

"No laboratório à conversa com..."

Laboratório de Fala

Faculdade de Psicologia e Ciências de Educação, Universidade do Porto

6 de Julho de 2011

FreP - Padrões de frequências na fonologia do Português

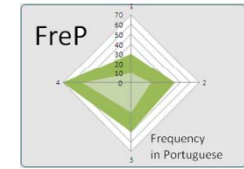
Marina Vigário

(trabalho em colaboração com
Fernando Martins & Sónia Frota)

Laboratório de Fonética – CLUL

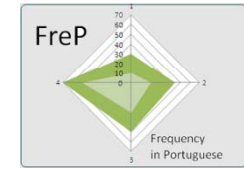
Universidade de Lisboa

Estrutura da apresentação



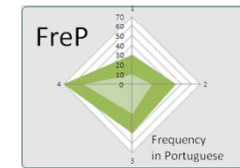
- Enquadramento
- Principais propriedades da ferramenta
- Avaliação
- Demo
- Em fase de conclusão
- Alguma da investigação feita utilizando o FreP

Enquadramento

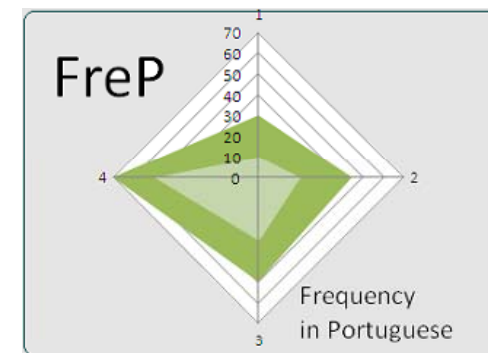


- ◆ É crescente o trabalho que toma em consideração informação de frequência em áreas como o processamento da linguagem, a produção linguística, ou a aquisição e desenvolvimento da linguagem
- ◆ Até recentemente, porém, não existiam ferramentas de extracção de informação de frequência de unidades e padrões fonológicos no Português
- ◆ Para além disso, era diminuta a informação sobre a frequência de unidades e padrões fonológicos na língua

Enquadramento



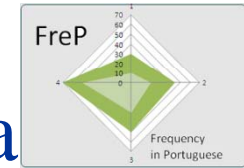
Desenvolvimento
de uma ferramenta electrónica
para a extracção automática
de informação de frequência
de unidades fonológicas
do nível da palavra e abaixo
a partir de texto escrito em
Português



- ❖ Projecto não-financiado iniciado em 2004
- ❖ Projecto financiado iniciado em Dezembro de 2007, terminado em Junho de 2011 (PTDC/LIN/70367/2006)

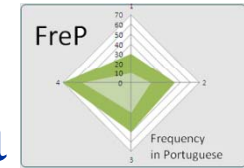
<http://www.fl.ul.pt/LaboratorioFonetica/frep>

Principais propriedades da ferramenta



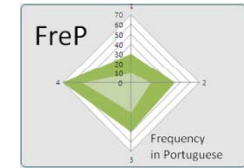
- ◆ Extrai informação sobre a frequência fonológica de unidades e padrões fonológicos desde o nível da palavra até ao traço; fonologia da palavra; fonologia obrigatória
- ◆ Corre sobre texto escrito, segundo as convenções ortográficas regulares
- ◆ Versão otimizada para o Português Europeu
- ◆ Disponibilizada livremente para uso científico ou para ensino/aprendizagem; sem fins comerciais
- ◆ Estatuto presente: *beta* (teste; pré-distribuição)

Principais propriedades da ferramenta



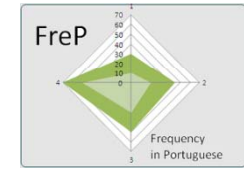
- ◆ Compilador: Visual C++ 2008
- ◆ Input: ficheiros não-formatados (.txt), baseado no ASCII ISO 8859-1
- ◆ Sistema operativo: Windows e Linux (usando *Wine emulator*)
- ◆ Output:
 - ✓ Display baseado em menus
 - ✓ Ficheiro com delimitação de campos de modo a ser lido por *spreadsheet/database* utility (e.g., Excel, OpenOffice Calc)

Avaliação



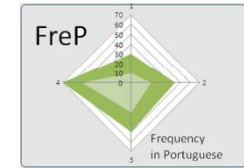
- ◆ Foi sendo feita investigação à medida que a ferramenta foi construída (verificação *in tandem*)
- ◆ Verificações manuais em várias ocasiões de modo não-sistemático
- ◆ Foi feita uma avaliação sistemática de todos os parâmetros desenvolvidos até 2009 – 2 investigadores juniores; *corpus* de 76.000 words

Avaliação



- ◆ Os resultados de fiabilidade foram acima dos 99% para todos os parâmetros (excluídos traços)
- ◆ Muitos erros provêm de palavras estrangeiras incluídas no *corpus*
- ◆ Outros erros decorrem de situações em que a grafia não espelha o som

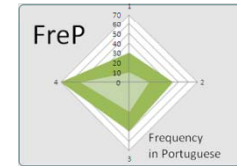
Avaliação



- ◆ Os resultados de fiabilidade foram acima dos 99% para todos os parâmetros (excluídos traços)
- ◆ Muitos erros provêm de palavras estrangeiras incluídas no *corpus*
- ◆ Outros erros decorrem de situações em que a grafia não espelha o som

Mas há formas semi-automáticas de modificar o texto e evitar muitos dos erros


Funcionalidades (demo)



(http://www.fl.ul.pt/LaboratorioFonetica/frep/frep_demo.htm)

Tudo_Léxico_270.000_corrigido_DEF_Cortado.txt - FreP

File Edit Export PW Clitics Syllables Segments Features Stress Lexical Frequency Help



Summary Info

Prosodic Words	19140
Clitics	7903
Syllables	46459
Segments	93612
Orthographic Words	29136
Orthographic Unique Words	3040
Orthographic Char	103151

List of Words

Aqui	é	palha	talvez	com
ele	já	é	não	catorze
usa	muito	que	tenha	fazia
-se	do	não	tido	-se
os	meu	faltava	ocasião	com
trajes	tempo	O	de	doze
novos	mas	chapéu	ter	mas
é	podiam	para	visto	não
pela	ser	ver	lá	era
festa	umas	qual	na	só
do	galochas	é	casa	virar
Espírito	E	que	mas	a
Santo	muita	fazia	lá	palha
E	gente	o	tem	e
no	ia	chapéu	o	fazer
dia	descalça	mais	mostruário	a
de	também	bonito	das	trança
São	ainda	com	tranças	Era
João	à	a	que	enfeitar
era	de	palha	se	Fazia
um	São	mais	faziam	-se
vestido	João	com	antiga	uns
de	Ou	tranças	-mente	arrendados
chita	poderia	Fazia	-se	na
ele	ser	-se	-se	trança
podia	Ele	tranças	trança	que
ser	mas	de	com	faziam
umas	então	oito	onze	o
galochas	o	palhas	palhas	chapéu
que	chapéu	O	fazia	mais
não	de	senhor	-se	bonito

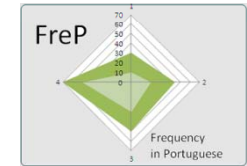
List of Words - SAMPA

```
#6.k'i#  
#e.l@#  
#u.z6#  
#s@#  
#u5#  
#tr'a.Z@5#  
#n'O.vu5#  
#E#  
#p@.l6#  
#f'eS.t6#  
#du#  
#V5.p'i.ri.tu#  
#s'6~.tu#  
#i#  
#nu#  
#d'i.6#  
#d@#  
#s'6~w~#  
#Zu."6~w~#  
#E.r6#  
#u~#  
#v@S.t'i.du#  
#d@#  
#S'i.t6#  
#e.l@#  
#pu.d'i.6#  
#s'er#  
#u.m65#  
#g6.l'o.S65#  
#k@#  
#n'6~w~#
```

Aqui, ele usa-se os trajes novos é pela festa do Espírito Santo. E no dia de São João, era um vestido de chita, ele podia ser umas galochas - que não é já muito do meu tempo, mas podiam ser umas galochas... E muita gente ia descalça também ainda à de São João! Ou poderia ser... Ele , mas então, o chapéu de palha é que não faltava! O chapéu para ver qual é que fazia o chapéu mais bonito, com a palha mais, com tranças... Fazia-se tranças de oito palhas. O senhor talvez não tenha tido ocasião de ter visto lá na casa, mas lá tem o mostruário das tranças que se faziam antigamente. Fazia-se

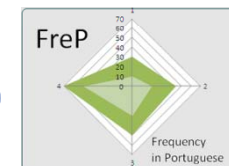
For Help, press F1

Em fase de conclusão



- ✓ Conclusão desta versão do FreP (correções, revisão do manual, ...)
- ✓ FreP_B – ferramenta adaptada à variedade brasileira do Português (com F. Fernandes Svartman; em fase de correção)
- ✓ FreP_LUDO – ferramenta simplificada, em Português, vocacionada para o ensino (utilizadores não-especializados) – permite trabalhar consciência fonológica, activação lexical.. (em conclusão)

Em fase de construção (*online*)

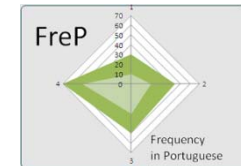


✓ **FrePOP – Frequency Patterns of Phonological Objects in Portuguese**

<http://frepop.fl.ul.pt/>

Base de dados de frequência fonológica

- em curso, mas já *online*
- presentemente, 1 milhão de palavras



Windows Internet Explorer window showing the FrePOP website. The address bar displays <http://frepop.fl.ul.pt/>. The page content includes a navigation menu, a login form, and a main text area.

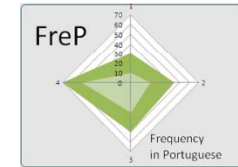
FrePOP
Frequency Patterns of Phonological Objects in Portuguese:
Research and Applications (PTDC/LIN/70367/2006)

Information on the frequency of phonological objects in the languages of the world, Portuguese included, is still scarce. This kind of information is relevant for fundamental research on areas such as the phonology of particular languages, linguistic universals/general trends, acquisition and development of phonology, L2 learning, and bilingual studies. Frequency is also an obvious source for establishing regular and deviating patterns in language use. Knowledge of such patterns is essential in domains like the diagnostic, evaluation and therapy of deviant/pathological speech, evaluation of proficiency of first and second language learners, or forensics. The project **Frequency Patterns of Phonological Objects in Portuguese - Research and Applications** aims at contributing to this field. One of the outputs of the project is the FrePOP (Frequency of Phonological Objects in Portuguese) database: a database of frequency information of phonological objects in different types of corpora, which may be used as reference information. Frequency values for different regions (including the varieties of Portuguese spoken in Brazil and in Africa) and for specific groups of population (by age, sex, education and profession) may be obtained, as well as for different historical periods (from the XVI century onwards) and types of corpora (please refer to the manual for details). The FrePOP includes information on prosodic words, clitics, syllables, segments and stress patterns. The basis of the FrePOP is a set of corpora with over 3.5 million orthographic words. The frequency data in the FrePOP were obtained with the FreP tool (Martins, F., M. Vigário & S. Frota. 2009. FreP - Frequências no Português. Software. IGAC, 5969/2009).

We gratefully acknowledge the collaboration of all institutions and individuals that contributed with their corpora to the FrePOP (see the full list in [About](#)). We also thank all the research assistants that made FrePOP possible, namely Nuno Matos, Nuno Paulino, and João Lourenço.

FrePOP (version 1.0) Copyright © 2010

Alguma da investigação feita utilizando o FreP



Vários trabalhos, várias áreas de aplicação

◆ Padrões de frequência na língua adulta

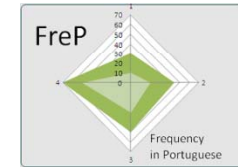
– Formatos de palavra, padrões acentuais, tipos silábicos, grandes classes de segmentos, propoção PW/CL

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.

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., 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.

Alguma da investigação feita utilizando o FreP



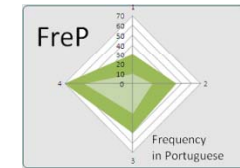
Vários trabalhos, várias áreas de aplicação

- ◆ Padrões de frequência na língua adulta
- Diferenciação sistemática entre PW e CL com base em dados de frequência

Vigário, M., S.Frota, F. Martins (2011) Asymmetries between prosodic words and clitics in the frequency of phonological patterns: implications for language acquisition. Poster presented at *PaPI 2011 – Phonetics and Phonology in Iberia*, Universitat Rovira i Virgili, June 21-22.

Vigário, M., S. Frota & F. Martins (no prelo). Para uma caracterização da distinção entre palavras prosódicas e clíticos com base em dados de frequência. A aparecer em *Textos Seleccionados. XXV Encontro Nacional da Associação Portuguesa de Linguística*.

Alguma da investigação feita utilizando o FreP

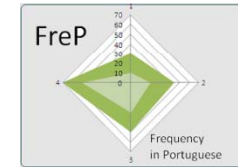


Vários trabalhos, várias áreas de aplicação

- ◆ Impacto da frequência no *input* na aquisição
- Aquisição de palavras prosódicas

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.

Alguma da investigação feita utilizando o FreP



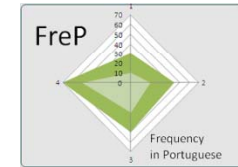
Vários trabalhos, várias áreas de aplicação

- ◆ Impacto da frequência no *input* na aquisição
- Desenvolvimento da estrutura silábica

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., 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.

Alguma da investigação feita utilizando o FreP



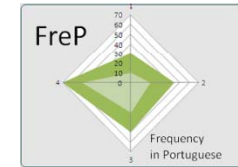
Vários trabalhos, várias áreas de aplicação

- ◆ Impacto da frequência no *input* na aquisição

- Aquisição dos traços de ponto de articulação

- ◆ 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.

Alguma da investigação feita utilizando o FreP

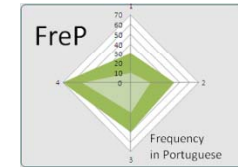


Vários trabalhos, várias áreas de aplicação

- ◆ Impacto da frequência no *input* na aquisição
- A frequência que conta: *types* ou *tokens*

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.

Alguma da investigação feita utilizando o FreP



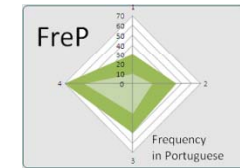
Vários trabalhos, várias áreas de aplicação

- ◆ Impacto da frequência no *input* na aquisição
- Desenvolvimento de palavras complexas

Vigário, M. & P. Garcia (2011) Aquisição da morfologia derivacional no Português Europeu: estudo de caso. Comunicação a apresentar no *XXVII Encontro Nacional da Associação Portuguesa de Linguística*. Universidade Nova de Lisboa, October 26-28.

Garcia, P. (2010). *Palavras complexas nas primeiras produções infantis (estudo de caso)*. Tese de Mestrado, Universidade Católica Portuguesa/Escola Superior de Saúde do Alcoitão.

Alguma da investigação feita utilizando o FreP

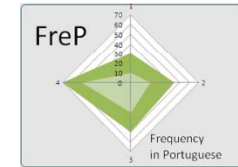


Vários trabalhos, várias áreas de aplicação

- ◆ Consciência fonológica e escrita

Cardoso, Ana (em curso). *Promoção da Consciência Fonológica e Desempenho na Escrita*. Tese de Mestrado, Universidade de Lisboa.

Alguma da investigação feita utilizando o FreP



Vários trabalhos, várias áreas de aplicação

◆ Forense – traçar perfis de falantes

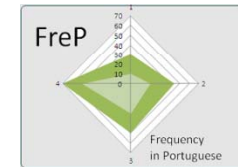
Frota, S., M. Vigário, F. Martins, M. Cruz, N. Matos & N. Paulino (2010).

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.

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.

Alguma da investigação feita utilizando o FreP



Vários trabalhos, várias áreas de aplicação

◆ Mudança linguística

Frota, S., M. Vigário, C. Galves, V. Gonzalez-Lopez & B. Abaurre (submetido). The phonology of rhythm from Classical to Modern Portuguese. *Journal of Historical Linguistics*.

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.

FreP

