Subjective frequency, imageability and concreteness norms for 3,800 European Portuguese words

Ana Paula **Soares**, Ana **Costa**, João **Machado**, Ana **Silva**, Joana **Oliveira**, Ana Margarida **Gonçalves**, & Montserrat **Comesaña**

Human Cognition Lab, School of Psychology, University of Minho, Braga, Portugal

Human Cognition Lab
Research Group in Psycholinguistics
School of Psychology, University of Minho

ESCOP 2013
EIGHTEENTH MEETING OF THE EUROPEAN SOCIETY FOR COGNITIVE PSYCHOLOGY
AUGUST 29 – SEPTEMBER 1, 2013 – BUDAPEST, HUNGARY

Words are widely used as stimuli in cognitive research. Because of their complexity, using words requires strict control of their objective and subjective properties (Soares et al., 2010). In this paper we present normative data for 3,800 European Portuguese words according to three subjective indices that are scarce in Portuguese in spite of being extensively used in the international research: subjective frequency (i.e., the estimation of how commonly a word is encountered in its written or spoken form), imageability (i.e., the ease with which a target word evokes a corresponding mental image) and concreteness (i.e., the degree to which words refer to objects, persons, places, or things that can be experienced by the senses) (Balota et al. 2001; Clark & Paivio, 2004; Paivio, Yuille, & Madigan, 1968). The ratings were obtained from a large sample of college students who were native speakers of European Portuguese. Each participant rated 100 words drawn randomly from the full set of words in the three subjective indices using a web survey procedure. Additionally, in order to assess the contribution of these variables in word recognition times, we collected lexical decision times for a subset of 1,920 words. The norms can be downloaded at **http://p-pal.di.uminho.pt/about/databases**.

Subjective ratings

Participants:

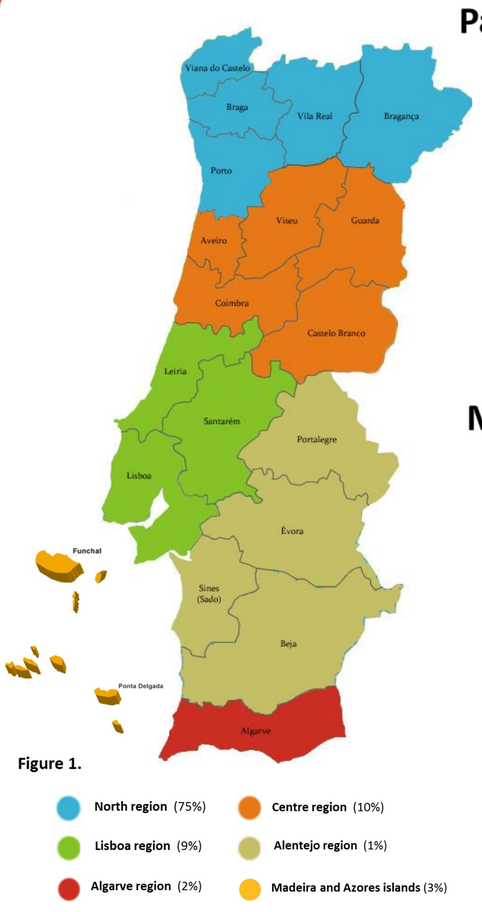

Figure 1.

- 2,352 undergraduate and graduate students (1,483 female. M_{age}=22.82, *SD*=5.41) from several public and private universities from all regions of Portugal participated in the study (see Fig. 1).
- All participants were native speakers of European Portuguese (EP) and had normal (54.6%) or corrected-to-normal vision (45.4%).

Materials:

- 3,800 EP words that vary in length ($M_{letters}$=7.16, *SD*=2.07, *range*: 2 to 12), and in *per million* word frequency of occurrence in the P-PAL *corpus* (M_{freq}=39.53, *SD*=85.49, *range*: 0.01 to 1,214.45) (http://p-pal.di.uminho.pt/tools) integrated this word dataset (see Fig. 2).
- The lexical material included words from the Portuguese adaptation of the Affective Norms for English Words (Soares et al., 2012), the Portuguese age-of-acquisition norms (Cameirão & Vicente, 2010; Marques, Fonseca, Morais, & Pinto, 2007), the P-PAL (Soares et al., *in press*) and ESCOLEX databases (Soares et al., 2013) and the EP translation of words from the Bristol norms words (Stadthagen-Gonzalez & Davis, 2006).

Procedure:

- Participants rated 100 words drawn randomly from the full set of words in each of the three subjective indices using a web survey procedure. An invitation with a hyperlink was sent via e- mail (http://palavras.no-ip.org/).
- After completing the registration, participants were asked to rate each word (one at a time) in each of the three subjective indices (see Fig. 3). The order of subjective index presentation was counterbalanced across participants.

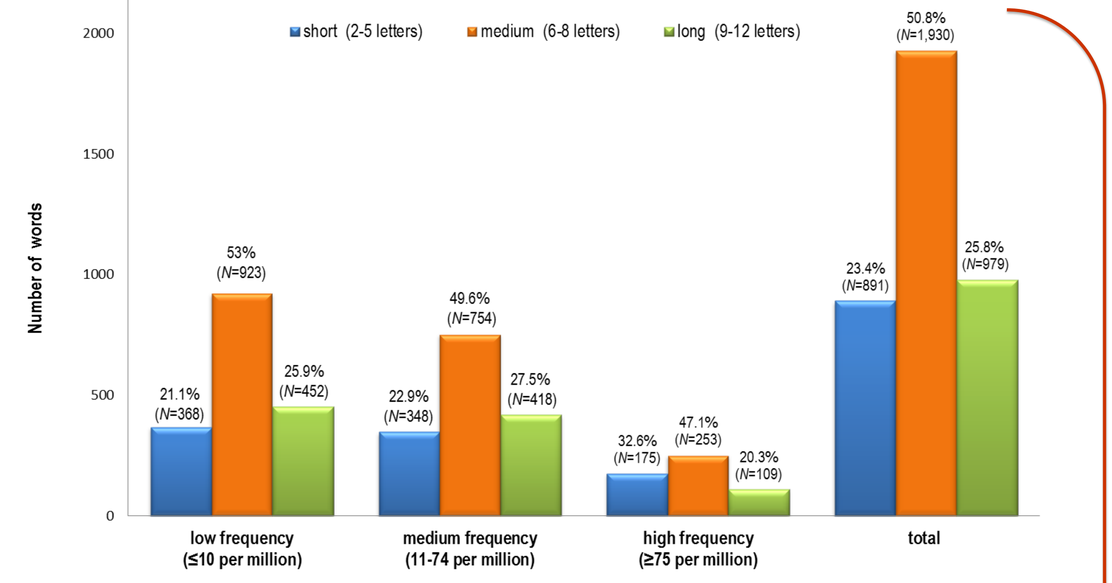

Figure 2. Distribution of the 3,800 words according to P-PAL word frequency intervals (low, medium and high) and word length intervals (short, medium and long words) and for the total word dataset.

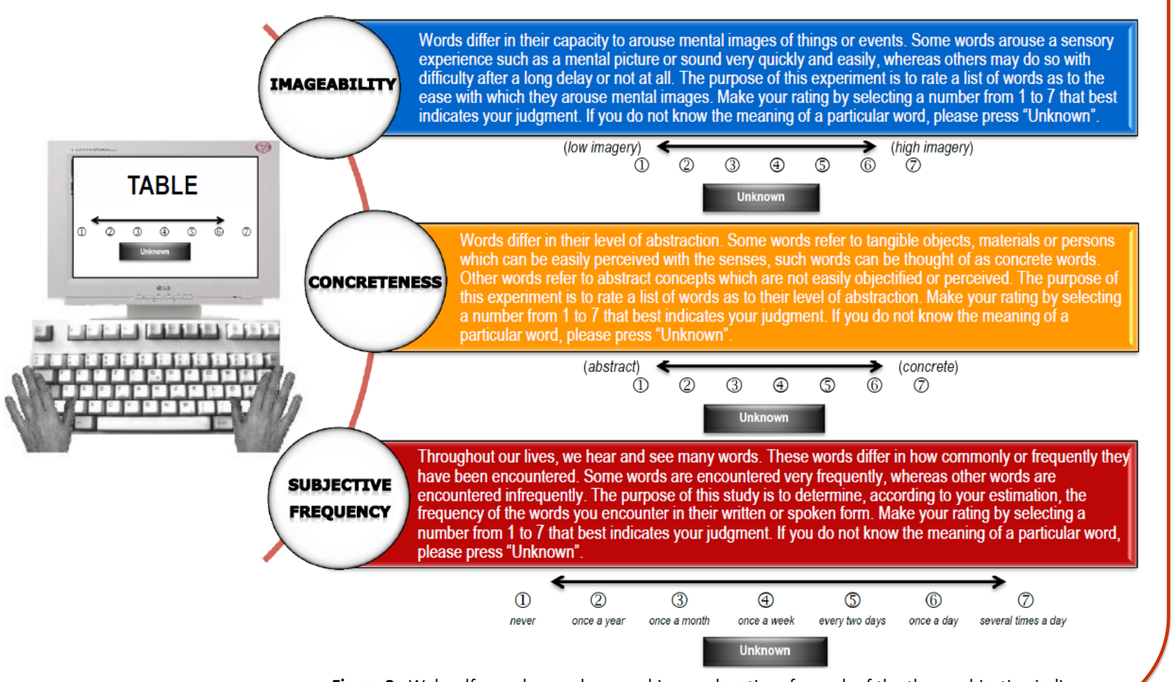

Figure 3. Web self-paced procedure used in words ratings for each of the three subjective indices.

Lexical decision

Materials:

- 1,920 out of the 3,800 Portuguese words that vary in length ($M_{letters}$=6.89, *SD*=2.10, *range*: 2 to 15), and in *per million* word frequency of occurrence (M_{freq}=67.33, *SD*=110.83, *range*: 0.44 to 1,214.45) from P-PAL database (http://p-pal.di.uminho.pt/tools) plus 1,920 orthographically legal nonwords due to task requirements.

Participants:

- 58 undergraduates students from the University of Minho (Portugal) participated in the experiment (52 females; M_{age}: 21.3, *SD*=3.06). All participants were native speakers of EP and had normal (60.1%) or corrected-to-normal vision (39.9%).

Task and Procedure:

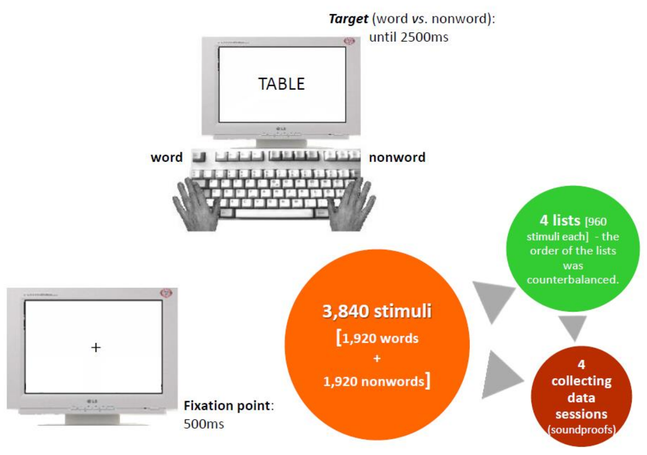

Figure 4. Example of an experimental trial of the lexical decision task.

Results:

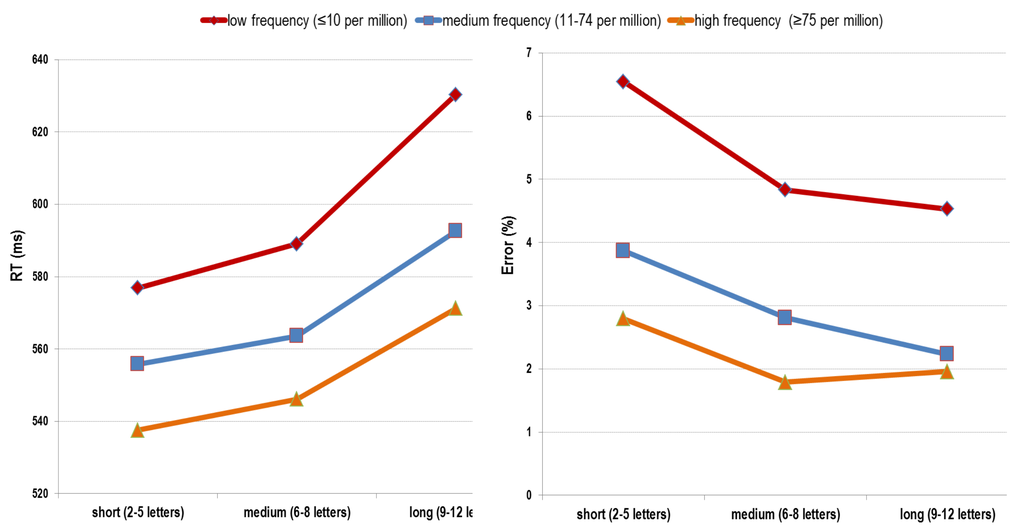

Figure 5. Reaction times (ms) of the correct trials (*N*=1,909) according to P-PAL word frequency intervals (low, medium and high) and word length intervals (short, medium and long words).

Figure 6. Error responses (%) according to P-PAL word frequency intervals (low, medium and high) and word length intervals (short, medium and long words).

Table 2: Correlations between Imageability (IMAG), Concreteness (CONC), Subjective Frequency (SUB_{freq}) and Lexical Decision Times (TR) and Accuracy (Acc) according to P-PAL word frequency intervals (low, medium and high) and word length intervals (short, medium and long words).

P-PAL word frequency intervals	Word length intervals	IMAG correlated with RT	CONC correlated with RT	$SUBJ_{freq}$ correlated with RT	IMAG correlated with Acc	CONC correlated with Acc	$SUBJ_{freq}$ correlated with Acc
All word dataset		**-.17****	**-.07****	**-.55****	**-.02**	**.04**	**-.39****
Low frequency (≤10)	**all**	**-.28****	**-.16****	**-.45****	**-.08**	**.01**	**-.47****
	short	-.35**	-.18	-.46**	-.42**	-.30**	-.43**
	medium	-.15*	-.04	-.57**	-.10	.01	-.46**
	long	-.11	.02	-.54**	.11	.17	-.51**
Medium frequency (11-74)	**all**	**-.28****	**-.18****	**-.50****	**-.08****	**-.03**	**-.33****
	short	-.22**	-.11	-.46**	-.18**	-.09	-.31**
	medium	-.23**	-.13**	-.52**	-.18**	-.10*	-.39**
	long	-.09	-.02	-.54**	.02	.03	-.36**
High frequency (≥75)	**all**	**-.23****	**-.12***	**-.42****	**-.11***	**-.04**	**-.06**
	short	-.23**	-.12	-.34**	-.15	-.09	-.06
	medium	-.21**	-.12	-.33**	-.18**	-.08	-.14*
	long	.05	.19	-.52**	-.06	.01	-.24*

*p<.05, **p<.001

Table 1: Correlations between Imageability (IMAG), Concreteness (CONC), Subjective Frequency (SUB_{freq}), P-PAL word frequency ($P\text{-}PAL_{freq}$) and Word Length in number of letters ($LENG_{lett}$).

	IMAG	CONC	$SUBJ_{freq}$	$P\text{-}PAL_{freq}$	$LENG_{lett}$
IMAG	-				
CONC	**.88****	-			
$SUBJ_{freq}$	**.05****	**-.07****	-		
$P\text{-}PAL_{freq}$	**-.09****	**-.11****	**.40****	-	
$LENG_{lett}$	**-.28****	**-.27****	**-.07****	**-.11****	-

** p<.001

Table 3: Percentages of variance in Lexical Decision Times (RT) and Accuracy (Acc) explained by Imageability (IMAG), Concreteness (CONC) and Subjective Frequency (SUB_{freq}) according to P-PAL word frequency intervals (low, medium and high) and word length intervals (short, medium and long words).

P-PAL word frequency intervals/ Subjective indices		Word length intervals	RT	Acc
All word dataset	SUBJfreq		**30.0****	**14.7****
	SUBJfreq +IMAG		**32.5****	-
	SUBJfreq + IMAG+CONC		**33.4****	-
Low frequency (≤10)	SUBJfreq	**all**	**20.2****	**21.5****
		short	20.5**	18.0**
		medium	32.3**	20.9**
		long	28.3**	24.8**
	SUBJfreq +IMAG	**all**	**27.6****	-
		short	28.9**	31.3**
		medium	33.6**	-
		long	-	-
	SUBJfreq + IMAG+CONC	**all**	**28.6****	-
		short	-	-
		medium	34.8**	-
		long	-	-
Medium frequency (11-74)	SUBJfreq	**all**	**25.2****	**11.1****
		short	21.3**	9.4**
		medium	27.3**	14.7**
		long	28.7**	12.6**
	SUBJfreq + IMAG	**all**	**30.0****	-
		short	24.8**	11.6**
		medium	30.5**	16.6**
		long	-	-
	SUBJfreq + IMAG+CONC	**all**	31.0**	-
		short	27.1**	12.5**
		medium	31.9**	17.3**
		long	-	-
High frequency (≥75)	SUBJfreq	**all**	**17.3****	**0.9***
		short	10.9**	-
		medium	10.2**	2.7*
		long	26.2**	4.5*
	SUBJfreq + IMAG	**all**	**20.2****	**1.6****
		short	13.8**	-
		medium	13.2**	4.3*
		long	-	-
	SUBJfreq + IMAG+CONC	**all**	**22.2****	-
		short	16.0**	-
		medium	-	-
		long	-	-

*p<.05, **p<.001

Research project [PTDC/PSI-PCO/104679/2008] funded by FCT (Fundação para a Ciência e Tecnologia), and FEDER (Fundo Europeu de Desenvolvimento Regional) through the European programs QREN (Quadro de Referência Estratégico Nacional), and COMPETE (Programa Operacional Factores de Competitividade).

asoares@psi.uminho.pt