



PLEX5 – A production lexicon of child speech for European Portuguese PLEX5 – Um léxico Infantil para o Português Europeu

Funding:

DEPE Project (PTDC/CLE-LIN/108722/2008)

Laboratório de Fonética & Lisbon Baby Lab

Centro de Linguística da Universidade de Lisboa

PLEX5 was based on the materials from two publicly available databases of child speech: **Correia, S. & T. Costa (2010)**. EP_Mono. Base de Dados de Aquisição do Português Europeu como Língua Materna (dados monolíngues). Laboratório de Psicolinguística, CLUL/Projecto PhonBank (<http://www.clul.pt/pt/investigacao/159-acquisition-in-european-portuguese-resources-and-results>); **LumaLiDa Resources for Child Language** (<http://www.fl.ul.pt/LaboratorioFonetica/LumaLiDa.htm> - Frota, S., M. Vigário & R. Jordão (2008). LumaLiDaOn (version 1). Lisboa: Laboratório de Fonética, CLUL/FLUL; Frota, S., M. Vigário, N. Matos, M. Cruz & R. Jordão. (2012). LumaLiDaAudy – Audio Child Speech Database with phonetic transcription and prosodic labeling. Lisboa: Lab.Fonética, CLUL/FLUL).

This child speech resource was built as part of the supporting documents for the development of the [McArthur Bates Communicative Inventories \(CDI\) for European Portuguese – Short forms for Level 1 \(8 to 18 months\) and Level 2 \(16 to 30 months\)](#). The main goal behind **PLEX5** was to provide information on the developmental course of lexical acquisition. Lexical frequency lists were obtained with the [FreP tool](#) (Martins, F., M. Vigário & S. Frota. 2011. FreP - Frequency in Portuguese. Version 3.0.).

PLEX5 contains the following information, as database fields:
Orthographic word list, together with frequency of occurrence
Morphosyntactic category
Mean age of emergence
Number of children that produced the word
Information on which children produced the word
Lemma list, together with frequency of occurrence
Lemma mean age of emergence
Number of children that produced the lemma

For the purposes of **PLEX5**, the following criteria were established:

Age of emergence corresponds to the date of the first of two consecutive sessions where the same word is produced

Lemma means the dictionary lexical entry (it collapses all regular flexional forms under a unique entry)

Ambiguous word forms were disambiguated according to context, by manual search on the text files, and frequency of occurrence was recalculated.



PLEX5 thus provides both group and individual lexical development data, assembled in a single database. Version 1.1, released in November 2012, includes the lexicon produced between 0;11 and 2;6, comprising 34.517 tokens and 2.166 word types. Specifically, it includes the following data:

Clara: 1;0 - 1;10

Inês: 0;11 - 2;6

Joana: 0;11 - 2;6

João: 1;0 - 2;0

Luma (video): 0;11 - 2;6

Luma (Audy) - 1;1 - 2;6

Luma (On) - 1;1 - 2;6 .

The database is currently being extended to include lexical information for 'João' between 2;0 and 2;6, based on new transcriptions by N. Matos extending Correia & Costa (2010). **PLEX5** (version 1.2) will thus comprise 36.426 tokens and 2.816 word types). We are also working on an extension of the lexical data covered until age 3;00 for 4 of the children (version 2.0, including 61.318 tokens and 7.960 word types):

Clara: 1;0 - 1;10

Inês: 0;11 - 3;0

Joana: 0;11 - 3;0

João: 1;0 - 3;0 (new data transcribed)

Luma (video): 0;11 - 2;6

Luma (Audy) - 1;1 - 3;0

Luma (On) - 1;1 - 3;0

For more information, please contact labfon@fl.ul.pt . To use PLEX5, contact labfon@fl.ul.pt by sending a message with the following information: Name, address, affiliation, purpose of using PLEX5.

PLEX5 is a publicly available resource for scientific, educational and clinical purposes.

User agreement

This database is protected by copyright law. Unauthorized reproduction or distribution may result in penalties. PLEX5 is a publicly available resource for scientific, educational and clinical purposes. By becoming a registered user you agree to comply with the following: (i) use the database exclusively for academic research, educational or clinical purposes, unless you have been given explicit written permission for other uses, namely for commercial, profit making purposes; (ii) mention the full reference of PLEX5 in all published work (including online publications and conference presentations) where data obtained by means of PLEX5 is used/referred to; (iii) your password to access the database is for your personal use only and thus you accept full responsibility for any misuse of this information.



How to cite **PLEX5**:

Frota, S., S. Correia, C. Severino, M. Cruz, M. Vigário & S. Cortês (2012) PLEX5 – A production lexicon of child speech for European Portuguese / Um léxico Infantil para o Português Europeu. Lisboa: Laboratório de Fonética CLUL/FLUL. ISBN 978-989-95713-6-5.

List of abbreviations used in the database:

N - Noun

V - Verb

Adj - Adjective

PP – Past Participle

Num - Numeral

Int/excl – Interjection/exclamation

Pron - Pronoun

Onom - Onomatopœia

Eng – English words

NE – No emergence

NO – No occurrence

TBA – To be added