

Frequência na Fonologia do Português: recursos e aplicações*

Marina Vigário, Sónia Frota, Fernando Martins & Marisa Cruz

Universidade de Lisboa

0. Introdução

Neste artigo sintetizamos algum do trabalho que tem vindo a ser desenvolvido nos últimos anos no domínio da frequência das unidades e padrões fonológicos do Português, fundamentalmente no/em colaboração com o Laboratório de Fonética, ou usando recursos disponibilizados por esta unidade.

Depois de um breve enquadramento evidenciando a importância do estudo da frequência (secção 1), apresentamos um conjunto de recursos desenvolvidos para o acesso a informação sobre frequência fonológica no Português (secção 2). Na secção 3, procedemos a uma revisão de trabalhos feitos neste domínio usando estes recursos, em diferentes áreas. Algumas direcções do trabalho futuro ou em curso são apontadas na secção 4. A secção 5 conclui o artigo com notas finais.

1. Enquadramento

Está amplamente demonstrado que a frequência de ocorrência de unidades ou padrões linguísticos desempenha um papel muito importante em diversas áreas relacionadas com a linguagem.

A frequência é desde há muito tomada como reveladora do que é ou não tendencialmente universal e do que é ou não marcado nas línguas do mundo ou numa língua particular (e.g., Hyman, 1975, 2008; Clements, 2009, entre muitos outros). Para além disso, é sabido que existem processos que dependem da frequência das palavras. Por exemplo, a elevada frequência das palavras pode potenciar processos de redução (e.g., Jurafsky, Bell & Girand 2001) ou inibir a regularização de formações irregulares (e.g., Bybee & Hooper, 2001). A emergência e sucesso de modelos formais relativamente recentes que incorporam informação probabilística, como o da Teoria da Optimalidade Estocástica de Boersma (1998), e trabalhos posteriores, mostram bem o reconhecimento da importância que hoje se confere à frequência, mesmo na área da fonologia formal.

São também inúmeros os trabalhos desenvolvidos nos últimos anos no domínio da aquisição da fonologia que mostram correlações entre a frequência no *input* e a ordem de emergência das categorias e padrões linguísticos (ver, por exemplo, a compilação em Demuth (2006) e referências aí incluídas e vários dos revistos na secção 3). Levelt, Schiller & Levelt (2000), por exemplo, sugerem explicitamente que o percurso de desenvolvimento da estrutura silábica é guiado pela frequência. E a

* Este artigo foi desenvolvido no âmbito do projecto *FreP: Padrões de Frequência na Fonologia do Português - Investigação e Aplicações*, PTDC/LIN/70367/2006.

mesma linha surgem modelos de aprendizagem que incorporam este factor (e.g., Boersma & Levelt, 2000; e a discussão em Jarosz, 2010). De resto, também ao nível da percepção/compreensão se sabe desde há muito que os bebés, em fases precoces de aquisição, mostram sensibilidade a informação estatística e distribucional no *input* (e.g., Jusczyk, Luce & Charles-Luce, 1994).

A frequência fonológica e de palavra são factores relevantes recorrentemente reportados nos estudos de processamento (e.g., Dell, 1990; Caramazza *et al.*, 2001; e as referências em Dell & Gordon, 2003). E, especificamente no âmbito do processamento na leitura, múltiplos estudos usando *eye-tracking* têm mostrado que, entre os factores com mais impacto no tempo de reconhecimento das palavras, está a sua frequência (e.g., Drieghe, Desmet & Brysbaert, 2007; Rastle, 2007).

Relacionada com a frequência, está também a previsibilidade. A importância deste factor é observada em áreas diversas: palavras altamente previsíveis são mais rapidamente reconhecidas e produzidas do que palavras com menor grau de previsibilidade (e.g., ver referências em Dell & Gordon, 2003); experiências de percepção com bebés mostram que as crianças usam informação sobre probabilidades de transição para a segmentação de palavras (Saffran, Aslin, & Newport, 1996); experiências com *eye-tracking* mostram que a previsibilidade das palavras se correlaciona com o comportamento ocular na leitura (e.g., Drieghe, Desmet & Brysbaert, 2007; Rastle, 2007).

Se bem que seja cada vez mais claro o importante papel da frequência de unidades e padrões linguísticos numa multiplicidade de áreas, até recentemente poucos dados estavam disponíveis sobre a frequência de objectos fonológicos no Português Europeu (Andrade & Viana, 1994; Vigário & Falé, 1994; Viana, Trancoso, Silva, Marques, Andrade & Oliveira, 1996). Este facto e a inexistência de recursos que permitissem extrair informação de frequência de *corpora* novos motivaram o desenvolvimento de um projecto iniciado informalmente em 2004, no Laboratório de Fonética da Faculdade de Letras de Lisboa. O objectivo do projecto era contribuir para o conhecimento sobre a frequência de unidades e padrões fonológicos no Português, desenvolver recursos nesta área e promover investigação nova em domínios em que a informação sobre frequência fonológica pode ser relevante. O projecto original deu lugar a um projecto financiado pela Fundação para a Ciência e Tecnologia, iniciado em 2007 (*FreP: Padrões de Frequência na Fonologia do Português - Investigação e Aplicações* – ver <http://www.fl.ul.pt/LaboratorioFonetica/frep/>). No âmbito desse projecto, (i) construíram-se vários recursos que permitem hoje o acesso a nova informação sobre frequência fonológica no Português; (ii) disponibilizaram-se novos dados de frequência de unidades e padrões fonológicos na língua, tendo em conta um conjunto de variáveis relevantes; e (iii) desenvolveu-se investigação sobre o papel da frequência fonológica em diversas áreas. Versões pré-publicação de vários dos trabalhos apresentados ou publicados no âmbito do projecto, ou fazendo uso dos recursos aí desenvolvidos, estão disponíveis na página do projecto. O presente artigo pretende fornecer uma panorâmica geral dos avanços recentes neste âmbito.

2. Recursos no domínio da frequência fonológica no Português

Central a todo o trabalho em curso é a ferramenta electrónica *FreP*, a partir da qual se construíram o *FreP_B* e o *FreP_LUDO*, e que alimentou a base de dados *FrePOP* e permitiu obter os dados usados no teste da ferramenta *TP – Transitional Probabilities*. Descreve-se abaixo os diversos recursos.

FreP – O *FreP* (Martins, Vigário & Frota, 2004-2010) é uma ferramenta electrónica que possibilita a extracção de informação de frequência de unidades e padrões fonológicos desde o nível da palavra, inclusive, até ao traço. As várias etapas do seu desenvolvimento aparecem descritas em Vigário, Martins & Frota (2005), Frota, Vigário & Martins (2006), Vigário, Frota & Martins (2006, 2010, submetido). Até onde sabemos, esta é uma ferramenta única. Destacaríamos entre as suas propriedades a quantidade de informações de frequência fonológica que fornece, a possibilidade de ser usada em diferentes *corpora* e a sua arquitectura.

Uma das características do *FreP* é extrair informação fonológica a partir do texto escrito (seguindo as convenções ortográficas). Isso permite obter informação de frequência quer a partir de texto originalmente escrito, quer a partir de dados orais transcritos ortograficamente. Tal é possível porque se assume-se a fonologia obrigatória (largamente coincidente com a fonologia lexical), o que significa que se capta essencialmente a fonologia invariante. Esta opção, com critérios bem definidos, permite (i) a extracção dos dados a partir de registos ortográficos; (ii) o uso de conceitos como *sílaba* sem o confronto com as situações de queda de vogal; (iii) a captação de um nível do conhecimento fonológico para o qual a gramática fornece evidência robusta (e.g., Vigário, 2003).

É uma ferramenta de uso intuitivo, que corre sobre ficheiros de texto não formatados, gravados como *plain text ANSI* (*.txt). A página do projecto inclui uma demonstração que possibilita uma introdução rápida à sua utilização.

A ferramenta é de domínio público, sendo o seu uso livre para investigação e para ensino/aprendizagem, com fins não comerciais, e pode ser obtida contactando o primeiro autor.

A última versão da ferramenta (em conclusão) contém o essencial das funcionalidades inicialmente previstas e vários acrescentos e modificações relativamente às versões anteriores.

A janela inicial, aberto o programa e corrido o *FreP* sobre o ficheiro de texto, contém os comandos *Export*, *PW*, *Clitics*, *Syllables*, *Segments*, *Features*, *Stress* e *Lexical Frequency*. No comando *Export* encontram-se submenus a partir dos quais se pode criar três tipos de ficheiros: em *Report* gera-se um ficheiro Excel, onde surge uma parte significativa dos dados disponibilizados de modo não-manipulável no *display* da ferramenta, e que auxilia no trabalho de extracção e tratamento dos dados; em *PW file* e *Clitics file* são gerados ficheiros não formatados (do mesmo tipo que o ficheiro de *input* do *FreP*) contendo apenas palavras prosódicas e clítics, respectivamente.

Os restantes comandos do menu inicial permitem aceder à identificação e contagem das palavras prosódicas (e.g., *sol*, *brinquedo*, *lindos*, *-zinho*, *pré-*), dos clíticos (e.g., *que*, *de*, *o*), dos tipos silábicos (e.g., CV, V, CVC), das sílabas concretas (e.g., [6S], [u] [n6~w~], em transcrição SAMPA), dos segmentos (e.g., [t], [S], [6]) e classes maiores de segmentos (C, V, G), dos traços fonológicos segmentais (e.g., [contínuo], [anterior], [nasal]) e dos padrões acentuais (final, penúltimo, antepenúltimo). Sobre as noções em causa e explicitação de opções importantes tomadas, veja-se em particular Vigário, Martins & Frota (2006), e referências aí incluídas. Em relação aos traços, uma das funcionalidades mais recentes, as opções feitas são compatíveis com a classificação dos segmentos que se encontram, designadamente, em Mateus, Andrade & Viana (1990) e Mateus & Andrade (2000).

Para além de valores de ocorrência total, para cada uma destas categorias/padrões fonológicos há um conjunto adicional de informações. Relativamente às palavras prosódicas e aos clíticos, o *FreP* fornece informação sobre a quantidade de palavras com diferentes tamanhos (em número de sílabas e em número de segmentos fonéticos). Para a palavra prosódica, é possível conhecer também a sua extensão incluindo enclíticos (como em *alegrar-se*) e/ou proclíticos (como em *a maravilha*). Relativamente aos clíticos, fornece ainda informação quantitativa a respeito da sua ocorrência proclítica (como em *de manhã*) ou enclítica (como em *rebolam-se*).

Para além da listagem e contagem das sílabas concretas (uma funcionalidade recente), permite a identificação e contagem dos tipos silábicos, independentemente do contexto, e/ou em função da posição na palavra e/ou da presença/ausência de acento. Pode assim conhecer-se, por exemplo, o número de ocorrências do tipo silábico CVC em posição átona final (como em *fácil*).

A contabilização dos segmentos fonéticos é feita nos mesmos moldes que a dos tipos silábicos, sendo ainda fornecida a frequência dos diversos segmentos em função da posição no interior da sílaba e em função da presença em sílaba inicial, medial, final ou em monossílabos. Interessa referir que os dados numéricos disponibilizados incluem as ocorrências de *V-slot*, isto é, da posição vocálica que pode ser instanciada com *schwa* quando há sequências consonânticas que não permitem gerar sílabas bem formadas no Português e que passam a ter um tratamento regular pela presença dessa posição vocálica. Exemplos deste tipo ocorrem em palavras como *admitia* ou *está* ([6.dV.mi.t'i.6], [VS.t'a]) (veja-se sobre o assunto, em particular, Mateus & Andrade, 2000, Freitas & Rodrigues, 2003 e, especificamente sobre esta opção no *FreP*, Vigário et al. 2006).

A contabilização dos traços é feita em função das mesmas variáveis que os segmentos: pode-se assim saber, por exemplo, o número de ocorrências do traço [-contínuo] em sílaba inicial acentuada.

Relativamente à distribuição do acento, para além de a ferramenta permitir contabilizar o número de palavras com as três localizações de acento admitidas no Português e em monossílabos (considerando ou não *V-slot*), fornece ainda dados sobre a localização do acento em função do tamanho de palavra (permitindo, por exemplo, excluir do número de palavras com acento antepenúltimo os dissílabos, que não podem instanciar esse acento por razões triviais de tamanho).

Pelo interesse desta informação, e porque o formato concreto das palavras também é informação fonológica, optou-se por introduzir informação sobre frequência lexical entre as funcionalidades directamente acessíveis a partir do *display* da ferramenta (em versões anteriores tal estava acessível apenas a partir do Relatório gerado pela ferramenta). Nesta janela, são mostradas todas as palavras, com o respectivo número de ocorrências e ordem de frequência.

Na janela inicial, para além do menu de comandos, surge uma caixa contendo os valores totais de palavras prosódicas, clítics, sílabas e segmentos, bem como de palavras ortográficas (correspondentes à noção de *token*, isto é, considerando as diferentes ocorrências de uma palavra), de palavras ortográficas únicas (correspondentes à noção de *types*, considerando apenas uma ocorrência de cada palavra) e de caracteres ortográficos. Para além desta, existem três outras caixas listando (i) as palavras extraídas do texto com a ferramenta para sua análise (palavras prosódicas e palavras clítics); (ii) o texto original; e (iii) as palavras extraídas do texto transcritas foneticamente (em SAMPA).

Em todas as funcionalidades relativas a categorias acima do segmento, e também no caso da distribuição do acento, a ferramenta permite visualizar as palavras em que ocorrem as categorias/padrões indicados – por exemplo, se se pretender saber em que palavra(s) ocorre o tipo silábico *VNC* existente num dado texto, seleccionando-se o tipo silábico pretendido em *Syllables>Types>General*, surge a lista de palavras onde o tipo silábico aparece (e.g., *uns*).

Encontra-se em finalização um número diminuto de correcções ou acrescentos pontuais.

Entre as limitações conhecidas da ferramenta está o tratamento dos algarismos (que têm de ser escritos por extenso para não serem ignorados) e das siglas (cada letra tem de ser escrita por extenso, de modo a não serem tratadas como acrónimos). Outros aspectos que decorrem da ausência de correlatos ortográficos para as unidades são geradores de erro: em muitos contextos as vogais abertas não se distinguem ortograficamente das vogais médias e/ou fechadas, o valor fonético associado ao grafema <x> nem sempre é previsível, e nem sempre se pode saber se a grafia <qu-> e <-gu> representa uma consoante labializada ou não (tal afecta os valores relativos a certos segmentos e traços); nem sempre é previsível quando as consoantes ortográficas são mudas (tal afecta a contagem de traços, segmentos, sílabas e tipos silábicos); também nem sempre existem correlatos escritos para a segmentação de palavras prosódicas (e.g., *monogâmico* é uma palavra ortográfica mas duas palavras prosódicas). Na generalidade dos casos é possível manipular o texto original de modo a eliminar estas fontes de erro.

Uma versão em funcionamento no final de 2009 foi objecto de uma verificação sistemática que insidiu sobre um *corpus* de 76 mil palavras.¹ A taxa de correcção obtida para todos os parâmetros então em funcionamento (que constituem a maior parte da ferramenta, mas não incluem os segmentos e os traços) situou-se acima dos 99%. Verificou-se que uma fatia significativa dos erros obtidos decorria da existência de palavras estrangeiras ou de origem estrangeira não integradas no sistema

¹ Esse trabalho foi feito por Marisa Cruz e Nuno Matos.

fonológico da língua (e.g., *shopping*, *print*). Diferentes versões da ferramenta foram também objecto de verificações mais ou menos pontuais em múltiplas ocasiões, em consequência da utilização da ferramenta para fins variados, o que permitiu a identificação de erros e sua correcção.

FreP_B – Encontra-se em fase de teste uma versão do *FreP* adaptada à variedade brasileira do Português (PB), correspondente à falada na região de São Paulo (*FreP_B*, versão 1, Martins, Vigário, Frota & Svartman-Fernandes 2011). Essa versão foi construída sobre a versão 3.0 do *FreP* e implicou a introdução de novas regras ou a alteração de regras existentes. Por exemplo, diferentemente do *FreP*, o *FreP_B* trata a consoante de *tia* como uma africada palatal, o segmento final de *mel* como [w] e assume que o [i] se realiza obrigatoriamente nos contextos em que no Português Europeu é assumido *V-slot*, como em *afta* [‘a.fi.ta]. Em relação ao conjunto de palavras contabilizadas como clíticas no PB, considerou-se que *pelo(s)/pela(s)* e *cada* formam palavras prosódicas, uma vez que nesta variedade não são conhecidas evidências claras para o seu estatuto de clíticos fonológicos, havendo no restante coincidência com o assumido para o Português Europeu (Vigário, 2003).

FreP_LUDO – O *FreP_LUDO* é uma versão simplificada da ferramenta originalmente construída para uso especializado. Foi concebida para utilização lúdica ou didáctica, em particular por alunos e professores de níveis de ensino pré-universitário.

Entre os trabalhos de adaptação da ferramenta encontra-se o uso sistemático da língua portuguesa e a eliminação de algumas funcionalidades que se entendeu serem úteis apenas em contextos mais especializados (e.g., a que permite obter informação de frequência sobre as combinações de palavras prosódicas e clíticos e muita da informação que consta do Relatório).

Por se tratar de uma versão que visa utilizadores não-especializados, o *FreP_LUDO* inclui notas explicativas sobre os termos usados na ferramenta, que estarão disponíveis através de um comando de pergunta e também aparecerão depois de o cursor ser colocado algum tempo sobre o termo relevante. As noções apresentadas não pretendem ser exaustivas, mas introduzir de modo simples os conceitos relevantes, tornando-os operacionais, fazendo uso de exemplificação abundante.

FrePOP – A base de dados *FrePOP* (Frota, Vigário, Martins & Cruz, 2010) foi criada com o objectivo de fornecer dados de referência sobre a frequência de unidades e padrões fonológicos no Português tendo por base um *corpus* de larga dimensão.²

A *FrePOP*, de consulta *online*, é de uso público, nas mesmas condições que o *FreP* (ver acima), sob registo (<http://frepop.fl.ul.pt/>). Instruções de utilização e informação geral sobre a base de dados podem ser encontradas no manual, fornecido no mesmo local.

² O preenchimento da Versão 1.0 da *FrePOP* foi feita com a colaboração de Nuno Matos, Nuno Paulino e João Lourenço.

Presentemente, a base de dados apresenta valores relativos a um *corpus* total de perto de um milhão de palavras (923 338). Entretanto, progride a bom ritmo o trabalho visando atingir mais de 3 milhões de palavras, a relativamente curto trecho.³

A *FrePOP* é baseada na extracção de dados de frequência com a ferramenta *FreP* a partir de textos escritos de tipo variado (as fontes são fornecidas na opção *About*) e permite, para um conjunto de variáveis, fazer escolhas no sentido de especificar o escopo da busca. As variáveis em causa são *tipo de texto* (escrito, oral); *sub-tipo de texto* (e.g., fala dirigida à criança (CDS), fala dirigida a adulto (CS), entrevista, jornalístico, espontâneo, histórico); *tipo de transcrição* (adaptado, ortográfico), ano (presentemente, desde 1970, mas brevemente desde o século XV); variedade (presentemente, apenas a Europeia, incluindo todas as variedades consideradas em Cintra 1971/1983); género; nível de escolarização (todos os níveis, desde o analfabeto ao universitário); ocupação (categorias segundo a *CNP – Classificação Nacional de Profissões*, às quais se adicionou ‘estudante’).

São fornecidos valores de frequência para as seguintes variáveis: (i) palavras ortográficas (*tokens*); (ii) palavras ortográficas únicas (*types*); (iii) caracteres ortográficos; (iv) palavras prosódicas; (v) clíticos; (vi) sílabas; (vi) segmentos fonéticos; (vii) padrões acentuais. Para cada uma das variáveis fonológicas, um conjunto adicional de informações está acessível. Em relação às palavras prosódicas e aos clíticos, pode conhecer-se o número de ocorrências em função do tamanho (em número de sílabas). No caso dos clíticos, o número de ocorrências de enclíticos e proclíticos é também dado. Relativamente à sílabas, a *FrePOP* fornece informação sobre o número de ocorrências dos 10 tipos silábicos mais frequentes no *corpus* (tendo ou não em conta a posição na palavra e/ou presença/ausência de acento). Quanto aos segmentos, fornece os valores totais dos segmentos em sílaba acentuada ou átona, e das classes maiores de segmentos (C, V, G), incluindo *V-slots*; fornece ainda o número de ocorrências de cada segmento fonético (e.g., [t], [u], [j]), em geral, ou em função da posição na palavra ou na sílaba (ataque, ataque ramificado, etc.), ou ainda da sua ocorrência em sílaba tónica ou átona. Relativamente ao padrão acentual, a *FrePOP* fornece, para cada uma das localizações do acento, os valores referentes à sua distribuição em função do tamanho da palavra.

2.5. *TP – Transitional Probabilities* – Como ficou dito na secção 1, entre as informações de frequência que parece ser altamente produtiva em termos de processamento e aquisição da linguagem encontram-se as probabilidades de transição. No sentido de dispormos de informação a este respeito no Português Europeu, foi construída uma ferramenta, a *TP – Transitional Probabilities* (Garcia & Gonzalez-Lopes, 2010), que permite conhecer, num dado *corpus* da língua, a probabilidade de um segmento fonético ou fronteira de palavra ou de sílaba prever o segmento fonético ou a fronteira de palavra ou de sílaba precedente ou seguinte.

³ Neste trabalho colaboram presentemente, adicionalmente, os tarefeiros Andreia Trindade e Gustavo Duarte.

3. Estudos de frequência no Português

Nesta secção fazemos uma breve descrição de um conjunto de estudos seleccionados, que investigam os padrões de frequência fonológica e o seu impacto em áreas como as da aquisição, a identificação do falante ou a mudança linguística.

Desde o início do projecto em curso no Laboratório de Fonética da Faculdade de Letras de Lisboa foram produzidos vários trabalhos onde são reportados valores de frequência para diversas unidades e padrões fonológicos na língua adulta. Vigário, Martins & Frota (2005) apresentam dados relativos à frequência dos diversos tamanhos de palavra e à frequência relativa de enclíticos e proclíticos. Entre outros, os dados mostram uma taxa elevada de palavras monossilábicas com sílaba aberta, que é considerada reveladora de que o Português não é sensível à restrição de palavra mínima, presente em inúmeras línguas do mundo (e.g., Kenstowicz, 1994). Neste estudo mostra-se ainda que a frequência fonológica não pode explicar a tendência recentemente observada no Português Europeu para os pronomes clíticos aparecerem em ênclise, uma vez que o padrão proclítico é imensamente mais frequente - 97% vs. 3% (valores replicados em Vigário, Frota & Martins 2010, baseado num *corpus* mais de dez vezes maior).

Vigário, Frota & Martins (2006) e Frota, Vigário & Martins (2006) apresentam valores de frequência relativos aos tipos silábicos, às grandes classes de segmentos (C, V, G) e à distribuição do acento. Considerando que o acento é atribuído no Português a partir do final da palavra, conjectura-se, nomeadamente, que os monossílabos podem contribuir para a contagem de palavras com acento final. Contando deste modo, a proporção entre palavras graves na língua não é tão diversa como habitualmente se presume (54.4% e 44%, respectivamente). Ao nível da sílaba, destacamos aqui a evidência de que os diferentes formatos silábicos e respectivas frequências são muito variáveis em função da distribuição na palavra (e.g., os tipos silábicos V e CCV predominam em início de palavra relativamente a outras posições, enquanto CVC e CVGN predominam em sílaba final). Quanto à distribuição relativa de C e V, os dados mostram que vogais são ligeiramente mais frequentes (46% v. 48%). Para além disso, a taxa de ocorrência de *V-slots* é muito baixa (0.2%).

Em Vigário, Frota & Martins (2009), os dados analisados aumentam muito significativamente, fruto de alterações introduzidas na ferramenta *FreP* (o número total de palavras analisadas é de 240.767 palavras ortográficas, cerca de dez vezes mais do que as analisadas nos estudos anteriores). Neste trabalho são mostrados os dados relativos às unidades e padrões apresentados anteriormente e também a segmentos e traços de ponto de articulação. Para além de apresentar dados computados sobre um *corpus* de dimensão muito superior e de possibilitar uma avaliação da importância da variável *dimensão do corpus*, ou simplesmente *corpus*, comparando os resultados com os anteriormente publicados, este artigo fornece uma descrição sistemática dos valores computados sobre *tokens* (i.e., considerando todas as ocorrências de cada palavra) e sobre *types* (i.e., considerando

apenas cada instância de uma palavra). Comparando ambos os tipos de dados, eles revelam diferenças importantes, que são exploradas depois do ponto de vista da natureza do *input* que é relevante para a aquisição da fonologia (o que retomamos abaixo). Entre as diferenças de frequência encontradas em *tokens* e *types*, destacamos as verificadas em (i) todos os formatos de palavra (e.g., 2/3 dos *types* têm 3 ou mais sílabas, mas as palavras mais frequentes considerados os *tokens* têm 1 ou 2 sílabas; os monossílabos e palavras submínimas não têm expressão nos *types*); (ii) ao nível do acento (e.g., considerados os formatos mais pequenos, mais frequentes, há uma proporção mais baixa de palavras graves nos *tokens* do que nos *types*); e (iii) nos formatos silábicos – a ordem de frequência entre os tipos mais frequentes é diversa (CV>V>CVC>CVN>CVGN>CCV, em *tokens*; e CV>CVC>CVN>V>CCV>CVGN, em *types*); os valores de frequência variam (mais CV e CVC em *types* do que *tokens* e mais V e CVGN em *tokens* do que *types*); contudo, é quando considerada a proporção dos diferentes tipos silábicos em função da posição na palavra e/ou presença/ausência de acento que as diferenças maiores surgem entre os dois tipos de contagem (e.g., o tipo V ocorre quase exclusivamente em posição inicial de palavra ou em monossílabos, se considerada a contagem sobre *tokens* mas não sobre *types* – 91% vs. 65%, respectivamente; e os tipos CVGN, VN e VG são muito mais frequentes em sílaba tónica na contagem sobre *tokens* do que na sobre *types*).

Em Vigário, Frota & Martins (submetido) são observadas as distribuições das várias categorias e padrões fonológicos em duas categorias de palavras distintas: palavras prosódicas (PW) e clíticos (CL). São encontradas diferenças muito evidentes em todos os parâmetros (e.g., nos formatos de palavra, já que CL são maximamente dissilábicos; nos tamanhos mais frequentes, uma vez que CL são muito predominantemente monossilábicos, enquanto PW são mais frequentemente dissilábicas e apresentam tamanhos diversificados; no inventário segmental que instanciam, reduzido a metade nos CL; no rácio *token/type*, com valores muito superiores nos CL do que nas PW, indiciando que CL-*types* são em pequeno número, mas ocorrem com elevada frequência). Dado o facto de todas as palavras lexicais serem PW e todos os CL serem palavras gramaticais (e mesmo que não seja total a coincidência entre as noções em causa), os autores sugerem que as diferenças encontradas entre PWs e CLs podem ser usadas pelas crianças para, mesmo numa fase pré-lexical, estabelecerem uma categorização rudimentar das palavras lexicais e das palavras gramaticais.

O impacto da frequência na aquisição da linguagem (em particular da fonologia e da morfologia) foi um dos tópicos mais explorados recentemente. Comparando os valores de frequência relativos aos diferentes formatos de palavra na fala adulta (AS), na fala adulta dirigida à criança (CDS) e nas primeiras produções (CS), Vigário, Freitas & Frota (2006) concluem que a frequência dos diversos formatos no *input* constitui um bom preditor para a ordem de emergência e frequência das palavras de diferentes tamanhos na criança: as palavras dissilábicas predominam na primeiras produções, mas os formatos mais pequenos e maiores também se encontram presentes com bastante incidência. Este facto afasta o comportamento das crianças a adquirir o Português do encontrado em

crianças a adquirir outras línguas, como o Inglês e o Catalão (e.g., em ambas, os formatos maiores tendem a não ocorrer nos dados iniciais, tal como não ocorre, no Inglês, o formato de palavra composto por uma sílaba aberta). Interessantemente, também nessas línguas parece existir uma forte correlação entre a frequência dos formatos de palavra no *input* e os dados exibidos pelas crianças. As autoras mostram ainda que, quando os valores de frequência encontrados no AS e no CDS não convergem, aqueles que melhor se ajustam aos dados da criança são os do AS.

Resultados em grande medida semelhantes são encontrados em Frota, Freitas, Vigário & Martins (2005) e Freitas, Frota, Vigário & Martins (2006) em relação à emergência e desenvolvimento dos tipos silábicos. Verificou-se que a ordem de frequência dos diferentes tipos silábicos se correlaciona largamente com a sua ordem de emergência nas primeiras produções (CV>V>CVC>CVG/CVN>CCV, dados de frequência no AS; CV / V > (C)VN > (C)VG, GV, ordem de emergência no SC). Contudo, verifica-se nos dados da criança um desenvolvimento precoce de V, juntamente com CV, e (C)VG/(C)VN surgem antes de (C)VC, embora no *input* sejam menos frequentes. O estudo avalia a distribuição dos diversos tipos silábicos em função da posição na palavra e da presença/ausência de acento. É observado que os tipos silábicos que emergem numa ordem não predita pela frequência relativa no *input* ocorrem mais em posições proeminentes: no caso de V, a posição inicial; no caso de CVG, CVN a posição acentuada. As autoras concluem que não apenas a frequência no *input* mas também factores de ordem estrutural/gramatical (limites de palavra, proeminência) podem desempenhar um papel na (ordem de) emergência dos tipos silábicos.

Costa, Freitas, Frota, Vigário & Martins (2007) e Costa (2010) avaliam o impacto da frequência dos traços de ponto de articulação (PoA) no *input* nas produções iniciais da criança. Tal como anteriormente, a frequência e distribuição nos dados do adulto parecem relevantes para explicar, por exemplo, por que as crianças adquirindo Português produzem em fases iniciais consoantes labiais e dorsais em início de palavra, contrariamente às que adquirem Holandês, que não produzem dorsais no mesmo estágio. Também aqui se coloca a hipótese de factores estruturais poderem desempenhar um papel de promoção do traço dorsal, uma vez que a ocorrência deste traço acontece predominantemente em início de palavra.

Considerando os estudos anteriores, sugestivos do papel da frequência na ordem de emergência das categorias e padrões fonológicos, Vigário, Frota & Martins (2010) investigam quais os dados do *input* que melhor se correlacionam com os da aquisição: os computados sobre *tokens* ou sobre *types*. Os autores observam que a ordem de emergência de categorias e/ou padrões fonológicos da criança acompanham em geral o que é predito pela frequência computada sobre *tokens*, mesmo quando há divergência nas predições feitas sobre *tokens* e sobre *types* (há diferenças na distribuição dos formatos de palavra, no padrão acentual, nos formatos silábicos e no ponto de articulação).

Outros estudos em aquisição em que a informação de frequência foi crucialmente usada, mas que por razões de espaço não podemos descrever mais detalhadamente incluem: Correia (2010), sobre

a aquisição dos padrões acentuais; Jordão (2009) e Jordão & Frota (2010) sobre a emergência de consoantes em coda, tendo em conta variáveis como a posição na estrutura prosódica e o acento; Frota & Matos (2009) e Matos (2010), onde são considerados aspectos como o formato da palavra e contagem lexical na relação com questões sobre a aquisição dos padrões temporais em fases iniciais; Guerreiro (2007) e Guerreiro & Frota (2010), na construção de um *corpus* balanceado para a avaliação fonológica da fala infantil; Garcia (2010) e Vigário & Garcia (submetido), onde se avalia, designadamente, o impacto da frequência dos formatos de palavra no *input* no desenvolvimento inicial da morfologia, e onde parte do tratamento dos dados passou pela extracção de informação sobre frequência lexical (*tokens* e *types*), tanto em dados da criança, como do adulto.

Uma área nova de aplicação dos dados de frequência é explorada em Aguiar (2008, 2009) e Aguiar & Vigário (2010). O objectivo é identificar indicadores de frequência que se correlacionem com variáveis geográficas e sociolinguísticas. Comparou-se as frequências de um conjunto de unidades e padrões fonológicos em vários pontos da região da Terra Quente Transmontana com os já publicados, baseados em *corpora* do Português Europeu de proveniência variada. Os resultados mostraram a existência de medidas com muito pouca variabilidade (e.g., tamanho de palavra e padrões acentuais). Noutros casos, existe variação em função da variável geográfica (a frequência de *V-Slots* e de C e V, bem como de diferentes tipos silábicos, em particular, V, VN e CVG, distingue os dois *corpora*) e em função de variáveis sociolinguísticas (e.g., indivíduos analfabetos apresentam menos *V-slots*, menos palavras esdrúxulas, mais tipos silábicos CV e V e mais palavras dissilábicas). O estudo contribui para perceber a amplitude e os limites da variação nos valores de frequência das unidades e padrões fonológicos e sugere que este tipo de dados pode concorrer para, em conjunto com outros elementos linguísticos, traçar o perfil de indivíduos, tendo assim aplicação forense.

Um outro estudo que pesquisa a extensão da variação nas diversas variáveis e os factores que se podem relacionar com essa variação incide sobre dados disponibilizados pela *FrePOP* (Frota, Vigário, Martins, Cruz, Matos & Paulino, 2010). Aqui, pesquisou-se o impacto de um conjunto mais alargado de variáveis na frequência de uso de diferentes unidades e padrões fonológicos (i.e., dimensão do *corpus*, tipo de texto, região, nível de escolarização). Entre vários resultados, a investigação revelou que existe uma diferença grande no uso de diferentes formatos de palavra prosódica em função das diversas variáveis. Por exemplo, a frequência dos formatos de palavra varia em função da extensão do *corpus* (contrariamente, por exemplo, às grandes classes de segmentos, C, V e G), tal como varia em função do tipo de texto (e.g., o uso de monossílabos tónicos é bastante superior no texto oral do que no escrito); há também grande variação entre regiões e entre populações com diferentes níveis de escolarização (por exemplo, na região da Grande Lisboa há mais palavras com mais de 2 sílabas e menos com 2 sílabas, quando comparado com outras regiões; e o mesmo padrão se observa na população universitária, contrariamente à iletrada, onde aumentam as palavras com uma única sílaba aberta e diminuem os formatos superiores a duas e sobretudo três sílabas).

Uma outra área pioneira de aplicação do conhecimento sobre a frequência de unidades e padrões fonológicos é a desenvolvida em Frota, Galves & Vigário (2008) e Frota, Galves, Vigário, Gonzalez & Abaurre (submetido). Neste trabalho são comparados os valores de frequência dos formatos de palavra e padrões acentuais em diferentes momentos da história do Português. Sustenta-se que estes aspectos são susceptíveis de captar eventuais mudanças rítmicas operadas no Português em fases anteriores da língua. Os dados revelam um ponto de mudança que é compatível com a hipótese de que uma mudança rítmica pode ter estado na origem da diferença entre a variedade europeia do Português e as restantes línguas/variedades Românicas na colocação dos clíticos pronominais.⁴

4. Outras áreas de aplicação em desenvolvimento

Entre os tópicos de investigação em curso está a exploração dos dados de frequência fonológica para determinar níveis de complexidade em textos, com eventual impacto nos tempos de leitura e no processamento da informação linguística. Um dos factores mais estáveis correlacionados com o tempo de leitura é a extensão da palavra, juntamente com a frequência lexical e a previsibilidade das palavras (e.g. Drieghe, Desmet & Brysbaert, 2007; Rastle, 2007). Este é um entre outros eventuais factores de natureza fonológica que pode influenciar na leitura. Com o pressuposto de que tanto a codificação fonológica como a frequência afectam o comportamento dos indivíduos na leitura silenciosa, encontra-se em curso um estudo onde se procede a uma avaliação da complexidade fonológica e da frequência de unidades fonológicas em textos informativos sobre remédios e procedimentos a seguir no quadro de um processo terapêutico (Cavaco, Vigário & Várzea, em curso).

Está também em desenvolvimento trabalho sistemático conducente à aplicação do *FreP_LUDO* com fins didácticos. Tem sido amplamente mostrado que a consciência fonológica é determinante para a aprendizagem da leitura e da escrita (e.g., Morais, 1991, entre muitos). Por esta razão, presume-se que actividades que promovam a consciência fonológica, como as em curso usando o *FreP_LUDO*, são também susceptíveis de promover a aprendizagem da leitura e da escrita. Assim, para além da ferramenta propriamente dita, está a ser construído um conjunto de materiais propondo actividades que se pretendem apelativas, e que deverão integrar instruções diferenciadas para professores/educadores/cuidadores e crianças ou jovens, em diferentes faixas etárias (inclui-se nesta linha de investigação Cardoso, em curso).

Encontra-se de igual modo em preparação a publicação dos valores das probabilidades de transição segmental no Português Europeu. Como vimos na secção 1, informação probabilística deste tipo pode desempenhar um papel importante no processo de segmentação do contínuo sonoro em

⁴ Outra investigação que recorreu a informação sobre frequência fonológica ou lexical de um modo mais pontual, usando o *FreP*, inclui, entre outros, Correia (2007), Frota, Vigário & Jordão (2008a, 2008b), Frota, Vigário, Cruz & Martins (2008), Nóbrega (2010) e Condelipes (2010).

palavras em fases precoces de aquisição (Saffran, Aslin & Newport, 1996). A título de exemplo, dada a sua distribuição, pode colocar-se a hipótese de a ocorrência de [S] poder ser usada como uma pista para o fim de palavra. Os dados obtidos com a *TP* são compatíveis com essa hipótese. Efectivamente, corrida a ferramenta sobre um *corpus* de 50 000 palavras observa-se que a probabilidade de [S] ocorrer em início de palavra é muito baixa (0.0062), valor que contrasta evidentemente com a probabilidade de ocorrência no final da palavra (0.644); e a diferença não é apenas explicada pela posição na sílaba, uma vez que a probabilidade de ocorrência em fim de sílaba não-final é muito inferior (0.2808).

5. Conclusão

No presente artigo pretendemos fornecer uma panorâmica sobre o estado da investigação e recursos desenvolvidos recentemente no domínio da frequência fonológica no Português.

Apesar das necessárias limitações de espaço, esperamos ter aqui contribuído para mostrar o interesse desta área de pesquisa e o progresso feito nos últimos anos, bem como algumas das linhas de trabalho que julgamos merecer exploração.

Referências

- Aguiar, J. & M. Vigário (2010). Contributos para o estudo da variação na frequência de ocorrência de unidades e padrões fonológicos. A. M. Brito, F. Silva & J. Veloso (Orgs.) *Textos Seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*. Porto: Associação Portuguesa de Linguística, 95-109.
- Aguiar, J. (2008). *Unidades e Processos Fonológicos no falar da região da Terra Quente: contributos para a Linguística Forense*. Tese de Mestrado, Universidade do Minho.
- Andrade, E. & M.C. Viana (1994). Sinérese, diérese e estrutura silábica. In *Actas do IX Encontro da Associação Portuguesa de Linguística*. Lisboa: APL/Colibri, pp. 31-42.
- Boersma, P. & C. Levelt (2000). Gradual constraint-ranking learning algorithm predicts acquisition order. In E. Clark (Org.) *Proceedings of the 30th Annual Child Language Forum*. Stanford, CA: CSLI, 229-237.
- Boersma, P. (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Tese de Doutoramento, Universidade de Amsterdão. LOT International Series 11. The Hague: Holland Academic Graphics.
- Bybee, J. e P. Hopper (2001) (Orgs.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.

- Caramazza, A., A. Costa, M. Miozzo & Y. Bi (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. In *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 1430-1450.
- Cardoso, Ana (em curso). *Promoção da Consciência Fonológica e Desempenho na Escrita*. Tese de Mestrado, Universidade de Lisboa.
- Cavaco, A.M., M. Vigário & D. Várzea (em curso). Phonetic contribution to assess the readability of patient information leaflets: application to irritable bowel syndrome.
- Cintra, L. F. (1971). Nova proposta de classificação dos dialectos galego-portugueses. *Boletim de Filologia* (Lisboa), XXII, (1964-1971); 81-116 + 2 mapas. Reeditado em *Estudos de Dialectologia Portuguesa*. Lisboa, Sá da Costa Editora, 1983, 117-163.
- Clements, N. (2009). The role of features in phonological inventories. In E. Raimy & C. Cairns (orgs.) *Contemporary Views on Architecture and Representations in Phonology*. Cambridge, Ma.: The MIT Press, 19-68.
- Condelipes, T. (2010). *Produção de fala por indivíduos gagos*. Tese de Mestrado, Universidade de Lisboa.
- Correia, P. (2007). *Sob o signo das emoções: expressões faciais e prosódia em indivíduos com perturbação vocal*. Tese de Mestrado, Universidade Católica Portuguesa/Escola Superior de Saúde do Alcoitão.
- Correia, S. (2010). *The Acquisition of Primary Word Stress in European Portuguese*. Tese de Doutoramento, Universidade de Lisboa.
- Costa, T. (2010). *The acquisition of the consonantal system in European Portuguese. Focus on place and manner features*. Tese de Doutoramento, Universidade de Lisboa.
- Costa, T., M. J. Freitas, S. Frota, M. Vigário & F. Martins (2007). Sobre o PA na periferia esquerda da palavra. In M. Lobo e M. A. Coutinho (Orgs.) *Textos Seleccionados. XXII Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 315-328.
- Dell, G. (1990). Effects of frequency and vocabulary type on phonological speech errors. In *Language and Cognitive Processes* 5, 313-349.
- Dell & Gordon (2003). Neighbors in the lexicon: Friends or foes? N. O. Schiller and A. S. Meyer (Orgs.) *Phonetics and phonology in language comprehension and production: Differences and similarities*. Berlin/New York: Mouton de Gruyter, 9-37.
- Demuth, K. (2006) (Org.) Crosslinguistic perspectives on the development of prosodic words. *Language and Speech* 49(2) (Special Issue).
- Drieghe, D., T. Desmet, M. Brysbaert (2007). How important are linguistic factors in word skipping during reading? *British Journal of Psychology* 98(1): 157-171.
- Freitas, M. J. & C. Rodrigues (2003). On the nature of sC-clusters in European Portuguese. *Journal of Portuguese Linguistics* 2(2): 55-85.

- Freitas, M. J., S. Frota, M. Vigário & F. Martins (2006). Efeitos prosódicos e efeitos de frequência no desenvolvimento silábico em Português Europeu. In F. Oliveira & J. Barbosa (Orgs.) *Textos Seleccionados XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 397-412.
- Frota, S. & N. Matos (2009). O tempo no tempo: um estudo do desenvolvimento das durações a partir das primeiras palavras. In A. Fiéis & M. A. Coutinho (Orgs.) *Textos Seleccionados. XXIV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 281-295.
- Frota, S., C. Galves & M. Vigário (2008). Ler a Fonologia: do Português Clássico ao Português Moderno. In A. Fiéis & M. A. Coutinho (Orgs.) *Textos Seleccionados. XXIV Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 193-206.
- Frota, S., M. J. Freitas, M. Vigário & F. Martins (2005). Prosody and frequency effects on the development of syllable structure in European Portuguese. Comunicação apresentada no Symposium “Exploring the Effects of Prosody, Morphology, Frequency and Representation on the Development of Syllable Structure in Romance Languages”, *Xth International Congress for the Study of Child Language*, Berlim, Julho.
- Frota, S., M. Vigário & F. Martins (2006). FreP: An Electronic Tool for Extracting Frequency Information of Phonological Units from Portuguese Written Text. *Proceedings of the 5th International Conference on Language Resources and Evaluation*, CD-ROM, Génova.
- Frota, S., M. Vigário & R. Jordão (2008a). *LumaLiDaOn* (version 1). Lisboa: Laboratório de Fonética da FLUL. <http://www.fl.ul.pt/laboratoriofonetica/lumalidaon.htm> [ISBN 978-989-95713-0-3]
- Frota, S., M. Vigário & R. Jordão (2008b). *LumaLiDaOnLexicon* (versão 1). Lisboa: Laboratório de Fonética da FLUL. <http://www.fl.ul.pt/laboratoriofonetica/lumalidaonlex.htm> [ISBN 978-989-95713-1-0]
- Frota, S., M. Vigário, C. Galves, V. Gonzalez-Lopez & B. Abaurre (submetido). The phonology of rhythm from Classical to Modern Portuguese. *Language Variation and Change*.
- Frota, S., M. Vigário, F. Martins & M. Cruz (2010). *FrePOP*. Version 1.0. Laboratório de Fonética (CLUL), Faculdade de Letras da Universidade de Lisboa. <http://frepop.fl.ul.pt/> [ISBN 978-989-95713-2-7]
- Frota, S., M. Vigário, F. Martins, M. Cruz, N. Matos & N. Paulino (2010). Frequency patterns of European Portuguese: corpus size, spoken/written language, region, and education. Comunicação apresentada no *Workshop on Frequency in Phonology*, Universidade de Lisboa, 26 de Outubro.
- Frota, S., M. Vigário, M. Cruz & F. Martins (2008). *LumaLiDaAudyLexicon* (versão 1), Lisboa: Laboratório de Fonética da FLUL. <http://www.fl.ul.pt/laboratoriofonetica/lumalidaudylex.htm>.
- Garcia J. & V. Gonzalez-Lopez (2010). *TP - Transitional Probabilities* (version 0.12).

- Garcia, P. (2010). *Palavras complexas nas primeiras produções infantis (estudo de caso)*. Tese de Mestrado, Universidade Católica Portuguesa/Escola Superior de Saúde do Alcoitão.
- Guerreiro, H. & S. Frota (2010). Processos fonológicos aos cinco anos de idade: tipologia e frequência. *Cadernos de Saúde* 3. Lisboa: Instituto de Ciências da Saúde, UCP.
- Guerreiro, H. (2007). *Processos Fonológicos na Fala da Criança de Cinco Anos*. Tese de Mestrado, Universidade Católica Portuguesa/Escola Superior de Saúde do Alcoitão.
- Hyman, L. (1975). *Phonology: theory and analysis*. New York: Holt, Rinehart & Winston.
- Hyman, L. (2008). Universals in phonology. *The Linguistic Review* 25(1-2): 83-137.
- Jarosz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning. *Journal of Child Language* 37: 565-606.
- Jordão, R. & S. Frota (2010). The intonational phrase constrains coda development. A. Castro, J. Costa, M. Lobo & F. Pratas (Orgs.) *Language Acquisition and Development. Proceedings of GALA 2009*. Cambridge: Cambridge Scholars.
- Jordão, R. (2009). *A estrutura prosódica e a emergência de segmentos em coda no PE: um estudo de caso*. Tese de Mestrado, Universidade de Lisboa.
- Jurafsky, D., A. Bell, M. Gregory & W.D. Raymond (2001). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In Bybee, J. and P. Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 229-254.
- Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33 (5): 630-645.
- Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.
- Levelt, C., N.O. Schiller & W.J. Levelt (2000). The acquisition of syllable types. *Language Acquisition* 8, 237-264.
- Martins, F., M. Vigário & S. Frota (2004-2010). *FreP - Frequências no Português*. Software in CD-ROM [v. 1.0 IGAC nº 3179/2007; v. 2.0 IGAC nº 5969/2009; v. 3.0 registada no IGAC em 2010].
- Martins, F., M. Vigário, S. Frota & F. Svartman-Fernandes (em curso) *FreP_B*, versão 1.
- Mateus, M.H. & E. Andrade (2000). *The Phonology of Portuguese*. Oxford: Oxford University Press.
- Mateus, M.H., A. Andrade, M.C. Viana & A. Villalva (1990). *Fonética, Fonologia e Morfologia do Português*. Lisboa: Universidade Aberta.
- Matos, N. (2010). *A métrica do tempo: um estudo sobre o desenvolvimento dos padrões de duração durante a aquisição do Português Europeu*. Tese de Mestrado, Universidade de Lisboa.
- Morais, J. (1991). Phonological awareness: A bridge between language & literacy. In D. Sawyer, & B. Fox (Eds.), *Phonological awareness in reading: The evolution of current perspectives*. New York Springer-Verlag, 31-71.

- Nóbrega, B. (2010). *Os lapsus linguae e o Léxico Mental*. Tese de Mestrado, Universidade Católica Portuguesa/Escola Superior de Saúde do Alcoitão.
- Rastle, Kathleen (2007). Visual word recognition. In M. Gareth (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press, 71-87.
- Saffran, J.R., R.N. Aslin & E.L. Newport (1996). Statistical Learning by 8-Month-Old Infants. *Science* 274(5294): 1926-1928.
- Viana, M.C., I.M. Trancoso, F.M. Silva, G. Marques, E., d'Andrade & L.C. Oliveira (1996). Sobre a pronúncia de nomes próprios, siglas e acrónimos em Português Europeu. In I. Duarte & I. Leiria (Orgs.), *Actas do Congresso Internacional sobre o Português*, vol. III. Lisboa: Colibri/APL, 481-517.
- Vigário, M. (2003) *The Prosodic Word in European Portuguese*. Berlin/New York: Mouton de Gruyter.
- Vigário, M., F. Martins & S. Frota (2005). Frequências no Português: a ferramenta FreP. In Inês Duarte & Isabel Leiria (Orgs.) *Actas do XX Encontro Nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 897-908.
- Vigário, M., F. Martins & S. Frota (2006). A ferramenta FreP e a frequência de tipos silábicos e classes de segmentos no Português. In *XXI Encontro da Associação Portuguesa de Linguística. Textos Seleccionados*. Porto: APL/Colibri, 675-687.
- Vigário, M., M. J. Freitas & S. Frota (2006). Grammar and frequency effects in the acquisition of prosodic words in European Portuguese. *Language and Speech* (Special Issue Crosslinguistic Perspectives on the Development of Prosodic Words, guest-edited by K. Demuth) 49(2), 175-203.
- Vigário, M. & P. Garcia (submetido) *Aquisição da morfologia derivacional no Português Europeu: estudo de caso*.
- Vigário, M., S. Frota & F. Martins (2010). A frequência que conta na aquisição da fonologia: types ou tokens? A. M. Brito, F. Silva & J. Veloso (Orgs.) *Textos Seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*. Porto: Associação Portuguesa de Linguística, 749-767.
- Vigário, M., S. Frota & F. Martins (submetido). Para uma caracterização da distinção entre palavras prosódicas e clínicos com base em dados de frequência. In *Textos Seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*.