

FrePOP manual

1. The database: short presentation	1
2. How to?	1
3. Variables available for selection	4
4. Parameters available for the search	5
Acknowledgements	8

1. The database: short presentation

FrePOP is one of the outputs of the project [Frequency Patterns of Phonological Objects in Portuguese - Research and Applications](#).

It is primarily a database of frequency information of phonological objects in different types of corpora, which may be used as reference information.

Its basis is a set of *corpora* with over 3.5 million orthographic words, and the frequency data in the FrePOP were obtained with the FreP tool¹.

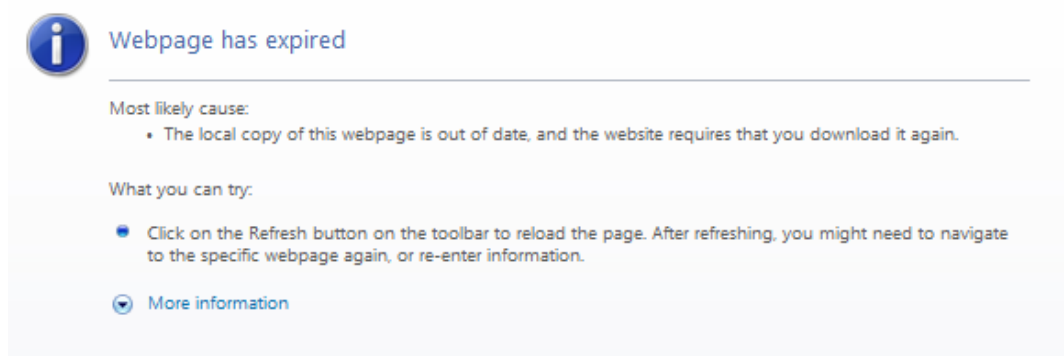
2. How to?

In order to become a user of FrePOP, you have to register in *New User* (follow the instructions given here and you will receive a username and password).

- After login, you have to define your **search scope** by selecting the variables related to the nature of the *corpora* included in the database.
- Before the selection, you may wish to see the list of *corpora* used to extract the frequency information in the database in [About](#).
- **Please notice:** inside each variable, you might find the options “irrelevant” or “n.a.” (not available). “Irrelevant” means that this variable is not relevant for specific subtypes of corpora (e.g., for the subtype “news”, the age group of its author is not relevant). Thus this is different from information which is not available, labelled here as “n.a.”.
- You may either choose the variables all at once or select only one or a combination of variables by using the **Control** key. Once selected, your search scope is defined.
- **Please notice:** when you do not select any option in a specific parameter (e.g., age), all options will be activated by default (i.e., the values given will refer to the data from all age groups).
- At the left side, below the green boxes, you will always find the list of options per variable that you have selected, defining your search scope.

¹ Martins, F., M. Vigário & S. Frota (2009) *FreP – Frequency in Portuguese v. 2* (IGAC nº 209/2010).

- In order to go back in your search, always use the button **Back**. The following error message will appear if you try to go back using another means:



- The possibility of selecting more than one option is restricted to the definition of the corpus search scope. Thus, if you want to know the values for combined options (e.g., Prosodic Words with more than 3 syllables), you must select each variable separately (i.e., Trisyllabic PWs and PWs with 4 and more syllables) and add their respective values, as signalled by the blue arrows and ellipses below.

The screenshot shows the FrePOP website interface. On the left, there is a navigation menu with links for HOME, HOW TO, ABOUT, CONTACTS, and LOGOUT. Below this is another menu with SEARCH SCOPE, DATABASE, and USERS. A 'Search Scope' section lists various filters: Text Type: all; Subtype: - interview, - spontaneous; Transcription: all; Year: - 1970, - 1970-1974, - 1970-1979; Age: all; Variety: all; Region: - GL; Gender: all; Level of Education: - high-secondary, - university; Occupation: all; and Prosodic Words.

The main content area features a header with the FrePOP logo and title: 'Frequency Patterns of Phonological Objects in Portuguese: Research and Applications (PTDC/LIN/70367/2006)'. Below the header, the 'Search' section displays the user 'frepop.admin' and level 'admin'. It shows the search path: Home - Search Scope - Search level 2 - [Search level 3](#). The query result is 'Prosodic Words: 8715'. A list of radio buttons allows filtering by syllable count: 'all', 'Monomoraic PWs', 'Monosyllabic PWs', 'Disyllabic PWs', 'Trisyllabic PWs', and 'PWs with 4 and more syllables'. The 'Trisyllabic PWs' option is selected. Below the list are 'Go Next' and 'Back' buttons. Two blue arrows point to the right from the 'Trisyllabic PWs' and 'PWs with 4 and more syllables' options.

Copyright © Frotz, Viegário, Martins & Cruz 2010 – All rights reserved | Developed by: Project Blue Book

This screenshot shows the search results for 'Trisyllabic PWs: 1769'. The navigation and search scope menus are identical to the first screenshot. The main content area shows the same header and search information, but the query result is 'Trisyllabic PWs: 1769', which is circled in blue. A 'Back' button is visible below the result.

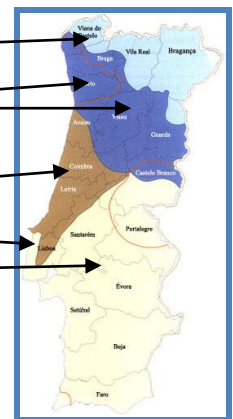
This screenshot shows the search results for 'PWs with 4 and more syllables: 928'. The navigation and search scope menus are identical to the first screenshot. The main content area shows the same header and search information, but the query result is 'PWs with 4 and more syllables: 928', which is circled in blue. A 'Back' button is visible below the result.

Prosodic Words with more than 3 syllables = 1769+928= 2697

3. Variables available for selection

- FrePOP provides frequency values for the following variables:
 - text type (spoken, written);
 - text subtype (child directed speech, child speech, biography, interview, news, spontaneous speech, technical/scientific, historical, literary);
 - transcription (adapted, orthography);
 - date (from the XVI century onwards – e.g., 1510, 1631, 1714, 1826, 1972, 1983, 1994, 2001);
 - different regions (including the varieties of Portuguese spoken in Brazil and in Africa), according to the classification in Cintra 1971², adapted by Segura & Saramago 2001³ (as shown below).

- N1: Trás-os-Montes and Alto Minho
- N2: Baixo Minho, Douro and Beiras
- GP: Oporto
- GL: Lisbon
- CS1: Littoral Centre
- CS2: Interior Centre and South
- AAM: Azores and Madeira
- BP: Brazilian Portuguese
- AP: African varieties of Portuguese



Cintra 1971, adapted by Segura & Saramago 2001 (<http://cvc.instituto-camoes.pt/conhecer/bases-tematicas/historia-da-lingua-portuguesa.html>).

- specific groups of population:
 - gender
 - } Female
 - } Male
 - age
 - } 1 year-old
 - } 2 years-old
 - } 3 years-old
 - } 4 years-old
 - } 5 years-old
 - } >5-10 years-old
 - } >10-20 years-old
 - } >20-40 years-old
 - } >40-60 years-old
 - } >60-80 years-old
 - } >80 years-old

² Cintra, Lindley. 1971. Nova proposta de classificação dos dialectos galego-portugueses. *Boletim de Filologia*, XXII, Lisboa, pp. 81-116.

³ Segura, Luísa & J. Saramago. 2001. Variedades dialectais portuguesas, *Caminhos do Português: Exposição Comemorativa do Ano Europeu das Línguas (Catálogo)*. Lisboa: Biblioteca Nacional, 221-237.

education	<ul style="list-style-type: none"> illiterate primary secondary high-secondary university
-----------	--

occupation [9 professional classes, according to CNP – National Classification of Occupations, available at <http://www.iefp.pt/formacao/CNP/Paginas/CNP.aspx> (info retrieved in October 2008), based on the *International Standard Classification of Occupations* for European Union purposes, 1998 (http://metaweb.ine.pt/sine/anexos/pdf/V00006-ISCO%2088_COM_.pdf) – plus the group of students].

Please find below the list of occupations (and its codes) according to CNP and its respective translation derived from the *International Standard Classification of Occupations*:

1	Quadros superiores da administração pública, dirigentes e quadros (Legislators, senior officials and managers)
2	Especialistas das profissões intelectuais e científicas (Professionals)
3	Técnicos e profissionais de nível intermédio (Technicians and associate professionals)
4	Pessoal administrativo e similares (Clerks)
5	Pessoal dos serviços e vendedores (Service workers and shop and market sales workers)
6	Agricultores e trabalhadores qualificados da agricultura e pescas (Skilled agricultural and fishery workers)
7	Operários, artífices e trabalhadores similares (Craft and related trades workers)
8	Operadores de instalações e máquinas e trabalhadores da montagem (Plant and machine operators and assemblers)
9	Trabalhadores não qualificados (Elementary occupations)
0 ⁴	Estudantes (Students)

4. Parameters available for the search⁵

FrePOP provides the frequency values for the units and patterns listed below.

- Orthographic units
 - orthographic words/tokens
 - orthographic types
 - orthographic characters

⁴ As said above, we have added the category 'student' to the CNP list.

⁵ Notes on the criteria for the identification of phonological units and patterns, and the notions involved may be found in FreP_Manual (see <http://www.fl.ul.pt/LaboratorioFonetica/frep/>).

- Phonological units and patterns
 - prosodic words
 - clitics
 - syllables
 - segments
 - stress patterns

Each phonological parameter is detailed in terms of size, position and type.

- For the variable **Prosodic Words** information is given on the number of prosodic words (PWs) by size condition:
 - monomoraic PWs
 - monosyllabic PWs
 - disyllabic PWs
 - trisyllabic PWs
 - PWs with 4 or more syllables

- For the variable **Clitics** information is given on the number of clitics (CL) according to their size and status in terms of directionality of phonological attachment:
 1. by **size**
 - monomoraic CLs
 - monosyllabic CLs
 - disyllabic CLs

 2. according to their status in **directionality** of attachment
 - proclitics
 - enclitics

- FrePOP provides the values for the most frequent **syllables types**:
 - CV
 - V
 - CVC
 - CVGN
 - CVN
 - VC
 - CVG
 - VN
 - CCV
 - VG

- For the each syllable type information is given on number of occurrences by position in the word and by stress condition:
 1. **position in the word**
 - #_: initial

- _#: final
- ..._...: internal
- mono: syllable in a monosyllable

2. **stress status** – status of the syllable with respect to stress

- stressed
- unstressed

▪ FrePOP provides the global frequency values for the following classes of segments:

- Consonants (C)
- Vowels (V)
- Glides (G)
- V- Slots
- all segments stressed
- all segments unstressed

▪ Information on the frequency of individual segments is also available, according to their position in the word, their position in the syllable and status of the syllable within which they appear with respect to stress:

- **position in the word:** #_ (initial); _# (final); ..._... (internal); mono (syllable in a monosyllable)
- **position in the syllable:** onset (whether or not branching), branching onset, coda (only for consonants); ambisyllabic (only for glides); pre- and post-vocalic (only for glides)
- status of the syllable within which segments appear: stressed, unstressed

▪ The frequency of the three **stress patterns** in the language is given, also in combination with the size of prosodic words (the values for monosyllables are shown separately):

- Final stress
 - PW2 final: oxytone disyllabic Prosodic Words
 - PW3 final: oxytone trisyllabic Prosodic Words
 - PW4 final: oxytone Prosodic Words with 4 or more syllables
- Penult stress
 - PW2 penult: paroxytone disyllabic Prosodic Words
 - PW3 penult: paroxytone trisyllabic Prosodic Words
 - PW4 penult: paroxytone Prosodic Words with 4 or more syllables
- Antepenult stress
 - PW3 antepenult: proparoxytone trisyllabic Prosodic Words
 - PW4 antepenult: proparoxytone Prosodic Words with 4 or more syllables
- Monosyllables

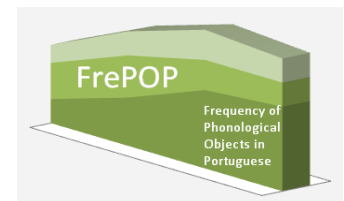
Acknowledgements

We would like to thank all the institutions and individuals that have made their corpora available, either for general public use, or just for the extraction of frequency values to be included in FrePOP (see the list of corpora in [About](#)) .

Thanks are also due to the research assistants Nuno Matos, Nuno Paulino and João Lourenço for their great help in filling in FrePOP database.

We also gratefully acknowledge the comments made by FrePOP users since the release of the database.

FrePOP (version 1.0)
Frota, S., M. Vigário, F. Martins & M. Cruz
Laboratório de Fonética (CLUL)
Faculdade de Letras da Universidade de Lisboa
ISBN: 978-989-95713-2-7
Copyright © 2010 – All rights reserved



Last update – January, 2011