Projeto de Mestrado em desenvolvimento, Mestrado em Psicologia, FAFICH/UFMG.
- MORTON, J. The interaction of information in word recognition. *Psychological Review*, v. 76, p. 165-178, 1969.
- MORTON, J. The logogen model and orthographic structure. In: FRITH, U. (Ed.). *Cognitive processes in spelling*. London: Academic Press, 1980.
- PARENTE, M. A. M. P.; SILVEIRA, A.; LECOURS A. R. As palavras do português. In: LECOURS, A. R.; PARENTE, M. A. M. P. (Org.). *Dislexia: Implicações do sistema de escrita do português*. Porto Alegre: Artes Médicas, 1997. p. 41-56.
- PEREA, M.; POLLATSEK, A. The effects of neighborhood frequency in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, v. 24, n. 3, p.767-779, 1998.

PINHEIRO, A. M. V. *Leitura e Escrita: Uma abordagem cognitiva*. Editorial Psy, Campinas, 1994.

PINHEIRO, A. M. V. Reading and spelling development in Brazilian Portuguese. *Reading & Writing*, special issue on Literacy Acquisition. 7/1 mar., p. 111-138, 1995.

PINHEIRO, A. M. V. *Contagem de frequência de ocorrência e análise psicolinguísticas de palavras expostas a crianças na faixa pré-escolar e séries iniciais do 1º grau*. São Paulo: Associação Brasileira de Dislexia, 1996.

PINHEIRO, A. M. V. Cognitive assessment of competent and impaired reading in scottish and brazilian children. *Reading & Writing*, 11/3 jun., p. 175-211, 1999.

PINHEIRO, A. M. V. Heterogeneidade entre leitores julgados competentes pelas professoras. *Psicologia Reflexão e Crítica*, v. 14, p. 3, p. 537-551, 2001.

SEYMOUR, P.H.K. *Cognitive analysis of dyslexia*. London: Routledge & Kegan Paul, 1986.

SEYMOUR, P.H.K. Individual cognitive analysis of competent and impaired reading. *British Journal of Psychology*, v. 78, p. 483-506, 1987.

SEYMOUR, P.H.K. The Assessment of Reading Disorders. In: PARKER, J. K., D. M.; MACKINLAY, W.W. (Ed.). *A Handbook of Neuropsychological Assessment*. London: Lawrence Erlbaum Associates, 1992.

ZIEGLER, J. ; PERRY, C. No more problems in Coltheart's neighborhood: resolving neighborhood conflicts in the lexical decision task. *Cognition*, 68, B53-B62, 1998.

AGRADECIMENTOS

Aos alunos do Centro Pedagógico da UFMG que participaram como sujeitos da pesquisa e às coordenadoras Renata Marink Medeiros e Viviane Marques Tomás por sua presteza e disponibilidade.

Às alunas que aplicaram as tarefas:

Cíntia Yoshihara - bolsista do projeto PAD;

Simone Eiterer Campos de Carvalho - bolsista do projeto PAD;

Samantha de Toledo Martin - bolsista do projeto PIBIC;

Sávia Duarte de Menezes - bolsista do projeto PIBIC;

Alícia Alcântara Barreto Araújo - aluna voluntária;

Graziele Alves Amaral - aluna voluntária.

À Pró-Reitoria de Pesquisa da UFMG (projeto PAD) e PIBIC (CNPq).

Nota sobre os autores

A primeira autora é Professora Titular do Departamento de Psicologia da UFMG. É doutora em Psicologia Cognitiva, pela Universidade de Dundee, Escócia. Desenvolve pesquisas na área de aquisição e desenvolvimento da leitura e da escrita, com ênfase no processo de avaliação destas habilidades e comparações translingüísticas entre português e inglês. Desenvolve também pesquisas na área de Psicolingüística, destacando-se, nessa área, uma contagem de frequência de ocorrência de palavras expostas a crianças mineiras cursando a pré-escola (3º período) e as séries iniciais do ensino fundamental.

A segunda autora, Professora Anna Edith Bellico da Costa, é Professora Titular aposentada da UFMG, sendo atualmente Professora Titular do Curso de Psicologia da Faculdade Metropolitana de Belo Horizonte, onde é coordenadora do Curso. É doutora em Psicologia Social, pelo ISOP/FGV-RJ e mestre em Psicologia Teórico-Experimental pela PUC-RJ. Desenvolve pesquisas sobre aprendizagem social e cognição social e atualmente integra um programa de validação de instrumentos desenvolvido pela Freie Universitat de Berlin, realizando pesquisas sobre a validação de auto-eficácia e construtos paralelos para o contexto brasileiro.

O terceiro autor cursa o programa de Mestrado em Psicologia da UFMG, participa do grupo de pesquisa AVACLE e é bolsista do CNPq.

ANEXO 1**Listas de palavras reais e de não-palavras**

PALAVRAS REAIS		NÃO-PALAVRAS	
CÓDIGO	ESTÍMULO	CÓDIGO	ESTÍMULO
01 Y AF4	BOLA	N (NP)L4	LABO
02 Y AF4	MEIO	N (NP)L4	JEIA
03 Y BF4	MOÇO	N (NP)L4	VOÇO
04 Y BF4	FEIO	N (NP)L4	REIO
05 Y AF5	VERDE	N (NP)L5	RERVE
06 Y AF5	FORTE	N (NP)L5	ZORMA
07 Y BF5	VESTE	N (NP)L5	CASGA
08 Y BF5	GORDO	N (NP)L5	SORPO
09. YAF6	PESSOA	N (NP)L6	DERROA
10. YAF6	CABEÇA	N (NP)L6	LEPEÇA
11 Y BF6	AMEIXA	N (NP)L6	AJEILA
12 Y BF6	FAXINA	N (NP)L6	BOXIMA

Nº total de PRs = 12

Nº total de NPs Legais = 12