

The input of *FrePOP*: list of *Corpora*

- **Hanna Batoreo's data**

Provided by: free access

Source: The CHILDES Database

<http://childes.psy.cmu.edu/data/Romance/Portuguese/>

- **Biographies**

Provided by: free access

Source: Instituto Camões, *História da Língua Portuguesa em linha* (directed by Ivo Castro; *Biographies* section coordinated by Luís Prista)

<http://cvc.instituto-camoes.pt/conhecer/bases-tematicas/historia-da-lingua-portuguesa.html>

- **Child Directed Speech**

Provided by: [Laboratório de Fonética](#) & [Lisbon Baby Lab](#) (transcribed by Cátia Severino, João Lourenço, Simão Cortês, and Susana Correia)

Source: Correia & Costa (2010), *LumaLiDaAudy*, *LumaLiDaOn*, *LumaLiDaVideo*.

- **Child Speech**

Provided by: [Laboratório de Fonética](#) & [Lisbon Baby Lab](#) (transcribed by Nuno Matos and Susana Correia)

Source: Correia & Costa (2010), *LumaLiDaVideo*.

- **C-ORAL-ROM – Integrated Reference *Corpora* for Spoken Romance Languages**

Provided by: free access

Source: C-ORAL-ROM Project (2001-2004) (CLUL – Centro de Linguística da Universidade de Lisboa), coordinated by Maria Fernanda Bacelar do Nascimento.

<http://lablita.dit.unifi.it/coralrom/>

http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_coralrom.php

- **CORDIAL-SIN – Syntax-oriented Corpus of Portuguese Dialects**

Provided by: free access

Source: CORDIAL-SIN Project (CLUL – Centro de Linguística da Universidade de Lisboa – and CLUNL – Centro de Linguística da Universidade Nova de Lisboa); funded by FCT – Fundação para a Ciência e a Tecnologia, coordinated by Ana Maria Martins (CORDIAL-SIN - PRAXIS XXI/P/PLP/13046/1998; CORDIAL-SIN, 2nd period - POSI/PLP/33275/1999; Sintaxe Dialectal - POCTI/LIN/46980/2002 ; DUPLEX - PTDC/LIN/71559/2006).

http://www.clul.ul.pt/sectores/variacao/cordialsin/projecto_cordialsin_corpus.php

- **CORP-ORAL – Corpus of Spontaneous Speech in European Portuguese**

Provided by: Tiago Freitas

Source: CORP-ORAL Project (2005-2008) (ILTEC – Instituto de Linguística Teórica e Computacional); funded by FCT – Fundação para a Ciência e a Tecnologia, coordinated by Maria Helena Mira Mateus (POCTI/LIN/60019/2004).

<http://www.iltec.pt/projectos/concluidos/corp-oral.html>

- **Correia & Costa (2010)**

Provided by: Teresa Costa

Source: Correia, Susana & Costa, Teresa (2010). *EP-Mono. Base de Dados de Aquisição do Português Europeu como Língua Materna (dados monolingues)*. Laboratório de Psicolinguística da FLUL – Universidade de Lisboa/Projecto PhonBank.

- **Correia, Susana (2008)**

Provided by: Susana Correia

Source: Correia, Susana (2008) Acoustic (lack of) evidence for word stress in the early disyllabic productions of two Portuguese children. In Anna Gavarro & M. João Freitas (eds.) *Language Acquisition and Development: Proceedings of GALA 2007*. Cambridge: Cambridge Scholars Publishing, pp. 123-132 (part of *The Acquisition of Primary Word Stress in European Portuguese*, PhD Thesis, Faculdade de Letras, Universidade de Lisboa, 2009).

- **CRPC – Reference Corpus of Contemporary Portuguese**

Provided by: free access

Source: CRPC Project (since 1988) (CLUL – Centro de Linguística da Universidade de Lisboa), coordinated by Maria Fernanda Bacelar do Nascimento.

http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_crpc.php

- **Cruz, Marisa (2009)**

Provided by: Marisa Cruz

Source: *Gaguez – Em busca de um padrão prosódico e entoacional (Stuttering – searching for a prosodic and intonational pattern)*, MA Thesis, Faculdade de Letras, Universidade de Lisboa, 2009 (available online).

http://www.fl.ul.pt/LaboratorioFonetica/English/marisa_cruz.htm

- **Diário de Notícias – newspaper**

Provided by: Director of *Diário de Notícias*, Artur Sardinha

Source: Online version from 5th to 11th May 2008, available at <http://dn.sapo.pt/Inicio/> [download: May 5th-11th]

- **Ferrari, Patrício (2011)**

Provided by: Patrício Ferrari

Source: Fernando Pessoa's Poetry analysed in *O Ritmo na Poesia de Fernando Pessoa*, PhD Dissertation, Faculdade de Letras, Universidade de Lisboa.

- **LumaLiDa – Linguistic Diary of Luma**

Provided by: Laboratório de Fonética (Faculdade de Letras da Universidade de Lisboa/Centro de Linguística da Universidade de Lisboa)

Source: *LumaLiDaOn* (Version 1), Laboratório de Fonética da Faculdade de Letras da Universidade de Lisboa [ISBN : 978-989-95713-0-3] and *LumaLiDaAudy*, in progress; both developed under the Project FreP (PTDC/LIN/70367/2006), coordinated by Sónia Frota, and funded by FCT – Fundação para a Ciência e a Tecnologia (PTDC/LIN/70367/2006 and POCTI-SFA-17-745).

<http://www.fl.ul.pt/LaboratorioFonetica/lumalidaon.htm>

<http://www.fl.ul.pt/LaboratorioFonetica/lumalidaaudy.htm>

- **Natura-PUBLICO**

Provided by: free access

Source: Natura Project (Universidade do Minho), coordinated by José João A. G. Dias de Almeida. Natura-PUBLICO contains data from the Portuguese newspaper *Público* (1991-1994).

<http://natura.di.uminho.pt/~jj/pln/pln.html#SECTION00021000000000000000>

- ***O Dia Seguinte* – debate about football**

Provided by:

Source:

- ***Português Falado* – Geographic and social varieties**

Provided by: free access

Source: Português Falado Project (1995-1997) (CLUL – Centro de Linguística da Universidade de Lisboa); funded by Programa LÍNGUA/SÓCRATES, Comissão Europeia DGXXII, coordinated by João Malaca Casteleiro.

http://www.clul.ul.pt/sectores/linguistica_de_corpus/projecto_portuguesfalado.php

- ***Público* – newspaper**

Provided by: free access

Source: Project AC/DC (Acesso a Corpora/Disponibilização de Corpora), *corpus* CHAVE, which includes 365 editions from 1994 and 1995 of the daily newspapers *Folha de São Paulo* (Brazil) and *Público* (Portugal). The sample consists of 80 articles randomly selected from 1994 and 1995 editions.

The *corpus* from *Público* used here, was organized by Vinícius Biscaro and Flaviane Fernande, coordinated by Nancy Garcia (Instituto de Matemática da Unicamp, Brazil).

<http://lusiadas.linguateca.pt/acesso>

- **Santos, Ana Lúcia (2006)**

Provided by: Ana Lúcia Santos

Source: *Minimal Answers. Ellipsis, syntax and discourse in the acquisition of European Portuguese*, PhD Thesis, Faculdade de Letras, Universidade de Lisboa, 2006; published in 2009 by John Benjamins.

http://www.clul.ul.pt/equipa/ana_santos.php

- **Santos, Isabel Joana Aguiar (2008)**

Provided by: Isabel Joana Aguiar Santos

Source: *Unidades e Processos Fonológicos no falar da região da Terra Quente: Contributos para a Linguística Forense*, MA Thesis, Universidade do Minho, 2008.

http://www.fl.ul.pt/LaboratorioFonetica/i_joana_aguiar.htm

- **Serra, Carolina (2009)**

Provided by: Carolina Serra

Source: *Realização e percepção de fronteiras prosódicas no português do Brasil: fala espontânea e leitura*, PhD Thesis, Universidade Federal do Rio de Janeiro, Brazil, 2009. Only the data from European Portuguese were considered here.

- **Soares, Carla (2006)**

Provided by: Carla Soares

Source: *La syntaxe de la périphérie gauche en portugais européen et son acquisition*, PhD Thesis, Université Paris VIII – Saint Denis, France.

- **Tycho Brahe – Parsed Corpus of Historical Portuguese**

Provided by: free access

Source: Project Rhythmic Patterns, Parameter Setting and Language Change (IEL, Unicamp, Brazil), coordinated by Charlotte Galves. Electronic corpus consisting of texts written in Portuguese, by authors born between 1380 and 1845.

<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/>

Last update – July, 2013