# Automating psycholinguistic statistics computation: Procura-PALavras

**[1]João Filipe Machado, [1]José João Almeida, [1]Alberto Simões & [2]Ana Soares**

[1] Departamento de Informática, & [2]Escola de Psicologia, Universidade do Minho, Braga, Portugal

**FALA 2010**
"VI Jornadas en Tecnología del Habla" and II Iberian SLTech Workshop, 10-12 November 2010, Vigo, Spain

## OBJECTIVE

In Psycholinguistics, the manipulation and/or control of words´ objective (e.g., word frequency), and subjective ratings, such as familiarity (i.e., how often we come in contact with a word), imageability (i.e., how easy it is for a word to elicit mental images), and age-of-acquisition (i.e., how early in life we learn a word) is critical in contemporary research.
The availability of these measures however is limited for European Portuguese (EP).
Because collecting norms for these subjective variables is a time-consuming process, we present an innovative triangulation method which aims at obtaining these indices in a fast and reliable way for EP researchers.
These estimated statistics will be available on the Procura-PALavras (P-PAL) on-line application.

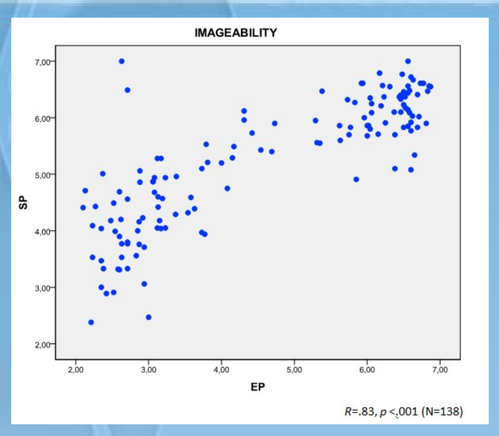
## MATERIALS

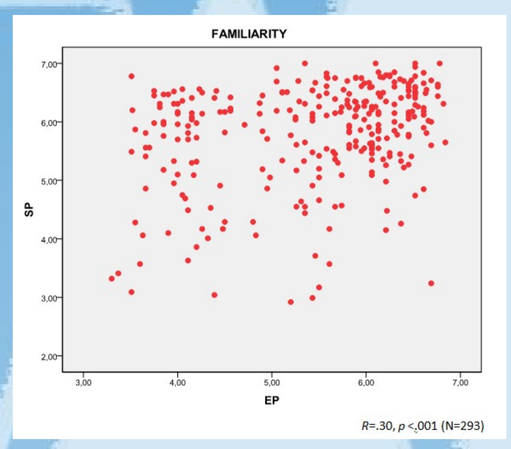
| | indexes | | | |
|---|---|---|---|---|
| Languages | Familiarity (FAM) | Imageability (IMAG) | Age-of-Acquisition (AoA) | Word frequency (Log WF) |
| European Portuguese (EP)[1] | 808 | 249 | 834 | 790 |
| English (ENG)[2] | 4944 | 4944 | 3136 | 30591 |
| Spanish (SP)[3] | 6223 | 6096 | 139 | 31491 |

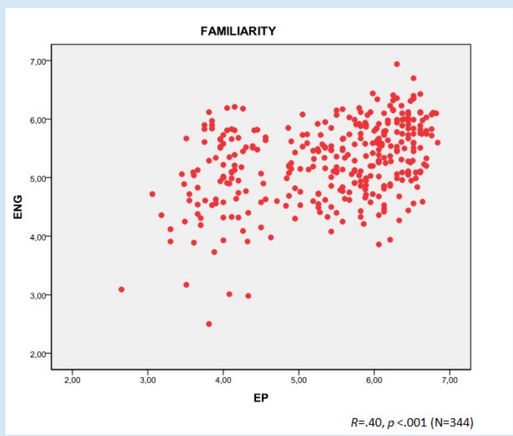
[1] taken from Marques (2004, 2005, 2007) and Bacelar do Nascimento et al. (2000)
[2] taken from N-Watch (Davis, 2005)
[3] taken from B-PAL (Davis & Perea, 2005)

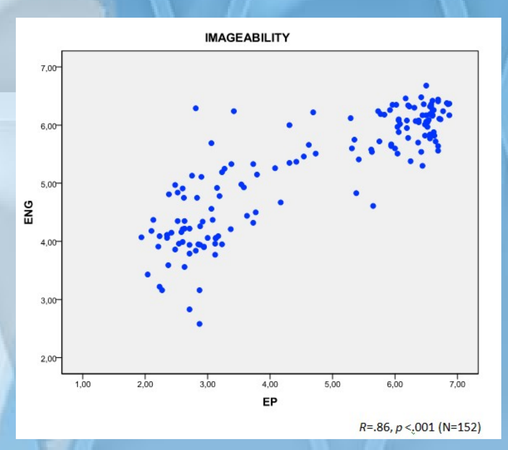
## PROCEDURE

- Normalizing the statistics between languages;
- Importing statistics to a single database;
- Connecting statistics by linking words from each language through its equivalent in EP ensuring that ENG-to-SP and SP-to-ENG translations match;
- Filtering out erroneous or low confidence translations;
- Determining Pearson correlations between languages in each index;
- Computing normative values for EP by ENG-SP means in each index.

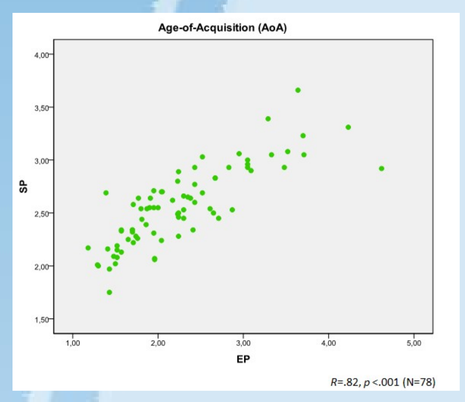


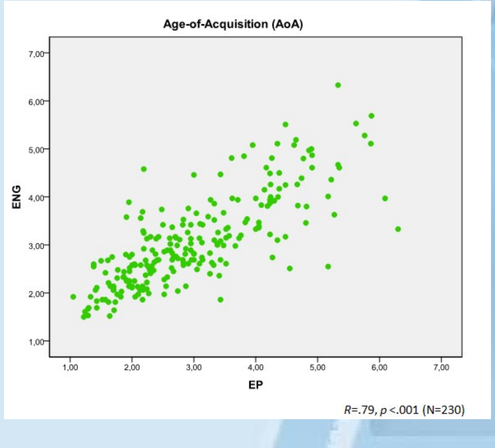



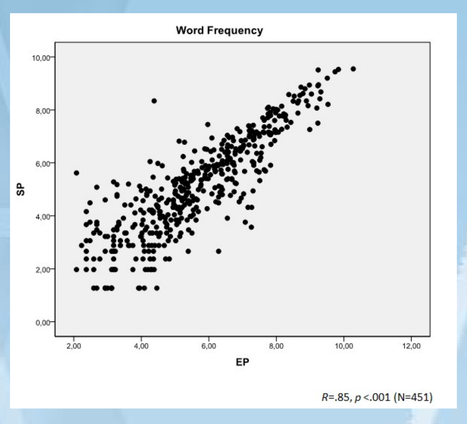




EP



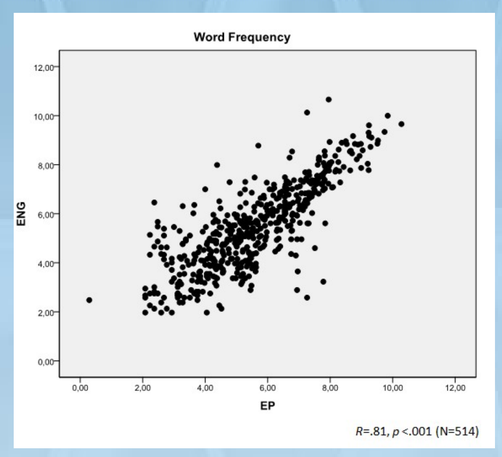



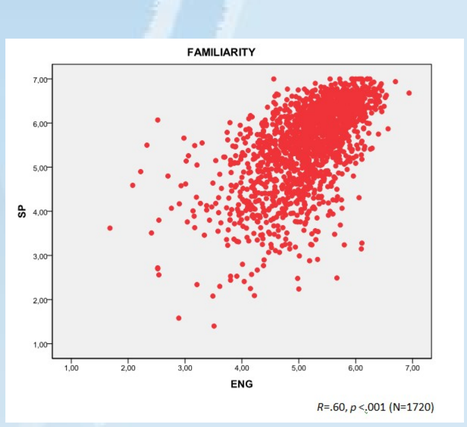



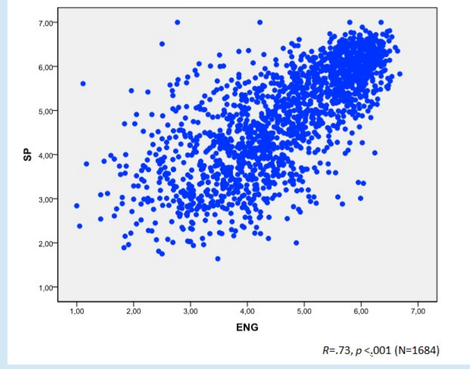

**TRIANGULATION METHOD**

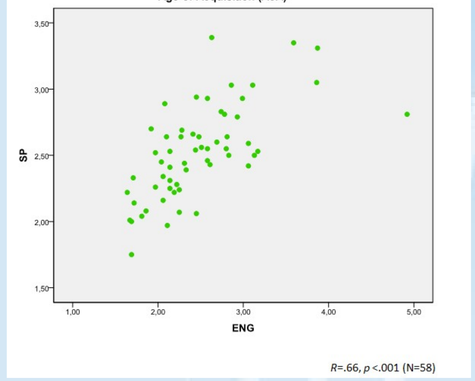
SP

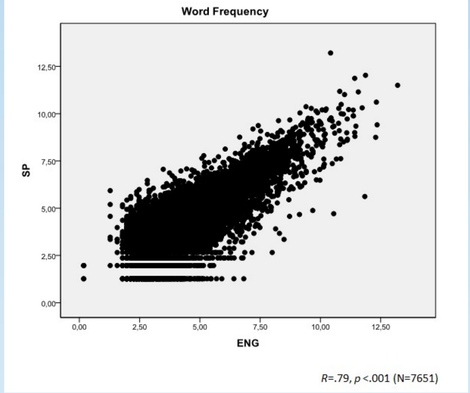
ENG




## CONCLUSIONS:

- Correlations between languages were positive (ranging between .30 and .86) and statistically significant (p<.001) for all indices, especially WF and IMAG.
- Triangulation seems to be a reliable method to obtain estimated word ratings for EP (results for FAM in SP-to-EP and ENG-to-EP may be less accurate because of the low correlations).
- Future research should compare data obtained from triangulation with those from normative studies as they become available for EP researchers.

## REFERENCES:

C. J. Davis, N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37(1), 65–70, 2005.
C. J. Davis & M. Perea, BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in spanish. *Behavior Research Methods*, 37(4), 665–671, 2005.
J. Marques, F. Fonseca, A. Morais & I. Pinto. Estimated age of acquisition norms for 834 portuguese nouns and their relation with other psycholinguistic variables. *Behavior Research Methods*, 39(3), 439–444, 2007.
J. Marques. Normas de familiaridade para substantivos comuns. *Laboratório de Psicologia*, 2, 5-19, 2004.
J. Marques. Normas de imagética e concreteza para substantivos comuns. *Laboratório de Psicologia*, 3, 65-75, 2005.

**For more information: joaoffm@gmail.com, jj@di.uminho.pt, ambs@di.uminho.pt, asoares@psi.uminho.pt**

Supported by Fundação para a Ciência e a Tecnologia (FCT), Quadro de Referência Estratégico Nacional (QREN), Programa Operacional Factores de Competetividade COMPETE and by Fundo Europeu de Desenvolvimento Regional (FEDER) .
Research project: PTDC/PSI-PCO/104679/2008.