

Supplementary Information: Data Dictionary

1.1 Related to the words/basewords

frq = Frequency of Use per Million. As measured by the ESPAL¹ database, it is defined as the number of times a word appears in the SUBTITLE corpus divided by the total count of the corpus words multiplied by one million.

zipf = Zipf Scale. Logarithmic scale with frequency values, computed as the $\log_{10}(\text{frq})+3$. Also, this variable was taken from ESPAL¹.

CDM_cnt = Contextual diversity (movie-based) count. The number of different movies the word appears in, measured by the ESPAL database¹.

CDM_frq = Contextual diversity (movie-based) frequency. The percentage of movies the word appears in, measured by the ESPAL database¹.

max_lem_cat = Most frequent part of speech category. ADJECTIVE, ADPOSITION, ADVERB, DATE, INTERJECTION, NOUN, NUMBER, VERB.

syll_structure = Syllabic Structure. For each word the syllabification is included, using “.” between each syllable.

Lev_N = Levenshtein N. This measure of similarity reflects the number of deletions, insertions, or substitutions required to transform one word into another. The cost of each operation is 1—note that the cost of transposing two letters would be 2. As usual, this metric was presented as the mean of the 20 nearest neighbors (i.e., OLD20²).

Familiarity. The degree of experience an individual has with a specific word. For 328 of the words, there is a familiarity rating average taken from the ESPAL database¹ (1-7 range), as well as using EmoFinder³, where the familiarity ratings from different databases⁴⁻⁶ can be retrieved, all converted to the same scale (i.e., 1–7).

Imaginability. The extent to which a word elicits a mental image, reflecting how easily it can be imagined. For 315 of the words, there is an imaginability rating average taken from the ESPAL database¹ (1-7 range), as well as using EmoFinder³, where the familiarity ratings from different databases⁴⁻⁶, can be retrieved, all converted to the same scale (i.e., 1–7).

Concreteness. The specificity of a word's content, ranging from concrete (tangible) to abstract (intangible). For 329 of the words, there is a concreteness rating average taken from the ESPAL database¹, as well as using EmoFinder³, where the concreteness ratings from different databases⁴⁻⁶ can be retrieved, all converted to the same scale (i.e., 1–7).

AoA = Age of acquisition; average chronological age at which individuals first learn a particular word, as estimated by adults. For 340 of the words, the age of acquisition rating average, taken from two databases^{7, 5}. Scale: 1-11.

Valence. The emotional value associated with a word, indicating whether the word conveys a pleasant (positive) or unpleasant (negative) emotional response. For 380 of the words, there is a valence rating average retrieved using EmoFinder³, where the valence ratings from different databases^{4, 8-11}, can be retrieved, all converted to the same scale (i.e., 1–9).

Arousal. The level of emotional activation or intensity that a word evokes, ranging from low (calming) to high (exciting). For 380 of the words, there is an arousal rating average retrieved using EmoFinder³, where the valence ratings from different databases^{4, 8-11}, can be retrieved, all converted to the same scale (i.e., 1–9).

Happiness. The extent to which a word elicits feelings of happiness, ranging from absolutely none to extremely high. For 379 of the words, there is a happiness rating average retrieved using EmoFinder³, where the ratings from different databases^{9, 8, 11} can be retrieved (scale 1-5).

Disgust. The extent to which a word elicits feelings of disgust, ranging from absolutely none to extremely high. For 379 of the words, there is a disgust rating average retrieved using EmoFinder³, where the ratings from different databases^{9, 8, 11} can be retrieved (scale 1-5).

Anger. The extent to which a word elicits feelings of anger, ranging from absolutely none to extremely high. For 379 of the words, there is an anger rating average retrieved using EmoFinder³, where the ratings from different databases^{9, 8, 11} can be retrieved (scale 1-5).

Fear. The extent to which a word elicits feelings of fear, ranging from absolutely none to extremely high. For 379 of the words, there is a fear rating average retrieved using EmoFinder³, where the ratings from different databases^{9, 8, 11} can be retrieved (scale 1-5).

Sadness. The extent to which a word elicits feelings of sadness, ranging from absolutely none to extremely high. For 379 of the words, there is a sadness rating average retrieved using EmoFinder³, where the ratings from different databases^{9, 8, 11} can be retrieved (scale 1-5).

Iconicity. The resemblance between word sound and its meaning, ranging from very arbitrary to very iconic. For 200 of the words, there is an iconicity rating average¹². Scale: 1-7.

SERs = Sensory experience ratings; the extent to which a word evokes a sensory and/or perceptual experience. For 322 of the words, there is a SERs rating average^{13, 5}. Scale: 1-7.

1.2 Related to the specific items (either words or pseudowords)

Rep_letters = Includes repeated letters. A flag indicating if there is at least one letter that is repeated.

MLBF = Mean of the logarithm of the frequency (token-type) of bigrams within the string, taken from BPAL¹⁴.

CV_O = Consonant-Vowel orthographic structure of the stimulus, taken from BPAL¹⁴.

Nplus = Sum of the number of replaced-letter orthographic neighbors (words that differ in one letter), nested (letter removal), nesting (letter addition), and by transposition, extracted from BPAL¹⁴.

1.3 Related to the pseudowords

CVchange = Consonant-Vowel orthographic structure of the pair of transposed letters.

CV_comp = Comparison of the consonant-vowel orthographic structure of the target letters between the baseword and the pseudoword (i.e., same[misma] vs. different[diferente]).

Reference list

1. Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A. & Carreiras, M. EsPal: One-stop shopping for Spanish word properties. *Behav Res Methods* 45, 1246–1258 (2013).
2. Yarkoni, T., Balota, D. & Yap, M. Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychon Bull Rev* 15, 971–979 (2008).
3. Fraga, I., Guasch, M., Haro, J., Padrón, I. & Ferré, P. EmoFinder: The meeting point for Spanish emotional words. *Behav Res Methods* 50, 84–93 (2018).
4. Guasch, M., Ferré, P. & Fraga, I. Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behav Res Methods* 48, 1358–1369 (2016).
5. Hinojosa, J. A. *et al.* The Madrid Affective Database for Spanish (MADS): Ratings of Dominance, Familiarity, Subjective Age of Acquisition and Sensory Experience. *PLoS One* 11, e0155866 (2016).
6. Ferré, P., Guasch, M., Moldovan, C. & Sánchez-Casas, R. Affective norms for 380 Spanish words belonging to three different semantic categories. *Behav Res Methods* 44, 395–403 (2012).
7. Alonso, M. A., Fernandez, A. & Díez, E. Subjective age-of-acquisition norms for 7,039 Spanish words. *Behav Res Methods* 47, 268–274 (2015).
8. Hinojosa, J. A. *et al.* Affective norms of 875 Spanish words for five discrete emotional categories and two emotional dimensions. *Behav Res Methods* 48, 272–284 (2016).
9. Ferré, P., Guasch, M., Martínez-García, N., Fraga, I. & Hinojosa, J. A. Moved by words: Affective ratings for a set of 2,266 Spanish words in five discrete emotion categories. *Behav Res Methods* 49, 1082–1094 (2017).
10. Ferré, P., Guasch, M., Moldovan, C. & Sánchez-Casas, R. Affective norms for 380 Spanish words belonging to three different semantic categories. *Behav Res Methods* 44, 395–403 (2012).
11. Stadthagen-Gonzalez, H., Imbault, C., Pérez Sánchez, M. A. & Brysbaert, M. Norms of valence and arousal for 14,031 Spanish words. *Behav Res Methods* 49, 111–123 (2017).
12. Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A. & Ferré, P. Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behav Res Methods* 53, 1262–1275 (2021).
13. Díez-Álamo, A. M., Díez, E., Wojcik, D. Z., Alonso, M. A. & Fernandez, A. Sensory experience ratings for 5,500 Spanish words. *Behav Res Methods* 51, 1205–1215 (2019).
14. Davis, C. J. & Perea, M. BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behav Res Methods* 37, 665–671 (2005).