Revista Electrónica de Metodología Aplicada
2003, Vol. 8 nº 2, pp. 1-10

Frecuencia, longitud y vecindad ortográfica de las palabras de 3 a 16 letras del Diccionario de la Lengua Española (RAE, 1992)

Miguel Ángel Pérez [1]
José Ramón Alameda[2]
Fernando Cuetos Vega[3]
[1] Departamento de Psicología Social y Metodología de las Ciencias del Comportamiento. Universidad de Murcia.
[2] Departamento de Psicología. Universidad de Huelva.
[3] Departamento de Psicología. Universidad de Oviedo.

RESUMEN

Frecuencia, longitud y vecindad ortográfica son tres variables sumamente influyentes en los tiempos de lectura y reconocimiento visual de palabras. De ahí la necesidad de contar con índices de estas variables para el mayor número posible de palabras. En este estudio se presenta una base de datos informatizada que incluye todas las palabras del español de entre tres y dieciséis letras, un total de 95.383, tomadas del Diccionario de la Lengua Española (RAE, 1992). Esta base de datos, en formato Microsoft Excel, permite buscar de manera rápida y automática la frecuencia, vecindad y longitud de cualquiera de esas palabras.

Palabras clave: base de datos, frecuencia léxica, vecindad ortográfica.

ABSTRACT

Frequency, length and orthographic neighbourhood are three extremely important determinants in reading and visual word recognition. For this reason is very important take into account scores of these variables for the greater possible number of words. In this study a data base computerized is presented that includes all the Spanish words, a total of 95,383, taken from the Dictionary of the Spanish Language (RAE, 1992). This data base, in format Microsoft Excel, allows to find in an automatic and fast way the frequency, neighbourhood and length of anyone of those words.

Keywords: database, lexical frequency, orthographic neighbourhood.

Correspondencia

Miguel Ángel Pérez
Univ. de Murcia. Dpto. de Psicología Básica y Metodología
Campus de Espinardo, s/n.
30100 – Murcia (España)
e-mail: maperez@um.es
Tlf: (+34) 968 36 34 82; Fax: (+34) 968 36 41 15

1.- Introducción

El reconocimiento visual de palabras, es sin duda, uno de los temas más investigados de la Psicolingüística. Han sido numerosos los estudios realizados con la metodología de tiempos de reacción dirigidos a averiguar cuáles son los procesos que intervienen en el reconocimiento de palabras y cuáles son las variables que determinan esos procesos. Y cualquiera que sea la técnica utilizada, presentación taquistoscópica, desenmascaramiento progresivo, lectura en voz alta, tarea de decisión léxica, etc, invariablemente se encuentra que la frecuencia de uso es una de las variables más influyentes en los tiempos de reconocimiento de las palabras escritas, en el sentido de que cuantas más veces vemos una palabra menos tiempo tardamos en reconocerla (Rubenstein, Garfield y Millikan, 1970; Foster y Chambers, 1973). De hecho, todos los modelos de reconocimiento de palabras disponen de alguna explicación de este efecto. Así, en el modelo Logogén (Morton, 1969), cada logogén, o representación de la palabra, tiene un umbral de activación que disminuye cada vez que se presenta una palabra. Los modelos conexionistas tienen explicación similar: cada vez que se procesa una palabra se fortalecen las conexiones correspondientes a esa palabra (McClelland y Rumelhart, 1981).

Otras variables que también influyen en el reconocimiento de las palabras escritas son la longitud, puesto que se tarda más tiempo en reconocer las palabras largas que las cortas (Cuetos, Dominguez y De Vega, 1997), la frecuencia silábica, a mayor frecuencia de las sílabas mayores son los tiempos de respuesta (Carreiras, Álvarez y De Vega, 1993), la imaginabilidad, la palabras más fáciles de imaginar se reconocen antes que las de baja imaginabilidad (Cuetos et al., 1997), la edad de adquisición, las palabras que se aprenden tempranamente a lo largo de la vida se reconocen antes que las que se aprenden a una edad tardía (Ellis y Morrison, 1998), la familiaridad, las palabras familiares se reconocen antes que las poco familiares (Connine, Mullennix, Shernoff y Yelen, 1990), la polisemia, cuantos más significados tiene una palabra menores son los tiempos de respuesta (Cuetos, et al 1997) o la vecindad ortográfica, cuando una palabra se parece ortográficamente a otras los tiempos de reacción aumentan porque se produce una competición entre ellas (Grainger y Seguí, 1990).

En consecuencia, cuando se realiza un experimento sobre reconocimiento de palabras hay que considerar un buen número de variables, si queremos hacer una interpretación correcta de los resultados. Obviamente no se pueden manipular todas a la vez, pero sí que se deben controlar. Pero eso exige tener datos de cada una de ellas, lo cuál no siempre es factible. Afortunadamente, en los últimos años se han publicado un buen número de artículos y bases de datos en español con información sobre las principales variables (en Pérez, Campoy y Navalón, 2001, se puede encontrar una relación de todos los estudios realizados en los últimos años).

La mayoría de las variables léxicas se recogen por medio de cuestionarios en los que se les pide a un grupo de sujetos que clasifiquen cada palabra sobre una escala (generalmente de 7 puntos) en la variable que se trata de cuantificar. Es el caso de la familiaridad o la imaginabilidad, se pide a los sujetos que indiquen lo familiar o imaginable que les resultan las palabras sobre una escala de 1 a 7, donde el 1 corresponde a palabras muy poco familiares o muy poco imaginables y el 7 muy familiares o muy imaginables. También la edad de adquisición se puede estimar mediante cuestionarios subjetivos en los que los sujetos tienen que indicar a la edad que creen que han aprendido cada palabra

Para la medida de la frecuencia, en cambio, aunque también se pueden utilizar cuestionarios, la forma más objetiva de hacerlo es tomando un corpus amplio y variado de material escrito y contando el número de veces que aparece cada palabra. Existen recuentos de frecuencia en prácticamente todos los idiomas. En español, por ejemplo, tenemos el de Juilland y Chang-Rodríguez (1964) sobre un corpus de 500.000 palabras, el de Alameda y Cuetos (1995) sobre un corpus de 2.000.000 de palabras y el de Sebastián, Martín, Carreiras y Cuetos (2000) sobre un corpus de 5.000.000 de palabras. Obviamente, cuanto mayor es el corpus más ajustada será la medida de la frecuencia. No obstante, si el corpus se hace con un buen muestreo de textos, no tiene por qué haber mucha variación en el porcentaje de frecuencia de cada palabra, por el hecho de que aumente el corpus, ya que la frecuencia se suele expresar por número de apariciones por millón de palabras. Pero ciertamente, el tamaño del corpus sobre el que se realiza el recuento puede producir variaciones en el valor de la frecuencia.

También la vecindad ortográfica depende de la lista de palabras que se utilice para hacer los cálculos. Según la definición de Coltheart, Davelaar, Jonasson y Vencer (1977) se consideran vecinas de una palabra a todas aquellas palabras de la misma longitud y que difieren en sólo una letra, conservando las demás en el mismo orden. Por ejemplo, la palabra "casa" tiene como vecinos a "masa", "cosa", "cama" y "caso" entre otros. Por consiguiente, la manera de cuantificar la vecindad ortográfica es comprobando cuántas palabras hay que difieren en sólo una letra de la que queremos medir. Y eso se suele hacer aplicando un programa que haga el recuento. Pero obviamente el número de vecinos que aparecen para cada palabra depende del total de palabras sobre las que se hace el recuento, de tal manera que a mayor número de palabras mayor es el número de vecinos que aparecen para cada palabra. Así, en un recuento hecho sobre un corpus de dos millones de palabras, Alameda (1996) encontró que había 86 palabras de cuatro letras que no tenían ningún vecino ortográfico. Sin embargo, sobre un corpus de cinco millones el número de palabras de cuatro letras que no tenían ningún vecino ortográfico era de 71. Lo ideal sería considerar todas las palabras del castellano para poder hablar realmente de todos los vecinos que tiene cada palabra.

Ese fue el objetivo de este trabajo, construir una base de datos que incluyera todas las palabras del castellano, para lo cual utilizamos el Diccionario de la Lengua Española (DLE) de la Real Academia Española, que contiene todas las formas del español. Y además, queríamos hacer el recuento para las palabras de todas las longitudes, pues aunque en la mayoría de los experimentos de reconocimiento de palabras se suelen utilizar palabras de cuatro y cinco letras, y por esa razón los recuentos de vecinos ortográficos se hacían sólo para palabras de cuatro letras (Alameda y Cuetos, 1996; Perea, 1993) y palabras de cinco letras (Alameda y Cuetos, 2001), algunos trabajos requieren conocer el número de vecinos de palabras más largas. Por ejemplo, en los estudios de escritura dónde se ha visto que los errores ortográficos dependen en buena medida del número de vecinos que tengan las palabras, es necesario disponer de bases de datos en las que aparezcan palabras de diferentes longitudes.

Por otra parte, quisimos contar, no sólo el número global de vecinos que tiene cada palabra, sino el número de vecinos por cada posición de las letras, ya que no todos los vecinos tienen los mismos efectos. Grainger y Seguí (1990) comprobaron que los sujetos nunca cometían errores con las letras inicial y final, sólo con las intermedias, lo que sugiere que los vecinos de letras intermedias producen mayor efecto de inhibición que los de las letras de los extremos. Además de la densidad de vecinos, también se ha utilizado la frecuencia de los

vecinos como variable dependiente en algunos estudios experimentales, y aunque se describen tanto efectos facilitadores (Alameda 1996; Alameda y Cuetos, 2000, Sears, Hino y Lupker, 1995, Sears, Lupker e Hino, 1999), como inhibidores (Alameda y Cuetos, 1997; Grainger, 1990; Grainger y Seguí, 1990; Seguí y Grainger, 1992), nos pareció interesante recoger la frecuencia relativa de los vecinos, es decir, el número de vecinos con frecuencia superior, igual e inferior de cada palabra, tanto posicionalmente como total.

Finalmente, también quisimos recoger la frecuencia acumulada por los vecinos (la suma acumulada de las frecuencias de los vecinos de cada palabra), que si bien es una variable mucho menos estudiada, no carece de interés, ya que puede aportar información sobre el nivel total de unidades competidoras que pueden estar compitiendo con una palabra determinada (Alameda y Cuetos, 1996, Perea, 1993). Aunque utilizando no-palabras, Alameda (1996) encontró un efecto inhibidor de la frecuencia acumulada sobre el tiempo de respuesta, es decir, a mayor frecuencia acumulada mayor tiempo de respuesta. También en este caso hemos querido recoger índices totales y posicionales.

2.- Método

2.1.- Materiales

Para la construcción de esta base de palabras se partió de las 100.544 palabras (entradas) que aparecen en el DLE (RAE, 1992; edición electrónica, 1995). La selección de las palabras que forman la base de datos final se realizó atendiendo a los siguientes criterios.

A) *Eliminación de las palabras menores de tres y mayores de dieciséis letras.* En la investigación psicolingüística habitual, el uso de palabras de 2 letras o de más de 16 es prácticamente nulo. Por lo tanto, desde este punto de vista, resulta injustificado el estudio de la vecindad léxica de dichas palabras.

B) *Omisión de las formas complejas.* El DLE recoge una serie de *formas complejas* (expresiones, dichos, locuciones, etc.) que se encuentran dentro de una entrada o artículo. Por ejemplo, la expresión ‘a tientas’ se encuentra dentro de la definición de ‘tienta’ y ‘pasar por los espinos de Santa Lucía’ dentro de ‘espino’. En muchas de esas formas complejas, hay nombres propios (de personas, personajes, lugares, etc.), diversas conjugaciones verbales y, sobre todo, muchos plurales de palabras. Sin embargo, ninguna de estas tres categorías es exhaustiva—no están todos los nombres propios, ni todas las conjugaciones verbales, ni todos los plurales—por lo que se decidió no incluir las palabras que formaban palabras complejas y no estuvieran en una entrada propia en el diccionario. Así, de los ejemplos anteriores, quedaron excluidas de la selección las palabras ‘tientas’, ‘espinos’ y ‘Lucía’.

C) *División de entradas compuestas*. Las entradas de voces, expresiones o locuciones que contenían más de una palabra fueron divididas, incorporándose en la base como palabras individuales todas aquellas que estuvieran en el rango de longitud establecido (criterio A). Así, por ejemplo, la locución latina ‘a priori’ fue dividida en ‘a’ y ‘priori’, entrando a formar parte del listado final únicamente el último componente. Lo mismo ocurrió con otras expresiones como ‘a mansalva’, ‘a la bartola’ o ‘de refilón’. En otros ejemplos, como en ‘urbi et orbi’ o en ‘ares y mares‘, se incluyó la primera y última palabra. Un caso particular fue el

de la locución adverbial ‘a trochemoche’ que también está aceptada como ‘a troche y moche’. Aunque no existe la palabra ‘troche’ como tal en el DLE, desde la lógica descrita anteriormente, se consideró que era pertinente introducirla en el corpus final de palabras.

D) *Eliminación de lemas repetidos*. Todas las palabras que tenían varias entradas en el DLE, debido a las distintas etimologías o significados, se contabilizaron solamente una vez, independientemente de su significado u origen (por ejemplo, ‘acotar’, que aparece tres veces en el listado de palabras del diccionario, sólo se computó una vez).

E) *Transformación de palabras con guiones*. Con las palabras compuestas que contenían guiones (19 en total) se actuó de la siguiente manera y en este orden: 1) las ocho palabras compuestas (‘judeo-española’, ‘judeo-español’, ‘astur-leonesa’, ‘astur-leonés’, ‘cha-cha-chá’, ‘cará-cará’, ‘caá-miní’ y ‘pil-pil’) que también están aceptadas sin guión en el DLE, fueron incluidas solamente de esta segunda forma; 2) de las 11 palabras que únicamente aparecen con guión (‘aovado-lanceolada’, ‘balto-eslava’, ‘balto-eslavo’, ‘tupí-guaraní’, ‘moro-moro’, ‘dalai-lama’, ‘maquil-ishuat’, ‘best-séller’, ‘agar-agar’, ‘ping-pong’ y ‘cri-cri’), las seis primeras no se seleccionaron porque todos sus componentes existen como palabras sencillas;[1] 3) en las cinco palabras restantes se eliminó en guión y se incluyeron tal y como sigue: ‘maquilishuat’, ‘bestséller’, ‘agaragar’, ‘pingpong’ y ‘cricri’.

F) *Eliminación de prefijos y sufijos*. Como último criterio de formación de la base de palabras para el cálculo de vecindad ortográfica, no se tomaron en cuenta los artículos de todos los prefijos y sufijos incluidos en el DLE.

El número final de palabras seleccionadas, una vez hechas esas eliminaciones, fue de 95.383. La Tabla 1 muestra la cantidad de palabras que hay en cada intervalo de longitud computado.

Tabla 1. Distribución por longitud de las palabras que integran la base de datos.

Longitud[a]	N[b]	Fa.[c]
3	392	392
4	1.914	2.306
5	5.126	7.432
6	9.505	16.937
7	13.618	30.555
8	15.792	46.347
9	15.428	61.775
10	12.635	74.410
11	8.861	83.271
12	5.528	88.799
13	3.312	92.111
14	1.798	93.919
15	1.010	94.919
16	464	95.383

[a] Número de letras.
[b] Número de palabras.
[c] Frecuencia acumulada de palabras.

[1] Concretamente, en la expresión ‘dalai-lama’, la primera palabra no existe como tal en el DLE (RAE, 1992) aunque la segunda sí. Por lo tanto, y siendo congruente con la inclusión de ‘troche’, también se incluyó la palabra ‘dalai’ en el corpus final.

Finalmente, respecto a la elección de la 21ª edición del DLE (RAE, 1992), cabría apuntar que, a pesar de que está publicada en papel la 22ª edición del DLE (RAE, 2001), no existe aun la versión electrónica[2], lo que hizo que fuese prácticamente imposible un estudio de estas características mediante un cálculo manual o, en su caso, la informatización de todas las palabras. Por lo tanto, hay que señalar que se ha trabajado sobre un léxico no actualizado. En este sentido, a título meramente informativo, en la Tabla 2 se presentan datos comparativos entre la 21ª y 22ª edición del DLE.

Tabla2. Comparación entre la 21ª (1992) y 22ª (2001) edición del DLE (RAE).

Variación (en la 22ª Ed.)	Cantidad
Lemas añadidos	11425ª
Lemas suprimidos	6008
Incremento de lemas	5417
Diferencia absoluta de lemas	17433

ª El 49% son lemas americanos.

2.2.- Procedimiento

Se realizó una base de datos en Excel con las 95.385 palabras seleccionadas, agrupando en distintas hojas de cálculo las palabras de la misma longitud. A continuación, se crearon matrices de cadenas de letras derivadas de las palabras[3] sustituyendo la letra de una determinada posición por un carácter comodín (*). De este modo, se calcularon, de forma automática, las coincidencias entre estas cadenas de letras (que originalmente diferían en un carácter de una misma posición), dando como resultado el número de vecinos posicionales. Se respetó la tonicidad de las palabras, es decir, las coincidencias eran sensibles a los acentos (por ejemplo, 'cajón' no se consideró vecino de 'canon'). Seguidamente, se sumaron todos los vecinos posicionales de cada palabra, obteniendo así el número total de vecinos. Además, se computó, en cada caso, si la frecuencia del vecino era igual, mayor o menor a la de la palabra, lo que nos permitía obtener el número total de vecinos de cada palabra en función de la frecuencia relativa. Y por último, se calculó la frecuencia acumulada de todos los vecinos de cada palabra, tanto total como por posición.

También, se obtuvo la frecuencia léxica de cada palabra a partir del diccionario de frecuencias de Alameda y Cuetos (1995). En este sentido, hay que reseñar que partimos de una base de más de 95300 palabras, y sin embargo el diccionario de Alameda y Cuetos está en torno a las 81300 unidades, unas 14000 palabras menos, pero, además debemos tener en cuenta que hay palabras del diccionario de Alameda y Cuetos que no aparecen en el diccionario de la Real Academia, como por ejemplo, *crepé*, *chute* o *modem*, por lo que al final nuestra base de datos recoge 26367 palabras cuya frecuencia no aparece en el diccionario de

[2] Este trabajo se realizó en marzo de 2003 y no habiendo anuncio alguno por parte de la Real Academia Española de la edición de la versión electrónica de la vigésimo segunda edición del DLE (2001), los autores decidieron realizar el estudio sobre la anterior versión que sí disponía de formato digital.

[3] Los vectores de longitud y anchura de las matrices venían determinados por la cantidad y número de letras de las palabras. Por ejemplo, para palabras de 4 letras, la matriz fue de 1.914x4 (= 7.656) y para 9 letras fue de 15.428x9 (=138.852).

Alameda y Cuetos. Por ello, se ha optado por incluir en la base de datos una diferenciación en función de la frecuencia de la palabra, es decir, todos los índices relativos a la vecindad se muestran "triplicados", primero los totales, después los calculados teniendo en cuenta a los vecinos que aparecen en el diccionario de Alameda y Cuetos, y por último los índices calculados teniendo en cuenta sólo a los vecinos con frecuencia cero.

3.- Resultados

A pesar de la automaticidad del proceso realizado, se llevaron a cabo diversas comprobaciones de la efectividad y exactitud de los cálculos obtenidos. El resultado final es una base de datos que contiene distintos índices de vecindad (total y por posición), longitud (número de letras) y frecuencia léxica (Alameda y Cuetos, 1995) de todas las palabras de 3 a 16 letras recogidas en el DLE (RAE, 1992). La base está completamente informatizada y se presenta en formato de Microsoft Excel. Se han diseñado varias hojas de cálculo (véanse anexos) atendiendo a las principales necesidades que podría tener el investigador psicolingüista. De este modo, en el Anexo I, mediante un formulario de búsqueda, se puede obtener para palabras de 3 a 9 letras: 1) la longitud de la palabra, 2) la frecuencia léxica, y 3) índices generales y posicionales de vecindad ortográfica, incluyendo el numero de vecinos con mayor, menor o igual frecuencia.

También se han incluido índices de vecindad para palabras con frecuencia mayor que cero, según el diccionario de frecuencias de Alameda y Cuetos. El Anexo II permite realizar lo mismo con palabras de 10 a 16 letras. Del Anexo III al XVI se hallan respectivamente las hojas de cálculo que permiten buscar cuáles son exactamente los vecinos de una determinada palabra de 3 a 16 letras. Con el Anexo XVII se puede calcular el número de vecinos de no-palabras de 3 a 16 letras, tomando como referencia todas las palabras del DLE. Con el Anexo XVIII se puede realizar lo mismo, pero teniendo en cuenta sólo las palabras del DLE que aparecen en el diccionario de frecuencias de Alameda y Cuetos (1995).

Por otro lado, en el Anexo XIX, se incluyen algunos datos complementarios que podrían interesar al investigador, como son las palabras de 17 o más letras (380 ítem), los prefijos (204 ítem) y los sufijos (372 ítem) que aparecen en el DLE (RAE, 1992).

Finalmente, en el Anexo XX existe más información y ayuda de uso de todos los ficheros que componen este estudio.

4.- Conclusiones

El objetivo de este estudio es poner a disposición de los investigadores que trabajan sobre reconocimiento visual de palabras, lectura y escritura una base de datos completa, con todas las palabras del castellano, que les permite buscar de una manera rápida y automática tres importantes variables: frecuencia léxica, longitud e índices de vecindad ortográfica

(Anexos I y II). También es posible localizar cuáles son exactamente los vecinos de una determinada palabra (Anexos del III al XVI), así como saber el número de vecinos de no-palabras (Anexos XVII y XVIII).

5.- Referencias

Alameda, J. R. (1996) *Incidencia de la vecindad léxica ortográfica en el reconocimiento visual de palabras*. Tesis doctoral no publicada. Universidad de Oviedo.

Alameda, J. R. & Cuetos, F. (1995). *Diccionario de frecuencias de las unidades lingüísticas del castellano*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.

Alameda, J. R. & Cuetos, F. (1996). Índices de frecuencia y vecindad ortográfica para un corpus de palabras de cuatro letras. *Revista Electrónica de Metodología Aplicada, 1*, 10-29. En http://www.psico.uniovi.es/REMA/

Alameda, J. R. y Cuetos (1997). Frecuencia y vecindad ortográfica: factores independientes o relacionados. *Psicológica 18(1)*, 10-29.

Alameda, J. R. y Cuetos (2000*). Incidencia de la vecindad ortográfica en el reconocimiento de palabras. Revista de Psicología General y Aplicada, 53(1)*, 85-107.

Alameda, J. R. & Cuetos, F. (2001). Índices de frecuencia y vecindad para palabras de cinco letras. *Revista Electrónica de Metodología Aplicada, 6,* 1-62. En http://www.psico.uniovi.es/REMA/

Carreiras, M., Álvarez, C. J., & de Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language, 32*, 766-780.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977) *Access to the internal lexicon*. En S. Dornic (Ed.), Attention and Performance VI. Nueva York, Academic Press.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1084-1096.

Cuetos, F., Dominguez, A., & de Vega, M. (1997). El efecto polisemia: Ahora lo ves otra vez. *Cognitiva, 9*, 175-194.

Ellis, A. W. & Morrison, C. M. (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 515-523.

Foster, K. I. & Chambers, S. M.(1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*, 627-635.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*, 228-244.

Grainger, J. & Seguí, J. (1990). Neighborhood frequency effects in visual word recognition: A comparison of lexical decision and masked identification latencies. *Perception and Psychophysics, 47*, 191-198.

Juilland, A. & Chang-Rodríguez, E. (1964). *Frequency dictionary of Spanish words*. La Haya: Mouton.

McClelland, J. & Rumelhart, D. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*, 375-407.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76*, 165-178.

Perea, M. (1993). Una base de palabras de cuatro letras: Índices de frecuencia, familiaridad y vecindad ortográfica. *Psicológica, 14*, 307-317.

Pérez, M. A., Campoy, G., & Navalón, C. (2001). Índice de estudios normativos en español. *Revista Electrónica de Metodología Aplicada, 6*, 85-105. En http://www.psico.uniovi.es/REMA/

Real Academia de la Lengua (1992). *Diccionario de la Lengua Española* (21ª edición). Madrid: Espasa Calpe.

Real Academia de la Lengua (1995). *Diccionario de la Lengua Española* (edición electrónica de la 21ª edición de 1992). Madrid: Espasa Calpe.

Real Academia de la Lengua (2001). *Diccionario de la Lengua Española* (22ª edición). Madrid: Espasa Calpe.

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9*, 487-494.

Sears, Ch.R., Hino, Y. y Lupker, S.J. (1995). Neighbourhood size and neighbourhood frecuency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21(4)*, 876-900.

Sears, Ch.R., Lupker, S.J. e Hino, Y. (1999). Ortographic Neighbourhood Effects in Perceptual Identification and Semantic Categorization Tasks: A Test of the Multiple Read-out Model. *Perception and Psichophisics, 61(8)*, 1537-1554.

Sebastián, N., Martí, M. A., Carreiras, M. F., & Cuetos, F. (2000). *LEXESP, léxico informatizado del Español*. Barcelona: Ediciones de la Universitat de Barcelona.

Seguí, J. y Grainger, J. (1992). Neighborhood frecuency and stimulus frecuency effects: Two different but related phenomena?. En J. Alegría, D. Hollander, J. Junça de Morais y M. Radeau (Eds). *Analytic approaches to human cognition*. Bruselas: Elsevier Science Publishers.