

XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional. II Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires, 2020.

¿Cuál es más frecuente?: cuenta palabras, diccionario de frecuencia léxica infantil para el español Rioplatense.

Sánchez, María Elina y Fumagalli, Julieta.

Cita:

Sánchez, María Elina y Fumagalli, Julieta (2020). *¿Cuál es más frecuente?: cuenta palabras, diccionario de frecuencia léxica infantil para el español Rioplatense. XII Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXVII Jornadas de Investigación. XVI Encuentro de Investigadores en Psicología del MERCOSUR. II Encuentro de Investigación de Terapia Ocupacional. II Encuentro de Musicoterapia. Facultad de Psicología - Universidad de Buenos Aires, Buenos Aires.*

Dirección estable: <https://www.aacademica.org/000-007/385>

ARK: <https://n2t.net/ark:/13683/etdS/nQm>

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

¿CUÁL ES MÁS FRECUENTE?: CUENTA PALABRAS, DICCIONARIO DE FRECUENCIA LÉXICA INFANTIL PARA EL ESPAÑOL RIOPLATENSE

Sánchez, María Elina; Fumagalli, Julieta

Universidad de Buenos Aires. Facultad de Filosofía y Letras. Instituto de Lingüística. Buenos Aires, Argentina.

RESUMEN

Introducción: La frecuencia léxica es una variable psicolingüística relevante en los estudios sobre comprensión y producción del lenguaje que explica por qué algunas palabras se procesan más rápido y de manera más eficaz que otras. Esta variable debe ser considerada durante la selección de estímulos en todos los trabajos que investiguen el procesamiento lingüístico. Sin embargo, en español rioplatense no contamos con diccionarios de frecuencia léxica para adultos ni para niños, por lo que las investigaciones y herramientas de evaluación clínica diseñadas para nuestra población utilizan bases generadas en otros dialectos -generalmente el español ibérico. Objetivo: En este trabajo se muestran los datos preliminares de un diccionario de frecuencia léxica infantil para el español en su variedad rioplatense. Materiales: El corpus actual se conformó con 810.799 palabras extraídas de 27 textos escolares de 1° a 7° grado de nivel primario. Resultados: Los datos se analizaron informáticamente para calcular la frecuencia total y la frecuencia por ítem según nivel escolar y fueron procesados con las herramientas Tokenizer y FreqDist (Python 3) y la versión para español del Stanford Log-linear PoS-Tagger. Para la lematización se utilizó SnowballStemmer y se eliminaron los nombres propios mediante NamedEntityRecognition.

Palabras clave

Psicolingüística - Léxico mental - Frecuencia léxica - Diccionario infantil

ABSTRACT

WHICH IS MOST FREQUENT?: CUENTA PALABRAS, DICTIONARY OF CHILDREN'S LEXICAL FREQUENCY FOR RIOPLATENSE SPANISH
Introduction: Lexical frequency is a relevant psycholinguistic variable in studies of language comprehension and production that explains why some words are processed faster and more efficiently than others. This variable must be considered during the selection of stimuli in all works that investigate linguistic processing. However, in Rioplatense Spanish we do not have lexical frequency dictionaries for adults or children, so the research and clinical evaluation tools designed for our population use bases generated in other dialects - generally Iberian Spanish. Objective: This work shows the preliminary data from a dictionary of children's lexical frequency for Spanish in its

Rioplatense variety. Materials: The current corpus was made up of 810,799 words extracted from 27 school texts from 1st to 7th grade of primary level. Results: The data were analyzed computerized to calculate the total frequency and the frequency by item according to school level and were processed with the Tokenizer and FreqDist tools (Python 3) and the Spanish version of the Stanford Log-linear PoS-Tagger. For the stemming, SnowballStemmer was used and the proper names were eliminated using NamedEntityRecognition.

Keywords

Psycholinguistics - Mental lexicon - Lexical frequency - Infantil dictionary

BIBLIOGRAFÍA

- Alameda, J.R. y Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.
- Corral, S.; Ferrero, M. y Goikoetxea, E. (2009) LEXIN: A lexical database from Spanish kindergarten and first-grade readers, *Behavior Research Methods*, 41 (4), 1009-1017.
- Davis, C. y Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665-671.
- Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods*, 37, 65-70
- Duchon, A., Perea, M., Sebastián-Gallés, N., Martín, A. y Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behav Res* 45: 1246.
- Juilland, A. y Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. La Haya: Mouton.
- Justicia, F. (1995). *El desarrollo del vocabulario: Diccionario de frecuencias*. Granada: Universidad de Granada.
- Martínez, J. A., y García, M. E. (2004). *Diccionario de frecuencias del castellano escrito en niños de 6 a 12 años*. Salamanca: Universidad Pontificia de Salamanca.
- Sebastián Gallés, N., Martí, M.A., Carreiras, M. y Cuetos, F. (2000). *Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.