

**Universidade do Porto**

Faculdade de Psicologia e de Ciências da Educação

# **Ler numa Ortografia de Profundidade Intermédia**

O Papel da Conversão Grafema-Fonema em Português

**César F. Lima**

Dissertação de Mestrado

Orientação de Prof. São Luís Castro

Janeiro de 2008

## **Agradecimentos**

Gostaria de tornar público o meu reconhecimento a um conjunto de pessoas que, de forma mais ou menos directa, constituíram um suporte importante para a realização do trabalho que se segue.

Em primeiro lugar, agradeço à Prof. São Luís Castro, pela orientação exigente e estimulante desde o primeiro momento e por ter despertado o meu interesse pela investigação.

Agradeço a todas as pessoas que aceitaram participar nas experiências apresentadas, a quem devo os resultados.

Agradeço à Prof. Selene Vicente, por ter possibilitado o acesso a parte dos participantes.

Por fim, agradeço aos colegas do Laboratório de Fala e do “Grupo de Linguagem”, pelas interessantes discussões científicas e apoio informal, que em muito contribuíram para o resultado final desta dissertação.

## Resumo

Vários estudos têm demonstrado que a conversão grafema-fonema é importante na leitura hábil em ortografias transparentes, enquanto nas profundas a sua utilização é reduzida e o conhecimento lexical assume maior preponderância. Neste trabalho examinámos o papel da conversão grafema-fonema numa ortografia de profundidade intermédia, o Português, através da análise do impacto da extensão (medida em número de letras) na leitura. Em dois estudos, jovens adultos leram palavras e não-palavras dissilábicas que variavam na frequência e na extensão. No primeiro, compararam-se itens de cinco e seis letras em tarefas de decisão lexical e de leitura em voz alta com listas mistas (design inter-sujeitos). No segundo estudo, itens entre quatro e seis letras foram apresentados em tarefas de decisão lexical e de leitura em voz alta, com listas mistas e com listas agrupadas (design inter-sujeitos). Os resultados evidenciaram efeitos de extensão robustos nas não-palavras. Nas palavras, a extensão teve impacto na decisão lexical e na leitura em voz alta com listas mistas, mas não quando as listas eram agrupadas. Observou-se assim que no Português os efeitos de extensão são modulados pela tarefa. A modulação dos efeitos de extensão pela tarefa pode ser mais aparente em ortografias de profundidade intermédia do que em ortografias transparentes, pois a codificação fonológica ao nível da correspondência grafema-fonema não é a estratégia óptima, nem exclusiva, de processamento fonológico.

Palavras-chave: conversão grafema-fonema; efeitos de extensão; leitura; Português.

## Abstract

Several studies have shown that grapheme-phoneme conversion mechanisms are important in skilled reading in shallow orthographies, whereas in deep ones their importance is minimal and lexical knowledge may take the lead. In this work we examined the role of grapheme-phoneme conversion in an orthography of intermediate depth, Portuguese, by determining the impact of length (as measured by the number of letters) in reading. In two studies, young adults read two-syllable words and non-words varying in frequency and length. In the first one, five- and six-letter items were compared in a lexical decision task and in a reading aloud task with mixed lists (between-subject design). In the second study, items ranging from four to six letters were presented in lexical decision and reading aloud tasks, with mixed and blocked lists (between-subject design). Our results evidenced a robust effect of length for non-words. For words, length determined latencies in lexical decision and in reading aloud with mixed lists, but not when the lists were blocked. Hence, in Portuguese, the effects of length are modulated by the task. This may be more apparent in orthographies of intermediate depth than in shallow ones, because phonological decoding at the grapheme-phoneme level is neither the optimal nor the exclusive strategy of phonological processing.

Keywords: grapheme-phoneme conversion; length effects; reading; Portuguese.

## Résumé

Plusieurs études ont démontré que la conversion graphème-phonème est importante pour la lecture habile dans les orthographies transparentes, pendant que dans les profondes son utilisation est réduite et la connaissance lexicale devient prépondérante. Dans ce travail nous avons étudié le rôle de la conversion graphème-phonème dans une orthographie de profondeur intermédiaire, le Portugais, en analysant de l'impact de l'extension (en nombre de lettres) sur la lecture. En deux études, des jeunes universitaires ont lu des mots et des non-mots dissyllabiques variant en fréquence et en extension. Dans la première, on a comparé des items avec cinq et six lettres dans des tâches de décision lexicale et de lecture en haute voix avec des listes mixtes (design inter-sujets). Dans la deuxième, des items avec quatre à six lettres ont été présentés en tâches de décision lexicale et de lecture en haute voix, avec des listes mixtes et avec des listes groupés (design inter-sujets). Les résultats ont montré des effets d'extension robustes pour les non-mots. Pour le mots, l'extension a eu un effet sur la décision lexicale et sur la lecture en haute voix avec des listes mixtes, mais non quand les listes étaient groupées. On a vu de cette manière que, en Portugais, les effets d'extension sont modulés par la tâche. La modulation des effets d'extension par la tâche peut être plus marquée dans les orthographies de profondeur intermédiaire que dans les transparentes, parce que la codification phonologique au niveau de la correspondance graphème-phonème n'est pas la stratégie optimale, ni exclusive, de lecture.

Mots-clés: conversion graphème-phonème; effets d'extension; lecture; Portugais.

## Índice

Introdução	7
Estudo 1	20
Método	20
Resultados	23
Discussão	28
Estudo 2	32
Método	33
Resultados e discussão	35
Discussão Geral	45
Conclusão	52
Referências	53
Anexos	58

## Introdução

A leitura é uma das capacidades cognitivas mais notáveis da espécie humana, possibilitando o acesso a uma quantidade virtualmente infinita de significados a partir de símbolos visuais. Para aceder ao significado contido na escrita, o leitor tem de conhecer o código utilizado pela sua cultura para representar a fala através de séries de símbolos escritos. Ler implica também o domínio de um sistema de mapeamento entre os símbolos visuais e as unidades de som, uma condição *sine qua non* para que o leitor possa transformar as palavras escritas em palavras faladas. O processo de emparelhamento entre ortografia e fonologia tem sido denominado *codificação fonológica* (*phonological recoding*; e.g., Share, 1995; Ziegler & Goswami, 2005). Um aspecto interessante da leitura é a concomitante invariância do seu objectivo, i.e., extrair informação lexical a partir da escrita, e a grande diversidade de línguas/ortografias existente, que podem diferir profundamente entre si. Será que os processos neurocognitivos envolvidos na leitura são universais ou, pelo contrário, dependem das características da língua que se lê?

Um conjunto crescente de estudos tem testado a hipótese de que as características ortográficas das línguas influenciam a leitura, tomando como ponto de partida o facto de as diferentes ortografias representarem a linguagem falada de modos diferentes. Uma dimensão importante que varia entre as ortografias alfabéticas é a complexidade e consistência do mapeamento entre letras e sons. Nas *ortografias transparentes*, como o Espanhol, o Italiano ou o Grego, os códigos ortográfico e fonémico são

isomórficos, i.e., as correspondências entre grafemas e fonemas são regulares e inequívocas. Pelo contrário, nas *ortografias profundas*, como o Inglês, existe uma elevada irregularidade e complexidade nas relações entre grafemas e fonemas: a mesma letra pode representar diferentes sons em diferentes contextos, e o mesmo som pode ser escrito com diferentes letras (e.g., Frost, Katz & Bentin, 1987; Frost, 1994). A questão de saber em que medida estas diferenças na profundidade ortográfica moldam a arquitectura do processamento de palavras escritas e a própria trajectória da aprendizagem da leitura tem sido objecto de intensa análise empírica, quer em estudos comportamentais, quer em estudos de neuro-imagem.

Paulesu e colaboradores (2000), por exemplo, compararam leitores adultos ingleses e italianos na leitura de palavras e de não-palavras, contrastando desta forma uma ortografia profunda com outra altamente transparente. Os italianos foram mais rápidos do que os ingleses, nas palavras e nas não-palavras, e tiveram maior activação cerebral em regiões temporais superiores esquerdas associadas ao processamento de fonemas. Nos leitores ingleses, a vantagem no processamento das palavras sobre as não-palavras (efeito de lexicalidade) foi superior à observada nos italianos, e registaram-se maiores activações na circunvolução temporal inferior posterior esquerda e na circunvolução frontal inferior, áreas associadas com o acesso às palavras (processamento lexical). Estes resultados mostram que as características da ortografia são passíveis de influenciar a organização neural e funcional do sistema mente-cérebro, que se adapta de forma a permitir um ajustamento óptimo à língua.

No âmbito da aprendizagem da leitura, Seymour, Aro e Erskine (2003) compararam várias ortografias europeias, verificando que o ritmo de



desenvolvimento é duas vezes mais lento no Inglês do que nas ortografias transparentes. Enquanto as crianças que estavam a aprender a ler em ortografias transparentes, como o Italiano e o Espanhol, tiveram desempenhos quase perfeitos na leitura de palavras e não-palavras no final do primeiro ano de escolaridade, as crianças inglesas tiveram resultados muito baixos (34% para as palavras e 29% para as não-palavras).

### *Hipóteses teóricas*

Em termos históricos, as diferenças entre línguas quanto aos processos envolvidos na leitura foram primeiramente analisadas no quadro da *hipótese da profundidade ortográfica (orthographic depth hypothesis; e.g., Frost, Katz & Bentin, 1987; Katz & Feldman, 1983)*, desenvolvida a partir dos modelos clássicos de dupla via. Os modelos de dupla via assumem genericamente a existência de dois procedimentos ou vias para a leitura: a via fonológica, sub-lexical, que opera de modo sequencial convertendo as letras em sons através da utilização de regras de correspondência grafema-fonema; e a via lexical, que opera de modo holístico, mapeando directamente a configuração ortográfica da palavra com a forma fonológica através do acesso ao conhecimento sobre a palavra armazenado no léxico. Estes dois procedimentos de reconhecimento de palavras funcionam em paralelo, ganhando aquele que for mais rápido e eficaz (e.g., Coltheart, Rastle, Perry, Langdon & Ziegler, 2001; Wydell, Vuorinen, Helenius & Salmelin, 2003). Segundo a hipótese da profundidade ortográfica, as duas vias de leitura podem ter uma importância diferencial de acordo com as exigências da ortografia a ser lida. Nas ortografias transparentes, os leitores são encorajados a usar a via fonológica, uma vez que o mapeamento entre letras

e sons é relativamente directo e consistente. Em contrapartida, nas ortografias profundas a via fonológica é pouco eficaz devido às ambiguidades nas correspondências grafema-fonema, pelo que os leitores têm de usar a via ortográfica directa, reconhecendo as palavras como um todo (e.g., Frost, 1994). Assim, os leitores ajustariam a estratégia de leitura, fundamentalmente “ortográfica” ou “fonológica”, de acordo com as características da língua. Esta hipótese foi até há pouco tempo o contributo teórico mais proeminente para a compreensão das diferenças translinguísticas na leitura. Contudo, a sua predição basilar de que os efeitos fonológicos são muito reduzidos em ortografias inconsistentes tem sido posta em causa por diversos estudos e em vários paradigmas experimentais (para uma revisão, Frost, 1998). O trabalho de Rayner, Sereno, Lesch e Pollatsek (1995), por exemplo, pôs em evidência efeitos robustos de *priming* fonológico na leitura em Inglês, visíveis quando o *prime* era apresentado durante apenas 36 milissegundos. Num paradigma de movimentos oculares, os autores observaram que as características fonológicas do material visto durante os primeiros 36 milissegundos de uma fixação influenciavam a rapidez do reconhecimento da palavra subsequente, o que demonstra que os códigos fonológicos são automática e precocemente activados na leitura, mesmo numa ortografia profunda.

No sentido de ultrapassar as críticas à hipótese da profundidade ortográfica, Ziegler e Goswami delinearam recentemente a *hipótese psicolinguística da granularidade* (*psycholinguistic grain size theory*, 2005; 2006). Esta proposta teórica partilha muitos aspectos com a anterior, contendo como principal vantagem o facto de conceber as unidades de codificação fonológica na leitura de modo contínuo e não de forma dicotómica,

como acontece na oposição entre fonologia lexical (reconhecimento da palavra como um todo) e processamento grafema-fonema, postulada pela hipótese da profundidade ortográfica. Procurando integrar os dados actualmente disponíveis sobre a aquisição da leitura, a dislexia desenvolvimental e a leitura hábil nas várias ortografias, a hipótese psicolinguística da granularidade sugere que as diferenças translinguísticas na leitura reflectem diferenças fundamentais nas estratégias de codificação fonológica que se desenvolvem em resposta às características ortográficas da língua. As crianças que aprendem a ler em línguas ortograficamente mais transparentes, como o Espanhol ou o Italiano, podem basear-se quase exclusivamente em estratégias de codificação ao nível do grafema-fonema porque as correspondências entre grafemas e fonemas são altamente consistentes. Já as crianças que aprendem a ler em ortografias profundas, como o Inglês, não podem usar tão eficazmente unidades linguísticas pequenas, na medida em que a ambiguidade nas correspondências grafema-fonema é muito elevada. Como consequência, as crianças inglesas têm de desenvolver estratégias de codificação fonológica que complementem a conversão grafema-fonema, como o reconhecimento de sílabas, de padrões de letras para rimas e as tentativas de reconhecimento da forma ortográfica da palavra como um todo. Assim, as ortografias inconsistentes forçariam os leitores a desenvolver múltiplas estratégias de codificação fonológica, variando na granularidade, entre unidades mais pequenas e unidades maiores. Nas ortografias transparentes, pelo contrário, a correspondência grafema-fonema seria geralmente suficiente para uma leitura bem sucedida. Estas diferenças na aprendizagem da leitura têm sido apoiadas por vários estudos (e.g., Ellis & Hooper, 2001; Ellis, Natsume, Stavropoulou, Hoxhallari,

Daal, Polyzoe, Tsipa & Pétalas, 2004; Hoxhallari, Daal & Ellis, 2004). Ellis e colaboradores documentaram em experiências translinguísticas (ibd.) a utilização preferencial de estratégias de conversão grafema-fonema pelas crianças leitoras de ortografias mais transparentes (Galês; Albanês; silabário Japonês Hiragana), visível nos efeitos de extensão mais pronunciados, e uma maior saliência das tentativas de reconhecimento das palavras como um todo nas ortografias mais profundas (Inglês; ideogramas Japoneses Kanji). Estas tentativas foram evidentes no padrão de erros de leitura, onde as substituições de palavras reais ou as ausências de resposta foram mais frequentes do que nas ortografias transparentes. Nestas, os erros foram em menor percentagem e envolviam sobretudo a pronúncia errada de não-palavras.

No que concerne à leitura hábil, a hipótese psicolinguística da granularidade sugere que o produto final da leitura não pode ser dissociado da trajectória desenvolvimental, pelo que percursos diferenciais na aprendizagem da leitura resultarão em diferenças no processamento de palavras escritas em adultos. Neste sentido, e considerando os dados desenvolvimentais existentes, esta hipótese prevê que os leitores hábeis utilizarão preferencialmente unidades de codificação fonológica pequenas nas ortografias transparentes, e recorrerão tendencialmente a unidades maiores (e.g., reconhecimento da palavra como um todo; rimas; sílabas), ou a múltiplas estratégias, nas ortografias mais profundas (Ziegler & Goswami, 2005; Goswami & Ziegler, 2006; Ziegler & Goswami, 2006). O trabalho de Ziegler, Perry, Jacobs e Braun (2001) constitui um suporte empírico para esta ideia, derivado da comparação entre o Inglês e o Alemão (uma ortografia transparente). Neste estudo, leitores adultos do Inglês e do Alemão leram

palavras e não-palavras idênticas nas suas respectivas línguas. Os efeitos de extensão, em número de letras, foram utilizados como marcadores da utilização de unidades de processamento fonológico pequenas, ao nível da correspondência grafema-fonema, e os efeitos de rima ortográfica como marcadores do recurso a unidades de processamento maiores. Observou-se que os efeitos de extensão foram mais robustos no Alemão do que no Inglês, apontando uma preferência por estratégias de conversão grafema-fonema nas ortografias consistentes. No Inglês, pelo contrário, os efeitos de rima foram mais robustos do que no Alemão, indicando o predomínio de estratégias de processamento fonológico baseadas em unidades maiores do que o fonema nas ortografias inconsistentes. Um estudo recente, de potenciais evocados, comparou também leitores adultos em duas línguas que variavam na consistência ortográfica, o Árabe e o Francês (Simon, Bernard, Lalonde & Rebai, 2006). O Árabe pode considerar-se como uma ortografia muito profunda, caracterizada pela complexidade da relação entre o código escrito e o falado, e o Francês como uma ortografia mais transparente. Os resultados mostraram que o N320, uma onda de negatividade com pico aos 320 milissegundos que marca a utilização da conversão grafema-fonema, está presente no Francês mas não no Árabe, sugerindo que a leitura hábil em Francês se baseia em maior medida do que a leitura em Árabe nas estratégias fonológicas de conversão grafema-fonema. Estes dois estudos colocam assim em evidência diferenças entre línguas quanto às unidades de codificação fonológica preferenciais no reconhecimento hábil de palavras escritas.

#### *Efeitos de extensão na leitura hábil*

Utilizando a análise comportamental dos efeitos de extensão, são vários os estudos que sustentam a hipótese de que nas ortografias altamente

transparentes a codificação fonológica ao nível da conversão grafema-fonema é efectivamente uma estratégia determinante e preferencial na leitura hábil. Os efeitos de extensão são um bom marcador comportamental para estudar o recurso a esta estratégia na leitura porque quando a unidade de processamento é pequena, ao nível do grafema-fonema, os tempos de reacção aumentam com o aumento do número de letras das palavras e das não-palavras, uma vez que são processadas mais unidades.

Peressotti e Mulatti (2005) compararam, no Italiano, a leitura de palavras e não-palavras de cinco e seis letras em tarefas de decisão lexical, de leitura em voz alta com listas mistas, compostas de palavras e de não-palavras, e também na leitura em voz alta com listas agrupadas, em que primeiro eram apresentadas todas as palavras e depois todas as não-palavras. Verificaram que, independentemente da tarefa, as não-palavras e as palavras de seis letras demoraram mais tempo a ser processadas do que as de cinco. Estes resultados demonstram que a diferença de apenas uma letra é suficiente para originar efeitos significativos de extensão numa variedade de tarefas, numa ortografia altamente transparente. Um outro estudo com o Italiano, de cariz desenvolvimental (Spinelli, De Luca, Di Filippo, Mancini, Martelli & Zoccolotti, 2005), comparou crianças de diferentes idades sem dislexia, crianças com dislexia, e jovens adultos, na leitura de palavras que variavam na extensão entre três e oito letras. Embora os efeitos de extensão tenham sido mais importantes nas crianças, sobretudo naquelas com dislexia, eles permaneceram visíveis nos adultos para as palavras entre cinco e oito letras. Ainda no Italiano, o estudo de Bates, Burani, D'Amico e Barca (2001) corrobora a ubiquidade dos efeitos de extensão na leitura hábil de palavras, demonstrando que o número de letras é preditor das latências de leitura em

voz alta. No Espanhol, que é também uma ortografia altamente transparente, Cuetos e Barbón (2006) conduziram um estudo com o objectivo de identificar que variáveis, e como, determinavam a velocidade da leitura de palavras em leitores hábeis adultos. Verificaram que os melhores preditores das latências de leitura eram a extensão, medida em número de letras e também em número de sílabas, a par da idade de aquisição subjectiva das palavras.

Se para as ortografias altamente transparentes os efeitos de extensão são robustos e facilmente observáveis, nas ortografias profundas, pelo contrário, não são evidentes. Weekes (1997), para o Inglês, numa tarefa de leitura em voz alta de palavras e não-palavras que variavam entre três e seis letras, não obteve qualquer efeito de extensão. As palavras de seis letras foram lidas tão rapidamente como as de três, sugerindo uma codificação fonológica da configuração ortográfica da palavra como um todo. A ideia de que no Inglês as palavras são processadas predominantemente em paralelo, por recurso à via lexical de leitura, está de resto incorporada em modelos computacionais de leitura como o Dual Route Cascaded (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001), que prevê uma interacção entre a lexicalidade e a extensão, no sentido em que o efeito de extensão é esperado na leitura de não-palavras mas não na leitura de palavras. Estudos mais recentes apontam também para a ausência de efeitos de extensão na tarefa de decisão lexical no Inglês (Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004; Baayen, Feldman & Schreuder, 2006; mas ver New, Ferrand, Pallier & Brysbaert, 2006). Na tarefa de leitura em voz alta, contudo, a extensão teve um impacto significativo mesmo numa ortografia profunda como o Inglês. Todavia, este efeito parece emergir apenas no contexto de designs

experimentais muito poderosos, em que a amplitude de variação da extensão e o número de itens testado são muito elevados.

Genericamente, verificamos que os efeitos de extensão são claros e ubíquos nas ortografias altamente transparentes, traduzindo uma utilização nitidamente preferencial e preponderante de estratégias de codificação fonológica grafema-fonema, e que a extensão tem um impacto nulo ou visível apenas em condições experimentais muito estritas nas ortografias profundas, denotando a necessidade de recorrer a unidades de processamento maiores do que grafema/fonema ou a múltiplas estratégias de codificação fonológica.

#### *O Português como ortografia de profundidade intermédia*

Como a maioria dos estudos sobre as diferenças na leitura entre línguas se centra em ortografias, ou altamente transparentes, ou muito profundas, o que se sabe sobre ortografias de profundidade intermédia é muito pouco. O presente trabalho pretende dar um contributo neste âmbito, colocando em análise o caso do Português Europeu. O Português Europeu, apesar de ser habitualmente considerado uma língua com uma ortografia relativamente transparente (Castro & Gomes, 2000; Gomes, 2001), exhibe um conjunto de traços ortográficos e fonéticos que permitem categorizá-lo como uma ortografia de profundidade intermédia. Entre estes traços inclui-se o facto de existirem apenas cinco letras de vogais para 14 fonemas vocálicos, o que coloca dificuldades no estabelecimento das correspondências grafema-fonema vocálicas. Outra dificuldade encontra-se ao nível da divisão silábica. Mesmo havendo uma predominância de sílabas pouco complexas de tipo CV, as fronteiras silábicas nem sempre são claras, devido a uma característica do Português falado que é a redução vocálica. Em quase todas as palavras que



terminam nas letras <e> ou <o>, por exemplo, a vogal final é suprimida, o que faz com que palavras ortograficamente dissilábicas como <men-te> sejam fonologicamente monossilábicas, /m~et/. Assim, no Português escrito, surgem vogais que não são pronunciadas e divisões silábicas não presentes na linguagem falada. As correspondências grafema-fonema consonantais também têm algumas complexidades, que incluem a existência de grafemas complexos (e.g., <lh> corresponde a /λ/) e de regularidades contextuais (e.g., <s> corresponde a /ʃ/ quando surge no fim de palavras ou sílabas, a /s/ quando aparece no início da palavra, e a /z/ quando está em posição intervocálica). A ideia de que o Português Europeu se afasta das ortografias altamente transparentes e se aproxima de uma posição mais intermédia parece ter correspondência psicológica nos resultados de investigação no âmbito da aprendizagem da leitura e da dislexia desenvolvimental. No estudo translinguístico de Seymour, Aro e Erskine (2003), os resultados das crianças portuguesas na leitura de palavras e não-palavras no final do primeiro ano de escolaridade situaram-se num ponto intermédio entre o desempenho quase perfeito das ortografias transparentes e os resultados muito baixos das crianças inglesas. Este estudo mostra que a aprendizagem da leitura no Português é mais lenta do que nas ortografias transparentes, como o Italiano, o Grego ou o Alemão. Numa análise da natureza da dislexia desenvolvimental em Português (Sucena, Castro & Seymour, em revisão), verificou-se que as crianças portuguesas com dislexia exibem um padrão específico, diferente do habitualmente descrito, quer para as ortografias transparentes (disfluência que se reflecte em maiores latências mas sem menor exactidão), quer para as profundas (défice de codificação que prejudica a exactidão e as latências). A dislexia em Português não inibe o desenvolvimento de capacidades de

descodificação ao nível dos leitores iniciantes normais, o que está de acordo com os dados de investigação em ortografias transparentes. Contudo, o desenvolvimento do léxico ortográfico, essencial para a leitura efectiva numa ortografia de profundidade intermédia, é deficitário, afectando os tempos de reacção e a exactidão. Tal como acontece em ortografias transparentes e profundas, a consciência fonológica das crianças portuguesas com dislexia é também deficitária.

### *Hipótese e objectivos*

No seu conjunto, os traços ortográficos descritos mostram que o Português não é tão consistente como as línguas com ortografias altamente transparentes e os resultados da investigação apontam para a hipótese de que esse estatuto tem impacto sobre os processos cognitivos envolvidos na aprendizagem da leitura e na dislexia. O nosso objectivo neste trabalho é o de averiguar se a posição intermédia do Português se reflecte também na leitura hábil, tendo como pano de fundo a hipótese de que as unidades de codificação fonológica dominantes variam de acordo com a profundidade ortográfica (e.g., Ziegler & Goswami, 2005; Frost, 2006). Será que no Português as estratégias de conversão grafema-fonema são preponderantes e preferencialmente recrutadas na leitura hábil, como acontece nas ortografias altamente transparentes? Utilizando os efeitos de extensão em número de letras como marcador da utilização destas estratégias, conduzimos dois estudos para determinar a importância da conversão grafema-fonema na leitura de palavras e não-palavras em Português em vários paradigmas experimentais. Para as não-palavras, a nossa predição é a de que efeitos de extensão deverão ser evidentes e robustos, uma vez que a inexistência de

uma representação lexical correspondente implica que leitura seja feita através da conversão grafema-fonema. Quanto às palavras, se no Português a unidade de processamento óptima for ao nível da correspondência grafema-fonema, como é típico das ortografias transparentes, os efeitos de extensão deverão ser visíveis nas várias tarefas experimentais. Se, pelo contrário, as complexidades da ortografia do Português implicarem o recurso a unidades de codificação fonológica maiores e a conversão grafema-fonema não for suficiente, como é comum nas ortografias inconsistentes, a extensão deverá ter um impacto mais reduzido ou nulo no processamento de palavras. Comparando directamente os nossos resultados com aqueles que têm sido obtidos para o Italiano (Peressotti & Mulatti, 2005) e para o Espanhol (Cuetos e Barbón, 2006), a questão de interesse é a de saber se a leitura hábil numa ortografia de profundidade intermédia é semelhante ao padrão das ortografias altamente transparentes, ou aos resultados descritos para as ortografias inconsistentes, ou se apresenta um padrão intermédio distinto de ambas. A existência de um padrão de resultados diferente do observado nas ortografias altamente transparentes, a verificar-se, será um suporte para a interessante hipótese de que os processos cognitivos básicos de reconhecimento de palavras escritas diferem entre as línguas, mesmo quando estas não ocupam lugares de oposição extrema no *continuum* transparência-profundidade.

## Estudo 1

O efeito de extensão na leitura de palavras e de não-palavras foi determinado utilizando como paradigmas experimentais as duas tarefas mais frequentemente consideradas no estudo da leitura: a leitura em voz alta e a decisão lexical. Dois grupos de jovens adultos, leitores hábeis, leram palavras e não-palavras dissilábicas que variavam quanto à extensão, cinco ou seis letras, e quanto à frequência, alta ou baixa. Um dos grupos realizou a tarefa de leitura em voz alta, e o outro a tarefa de decisão lexical. O design utilizado foi semelhante ao do estudo conduzido por Peressotti e Mulatti (2005) para o Italiano, que revelou efeitos de extensão robustos para as palavras e para as não-palavras em ambas as tarefas.

### Método

#### *Participantes*

Participaram nesta experiência 60 jovens adultos (média de idades = 23.2 anos; DP = 7.6). Trinta e um realizaram a tarefa de leitura em voz alta (média de idades = 25.7 anos; DP = 8.75), e 29 foram recrutados para a tarefa de decisão lexical (média de idades = 20.3 anos; DP = 5). Eram todos de língua materna portuguesa, leitores hábeis, sem história conhecida de perturbações da leitura e com visão normal ou corrigida. A participação foi voluntária.

### *Materiais*

Foram utilizados 200 estímulos, metade palavras e metade não-palavras (cf. Anexo 1). Todos os itens começavam com uma consoante e terminavam com uma vogal.

As palavras, ortograficamente dissilábicas, foram seleccionadas da base de dados lexical Porlex (Gomes & Castro, 2003). Foram divididas em quatro condições experimentais, cada uma contendo 25 itens, formadas através da combinação ortogonal de dois níveis da frequência escrita da palavra (alta vs. baixa;  $M = 858.6$  e  $M = 46.8$  por milhão, respectivamente) e dois níveis de extensão, medida pelo número de letras (cinco vs. seis letras; ver dados descritivos das palavras no Quadro 1). Todas as palavras, incluindo as de frequência baixa, eram relativamente comuns no Português. Os valores de frequência foram obtidos da base de frequências escritas do Português Corlex (Bacelar do Nascimento, Casteleiro, Marques, Barreto, Amaro & Veloso, sem data). As várias condições de palavras tinham um número semelhante de vizinhos ortográficos, que era muito reduzido (número médio de vizinhos = 2.81; Gomes & Castro, 2003). As palavras de cinco e seis letras eram igualmente complexas no que respeita à estrutura silábica e às regularidades nas correspondências grafema-fonema.

As não-palavras consistiam em segmentos pronunciáveis de letras, formados através da substituição da quarta letra de cada uma das palavras. Tal como as palavras, foram divididas em quatro condições experimentais variando na frequência (frequência da palavra-base) e na extensão.

Palavras e não-palavras foram pseudo-aleatorizadas e divididas em dois blocos de ensaios mistos (que incluíam palavras e não-palavras), cada um contendo 100 itens. Eliminou-se assim a possibilidade de os participantes

preverem se o ensaio seguinte era uma palavra ou uma não-palavra. A ordem de apresentação dos blocos foi contra-balanceada entre os sujeitos.

*Quadro 1.* Dados descritivos (valores médios) das palavras utilizadas como estímulos no Estudo 1

	Frequência Alta		Frequência Baixa	
	5 letras	6 letras	5 letras	6 letras
Frequência escrita	844	874	48	46
Vizinhos ortográficos	3.12	2.84	2.72	2.56
Número de fonemas	4.92	5.2	4.68	5.36

### *Procedimento*

Os participantes foram testados individualmente numa sessão experimental única numa sala com condições adequadas de iluminação e silêncio.

A apresentação dos estímulos e o registo dos tempos de reacção (TR) foram controlados pelo SuperLab V4.0 para Macintosh (Cedrus Corporation, 2007) e utilizou-se um computador PowerBook G4 da Apple com um monitor colorido de 15 polegadas. Na tarefa de leitura em voz alta, a medição dos TR vocais (em milissegundos, ms) foi feita com recurso a uma VoiceKey da Cedrus. Na decisão lexical, os TR foram medidos através do teclado do computador, utilizando-se as teclas "<" e "-" com as etiquetas "SIM" e "NÃO".

Os participantes eram confortavelmente sentados em frente ao computador e era-lhes dito que iriam aparecer no centro do monitor conjuntos de letras, um de cada vez, alguns dos quais constituíam palavras reais e outros palavras inventadas. Na experiência de leitura em voz alta, era dito aos participantes que a tarefa consistia em ler em voz alta cada um dos conjuntos. Na decisão lexical, era dito que, perante os conjuntos de letras que iriam aparecer no ecrã do computador, deveriam decidir se se tratava de uma

palavra real (pressionar a tecla "SIM") ou de uma palavra inventada (pressionar a tecla "NÃO"). A localização das respostas "SIM" e "NÃO" no teclado do computador foi contra-balanceada entre os participantes. Em ambas as tarefas, a resposta devia ser a mais rápida e precisa possível.

Os estímulos eram apresentados no centro do ecrã em letras minúsculas, na fonte Gill Sans, tamanho 110. Os TR foram medidos desde o aparecimento de cada estímulo até à resposta do participante (início da resposta vocal, na leitura em voz alta). Cada estímulo permanecia visível até ao início da resposta e o intervalo inter-estímulos era de 2000 ms.

Os dois blocos de ensaios experimentais eram precedidos por 20 ensaios de treino e era feita uma pequena pausa entre os blocos. As respostas erradas foram registadas manualmente pelo experimentador.

A sessão experimental durava aproximadamente 15 minutos.

## **Resultados**

### *Leitura em voz alta*

As latências médias de nomeação (TR) e a percentagem de erros para cada uma das condições experimentais são apresentados no Quadro 2. Foram excluídas das análises da latência as respostas incorrectas (2.2%) e também os TR mais rápidos do que 200 ms ou acima dos 2000 ms, os erros da VoiceKey e as latências que excediam os três desvios-padrão da média total de cada participante (outliers; 2.5% das respostas). Os TR e a percentagem de erros foram analisados separadamente e submetidos a análises por sujeito ( $F_1$ ) e por item ( $F_2$ ). Determinaram-se os efeitos da frequência e da extensão nas palavras e nas não-palavras através de análises de variância (ANOVAs)

com os factores intra-sujeito Frequência (alta e baixa; Frequência da palavra-base, no caso das não-palavras) x Extensão (5 e 6 letras). Foi também realizada uma ANOVA incluindo os factores Lexicalidade (palavras e não-palavras) x Frequência (alta e baixa) x Extensão (5 e 6 letras) para averiguar os efeitos clássicos de lexicalidade e a interacção lexicalidade x extensão.

Quadro 2. Latências médias na leitura em voz alta por condição (em ms). Os valores entre parêntesis representam os erros (em percentagem).

Frequência/Extensão	Lexicalidade			
	Palavras		Não-palavras	
	5 letras	6 letras	5 letras	6 letras
Alta	662 (.1)	676 (.3)	705 (2.9)	725 (4.4)
Baixa	681 (.8)	691 (1.3)	699 (2.6)	731 (5.2)

Palavras. Na análise das latências de resposta, observaram-se efeitos significativos da extensão [ $F_1(1,30) = 24.03, p < .0001, \eta^2 = .45$ ; efeito marginal na análise por item,  $F_2(1,96) = 2.84, p = .09, \eta^2 = .03$ ] e da frequência [ $F_1(1,30) = 27.63, p < .0001, \eta^2 = .48$ ; marginal na análise por item,  $F_2(1,96) = 2.78, p = .099, \eta^2 = .03$ ]. As palavras curtas foram lidas mais rapidamente (671 ms) do que as longas (684 ms), e houve uma vantagem das palavras de frequência alta (669 ms) em relação às de frequência baixa (684 ms). A interacção frequência x extensão não foi significativa ( $F_s < 1$ ).

A análise dos erros revelou apenas um efeito significativo da frequência [ $F_1(1,30) = 11.9, p < .003, \eta^2 = .28$ ;  $F_2(1,96) = 7.90, p < .007, \eta^2 = .08$ ]. As palavras de frequência alta tiveram uma menor percentagem de erros (0.2%) do que as de frequência baixa (1%).



Não-palavras. A extensão teve um impacto elevado nas latências das não-palavras [ $F_1(1,30) = 30.02, p < .0001, \eta^2 = .5; F_2(1,96) = 10.31, p < .003, \eta^2 = .08$ ], com vantagem das palavras curtas (702 ms) em relação às longas (728 ms). O efeito da frequência da palavra-base não foi significativo ( $F_s < 1$ ).

A análise da percentagem de erros corrobora os resultados das latências, verificando-se um efeito significativo da extensão [ $F_1(1,30) = 5.52, p < .03, \eta^2 = .16; F_2(1,96) = 4.75, p < .04, \eta^2 = .05$ ] e ausência de efeito de frequência da palavra-base ( $F_s < 1$ ). As não-palavras curtas obtiveram menos respostas erradas (2.7%) do que as longas (4.8%).

Por fim, a ANOVA global, que incluía o factor lexicalidade, evidenciou uma vantagem das palavras (677 ms; 0.6%) sobre as não-palavras (715 ms; 3.8%), visível na análise da latência [ $F_1(1,30) = 73.63, p < .0001, \eta^2 = .71; F_2(1,192) = 35.83, p < .0001, \eta^2 = .16$ ] e na análise dos erros [ $F_1(1,30) = 40.18, p < .0001, \eta^2 = .57; F_2(1,192) = 40.34, p < .0001, \eta^2 = .17$ ]. A interacção lexicalidade x extensão foi também obtida [latências:  $F_1(1,30) = 8.56, p < .007, \eta^2 = .22$ ; ns na análise por item,  $F_2 < 1$ ; marginal na análise dos erros:  $F_1(1,30) = 3.78, p = .06, \eta^2 = .11; F_2(1,192) = 3.03, p = .08, \eta^2 = .02$ ], e é explicada pelo facto do efeito de extensão ter sido mais determinante nas não-palavras (26 ms) do que nas palavras (13 ms).

#### *Decisão lexical*

Os TR e a percentagem de erros para cada uma das condições experimentais são apresentados no Quadro 3. Foram removidas das análises da latência as respostas incorrectas (8.3%), os tempos de reacção mais

rápidos do que 300 ms ou acima dos 3500 ms e os outliers (1.3% das respostas).

*Quadro 3.* Latências médias na decisão lexical por condição (em ms). Os valores entre parêntesis representam os erros (em percentagem).

Frequência/Extensão	Lexicalidade			
	Palavras		Não-palavras	
	5 letras	6 letras	5 letras	6 letras
Alta	661 (1.9)	662 (1.7)	896 (7.5)	952 (8.2)
Baixa	809 (15.7)	826 (17)	900 (6.1)	1008 (9.8)

Palavras. A extensão não teve um impacto significativo nas latências ( $F_s < 1$ ). Contudo, nas palavras de frequência baixa, houve uma ligeira tendência para os itens mais curtos serem respondidos mais rapidamente (809 ms) do que os mais longos (826 ms). O efeito de frequência na análise dos TR foi significativo [ $F_1(1,28) = 67.86, p < .0001, \eta^2 = .71; F_2(1,96) = 97.73, p < .0001, \eta^2 = .52$ ], com vantagem das palavras de frequência alta (662 ms) em relação às de frequência baixa (818 ms). A interação frequência x extensão não foi significativa ( $F_s < 1$ ).

A análise dos erros é concordante com os resultados dos TR, revelando apenas uma menor percentagem de erros nas palavras de frequência alta (1.8%) do que nas de frequência baixa [16.3%;  $F_1(1,28) = 117.99, p < .0001, \eta^2 = .81; F_2(1,96) = 25.22, p < .0001, \eta^2 = .21$ ].

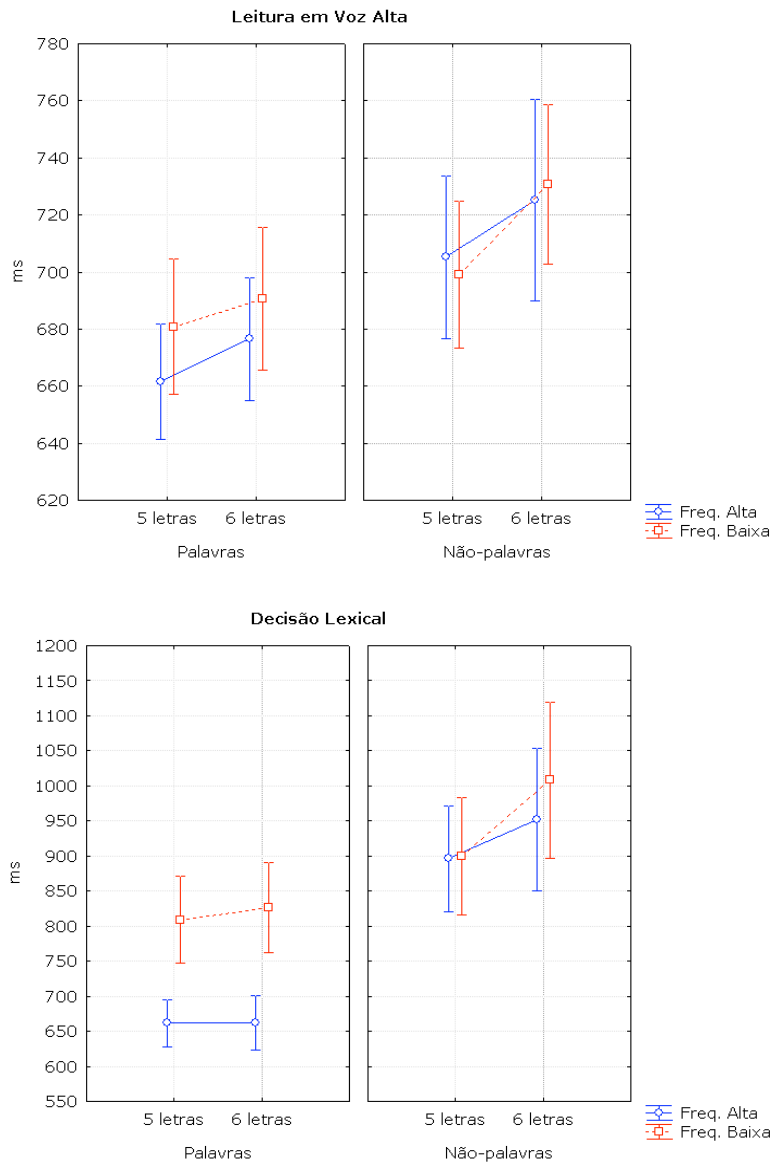
Não-palavras. Na análise da latência observaram-se efeitos significativos da extensão [ $F_1(1,28) = 22.75, p < .0001, \eta^2 = .45; F_2(1,96) = 23.66, p < .0001, \eta^2 = .20$ ] e da frequência da palavra-base [ $F_1(1,28) = 11.88, p < .003, \eta^2 = .30$ ; marginal na análise por item,  $F_2(1,96) = 2.94, p = .09, \eta^2 = .03$ ]. As não-palavras curtas foram processadas mais rapidamente

(898 ms) do que as longas (980 ms) e houve uma vantagem das não-palavras com frequência da palavra-base alta (924 ms) em relação às aquelas com frequência da palavra-base baixa (954 ms). De notar ainda que a interacção extensão x frequência da palavra-base foi significativa [ $F_1(1,28) = 9.10, p < .006, \eta^2 = .25; F_2(1,96) = 3.85, p = .053, \eta^2 = .04$ ]. O efeito de extensão foi mais pronunciado para as não-palavras derivadas de palavras de frequência baixa (108 ms) do que para as derivadas de palavras de frequência alta (55 ms).

A análise dos erros revelou um efeito significativo da extensão na análise por sujeito [ $F_1(1,28) = 7.12, p < .02, \eta^2 = .20$ ; ns na análise dos itens,  $F_2 < 1$ ], tendo havido menos erros nas não-palavras curtas (6.8%) do que nas longas (9%). Embora o efeito principal de frequência da palavra-base não tenha sido significativo ( $F_s < 1$ ), a interacção entre a extensão e a frequência da palavra-base foi marginal na análise por sujeito [ $F_1(1,28) = 3.84, p = .06, \eta^2 = .12$ ; ns na análise por item,  $F_2 < 1$ ]. Corroborando os resultados da análise da latência, o efeito de extensão foi ligeiramente mais pronunciado para as não-palavras derivadas de palavras de frequência baixa (3.7%) do que para as derivadas de palavras de frequência alta (0.8%).

Na ANOVA global, o efeito de lexicalidade [ $F_1(1,28) = 49.06, p < .0001, \eta^2 = .64; F_2(1,192) = 279.62, p < .0001, \eta^2 = .59$ ] e a interacção lexicalidade x extensão [ $F_1(1,28) = 11.06, p < .003, \eta^2 = .28; F_2(1,192) = 8.06, p < .006, \eta^2 = .04$ ] foram obtidos na análise da latência mas não na análise dos erros ( $F_s < 1$ ). Observou-se uma vantagem na velocidade de processamento das palavras (740 ms) em relação às não-palavras (939 ms) e a extensão foi determinante nas não-palavras (82 ms) mas não teve impacto nas palavras (9 ms).

Figura 1. Latências médias dos participantes, em ms, em função da lexicalidade, da frequência e da extensão no Estudo 1. O painel de cima representa a experiência de leitura em voz alta e o de baixo a experiência de decisão lexical. As barras de erro representam um erro-padrão da média.



## Discussão

Neste estudo analisou-se se apenas uma letra de diferença seria suficiente para dar origem a efeitos de extensão significativos na leitura de palavras e não-palavras em Português. No que respeita às não-palavras, tal

como era esperado, as de cinco letras foram processadas mais rapidamente do que as de seis, tanto na leitura em voz alta como na decisão lexical, reflectindo a utilização de estratégias fonológicas de conversão grafema-fonema na leitura destes itens (ver Figura 1). Este resultado, em conjunto com o efeito de lexicalidade, a interacção entre lexicalidade e extensão e o efeito de frequência para as palavras, corrobora o padrão que tem sido descrito para outras línguas, independentemente da profundidade ortográfica (e.g., Weekes, 1997; Juphard, Carbonnel & Valdois, 2004; Valdois, Carbonnel, Juphard, Baciú, Ans, Peyrin & Segebarth, 2006; Peressotti & Mulatti, 2005), e coincide com as predições de modelos de leitura actuais como o Dual Route Cascaded (DRC; Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) ou o Connectionist Dual Process Model (CDP+; Perry, Ziegler & Zorzi, 2007).

No que respeita à questão crítica de saber se apenas uma letra a mais tem impacto no processamento de palavras, os resultados foram diferentes na tarefa de leitura e na de decisão lexical. Na leitura, a vantagem dos itens curtos em relação aos longos foi observada. Na decisão lexical, este efeito não foi estatisticamente significativo, apesar de nas palavras de frequência baixa ter havido uma ligeira tendência para os itens de cinco letras terem sido processados mais rapidamente do que os de seis (ver Figura 1). Assim, se na tarefa de leitura em voz alta os leitores portugueses parecem ter utilizado como estratégia de leitura preferencial a conversão grafema-fonema, como é típico das ortografias transparentes (Ziegler, Perry, Jacobs & Braun, 2001), o mesmo não foi evidente na decisão lexical. Nesta tarefa, pelo contrário, o conhecimento lexical parece ter tido um peso preponderante, como se deduz da ausência de efeito de extensão e dos efeitos da frequência da palavra-base

na leitura de não-palavras. A variável frequência, associada ao conhecimento lexical, foi relevante mesmo quando os itens não eram lexicais, i.e., quando eram não-palavras. Este efeito deve-se provavelmente ao facto de as não-palavras terem sido formadas através da substituição de apenas uma letra das palavras, mantendo assim muitas semelhanças com os itens lexicais originais. Efeitos de frequência em não-palavras na decisão lexical foram também descritos por Perea, Rosa e Gómez (2005) quando as não-palavras mantinham muitas semelhanças com as palavras originais.

Afastando-se dos resultados obtidos para o Italiano com um design experimental semelhante (Peressotti & Mulatti, 2005), em que o efeito de extensão emergiu claramente na leitura em voz alta e na decisão lexical, observámos neste primeiro estudo um efeito de extensão modulado pela tarefa. Uma possível interpretação é a de que, no Português, a estratégia de conversão grafema-fonema não é tão preponderante na leitura como nas ortografias altamente transparentes, sendo utilizada a par de outras estratégias de codificação fonológica maiores, eventualmente ao nível da configuração ortográfica/fonológica da palavra. Os leitores de ortografias de profundidade intermédia utilizariam preferencialmente a conversão grafema-fonema em situações em que a fonologia é mais saliente, como é o caso da tarefa de leitura em voz alta (é possível ler uma palavra rapidamente sem um processamento lexical profundo), e recorreriam a outras unidades de processamento maiores quando a tarefa é em larga medida dependente do conhecimento lexical (na decisão lexical, não é possível decidir se um item é palavra ou não sem aceder ao léxico). Esta interpretação remete para a hipótese de que em ortografias que se afastam das altamente transparentes, as estratégias de leitura utilizadas podem ser mais sensíveis às características

da tarefa. Contudo, uma possível crítica aos resultados deste estudo deve ser apreciada antes de um aprofundamento desta hipótese. No nosso design, as várias condições de palavras e de não-palavras não foram emparelhadas quanto ao grafema/fonema iniciais. No estudo de Balota, Cortese, Sergeant-Marshall, Spieler e Yap (2004) foram observados efeitos diferenciais da consistência do ataque na leitura em voz alta e na decisão lexical. Assim, existe a possibilidade de o padrão de resultados que obtivemos se dever, não só à manipulação da extensão, mas a diferenças não controladas entre as condições experimentais na consistência do ataque. Embora tenhamos repetido as análises com um sub-grupo de itens emparelhados quanto ao grafema e fonema iniciais (104 no total) e o padrão de resultados tenha sido *grosso modo* semelhante ao inicial, no Estudo 2 pretendeu-se estabelecer a robustez dos resultados do Estudo 1 controlando esta variável.

## Estudo 2

O segundo estudo teve três objectivos. O primeiro foi o de verificar se os resultados obtidos no Estudo 1 podem ser replicados num novo conjunto de estímulos e de participantes, e quando os itens são emparelhados quanto ao grafema/fonema iniciais. O segundo foi o de analisar os efeitos de extensão com uma amplitude de variação ligeiramente maior, entre quatro e seis letras. Esta manipulação foi introduzida para diminuir a possibilidade de os efeitos de extensão não serem visíveis devido a considerar-se apenas uma letra de diferença. Se apenas uma letra de diferença pode ser suficiente para evidenciar efeitos de extensão independentes da tarefa em ortografias altamente transparentes (Peressotti & Mulatti, 2005), numa ortografia de profundidade intermédia pode ser necessário aumentar a amplitude de variação para obter efeitos semelhantes. O terceiro objectivo foi o de determinar os efeitos de extensão, não só na leitura em voz alta e na decisão lexical, como no primeiro estudo, mas também numa terceira tarefa de leitura em voz alta com listas agrupadas, i.e., listas que contêm apenas um tipo de itens, palavras ou não-palavras. É possível argumentar que a utilização de listas mistas, compostas de palavras e não-palavras, encoraja o processamento não lexical e o recurso à conversão grafema-fonema (Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004). Este enviesamento em favor do processamento fonológico ao nível do grafema/fonema pode assim dar origem à utilização de estratégias de leitura que em situações mais ecológicas, apenas com palavras, não seriam adoptadas. Neste sentido, a comparação dos efeitos de extensão em listas agrupadas e mistas é um



elemento crítico para determinar de forma mais neutra e sólida a importância e a generalidade das estratégias sequenciais de codificação fonológica ao nível da conversão grafema-fonema na leitura. Se estas assumirem o carácter de unidade fonológica óptima no processamento de palavras em Português, é de esperar que os efeitos de extensão sejam visíveis independentemente do tipo de listas, mistas ou agrupadas, como acontece nas ortografias altamente transparentes (Bates, Burani, D'Amico & Barca, 2001; Cuetos & Barbón, 2006; Spinelli, De Luca, Di Filippo, Mancini, Martelli & Zoccolotti, 2005; Peressotti & Mulatti, 2005). Se, pelo contrário, a conversão grafema-fonema não for claramente preferencial e os leitores portugueses mudarem de estratégia de codificação fonológica em resposta adaptativa às condições da tarefa, como indicou o Estudo 1, prevêem-se diferenças entre as tarefas no impacto da extensão.

## **Método**

### *Participantes*

Participaram no segundo estudo 75 estudantes de Psicologia da Universidade do Porto (média de idades = 20.79 anos; DP = 3.7), que receberam créditos lectivos pela sua participação. Trinta e quatro realizaram a tarefa de leitura em voz alta com listas mistas (média de idades = 20.8 anos; DP = 3.84), 21 realizaram a tarefa de decisão lexical (média de idades = 20.1 anos; DP = 4.1), e 20 a tarefa de leitura em voz alta com listas agrupadas (média de idades = 21.25 anos; DP = 3.29).

### *Materiais*

Os estímulos consistiram em 300 itens, compreendendo 150 palavras e 150 não-palavras (cf. Anexo 2). Tal como no Estudo 1, todos os estímulos tinham consoante inicial e vogal final.

As palavras, ortograficamente dissilábicas, foram divididas em seis condições experimentais, variando na frequência (alta vs. baixa;  $M = 757.8$  e  $M = 46.8$ , respectivamente) e na extensão, com três níveis (quatro, cinco e seis letras; ver dados descritivos das palavras no Quadro 4). Todas as condições tinham um número semelhante de vizinhos ortográficos (número médio de vizinhos = 2.89) e as palavras nas diferentes categorias foram emparelhadas quanto ao fonema e grafema iniciais. Em termos de propriedades fonéticas do ataque da primeira sílaba, em todas as categorias de palavras existiam 16 oclusivas, seis fricativas, duas aproximantes e uma vibrante. Os itens de diferentes extensões eram igualmente complexos no que respeita à estrutura silábica e às regularidades grafema-fonema.

As não-palavras foram formadas através da substituição de uma letra da segunda sílaba das palavras. As derivações foram feitas de forma a não alterar a estrutura silábica da palavra-base, e substituindo consoantes por consoantes e vogais por vogais.

Nas tarefas de leitura em voz alta com listas mistas e de decisão lexical, os itens foram pseudo-aleatorizados e divididos em dois blocos de 150 itens cada, incluindo palavras e não-palavras. Na tarefa de leitura em voz alta com listas agrupadas, os 300 itens foram divididos em dois blocos, um de palavras e outro de não-palavras. Dentro de cada bloco, os estímulos foram pseudo-aleatorizados.

Quadro 4. Dados descritivos (valores médios) das palavras utilizadas como estímulos no Estudo 2

	Frequência Alta			Frequência Baixa		
	4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
Frequência escrita	764	815	694	49	47	43
Vizinhos ortográficos	3.08	3.13	2.6	3.24	3	2.32
Número de fonemas	4	4.52	5.32	4.04	4.72	5.6

### *Procedimento*

O procedimento foi semelhante ao do Estudo 1 e a sessão experimental durava cerca de 20 minutos.

Na tarefa de leitura em voz alta com listas agrupadas, foi dito aos participantes que as palavras e as não-palavras iriam aparecer todas juntas. A ordem de apresentação dos blocos foi contra-balanceada entre os participantes.

## **Resultados e Discussão**

### *Leitura em voz alta com listas mistas*

Os TR e a percentagem de erros para cada uma das condições experimentais são apresentados no Quadro 5. Os erros foram excluídos das análises (2.2% das respostas), assim como as respostas abaixo de 200 ms e acima de 2000 ms, as falhas da VoiceKey e os outliers (5.3% das respostas).

Quadro 5. Latências médias na leitura em voz alta com listas mistas por condição (em ms). Os valores entre parêntesis representam os erros (em percentagem).

Freq./Ext.	Lexicalidade					
	Palavras			Não-palavras		
	4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
Alta	587 (.0)	592 (.1)	594 (.8)	635 (3.1)	635 (4.3)	651 (5.5)
Baixa	602 (2)	601 (1.2)	613 (1.6)	621 (1.5)	626 (2.1)	651 (5.7)

Palavras. Tal como na experiência de leitura em voz alta do Estudo 1, o efeito de extensão [ $F_1(2,66) = 4.14, p < .03, \eta^2 = .11$ ; ns na análise por item,  $F_2 < 1$ ] e o efeito de frequência [ $F_1(1,33) = 20.09, p < .0001, \eta^2 = .38$ ;  $F_2(1,144) = 7.96, p < .006, \eta^2 = .05$ ] foram significativos na análise da latência. Houve uma vantagem no processamento das palavras curtas (594 ms) em relação às longas (603 ms) e das palavras de frequência alta (591 ms) em relação às de frequência baixa (605 ms). Análises post-hoc mostraram que o efeito de extensão foi significativo entre as palavras de cinco e seis letras ( $p < .03$ ) mas não entre as de quatro e cinco letras ( $p = .45$ ). A interacção frequência x extensão não foi significativa ( $F_s < 1$ ).

A análise dos erros revelou apenas um efeito significativo de frequência [ $F_1(1,33) = 10.79, p < .003, \eta^2 = .25$ ;  $F_2(1,144) = 8.4, p < .005, \eta^2 = .06$ ]. As palavras de frequência alta foram lidas com uma menor percentagem de erros (0.3%) do que as de frequência baixa (1.6%).

Não-palavras. Na análise da latência, o impacto da extensão [ $F_1(2,66) = 12.33, p < .0001, \eta^2 = .27$ ;  $F_2(2,144) = 5.81, p < .005, \eta^2 = .08$ ] e da frequência da palavra-base [ $F_1(1,33) = 10.5, p < .004, \eta^2 = .24$ ; ns na análise por item,  $F_2 < 1$ ] foram significativos. As palavras curtas foram lidas mais rapidamente (628 ms) do que as longas (651 ms), e as não-palavras derivadas de palavras de frequência baixa foram lidas ligeiramente mais rapidamente (633 ms) do que as derivadas de palavras de frequência alta (640 ms). Testes post-hoc mostraram que, tal como nas palavras, a extensão foi determinante entre as não-palavras de cinco e seis letras ( $p < .0001$ ) mas não entre as de quatro e cinco letras ( $p = .58$ ). A interacção extensão x frequência da palavra-base não foi significativa ( $F_s < 1$ ).

Os resultados dos erros foram concordantes com os das latências, revelando um efeito de extensão [ $F_1(2,66) = 13.11, p < .0001, \eta^2 = .28$ ;  $F_2(2,144) = 3.05, p = .051, \eta^2 = .04$ ] e um efeito de frequência da palavra-base significativo na análise por sujeito [ $F_1(1,33) = 4.00, p < .04, \eta^2 = .13$ ; ns na análise por item,  $F_2 < 1$ ]. Os itens curtos foram lidos com uma maior exactidão (2.3%) do que os longos (5.6%; diferença significativa entre os de cinco e seis letras,  $p < .0001$ , mas não entre os de quatro e cinco,  $p = .18$ ), e as não-palavras com frequência da palavra-base baixa tenderam a ser respondidas com menor percentagem de erros (3.1%) do que as de frequência da palavra-base alta (4.3%).

Na ANOVA global obteve-se um efeito de lexicalidade muito robusto [latências:  $F_1(1,33) = 60.98, p < .0001, \eta^2 = .65$ ;  $F_2(1,288) = 80.44, p < .0001, \eta^2 = .22$ ; erros:  $F_1(1,33) = 36.04, p < .0001, \eta^2 = .52$ ;  $F_2(1,288) = 19.56, p < .0001, \eta^2 = .06$ ] e a interacção lexicalidade x extensão foi significativa nas análises por sujeito [latências:  $F_1(2,66) = 4.90, p < .02, \eta^2 = .13$ ; ns na análise por item,  $F_2 < 1$ ; erros:  $F_1(2,66) = 7.44, p < .002, \eta^2 = .18$ ; ns na análise por item,  $F_2 < 1$ ]. Houve assim uma vantagem no processamento de palavras (598 ms; 0.9%) em relação ao de não-palavras (636 ms; 3.7%), e a extensão teve um impacto maior nas não-palavras (23 ms; 3.3%) do que as palavras (9 ms; 0.3%).

Este padrão de resultados replica globalmente o do primeiro estudo. Os dados sugerem que a estratégia de codificação fonológica grafema-fonema é utilizada pelos leitores portugueses na tarefa de leitura em voz alta com listas mistas, não só na leitura de não-palavras, mas também, embora com menor impacto, na de palavras.

*Decisão lexical*

Os TR e a percentagem de erros para cada uma das condições experimentais são apresentados no Quadro 6. Foram excluídos das análises os erros (7% das respostas), os TR mais rápidos do que 300 ms ou acima dos 3500 ms e os outliers (1.5% das respostas).

Quadro 6. Latências médias na decisão lexical por condição (em ms). Os valores entre parêntesis representam os erros (em percentagem).

Freq./Ext.	Lexicalidade					
	Palavras			Não-palavras		
	4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
Alta	641 (1.1)	653 (1)	696 (1.3)	910 (6)	896 (4.9)	958 (5.4)
Baixa	774 (18.3)	770 (13.9)	800 (13.5)	906 (3.2)	908 (6)	956 (8.1)

Palavras. O efeito de extensão foi visível e elevado na análise da latência [ $F_1(2,40) = 14.47, p < .0001, \eta^2 = .42$ ; marginal na análise por item,  $F_2(2,144) = 2.89, p = .06, \eta^2 = .04$ ], tendo as palavras curtas sido processadas mais rapidamente (708 ms) do que as longas (748 ms). Esta diferença foi significativa apenas entre as palavras de cinco e seis letras ( $p < .0001$ ; ns entre as de quatro e cinco,  $p = .57$ ). Observou-se também um efeito de frequência [ $F_1(1,20) = 134.15, p < .0001, \eta^2 = .87$ ;  $F_2(1,144) = 68.91, p < .0001, \eta^2 = .32$ ]. As palavras de frequência alta foram respondidas mais rapidamente (664 ms) do que as de frequência baixa (781 ms). A interacção frequência x extensão não foi significativa ( $F_s < 1$ ).

A análise dos erros revelou um efeito da extensão pequeno mas significativo na análise por sujeito [ $F_1(2,40) = 4.12, p < .03, \eta^2 = .17$ ; ns na análise por item,  $F_2 < 1$ ], que deve ser interpretado no quadro da interacção frequência x extensão [ $F_1(2,40) = 72.73, p < .05, \eta^2 = .15$ ; ns na análise por

tem,  $F_2 < 1$ ]. Nas palavras de frequência baixa, os itens de quatro letras tiveram mais erros (18.3%) do que os de seis (13.5%). O efeito principal de frequência também foi significativo [ $F_1(1,20) = 68.12, p < .0001, \eta^2 = .77$ ;  $F_2(1,144) = 49.75, p < .0001, \eta^2 = .26$ ], tendo havido menos erros nas palavras de frequência alta (1.1%) do que nas de frequência baixa (15.2%).

Não-palavras. Na análise da latência, apenas o efeito de extensão foi significativo [ $F_1(2,40) = 9.71, p < .0001, \eta^2 = .33$ ;  $F_2(2,144) = 3.1, p < .05, \eta^2 = .04$ ]. Os itens curtos foram processados mais rapidamente (908 ms) do que os longos (957 ms). Tal como nas palavras, o efeito de extensão foi significativo apenas entre os itens de cinco e seis letras ( $p < .0001$ ; ns entre os de quatro e cinco,  $p = .65$ ).

A análise dos erros não revelou nenhum efeito significativo ( $F_s < 1$ ), excepto a interacção entre frequência e extensão na análise por sujeito [ $F_1(2,40) = 4.63, p < .02, \eta^2 = .19$ ; ns na análise por item,  $F_2 < 1$ ]. Esta interacção é explicada pelo facto do efeito de extensão ter sido significativo nas não-palavras com frequência da palavra-base baixa (3.2% nos itens curtos vs. 8.1% nos longos) mas não naquelas com frequência da palavra-base alta (6% nos itens curtos vs. 5.4% nos longos).

Na ANOVA global obteve-se um efeito de lexicalidade muito robusto na análise da latência [ $F_1(1,20) = 39.29, p < .0001, \eta^2 = .63$ ;  $F_2(1,288) = 287.57, p < .0001, \eta^2 = .5$ ; ns na análise dos erros,  $F_s < 1$ ], denotando a vantagem do processamento das palavras (723 ms) em relação às não-palavras (922 ms). A interacção entre lexicalidade e extensão não foi significativa, excepto na análise dos erros por sujeito [ $F_1(2,40) = 4.02, p < .03, \eta^2 = .17$ ].

O impacto da extensão, sobretudo nas palavras, foi mais importante nesta experiência do que no Estudo 1. Esta diferença nos resultados deve-se provavelmente, não só ao aumento da amplitude de variação da extensão, mas sobretudo a diferenças ao nível dos controlos na selecção dos itens. No Estudo 2 houve cuidados mais estritos, quer na selecção das palavras (que foram emparelhadas quanto ao fonema e grafema iniciais), quer na derivação das não-palavras (as derivações foram feitas de modo a manter a estrutura silábica da palavra original e foram substituídas consoantes por consoantes e vogais por vogais). Os efeitos de extensão no processamento de palavras aqui descritos, em conjunto com os dados da tarefa de leitura em voz alta, sublinham a importância das estratégias fonológicas de conversão grafema-fonema na leitura hábil em Português. A extensão em número de letras é uma variável relevante na leitura de palavras, quer na leitura em voz alta, quer na decisão lexical, no contexto de listas mistas.

#### *Leitura em voz alta com listas agrupadas*

Os TR e a percentagem de erros para cada uma das condições experimentais são apresentados no Quadro 7. Os erros foram excluídos das análises (1.9% das respostas), assim como as respostas abaixo de 200 ms e acima de 2000 ms, as falhas da VoiceKey e os outliers (4% das respostas).

*Quadro 7.* Latências médias na leitura em voz alta com listas agrupadas por condição (em ms). Os valores entre parêntesis representam os erros (em percentagem).

Freq./Ext.	Lexicalidade					
	Palavras			Não-palavras		
	4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
Alta	558 (.0)	558 (.4)	551 (.8)	615 (1.4)	619 (4.4)	639 (8)
Baixa	565 (.8)	565 (.6)	570 (.6)	606 (.8)	618 (1)	636 (4.6)



Palavras. Contrariamente ao observado nas experiências de leitura em voz alta e de decisão lexical com listas mistas, na análise dos TR a extensão não teve qualquer impacto. As palavras longas foram lidas tão rapidamente como as curtas ( $F_s < 1$ ). No que respeita à frequência, o efeito foi significativo [ $F_1(1,19) = 15.10, p < .002, \eta^2 = .44$ ;  $F_2(1,144) = 4.26, p < .05, \eta^2 = .03$ ], indicando maior rapidez na leitura de palavras de frequência alta (556 ms) do que na leitura de palavras de frequência baixa (567 ms).

Na análise dos erros nenhum efeito foi observado ( $F_s < 1$ ).

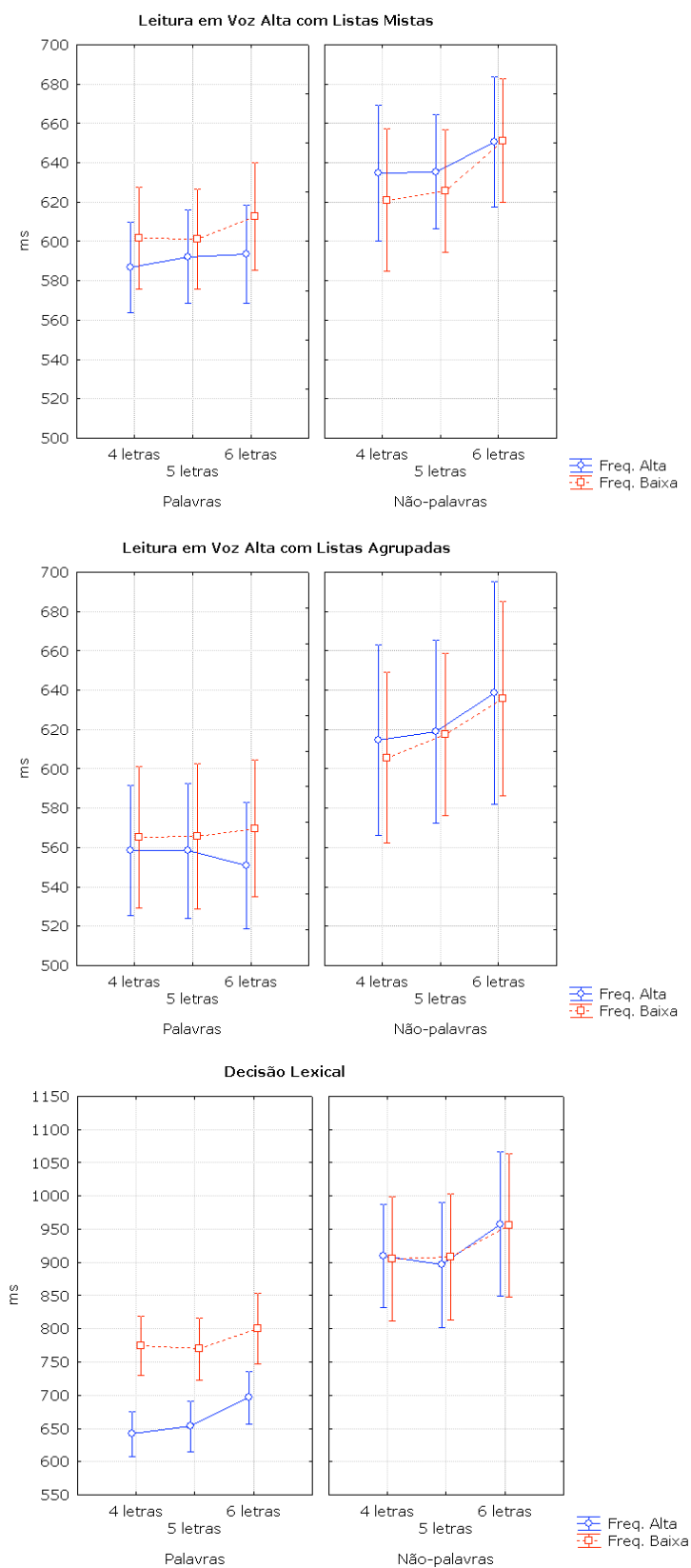
Não-palavras. O efeito de extensão foi significativo [ $F_1(2,38) = 9.34, p < .002, \eta^2 = .33$ ;  $F_2(2,144) = 6.79, p < .003, \eta^2 = .09$ ], indicando uma vantagem dos itens curtos (610 ms) em comparação com os longos (637 ms). Esta vantagem foi visível entre as não-palavras de quatro e cinco letras ( $p < .03$ ) e também entre as de cinco e seis letras ( $p < .02$ ). O efeito da frequência da palavra-base não foi significativo ( $F_s < 1$ ).

As análises dos erros revelaram um efeito significativo da extensão [ $F_1(2,38) = 22.05, p < .0001, \eta^2 = .54$ ;  $F_2(1,144) = 7.32, p < .002, \eta^2 = .09$ ] e da frequência da palavra-base [ $F_1(1,19) = 19.61, p < .0001, \eta^2 = .51$ ;  $F_2(1,144) = 3.99, p < .05, \eta^2 = .03$ ]. As não-palavras curtas foram respondidas com maior exactidão (1.1%) do que as longas (6.3%), sendo que a diferença foi significativa entre os itens de quatro e cinco letras ( $p < .03$ ) e entre os de cinco e seis ( $p < .0001$ ), e houve uma vantagem das palavras com frequência da palavra-base baixa (2.1%) em relação às de frequência da palavra-base alta (4.6%).

Na ANOVA geral, incluindo o factor lexicalidade, a vantagem das palavras (561 ms; 0.5%) sobre as não-palavras (622 ms; 3.4%) foi significativa [latências:  $F_1(1,19) = 27.43, p < .0001, \eta^2 = .59$ ;  $F_2(1,288) =$

213.13,  $p < .0001$ ,  $\eta^2 = .43$ ; erros:  $F_1(1,19) = 60.15$ ,  $p < .0001$ ,  $\eta^2 = .76$ ;  $F_2(1,288) = 25.1$ ,  $p < .0001$ ,  $\eta^2 = .08$ ]. A interacção entre lexicalidade e extensão foi também significativa [latências:  $F_1(2,38) = 8.5$ ,  $p < .003$ ,  $\eta^2 = .31$ ;  $F_2(2,288) = 4.89$ ,  $p < .04$ ,  $\eta^2 = .03$ ; erros:  $F_1(2,38) = 16.2$ ,  $p < .0001$ ,  $\eta^2 = .46$ ;  $F_2(1,288) = 6.73$ ,  $p < .002$ ,  $\eta^2 = .05$ ], devendo-se ao facto da extensão ter sido determinante nas não-palavras (27 ms; 5.2%) mas não ter tido qualquer impacto nas palavras (-2 ms; 0.3%).

Figura 2. Latências médias dos participantes, em ms, em função da lexicalidade, da frequência e da extensão no Estudo 2. O painel de cima representa a experiência de leitura com listas mistas, o do meio a experiência de leitura com listas agrupadas, e o painel de baixo representa a experiência de decisão lexical. As barras de erro representam 1 erro-padrão da média.



Comparando as três experiências do Estudo 2, verificámos que a extensão teve impacto no processamento de palavras na decisão lexical e na leitura em voz alta no contexto de listas mistas mas não no contexto de listas agrupadas (ver Figura 2). Estes resultados sugerem que no Português a utilização de estratégias de leitura de conversão grafema-fonema é suficientemente importante para ser visível em diferentes tarefas, mesmo utilizando uma amplitude reduzida de variação da extensão. Não obstante, a ausência de efeito de extensão no contexto de listas agrupadas é indicativa de que a utilização desta estratégia não é tão preferencial nem assume o carácter de processo de leitura óptimo, quase exclusivo, característico das ortografias altamente transparentes (Bates, Burani, D'Amico & Barca, 2001; Cuetos & Barbón, 2006; Peressotti & Mulatti, 2005; Spinelli, De Luca, Di Filippo, Mancini, Martelli & Zoccolotti, 2005). Os resultados deste estudo puseram assim em evidência um efeito de extensão modulado pelas características das listas. Quanto as listas envolviam palavras e não-palavras, promovendo a codificação de unidades fonológicas pequenas, os leitores portugueses tenderam a recrutar estratégias de conversão grafema-fonema. Quando as listas eram agrupadas, criando um contexto exclusivamente lexical (ou não-lexical, no caso da lista de não-palavras) a utilização de unidades de codificação fonológica maiores do que o grafema/fonema, foi preferido.

## Discussão Geral

O objectivo do conjunto de experiências aqui apresentadas foi o de averiguar a importância que as estratégias fonológicas de conversão grafema-fonema têm na leitura hábil em Português. Utilizaram-se os efeitos de extensão no processamento de palavras e não-palavras como marcador comportamental do recurso a estas estratégias. Para as não-palavras, a predição foi a de que os efeitos de extensão deviam observar-se, traduzindo a necessidade de recorrer às unidades grafema/fonema para processar itens não-lexicais. Para as palavras, previu-se que, se no Português a codificação fonológica ao nível do grafema-fonema for a estratégia óptima utilizada na leitura, os efeitos de extensão deveriam ser ubíquos, como acontece nas ortografias transparentes. Se, pelo contrário, esta estratégia não for tão eficaz e preferencial, o impacto da extensão deveria ser reduzido ou nulo, reflectindo a codificação de unidades fonológicas maiores do que o fonema, ou mesmo a tentativa de emparelhar directamente a forma escrita com a forma falada das palavras, o que é típico das ortografias profundas.

Os principais factos empíricos estabelecidos nos dois estudos podem resumir-se da forma que se segue. Nas não-palavras, como previsto, observou-se um efeito robusto da extensão na decisão lexical e na leitura em voz alta com listas mistas e agrupadas, com vantagem no processamento dos itens curtos em relação aos longos. Quanto às palavras, o impacto da extensão foi dependente do tipo e condições da tarefa. Na leitura em voz alta, as palavras curtas foram lidas mais rapidamente do que as longas nas duas experiências quando as listas eram mistas. Contudo, quando as listas eram

agrupadas, a extensão não teve qualquer impacto na velocidade e exactidão de leitura. Quanto à decisão lexical, embora no primeiro estudo a vantagem das palavras curtas em relação às longas tenha sido apenas tendencial, no segundo estudo esta diferença tornou-se evidente entre as palavras de cinco e seis letras. Assim, genericamente, depois de controladas outras variáveis relevantes como a frequência, o número de sílabas, o número de vizinhos ortográficos e a consistência do ataque da primeira sílaba, os nossos resultados demonstram que a extensão tem um impacto significativo no reconhecimento de palavras escritas em Português quando as palavras e as não-palavras são apresentadas juntas, mas não quando as palavras e as não-palavras são apresentadas agrupadas, em listas separadas.

Os efeitos de lexicalidade, de frequência em palavras, e a interacção entre lexicalidade e extensão foram também obtidos, replicando para o Português os resultados que têm sido descritos para outras línguas e que são visíveis independentemente da profundidade ortográfica (e.g., Weekes, 1997; Juphard, Carbonnel & Valdois, 2004; Valdois, Carbonnel, Juphard, Baciú, Ans, Peyrin & Segebarth, 2006; Peressotti & Mulatti, 2005).

#### *Efeitos de extensão dependentes da tarefa*

O facto do efeito de extensão em palavras ter sido dependente da tarefa sugere fortemente que ler em Português não é como ler em ortografias altamente transparentes. Estudos com o Italiano (Peressotti & Mulatti, 2005) e com o Espanhol (Cuetos & Barbón, 2006), ambas ortografias altamente transparentes, demonstram que o efeito de extensão é robusto independentemente da tarefa ser de leitura em voz alta ou de decisão lexical e das listas serem mistas ou agrupadas. De notar que nestes estudos os

efeitos de extensão foram significativos com estímulos semelhantes aos que utilizámos neste trabalho e com amplitudes de variação da extensão similares (cinco e seis letras no Italiano; entre quatro e seis letras no Espanhol), o que aumenta a confiança nas comparações directas. A ubiquidade dos efeitos de extensão encontrada nestas línguas é um marcador comportamental de que nas ortografias altamente transparentes as estratégias de conversão grafema-fonema são preponderantes no reconhecimento de palavras escritas. No âmbito da hipótese psicolinguística da granularidade (Ziegler & Goswami, 2005; Goswami & Ziegler, 2006; Ziegler & Goswami, 2006), pode dizer-se que a unidade granular mais pequena, ao nível do grafema/fonema, é a unidade de processamento fonológico óptima. No Português, pelo contrário, os efeitos de extensão em palavras não foram ubíquos.

O número de letras teve um impacto pequeno, ainda que significativo, em tarefas de decisão lexical e de leitura com listas mistas, sugerindo que a conversão grafema-fonema está patente neste contexto. Contudo, quando as palavras foram apresentadas de modo agrupado, sem não-palavras, a extensão não teve qualquer impacto na velocidade e na exactidão de leitura, traduzindo a preponderância do conhecimento lexical na leitura e o recurso a unidades de processamento fonológico maiores do que o grafema/fonema. Este padrão de resultados sugere que, no Português, mais do que acontece no Italiano ou no Espanhol, os leitores mudam adaptativamente de estratégia de leitura de acordo com as exigências da tarefa. Quando a tarefa encoraja a utilização de unidades de processamento mais pequenas, como é o caso das listas mistas (as não-palavras não podem ser lidas sem recurso à conversão grafema-fonema e o leitor não pode prever se o item seguinte é palavra ou não-palavra), a unidade fonológica preferencial é o grafema/fonema. Quando

a tarefa encoraja o processamento lexical, como é o caso das listas agrupadas, a unidade preferida é maior, eventualmente ao nível da palavra. Deste modo, na leitura hábil em Português parecem existir diferentes unidades de processamento fonológico igualmente efectivas, que são recrutadas de modo flexível em resposta adaptativa a diferenças nas características da tarefa, mesmo quando essas diferenças são subtis, como é o caso das listas mistas vs. agrupadas. A conversão grafema-fonema é assim importante na leitura hábil em ortografias de profundidade intermédia mas não assume o carácter de unidade de codificação fonológica óptima, sendo utilizada a par de outras unidades maiores. Comportamentalmente, a modulação dos efeitos de extensão pelas características da tarefa pode ser mais evidente nestas ortografias, onde o Português se inclui, precisamente porque a leitura não depende quase exclusivamente de estratégias de conversão grafema-fonema, como acontece nas altamente transparentes.

Na sequência dos resultados sobre a aprendizagem da leitura (Seymour, Aro & Erskine, 2003) e a dislexia desenvolvimental em Português (Sucena, Castro & Seymour, em revisão), os nossos dados estendem para a leitura hábil a ideia de que os processos cognitivos envolvidos na leitura em Português diferem das ortografias altamente transparentes. É assim reforçada a posição intermédia do Português Europeu no *continuum* transparência-profundidade e a hipótese de que essa posição tem impacto nos processos envolvidos na leitura. Ziegler e colaboradores (2001) referem que as diferenças entre línguas na leitura têm sobretudo que ver com o tamanho das unidades de processamento fonológico predominantes, o número de níveis diferentes de unidades granulares utilizados, e também com a flexibilidade dos leitores para mudar entre esses níveis. O nossos resultados apontam para



a hipótese de que nas ortografias de profundidade intermédia existem mais níveis de unidades granulares utilizadas e os leitores têm maior flexibilidade na mudança entre eles do que os leitores das ortografias altamente transparentes.

*Implicações para a literatura sobre efeitos de extensão*

Os nossos resultados podem também ser relevantes, cremos, para compreender alguma da confusão que caracteriza os resultados que têm sido descritos na literatura sobre efeitos de extensão. New, Ferrand, Pallier e Brysbaert (2006) chamam atenção para a dificuldade em extrair um padrão de resultados claro dos estudos que pretendem averiguar em que medida a extensão das palavras determina a leitura. Parece-nos que parte desta variabilidade pode dever-se ao facto de os efeitos de extensão serem dependentes da tarefa, por um lado, e da língua, por outro. No que respeita à tarefa, o impacto da extensão poderá ser tanto maior quanto mais a tarefa dependa do processamento fonológico e estimule a utilização de unidades de processamento fonológico pequenas. No que respeita à língua, os efeitos de extensão deverão ser tanto mais evidentes quanto mais as correspondências entre letras e sons da ortografia são consistentes. Os nossos resultados ilustram estes aspectos, tendo-se observado que apenas uma manipulação subtil na tarefa, listas mistas ou agrupadas, é passível de conduzir a conclusões diversas. Esta aparente contradição nos resultados pode, contudo, ser compreendida à luz das características da língua e da tarefa: o cruzamento das complexidades que fazem do Português uma ortografia não altamente transparente com um paradigma experimental que não encoraja a utilização de unidades de processamento fonológico pequenas (listas

agrupadas), conduziu a um efeito nulo da extensão; quando se utilizaram tarefas que promoviam a utilização de unidades de processamento fonológico pequenas (listas mistas), os efeitos de extensão emergiram apesar das complexidades da ortografia.

Um terceiro eixo a ter em conta, além da tarefa e da língua, deverá ser o poder do design experimental. Estudos com o Inglês, uma ortografia altamente inconsistente, na qual os efeitos de extensão tendem a ter um impacto muito reduzido ou nulo (Weekes, 1997), documentaram um impacto significativo da extensão na leitura quando era utilizado um design que testava milhares de itens que variavam na extensão entre pelo menos duas e sete letras e numa tarefa que promovia o processamento fonológico, a leitura em voz alta (Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004; Baayen, Feldman & Schreuder, 2006). Com efeito, as unidades de codificação fonológica pequenas, ao nível do grafema/fonema, podem ser utilizadas mesmo nas ortografias inconsistentes. Não obstante, estas estratégias são pouco relevantes nestas ortografias, porque ineficazes, o que explica a necessidade de designs experimentais muito poderosos para as identificar. Nas ortografias mais consistentes, pelo contrário, os efeitos de extensão são mais facilmente observáveis com muito menos itens e menor amplitude de variação da extensão, demonstrando que as estratégias de conversão grafema-fonema são mais importantes do que no Inglês. Este é o caso do Italiano (Peressotti & Mulatti, 2005) e do Espanhol (Cuetos & Barbón, 2006), em que a conversão grafema-fonema é dominante na leitura e os efeitos de extensão são ubíquos, mas é também a situação do Português, em que os efeitos de extensão foram visíveis em algumas tarefas com amplitudes de variação da extensão reduzidas.

*Implicações para os modelos de leitura*

O padrão de resultados aqui apresentado não pode ser totalmente explicado por modelos que fazem predições específicas quanto aos efeitos da extensão em número de letras na leitura, como o DRC (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) ou o CDP+ (Perry, Ziegler & Zorzi, 2007). O DRC (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) é uma realização computacional dos modelos de dupla via, simulando um conjunto de efeitos que têm sido observados em humanos em tarefas de leitura em voz alta e de decisão lexical. Este modelo prevê que os efeitos de extensão sejam visíveis nas não-palavras, devido a estas serem processadas por meio da via fonológica de leitura, mas não nas palavras, que deverão ser lidas através da via lexical. Desta forma, o modelo não explica, nem os efeitos de extensão que obtivemos em palavras (visíveis igualmente para as de frequência alta e para as de frequência baixa), nem a modulação dos efeitos de extensão pelo tipo de listas, mistas ou agrupadas. O modelo conexionista CDP+ (Perry, Ziegler & Zorzi, 2007) também distingue entre dois processos gerais de leitura, um mais baseado na codificação fonológica grafema-fonema e outro no conhecimento lexical. Contudo, ao contrário do DRC, que é totalmente definido *a priori*, o CDP+ recorre a algoritmos de aprendizagem. Embora o CDP+ seja capaz de simular com sucesso a interacção entre lexicalidade e extensão, também não faz qualquer previsão quanto aos efeitos diferenciais da extensão em listas mistas e em listas agrupadas. Os modelos actuais de leitura podem assim necessitar de ser ajustados para serem capazes de explicar com sucesso as diferenças translinguísticas no processamento de palavras escritas.

## Conclusão

Num conjunto de estudos no Português Europeu, demonstrámos que o efeito de extensão em palavras é dependente da tarefa. As palavras curtas são processadas mais rapidamente do que as longas nas tarefas de decisão lexical e de leitura em voz alta com listas mistas, mas a extensão não tem qualquer impacto na leitura em voz alta com listas agrupadas. Estes resultados sugerem que os leitores hábeis portugueses implementam estratégias de conversão grafema-fonema quando estas são úteis no contexto da tarefa, mas usam unidades de codificação fonológica maiores do que o fonema quando a tarefa promove o processamento lexical. Assim, os leitores portugueses evidenciam maior flexibilidade no seu uso estratégico de diferentes unidades de processamento fonológico na leitura do que os leitores de ortografias altamente transparentes, para quem as estratégias de conversão grafema-fonema são predominantes e quase exclusivas (Spinelli, De Luca, Di Filippo, Mancini, Martelli & Zoccolotti, 2005; Peressotti & Mulatti, 2005; Cuetos & Barbón, 2006). Por outro lado, a leitura em Português parece também afastar-se das ortografias altamente inconsistentes, nas quais a utilização de unidades de codificação fonológica maiores, nomeadamente ao nível da palavra, são preponderantes, e a conversão grafema-fonema tem um papel reduzido ou mesmo nulo (Weekes, 1997).

## Referências

- Bacelar do Nascimento, M. F., Casteleiro, J. M., Marques, M. L. G., Barreto, F., Amaro, R., & Veloso, R. (s.d.). *Corlex: Léxico de frequências do português* [Base lexical]. Consultado em <http://www.clul.ul.pt> (30 de Outubro de 2007).
- Baayen, R., Feldman, L., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, *55*, 290-313.
- Balota, D., Cortese, M., Sergent-Marshall, S., Spieler, D., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283-316.
- Bates, E., Burani, C., D'Amico, S., & Barca, S. (2001). Word reading and picture naming in Italian. *Memory & Cognition*, *29*(7), 986-999.
- Castro, S. L., & Gomes, I. (2000). *Dificuldades de aprendizagem da língua materna*. Lisboa: Universidade Aberta.
- Cedrus Corporation (2007). *SuperLab*. Consultado em <http://www.superlab.com> (3 de Dezembro de 2007).
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204 - 256.
- Cuetos, F., & Barbón, A. (2006). Word naming in Spanish. *European Journal of Cognitive Psychology*, *18*(3), 415-436.

- Ellis, N., & Hooper, M. (2001), Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics*, 22, 571-599.
- Ellis, N., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V., Polyzoe, N., Tsipa, M., & Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, 39(4), 438-460.
- Frost, R. (1994). Prelexical and postlexical strategies in reading: Evidence from a deep and a shallow orthography. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 116-129.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123(1), 71-99.
- Frost, R. (2006). Becoming literate in Hebrew: the grain size hypothesis and Semitic orthographic systems. *Developmental Science*, 9(5), 439-440.
- Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 104-115.
- Gomes, I. (2001). *Ler e escrever em Português Europeu*. Tese de doutoramento não publicada, Universidade do Porto, Portugal.
- Gomes, I., & Castro, S. L. (2003). Porlex, a lexical database in European Portuguese. *Psychologica*, 32, 91-108.
- Goswami, U., & Ziegler, J. (2006). Fluency, phonology and morphology: A response to the commentaries on becoming literate in different languages. *Developmental Science*, 9(5), 451-453.

- Hoxhallari, L., Daal, V., & Ellis, N. (2004). Learning to read words in Albanian: A skill easily acquired. *Scientific Studies of Reading, 8*(2), 153-166.
- Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition, 55*, 332-340.
- Katz, L., & Feldman, L. B. (1983). Relation between pronunciation and recognition of printed words in deep and shallow orthographies. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*(1), 157-166.
- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review, 13*(1), 45-52.
- Perea, M., Rosa, E., & Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics, 67*(2), 301-314.
- Perry, C., Ziegler, J., & Zorzi, M. (2007). Nested incremental modelling in the development of computational theories: The CDP+ Model of reading aloud. *Psychological Review, 114*(2), 273-315.
- Peressotti, F., & Mulatti, C. (Setembro, 2005). Reading pseudowords: base-word frequency effect and length effect. *Proceedings of the XIV Conference of The European Society for Cognitive Psychology, 99-99*.
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Pesenti, S., Gallagher, A., Perani, D., Price, C., Frith, C., & Frith, U. (2000). A cultural effect on brain function. *Nature Neuroscience, 3*(1), 91-96.

- Rayner, K., Sereno, S., Lesch, M., & Pollatsek, A. (1995). Phonological codes are automatically activated during reading: Evidence from an eye movement priming paradigm. *Psychological Science, 6*(1), 26-32.
- Seymour, P. H. K., Aro, M., Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143-174.
- Share, D. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.
- Simon, G., Bernard, C., Lalonde, R., & Rebai, M. (2006). Orthographic transparency and grapheme-phoneme conversion: An ERP study in Arabic and French readers. *Brain Research, 1104*(1), 141-152.
- Spinelli, D., De Luca, M., Di Filippo, G., Mancini, M., Martelli, M., & Zoccolotti, P. (2005). Length effect in word naming in reading: Role of reading experience and reading deficit in Italian readers. *Developmental Neuropsychology, 27*(2), 217-235.
- Sucena, A., Castro, S. L., & Seymour, P. (em revisão). Developmental dyslexia in an orthography of intermediate depth: The case of European Portuguese.
- Valdois, S., Carbonnel, S., Juphard, A., Baciú, M., Ans, B., Peyrin, C., & Segebarth, C., (2006). Polysyllabic pseudo-word processing in reading and lexical decision: Converging evidence from behavioural data, connectionist simulations and functional MRI. *Brain Research, 149*-162.
- Weekes, B. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology, 50A*(2), 439-456.



- Wydell, T., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural correlates of letter-string length and lexicality during reading in a regular orthography. *Journal of Cognitive Neuroscience, 15*(7), 1052-1062.
- Ziegler, J., Perry, C., Jacobs, A., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science, 12*(5), 379-384.
- Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3-29.
- Ziegler, J., & Goswami, U. (2006). Becoming literate in different languages: Similar problems, different solutions. *Developmental Science, 9*(5), 429-436.

## **Anexos**

Anexo 1. *Estímulos utilizados no Estudo 1*

Palavras				Não-palavras			
Frequência Alta		Frequência Baixa		Freq. Palavra-base Alta		Freq. Palavra-base Baixa	
5 letras	6 letras	5 letras	6 letras	5 letras	6 letras	5 letras	6 letras
berço	branco	banha	bastão	berpo	bralco	banfa	bascão
bispo	brinco	bilha	brecha	bismo	brilco	bilua	brelha
caixa	caixão	boina	bronco	caiva	caitão	boipa	broaco
circo	cancro	caule	calhau	cirto	canfro	caute	caliau
cobre	centro	ceifa	caução	coble	cengro	ceica	cautão
dança	ficção	dogma	charro	danta	ficlão	dogna	chabro
doido	filtro	farsa	crente	doigo	filbro	farla	crelte
febre	franco	feudo	crosta	febne	fraico	feuto	cronta
filme	grelha	fisga	desvão	filve	grenha	fisna	desfão
golfe	leilão	ganso	fausto	golte	leicão	ganvo	faucto
golpe	língua	genro	fresta	golme	línua	genco	frenta
gordo	mancha	jarro	gralha	gorno	manlha	jarbo	granha
justo	montra	lacre	grémio	jusbo	monfra	lacte	grélio
lábio	padrão	leigo	guelra	lábua	padião	leimo	gueira
lesão	parque	manco	lastro	lesio	pargue	manfo	lascro
marco	pátria	melga	mansão	marno	pátua	melta	mantão
metro	planta	naipe	neutro	metso	plasta	naire	neucro
ninho	quinta	pauta	pinhão	ninto	quiata	pauda	pinião
pátio	rancho	plebe	plasma	pátua	ranlho	plete	plaima
pausa	sombra	quina	prenda	pauna	sompra	quiba	presda
regra	tanque	rasgo	quelha	regma	tangue	rasdo	quenha
relva	tensão	repto	quisto	relta	tenlão	repto	quilito
salto	treino	sacro	rasgão	salco	treuno	sacno	rasbão
texto	triste	tango	térreo	texpo	trinte	tanco	térmeo
verde	versão	vulgo	tralha	verfe	verbão	vulfo	tranha

## Anexo 2. Estímulos utilizados no Estudo 2

Palavras					
Frequência Alta			Frequência Baixa		
4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
base	balão	bosque	bege	banha	bastão
bebé	berço	branco	bibe	bilha	brando
bife	bispo	brinco	bode	boina	brecha
café	caixa	caixão	caju	caule	calhau
cego	circo	centro	cepo	ceifa	cifrão
cume	cobre	contra	cone	cloro	caução
doce	dança	dragão	demo	dogma	dicção
face	febre	ficção	feno	farsa	franja
fome	filme	filtro	figo	feudo	fraque
fumo	fruto	franco	fole	fisga	fresta
gota	golfe	grelha	galã	ganso	gralha
gozo	golpe	grosso	gare	gorro	grémio
guia	gordo	guarda	grua	grilo	guelra
leão	lábio	leilão	laje	lacre	lastro
lixo	lesão	língua	lupa	leigo	lustro
maçã	marco	mancha	muco	manco	mansão
maré	metro	montra	musa	melga	mescla
nulo	ninho	núcleo	nexo	naipe	neutro
pêra	pátio	padrão	peru	pauta	plasma
puré	pausa	pátria	poça	plebe	prensa
rede	relva	rancho	rubi	rasgo	rasgão
sete	sócio	sombra	sapo	sacro	salmão
táxi	texto	tanque	tabu	tampo	térreo
tese	tigre	triste	tule	tango	tralha
vale	verde	versão	veto	vulgo	vítreo

Não-palavras					
Frequência da Palavra-base Alta			Frequência da Palavra-base Baixa		
4 letras	5 letras	6 letras	4 letras	5 letras	6 letras
bafe	baluo	bosgue	beze	balha	basdão
beté	berpo	branto	bine	binha	branfo
bifo	bismo	brinfo	bope	boipa	brenha
cané	caiva	caitão	caji	caute	calhiu
cefo	cirto	cendro	ceto	ceiga	ciftão
cuma	coble	confra	cove	cloto	cautão
dole	danta	dranão	dema	dogna	dicfão
fabe	febne	fictão	fene	farna	franta
foma	filve	filbro	fizo	feuco	frague
funo	frupo	frando	fope	fista	frelta
gote	golte	grenha	ganã	ganfo	granha
gobo	golme	grotso	garo	gorre	grébio
guea	gorno	guarma	grue	gripo	guelpa
leuo	lábuo	leifão	laja	lacde	lasbro
libo	lesio	línua	lupe	leibo	luspro
matã	marno	manlha	muca	manzo	mantão
mafé	metso	mondra	musi	melfa	mesfla
nule	ninhe	núcteo	nefo	naile	neucro
pêma	pátuo	padlão	penu	pauda	plasta
pufé	pauna	pátnia	pola	plefe	prenfa
redo	relta	ranlho	ruli	rasno	rasfão
seto	sólio	sompra	salo	saclo	salgão
tápi	texco	tangue	tadu	tambo	térleo
tepe	tigle	trisve	tuve	tanfo	tranha
valu	verpe	verfão	vepo	vuldo	vítneo