

***FreP*: An Electronic Tool for Extracting Frequency Information of Phonological Units from Portuguese Written Text**

S. Frota, M. Vigário and F. Martins

Univ. Lisboa (DLGR/ Onset-CEL), Univ. Minho (CEHUM), Univ. Lisboa/ILTEC

DLGR & Lab. Fonética, Faculdade de Letras, Univ. Lisboa,

Alameda da Universidade, 1600-214 Lisboa, Portugal

E-mail: sonia-frota@mail.telepac.pt, marina.vigario@mail.telepac.pt, fmartins@fl.ul.pt

Abstract

The importance of frequency for phonological phenomena has long been noticed in the literature. However, frequency information available for phonological units in Portuguese is scarce, non-replicable, corpus dependent, and hard to obtain due to the non-existence of a free tool for public use. This paper describes *FreP*, a new electronic tool that provides frequency counts of phonological units at the word-level and below from Portuguese written text: namely, major classes of segments, syllables and syllable types, phonological clitics, clitic types and size, prosodic words and their shape, word stress location, and syllable type by position within the word and/or status relative to word stress. Useful applications of *FreP* in general linguistics, phonology, language acquisition and development, speech evaluation and therapy are also described. Forthcoming extensions of the tool include the ability to extract frequency information for different varieties of Portuguese, Brazilian Portuguese in particular, and the ability to provide a SAMPA output from the written text, together with the frequency of segmental features, like manner, place of articulation and laryngeal features. Updated information on *FreP* can be found at <http://www.fl.ul.pt/LaboratorioFonetica/FreP>.

1. Introduction

The importance of frequency for phonological phenomena has long been noticed in the literature. However, in recent years there has been a growing interest in the frequency of grammatical objects, as shown by the body of research devoted to or taking into account frequency issues (inter alia, Bybee, 2001; Bybee & Hopper, 2001; Pierrehumbert, 2002; Moates, Bond & Stockmal, 2002). Frequency is argued to play a role in phonetic reduction, in the regularization of irregular paradigms, in the activation of constraints, or in the emergence and development of grammatical units such as the syllable and the word (to mention only a few of the areas covered in recent work – e.g. Jurafsky, Bell & Girand, 2002; Vigário, 2003, Demuth & Johnson, 2003, Prieto, to appear).

Despite the growing interest in frequency information, frequency data for phonological units in Portuguese is still scarce, and in most cases non-replicable and corpus dependent. Some frequency results have already been reported for syllable structure (Andrade & Viana, 1994; Vigário & Falé, 1994, Viana et al, 1996), segments (Viana et al, 1996) and word stress distribution (Viana et al, 1996). On the one hand, the data reported are either the result of manual counts or (semi)automatic procedures using tools integrated in speech synthesis systems and thus not freely available for research purposes. On the other hand, most reports are based on the same corpus (*Português Fundamental*, cf. Bacelar, Marques & Segura da Cruz, 1987). In addition, to our knowledge, there is no frequency data available on other phonological units, such as clitics and clitic types or prosodic words and their shape, on syllables by position and status relative to stress, nor comparable frequency data for corpora of adult and

children speech, of different speech styles, etc.

In this paper, a new electronic tool for extracting frequency information of phonological units from Portuguese written text is described: *FreP*, an acronym of “Frequency in Portuguese”. The need for this tool arose in the absence of a simple method to obtain frequency data for Portuguese useful to linguistic research on the phonology of the language. The tool was thus designed to fulfil this gap, as well as to be useful to other areas in the study of speech, namely speech evaluation and therapy, and language teaching and learning.

2. The Tool

This section describes the main properties of *FreP*. Updated information and an user manual can be found at <http://www.fl.ul.pt/LaboratorioFonetica/FreP>.

2.1. Introducing *FreP*

FreP for Windows is a tool developed in the course of research on the phonology and phonetics of European Portuguese conducted by the authors of this paper. It runs on Windows XP, Windows Millennium, Windows 98, and Windows Server 2003. It is written on C++ and uses the Microsoft Visual C++ 2005 compiler.

FreP takes as input non-formatted, plain text files, based on ASCII ISO 8859-1 (.txt files or similar). It provides two kinds of outputs: a menu based display, and a formatted text file designed to be read by a spreadsheet or database utility. The output display includes the frequency counts of the units selected plus a list of the first 87 cases found in the input file. The output text file contains all the information obtained by running *FreP* on a designated input file, both the frequency data for all types of units and the intermediate outputs of the *FreP* functions that

underlie the analysis of the input file. The latter are organized in columns as described below (see *Functions*). FreP is freely available to users for scientific, research, or teaching and learning non-commercial purposes. The tool is presently in Beta mode and still in progress.

2.2. From Orthography to Phonology

FreP takes advantage of a fairly predictable relation between Portuguese orthography and the (lexical) phonology of the language. This relation underlies the procedure of identification and segmentation of phonological units, together with the phonological knowledge of the obligatory segmental phenomena that characterize the language. There are, however, cases where phonology cannot be predicted from the orthographic conventions. These cases are limited in number and have been dealt with using two kinds of procedure: (i) based on a small lexicon of ‘exceptional’ words, a word may be tagged to (a) be rewritten so that it becomes phonologically transparent, or (b) be excluded from the computation of a specific function of analysis so that it does not feed the process and is analysed in an adequate manner; (ii) (semi)automatic interventions in the input through the exclusion of phonologically non-interpretable orthographic characters. Examples of each type of procedure are provided below.

FreP starts by taking all sequences of orthographic characters between blank spaces as words, and then analysis these sequences to map each orthographic object within them into a phonological object. The only orthographic characters not read by *FreP*, and thus excluded from the analysis procedures, are punctuation marks, digits (both Arabic and Roman numerals), enumeration signs like (a), (b), (c), or (i), (ii), (iii), and sequences adjacent to and between square brackets (as transcribed corpora frequently make use of square brackets to indicate incomplete sequences or sequences not understandable to the transcriber). The examples in (1) illustrate a partial analysis of regular cases where phonology can be predicted from orthography. The examples in (2) illustrate a partial analysis of ‘exceptional’ cases: (2a) a rewriting process, where the Nasal, not predictable from orthography, is added; (2b) an exclusion process, to avoid the form ending with ‘zinho’ to be parsed in two prosodic words (cf. 1d); (2c) an input intervention, where the first orthographic consonant is deleted as it is not pronounced (word stressed vowels appear preceded by ‘˘’; C, V, G, and N stand, respectively, for consonant, vowel, glide, and nasal (non-consonantal) segment; syllables are separated by dots; orthographic word identification signalled by #, orthographic routines for phonological transparency (if applicable), word stress and prosodic word identification, and major classes of segments and syllable division are given in this order).

- (1) a. #família# > fam˘ília > CV.C˘V.CGV ‘family’
 b. #bem# > b˘em > CVGN ‘good’
 c. #guarda# > garda > g˘arda > C˘VC.CV ‘guard’

- d. #Marizinha# > Mar˘ia > CV.C˘V.V
 > zina > z˘ina > C˘VGN.CV
 ‘little Mary’

- (2) a. #muito# > m˘uito > m˘uiNto > C˘VGN.CV
 ‘very’
 b. #vizinho# > vizino > viz˘ino > CV.C˘V.CV
 ‘neighbour’
 c. acção > açãõ > #açãõ# > aç˘ãõ > V.C˘VGN
 ‘action’

The conjoint implementation of the relation between Portuguese orthography and Portuguese phonology and the understanding of the workings of the language specific phonology allow the automatic extraction (identification and count) of the following phonological units: classes of segments (e.g. consonants, vowels, glides), syllables, phonological clitics and prosodic words (PWs). In addition, the tool locates word stress, provides information on the distribution of stress within words (i.e. final, penult and antepenult stress), on the frequency of different syllable types (CV, V, CVC...), by position in the word (initial, internal and final) and/or taking into account the presence/absence of word stress, provides information on the size of prosodic words or clitics (number of words/clitics with one, two, N syllables/segments), and within the class of phonological clitics sets enclitics and proclitics apart, giving frequency information for both types of units separately. The tool also provides information on orthographic objects, namely number of orthographic words and characters. All the frequency data can be obtained on screen via a summary information box with the total number of PWs, clitics, syllables, segments and orthographic objects, and a set of Menus with the structure shown in Figure 1.

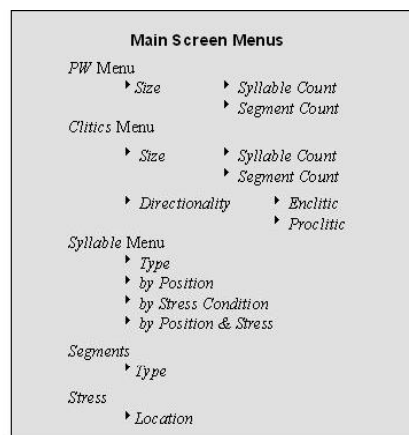


Figure 1: Menu structure of *FreP*

The frequency data, together with all the information provided by the tool, including the outputs of the main functions of analysis, are shown in the output text file.

2.3. Functions

FreP starts by reading the input text file and wiping off

the orthographic objects to be ignored (listed in the section above). The output of the wiping process is presented in column A of the output text file, where the word units to be analysed are already divided and listed. The units of column A are subject to a set of functions that prepare the application of the stress assignment algorithm, such as word tagging for rewriting purposes, or word tagging for exclusion from the computation of a specific function. The result of these processes is given in columns B and C of the output text file. The stress assignment algorithm applies to the units in column C which are able to feed this process and the result is the division of all units into stressed and unstressed, that is prosodic words and clitics (based on insights from Vigário, 2003). The output of stress assignment is given in column D of the output text file.

The next set of functions prepares the application of the syllabification algorithm, namely by differentiating vowels from glides and identifying the presence of the nasal (non-consonantal) segment responsible by nasal vowels and nasal diphthongs in Portuguese. The output form given in column E constitutes the sequence to be syllabified. The operation of the syllabification algorithm yields the result given in column F of the output file, which lists all the PWs and clitics divided into syllables (the treatment of glides and the identification of syllable boundaries essentially follows proposals by Mateus, 1975; Vigário & Falé, 1994; Andrade & Viana, 1994; Mateus & Andrade, 2000).

Finally, a function of analysis applies to translate the sequences of objects in column F into major classes of segments, that is C, V, G, and N. The output presented in column G is thus a phonological template of the orthographic word units initially listed in column A. This template identifies the number of syllables, syllable structure, the absence or presence and location of stress, the number and type of segments.

The schema in Table 1 summarises the main functions of analysis and illustrates their outputs as given in the output text file, for the orthographic word forms *sozinho* ["sO"ziJu] 'alone' (a morphosyntactic word that contains two prosodic words, and exemplifies word tagging for rewriting purposes) and *praia* ["praj6] 'beach' (a case that exemplifies glide ambisyllabicity in Portuguese).

<i>Read & Wipe</i>	<i>Prepare Stress</i>	<i>Assign Stress</i>
#so#	só	s'ó
#zinho#	zino	z'ino
#praia#	praia	pr'aia
<i>Prepare Syllabific.</i>	<i>Syllabify</i>	<i>Show Major Classes</i>
s'ó	s'ó	C'V
z'ino	z'í . no	C'V . CV
pr'aGa	pr'aG . Ga	CC'VG . GV

Table 1: Main functions and respective outputs.

2.4. Evaluation of the Tool

A global evaluation of the tool is still in progress, as the different beta versions are being tested by the authors, as well as by various users. However, some data is already available, based on the TA90PE corpus, a section of the *Português Falado* corpus with 22994 orthographic words, and on a child-directed speech corpus with 23674 orthographic words (Vigário, Martins & Frota, 2005; Vigário, Freitas & Frota, to appear). The automatic calculations generated by FreP were checked by hand with regard to prosodic word and phonological clitic identification, stress location, and syllable count, with the following reliability results: 99,935% for PW/clitic identification; 99,930% for stress location; and 99,709% for syllable count.

Up to now, syllable structure and major classes of segments have only been systematically checked on a subsection of TA90PE comprising 2796 orthographic words, a subsection of the child-directed speech corpus with 4000 orthographic words, and a subsection of a child speech corpus containing 1003 orthographic words (Frota et al, 2005), with the following reliability result: 99,746% for both syllable structure and major classes of segments. The errors generated by *FreP* belong to three classes: (i) unpredictable pronunciation of sequences of orthographic consonants, like <ct> where the first grapheme is pronounced as [k] in some words (e.g. *facto* ["faktu] 'fact'), but is not pronounced at all in others (e.g. *acto* ["atu] 'act'); (ii) cases of illegal V-insertion in sequences of consonants that do not conform to the general principles of syllable construction in Portuguese, when they appear in post-tonic position (e.g. #óbvio# > 'ó.bV.vGo 'obvious', which is illegal, *versus* #facto# > f'a.cV.to 'fact', which is legal); (iii) instances of prosodic words internal to morphosyntactic compounds or derived words that show erroneous identification, and/or misplaced stress location due to exceptional stress assignment (e.g. *pezinho* 'little foot' is parsed into #pe# and #z'ino#, i.e. a clitic and a PW, and not into #p'é# ["pE] and #z'ino# ["ziJu], i.e. two PWs, as it should be). A solution for at least some cases belonging to classes (ii) and (iii) is presently being envisaged. Updated information on the errors generated by *FreP*, together with tips for avoiding them, is found in the *FreP User Manual* available online at the *FreP* website. It is crucial to note that *FreP* is optimized for (European) Portuguese, and thus any text not following Portuguese orthography will not be adequately parsed into the relevant phonological units.

3. Applications

In this section, some useful applications of *FreP* in general linguistics and phonology, language acquisition and development, speech evaluation and therapy are described. Other potential applications are also suggested. The frequency data reported is always based on token frequency.

3.1. General Linguistics and Phonology

3.1.1. Constraints on Word Size

The frequency of monosyllabic lexical words of the CV, CCV or V shape (monomoraic words) and the frequency of trisyllabic and longer word shapes in the lexicon of a language may be respectively used to argue for/against the sensitivity to *Minimality* and *Maximality* word requirements. Frequency data on prosodic word shapes obtained with *FreP* shows that monosyllabic words and trisyllabic and longer words appear in similar proportions (28.6% and 27%, respectively), and the same is also true for monosyllabic CV shapes and words with more than 3 syllables (7.4% and 8%, respectively – Vigário, Freitas & Frota, to appear). These data support the argument that European Portuguese (EP) is not sensitive to minimality (or maximality) requirements (cf. Vigário, 2003; Vigário, Martins & Frota, 2005).

3.1.2. Placement of Pronominal Clitics

Frequency information on the use of phonological clitics shows that these units amount to 29.5% of all word units. Clitics are divided into proclitics and enclitics, according to the direction of cliticization they display. Of all clitics, 97% are proclitics and only 3% are enclitics. However, at least since Frota (1994) a tendency to place verbal pronominal clitics, which were proclitic to the verb, in an enclitic position (i.e. after the verb), has been reported in the literature. It is thus clear that this tendency cannot be promoted by a frequency effect, and must be explained by other factors (cf. Vigário, Martins & Frota, 2005).

3.1.3. Word Stress Distribution

Speakers of EP have the intuition that most words in the languages exhibit penultimate stress. Grammars usually present this stress pattern as the ‘regular’ pattern in the language, and frequency information already available confirms that this pattern is dominant (Viana et al, 1996). Data obtained with *FreP* supports these observations, as long as monosyllabic prosodic words are excluded from the frequency count: among polysyllabic PWs, 76.4% show penult stress, 21.6% final stress, and 2% antepenult stress (Vigário, Martins & Frota, submitted). However, monosyllabic PWs, which are frequent in the language (see above), add to the final stress pattern. All PW shapes included, word stress distribution is as follows: 54.4% for penult, 44% for final, and 1.5% for antepenult stress. Therefore, in frequency terms final stress is far from being exceptional or rare. This fact may have implications for the understanding of how stress is assigned in EP, as well as for the acquisition of word stress.

3.1.4. Distribution of Syllable Types

In accordance with previous results available in the literature, the CV syllable type is dominant in EP (46.4%). Only two other types are above 10%: V with 15.8% and CVC with 11%. Apart from CV, most types are not equally distributed by position within the word: V and

CCV prevail word-initially; CVC is more common word-finally; types with N(asal) appear mainly in word initial and final position, as well as in monosyllabic words; types with G(lide) are more frequent in monosyllabic words. Likewise, syllable types are also not equally distributed by stressed/unstressed position. CV, V, CVC and CCV are much more frequent in unstressed position, whereas syllable types with N and/or G are prevalent in stressed position (detailed data on the distribution of syllable types is found in Vigário, Martins & Frota, submitted). As word-edges are prominent positions in EP (Vigário, 2003), and the syllable bearing stress is naturally prominent, syllable types that prevail in such positions may stand out even though their overall frequency is low, or lower than the frequency of types appearing in non-prominent positions.

3.1.5. Major Classes of Segments

Frequency data on the occurrence of consonants (C), vowels (V), glides (G), and the nasal segment (N) responsible for V and G nasalization show that C and V are equally distributed in EP, respectively with 46% and 48%. The proportion of G amounts to 6%. Of all vowels, 16% are nasal vowels. *FreP* also provides the frequency of V-insertion in sequences of consonants that do not conform to the general principles of syllable construction in Portuguese. V-insertion occurs only in 0.2% of the total number of syllables of the TA90PE corpus, showing that such sequences of consonants are extremely rare in the language (Vigário, Martins & Frota, submitted).

3.2. Language Acquisition and Development

3.2.1. Shape of Early Words

In recent studies on language acquisition, language-specific frequency distributions of prosodic word shapes in the input have been argued to constrain prosodic word development. In particular, the frequency patterns may constrain the timing of emergence and the course of development of subminimal words and/or words with more than a binary foot. Using *FreP* to extract prosodic word shape frequencies both on an adult speech and a child-directed speech corpus, Vigário, Freitas & Frota (to appear) show that the statistical properties of the input do constrain the shape of early words in Portuguese: namely, a fairly high frequency of monosyllabic shapes, especially monosyllabic CV-like shapes, in the input agrees with the production of subminimal words in child speech; a fairly high frequency of trisyllabic and larger shapes in the input (adult speech in particular) matches the early development of words larger than a binary foot. The distribution of the different prosodic word shapes in adult speech (AS), child-directed speech (CDS) and child speech (CS), based on data from Vigário, Freitas & Frota (to appear), is shown in Figure 2.

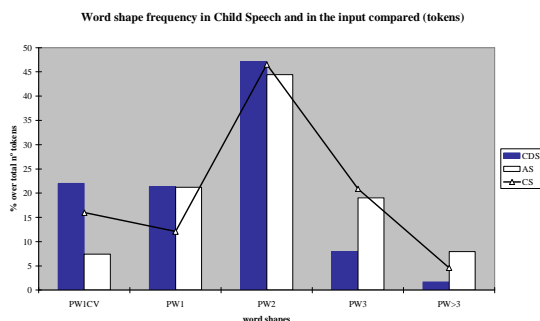


Figure 2: Input (CDS, AD) and child speech (CS) frequencies of word shapes.

3.2.2. Emergence of Syllable Types

The relevance of syllable type frequency in the input to the emergence of syllable types in child speech has been observed in Frota et al. (2005). Using *FreP* to obtain overall syllable type frequency, as well as syllable type distribution by position within the word and by position relative to stress (e.g. stressed/unstressed), both in adult speech corpora (AS and CDS) and in child speech corpora, it is shown that overall input frequency is not a good predictor of the order of emergence of syllable types: overall frequency predicts the order CV > V > CVC > CVG/N, whereas the development pattern in child speech displays both CV and V at the beginning, and CVG/N before CVC. However, the two mismatches can be explained if the distribution of syllable types by position within the word and by position relative to stress is taken into account. V and CVG/N are syllable types that appear mainly in prominent positions in the input (i.e. at word-edges and/or in stressed position). It is thus hypothesized that both frequency and structural information (namely, prosodic prominence) are required to understand the order of emergence of syllable types in European Portuguese.

3.3. Speech Evaluation and Therapy

Tools for the phonological evaluation of EP children speech are scarce. *FreP* has been used in the design of a phonological evaluation procedure to be applied to children between 3 and 6 years of age. The goal is to ensure that the chosen lexicon meets the language-specific syllable type and segment type distributions, as well as stress patterns (Guerreiro, in progress; Silva, in progress).

FreP has also proved useful to another area of speech evaluation and therapy, specifically the research of phonetic and phonological correlates of emotions in speech. As one of such correlates is speech rate and articulation rate, *FreP* has been used to obtain the number of syllables over time present in speech corpora produced by different groups of subjects (Correia, in progress).

3.4. Other Applications

Among other possible applications of the tool, we envisage its use as a means for teaching/learning explicit grammatical concepts, such as the difference between prosodic words and phonological clitics, the types of clitics that Portuguese shows, or the different syllable types present in the language. *FreP* can also be used as a means for teaching/learning how orthography and speech are related in Portuguese. A further use of the tool may be its contribution to performance-based assessment procedures that measure first or second language proficiency of students. This may be particularly useful in the present context of second language learning of Portuguese by adults from various emigrant communities.

4. Extensions

FreP was conceived as a tool open to include new facilities, as planned by the authors and as requested by *FreP* users. Planned extensions include: (i) the ability to extract frequency information for different varieties of Portuguese, namely Brazilian Portuguese; (ii) the ability to provide a SAMPA output from the written text; (iii) the frequency count of segmental features, like manner, place of articulation, or laryngeal features.

The adaptation of *FreP* to Brazilian Portuguese (or any other variety) involves some changes in the orthography to phonology mapping assumed, as well as slight changes in the phonological rules implemented in *FreP*. The user will be able to select the variety of Portuguese he/she wants to work with.

Besides the major classes of segments, the new output in SAMPA to be included in *FreP* will allow a direct measure of the grapheme to phoneme ambiguity in Portuguese. The SAMPA output will also be helpful as a first phonetic transcription from written text.

Once the SAMPA output is achieved, it can be used to provide the frequency of segmental features, and its distribution by position within the word, position within the syllable, and position relative to stress.

As it is based both on Portuguese orthography and Portuguese phonology, *FreP* is not adaptable to other languages. However, languages with alphabetic writing systems, and thus a grapheme – phoneme relation, can in principle take advantage of the *FreP* concept.

5. Acknowledgments

This research was funded by Onset-CEL from Universidade de Lisboa, and CEHUM from Universidade do Minho. We are grateful to the audience of the XX Encontro Nacional da Associação Portuguesa de Linguística, 2004, XXI Encontro Nacional da Associação Portuguesa de Linguística, 2005, and the IASCL Symposium 'Exploring the Effects of Prosody, Morphology, Frequency and Representation on the Development of Syllable Structure in Romance Languages', 2005. Thanks are also due to our colleagues Maria João Freitas, Ana Lúcia Santos and Carla Soares, who allowed *FreP* to be tested on their speech databases, and to our students and all other *FreP* users.

6. References

- Andrade, E. & Viana, M. C. (1994) Sinérese, diérese e estrutura silábica. In Actas do IX Encontro da Associação Portuguesa de Linguística (pp. 31--42). Lisboa: APL/Colibri.
- Bacelar, M. F., Marques, M. L. & Segura da Cruz, M. L. (1987) Português Fundamental. Vol. II. Métodos e Documentos, Tomo 2: Inquérito de Disponibilidade. Lisboa: INIC/CLUL.
- Bybee, J. (2001) Phonology and Language Use. Cambridge: Cambridge University Press.
- Bybee, J. & Hopper, P. (2001) (Eds.) Frequency and the Emergence of Linguistic Structure. Amsterdam: John Benjamins.
- Correia, P. (in progress) No signo das emoções: comunicação não verbal em indivíduos com perturbação vocal. Tese de Mestrado em Ciências da Fala. ESSA/Universidade Católica.
- Demuth, K. & Johnson, M. (2003) Truncation to subminimal words in Early French. *Canadian Journal of Linguistics*, 48, 211—241.
- Frota, S. (1994) Is Focus a phonological category in Portuguese? In P. Ackema & M. Schoorlemmer (Eds.) *Proceedings of ConSole 1* (pp. 69--86). The Hague: Holland Academic Graphics.
- Frota, S., Freitas, M. J., Vigário, M. & Martins, F. (2005) Prosody and frequency effects on the development of syllable structure in European Portuguese. Paper given at the Xth International Congress for the Study of Child Language, Berlin.
- Guerreiro, H. (in progress) Processos fonológicos na crianças de 5 anos. Tese de Mestrado em Ciências da Fala. ESSA/Universidade Católica.
- Jurafsky, D. A., Bell, A. & Girand, C. (2002) The Role of Lemma in Form Variation. In C. Gussenhoven & N. Warner (Eds.) *Laboratory Phonology 7* (pp. 3—34). Berlin/New York: Mouton de Gruyter.
- Mateus, M. H. (1975) A silabificação de base em Português. In Actas X Encontro Nacional da Associação Portuguesa de Linguística (pp. 289--300). Lisboa: Colibri/APL
- Mateus, M. H. & Andrade, E. (2000) *The Phonology of Portuguese*. Oxford: Oxford University Press.
- Moates, D. R., Bond, Z.S., & Stockmal, V. (2002) Phoneme frequency in spoken word reconstruction. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 141--169). Berlin/New York: Mouton de Gruyter.
- Pierrehumbert, J.B. (2002) Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101--139). Berlin/New York: Mouton de Gruyter.
- Prieto, P. (to appear) Early prosodic word acquisition in Catalan. *Language and Speech* (Special Issue on the Acquisition of the Prosodic Word, guest-edited by K. Demuth).
- Silva, C. (in progress) Processos fonológicos em crianças de 4 anos. Tese de Mestrado em Ciências da Fala. ESSA/Universidade Católica.
- Viana, M. C., Trancoso, I. M., Silva, F. M., Marques, G., Andrade, E. & Oliveira, L. C. (1996) Sobre a pronúncia de nomes próprios, siglas e acrónimos em Português Europeu. In Actas do Congresso Internacional sobre o Português, Vol. III (pp. 481--517). Lisboa: Colibri/APL.
- Vigário, M. (2003) *The Prosodic Word in European Portuguese*. Berlin/New York: Mouton de Gruyter.
- Vigário, Freitas, M. J. & Frota, S. (to appear) Grammar and frequency effects in the acquisition of the prosodic word in European Portuguese. *Language and Speech* (Special Issue on the Acquisition of the Prosodic Word, guest-edited by K. Demuth).
- Vigário, M., Martins, F. & Frota, S. (2005) Frequências no Português: a ferramenta *FreP*. In Actas do XX Encontro Nacional da Associação Portuguesa de Linguística (pp. 897—908). Lisboa: Colibri/APL.
- Vigário, M., Martins, F. & Frota, S. (submitted) A ferramenta *FreP* e a frequência de tipos silábicos e classes de segmentos no Português.
- Vigário, M. & Falé, I. (1994) A Sílabas no Português Fundamental: uma descrição e algumas considerações de ordem teórica. In Actas do IX Encontro da Associação Portuguesa de Linguística (pp. 465--477). Lisboa: APL/Colibri.

