

Desarrollo de un algoritmo para la medición del grado de similitud fonológica entre formas escritas

Hohendahl, A.¹, Zanutto, B. S.^{1,2}, Wainselboim, A. J.²

1- Instituto de Ingeniería Biomédica, Facultad de Ingeniería (UBA). Paseo Colón 850, Buenos Aires, Argentina.

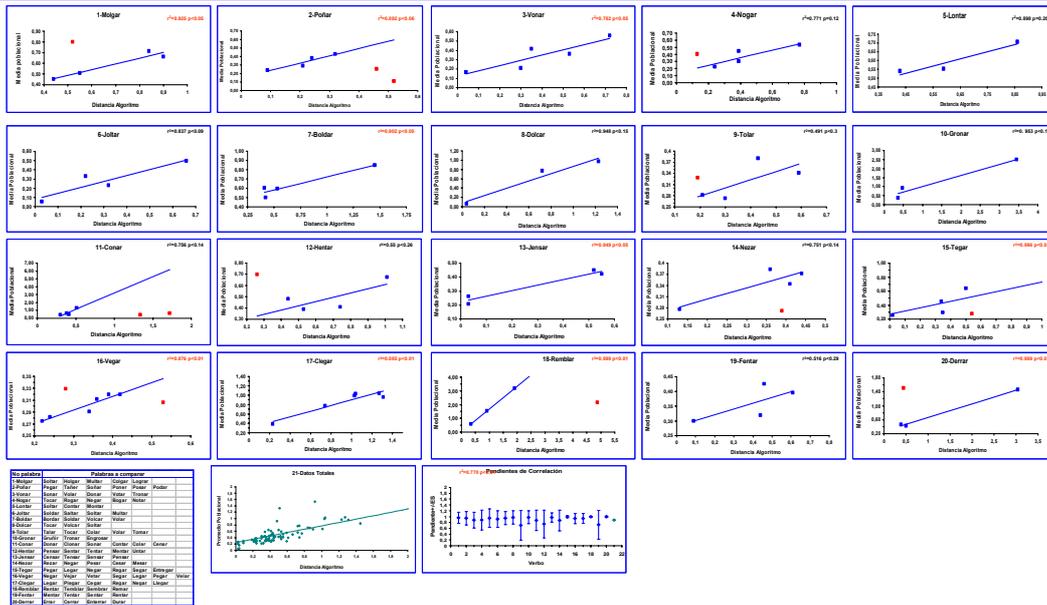
2- Instituto de Biología y Medicina Experimental (Conicet). Vuelta de Obligado 2490 (1428), Bs. AS. Argentina

Introducción. Se define como vecino ortográfico a toda palabra que presenta un cambio, inserción o deletión de grafema con respecto a la forma analizada, manteniendo el resto de las posiciones inalteradas (1). Múltiples estudios muestran que el número de vecinos y su frecuencia influye el procesamiento de las formas analizadas (2, 3, 4, 5). Adicionalmente, en tareas como lectura y repetición de palabras se ve una influencia del número de vecinos fonológicos y del grado de solapamiento o similitud fonológica entre formas (6), sugiriendo una influencia de los aspectos fonológicos de la forma escrita durante su análisis.

Objetivo. A partir del interés en estudiar la influencia de aspectos fonológicos en el análisis de formas escritas del español, nos planteamos el objetivo de desarrollar un algoritmo que calculara el grado de distancia fonética entre formas escritas del español rioplatense, reflejando las distancias perceptuales reportadas por los sujetos.

Metodología. Medidas experimentales: 92 adultos, nativos de Capital de nivel terciario, realizaron una tarea escrita con 20 grupos distintos de palabras. Ordenaron en una escala numérica las palabras de cada grupo (ej. "sonar", "volar", "donar", "votar") en función de su diferencia fonética con una no-palabra a comparar (ej. "vonar"). A partir de la población de datos se calculó luego para cada palabra el valor de ordenación medio. Estos valores fueron correlacionados luego con las distancias fonéticas dadas por el algoritmo. **Algoritmo:** evalúa los parámetros *Posición*, *Tamaño (volumen y duración)*, *Ataque (plosividad)*, *Glótica (expresa)*, *Fricativa* y *Vibrante* en cada fonema. *El algoritmo de posee varias etapas: conversión de fonemas en grafemas, cálculo de distancia entre los fonemas homónimos, hallando la alineación para menor distancia y finalmente una penalización por diferencia en el número de fonemas. La etapa consistente en convertir fonemas a grafemas, adoptamos un subconjunto de 30 fonemas del IPA, (castellano rioplatense), codificados como SAMPA (Gurlekian, LIS). Una vez obtenida la secuencia de fonemas, se realiza una comparación de distancia inter-fonemas, analizando todas las posibles alineaciones (en caso de tener diferente número de fonemas) tomándose como distancia resultante la suma de las distancias que resulte menor. Las distancias entre fonemas, se calculan mediante vectores. Se asignan valores de penalización alineal (función potencial) por la diferencia entre el número de fonemas comparados, sumando la distancia al silencio de cada fonema omitido de la palabra mas larga. La comparación se realiza por fonemas enfrentados para cada posible posición enfrentada y se compara con la posibilidad de que la distancia entre fonemas adyacentes sea menos que la suma de las distancias entre los enfrentados, resultando un mecanismo de absorción similar a cuando una persona percibe letras mas largas o dos fonemas seguidos, mejorando la efectividad del algoritmo a expensas de aumentar su complejidad y tiempo de ejecución.*

Resultados. Figuras 1 a 20: correlaciones entre los valores experimentales medios y las distancias obtenidas mediante el algoritmo, para cada grupo de palabras/no palabra presentado (ver grupos en Tabla 1). En rojo, datos experimentales que no ajustan a cada correlación (14 de 95 comparaciones) **Figura 21:** correlación entre el total de valores experimentales ajustados y las distancias del algoritmo. **Figura 22:** Pendientes +/- ES para cada una de las correlaciones anteriores.



Palabras	Palabras a comparar
1- Mijor	Mijor - Pijor
2- Poliar	Poliar - Piliar
3- Volar	Volar - Pular
4- Niggar	Niggar - Pigar
5- Lonlar	Lonlar - Pular
6- Jolar	Jolar - Pular
7- Bollar	Bollar - Pular
8- Dolar	Dolar - Pular
9- Tolar	Tolar - Pular
10- Golar	Golar - Pular
11- Conlar	Conlar - Pular
12- Holar	Holar - Pular
13- Jolar	Jolar - Pular
14- Nolar	Nolar - Pular
15- Tolar	Tolar - Pular
16- Volar	Volar - Pular
17- Colar	Colar - Pular
18- Rolar	Rolar - Pular
19- Folar	Folar - Pular
20- Oolar	Oolar - Pular

Tabla 1. Grupos de palabras/no palabras a comparar por los sujetos.

Discusión. Las distancias calculadas por medio del algoritmo mostraron una importante correlación con los valores de distancia fonética obtenidos experimentalmente en aproximadamente 85,25% de los casos analizados (81 de 95 comparaciones). La ausencia de significancia estadística en varias de las pendientes de correlación calculadas, se debió principalmente al bajo número de datos obtenidos dentro de cada grupo a comparar.

En comparación con lo que sucede con los hablantes reales, el algoritmo parece otorgarle un *menor* peso a la diferencia entre vocales (Figs. 1, 4 y 20), y un *mayor* peso a las inserciones consonánticas (Figs 11 y 18).

Creemos que aunque aún se encuentra en una etapa preliminar, el desarrollo del presente algoritmo podría servir en un futuro como herramienta en la selección, diseño y análisis de estímulos en experimentos que estudiaran relaciones entre fonología y ortografía, permitiendo estudiar la influencia del grado de semejanza formal entre estímulos lingüísticos.

Referencias

1. Coltheart M., Davelaar E., Jonasson J. F. & Besner D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.
2. Alameda y Cuetos 1997. Alameda, J. R., & Cuetos, F. (1997). Frecuencia y vecindad ortográfica: ¿Factores independientes o relacionados? *Psicológica* 18: 43-68.
3. Pollatsek A., Perea M. & Binder K. (1999) The effects of neighborhood size in reading and lexical decision. *Journal of Experimental Psychology: Human Perception and Performance*, 25: 1142-1158.
4. Siakaluk P.D., Sears C. R. & Lupker S. J. (2002) Orthographic Neighborhood Effects in Lexical Decision: The Effects of Nonword Orthographic Neighborhood Size. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 28, No. 3: 661-681.
5. Arduino L. S. & Burani C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research* 33: 75-95.
6. Peereman, R. & Content A. (1997) Orthographic and Phonological Neighborhoods in Naming: Not All Neighbors Are Equally Influential in Orthographic Space. *Journal of Memory and Language* 37: 382-410.