

FRECUENCIA Y PROCESAMIENTO LÉXICO

BENILDE GRAÑA LÓPEZ
Universidad de Oviedo

RESUMEN. *Este artículo es un estudio sobre el modo en que el nivel de frecuencia de las palabras complejas y de los morfemas que las constituyen incide sobre la organización del léxico mental (LM) de los hablantes y sobre la manera en que discurre el procesamiento de tales palabras, es decir, las operaciones que conducen a su comprensión/producción en la comunicación lingüística. Se describen los efectos (de orden tanto psicolingüístico como lingüístico) que se derivan de la frecuencia y se adopta un modelo explícito de procesamiento morfológico que hace posible el registro de la frecuencia en el LM a través del propio procesamiento. De este modo el procesamiento es agente y paciente de la frecuencia: agente en cuanto que hace posible la codificación de niveles de frecuencia en el léxico mental y paciente en cuanto que se ve condicionado por ella.*

PALABRAS CLAVE. *Procesamiento del lenguaje, morfología, frecuencia léxica, léxico mental.*

ABSTRACT. *This article is about the way the frequency level of complex words and of the morphemes that make them up affects the organization of the mental lexicon of speakers and the processing of such words (that is, the operations that lead to their comprehension/production in linguistic communication). Several frequency effects, both psycholinguistic and linguistic, are described, and an explicit model of morphological processing is defended in which processing itself makes it possible for frequency to be registered in the mental lexicon. Thus, processing is both the agent and the patient of frequency: it is the agent in that it determines the codification of frequency levels in the mental lexicon, and the patient in that it is constrained by frequency.*

KEYWORDS. *Language processing, morphology, word frequency, mental lexicon.*

1. INTRODUCCIÓN

La psicolingüística, que constituye un área de conocimiento comúnmente reconocida dentro del ámbito de la lingüística aplicada, cuenta entre sus mayores atractivos el hecho de que explora la relación entre lenguaje y cerebro. Esta relación se instancia en dos procesos bien diferenciados (aunque innegablemente relacionados, sobre todo en la edad infantil): la adquisición del lenguaje y el procesamiento del mismo, es decir, el mecanismo cognitivo de comprensión y producción del lenguaje que caracteriza

a los humanos. Este trabajo se centra en este segundo aspecto, y más precisamente en el procesamiento dentro del ámbito léxico-morfológico, o sea, el de las palabras complejas.

Dentro del terreno de la lingüística teórica, pocos dudarían que ha sido el paradigma generativista el que ha mostrado un interés más claro por la vertiente psicológica de los lenguajes naturales. Basta recordar la noción de Gramática Universal (Chomsky 1986), a la que se le supone un fundamento biológico e innato, o el debate en torno a si la Facultad del Lenguaje es autónoma con respecto al resto de las capacidades cognitivas o forma parte de un entramado cognitivo general (Escribano 1991: 212). No conviene olvidar, sin embargo, que ya desde sus orígenes el generativismo se ha ocupado de separar dos niveles en el estudio de los lenguajes naturales: el de la “competencia”, o conocimiento abstracto e internalizado del lenguaje, y el de la “actuación”, es decir, el uso, o ejecución externa de ese conocimiento en la comunicación verbal. En este sentido, las aportaciones de los generativistas se han limitado en general al ámbito de la competencia, si bien el deseo de reducir distancias entre ambos planos se deja ver en diversos trabajos (por ejemplo, Hall 1992 o Escribano 1991¹).

En el campo del léxico las cosas no han sido demasiado distintas, pero hay que reconocer que, si bien la mayoría de los estudios abordan cuestiones de competencia, se han ido acumulando en los últimos años numerosos trabajos sobre el procesamiento que se enmarcan dentro de la órbita del generativismo. Cualquier consideración sobre el procesamiento léxico ha de partir del presupuesto difícilmente rebatible de que las palabras, al contrario que la mayoría de los sintagmas y oraciones, pertenecen al banco de datos de la memoria, y ello determina que haya dos aspectos que resulta necesario investigar: primero, cómo están organizadas o almacenadas las palabras en ese banco de datos, y segundo, cómo se usan, o de manera más precisa, cómo se recuperan, en las tareas de comprensión y producción del habla. Pues bien, este trabajo analiza esos dos aspectos en relación con las palabras complejas, en particular, aquellas que llevan afijos derivativos y flexivos, y aborda detenidamente la incidencia del factor frecuencia sobre la organización y el procesamiento de tales palabras. Así, el apartado 2 se centra en la descripción de lo que se ha dado en llamar *frequency effects* (efectos de frecuencia), mientras que en el tercer apartado se reseñan brevemente los principales modelos de procesamiento que se han propuesto en relación con las palabras complejas y se defiende un modelo reciente (el de Schreuder y Baayen 1995), entre otras razones, porque permite acomodar de manera más convincente que otros los hechos de frecuencia. Casi todos los ejemplos que usaremos están tomados del inglés pues es esta la lengua más ampliamente estudiada en los trabajos sobre procesamiento léxico.

2. DESCRIPCIÓN DE HECHOS DE FRECUENCIA

En el ámbito de las palabras complejas el término *frecuencia* tiene dos acepciones. Lo que llamaremos frecuencia de uso o frecuencia textual (*token frequency*) se refiere al

grado de recurrencia o repetición en el uso tanto oral como escrito de una determinada forma; así, *curiosity* tiene mayor frecuencia de uso que *mendacity*. En cambio, hablaremos de frecuencia de tipo o frecuencia afijal (*type frequency*)² para designar el nivel de productividad de un afijo, es decir, la cantidad de palabras diferentes en que se repite el mismo afijo: no cabe duda de que *-ity*, por ejemplo, tiene mucha mayor frecuencia afijal que *-th*, que sólo aparece en un puñado de palabras (*length, warmth, depth...*). Que se trata de cuestiones independientes lo demuestra el hecho de que una misma palabra, como *warmth*, puede ser muy frecuente en el uso y representar, sin embargo, un patrón morfológico de baja frecuencia (la sufijación de *-th*). Naturalmente la frecuencia de uso es independiente de si la palabra tiene o no estructura interna; se aplica en consecuencia también a las palabras simples.

La frecuencia de uso se ha revelado como un condicionante de primer orden en el procesamiento de las palabras complejas: en el proceso de producción del habla se ha observado que el acceso a la palabra es más rápido y el nivel de errores más bajo cuanto mayor es el índice de frecuencia de la misma. La correlación entre alto índice de frecuencia y bajo nivel de error se describe, entre otros, en el trabajo de Stemmer y MacWhinney 1988 en relación con las formas de pasado y participio pasivo de los verbos ingleses³, y se desprende del análisis de los errores (tanto espontáneos como obtenidos experimentalmente) resultantes de la no utilización de la necesaria marca flexiva (se trata de lo que los autores (Stemmer y MacWhinney 1998:103) llaman *no marking errors*), como la producción de *that's what I need to do*, cuando la forma correcta del verbo sería *needed*.

Para Stemmer y MacWhinney el hecho de que los pasados y participios más frecuentes den lugar a menos errores que los de menor frecuencia se debe a que los primeros, pero no los segundos, se encuentran almacenados en el léxico mental (LM) y el hablante sólo tiene que recuperarlos como si se tratara de palabras simples, es decir, no necesita construirlos a partir de una base y un afijo, lo cual constituiría una operación más compleja y por tanto más duradera y susceptible de error. Parece claro, pues, que la frecuencia es un factor a tener en cuenta a la hora de diseñar modelos de organización y procesamiento léxico.

En el terreno de la derivación la incidencia de la frecuencia sobre el procesamiento se estudia en Burani *et al.* (1997). Estos autores llevan a cabo dos experimentos (con ejemplos del italiano) en los que usan pseudo-palabras (*pseudo-words*), o sea, palabras no existentes, en este caso compuestas de dos morfemas existentes, una raíz verbal y un sufijo derivativo⁴. El primero es un experimento de decisión léxica en el que los informantes tienen que decidir si esas formas son o no palabras de su lengua; en el segundo (al que los autores denominan *naming experiment*, o experimento de nombramiento) se pide que los informantes lean en voz alta secuencias de ese tipo (para más detalles en relación con la mecánica y el diseño de los experimentos *viz.* Burani *et al.* 1997). Al tratarse de palabras que no existen resulta imposible que se activen unidades equivalentes a palabras en el LM de los sujetos, pero sí se pueden activar, en cambio, unidades equivalentes a morfemas dado que las palabras están hechas de

morfemas reales. Lo que descubren los autores es que el valor cuantitativo del sufijo (entendido como la combinación de su frecuencia afijal y la frecuencia de uso de las palabras que lo llevan, en el léxico real del italiano) influye en el procesamiento del siguiente modo: a mayor valor (es decir, mayor frecuencia global), mayor vacilación y lentitud se produce en la tarea de decisión léxica, y mayor rapidez y exactitud en la de lectura. Abordaremos la interpretación de estos resultados en el apartado 3.4.

Además de condicionar el almacenamiento y procesamiento léxico, a la frecuencia de uso también se la ha visto como responsable de hechos lingüísticos que tienen en último término su fundamento en el LM. Bybee (1988, 1995) desarrolla toda una noción de fuerza léxica (*lexical strength*) basada en la frecuencia: la fuerza léxica acumulada por una palabra está en función de la frecuencia de su procesamiento, de modo que cada vez que se procesa la palabra, se proyecta sobre una representación mental ya existente y cuantas más veces se procese más se refuerza esa representación (1995: 232). Según esta autora, la fuerza léxica explica hechos evolutivos, así como determinados patrones de organización de palabras morfológicamente relacionadas.

En cuanto a los primeros, sostiene Bybee que la perseverancia en la lengua de formas morfológicamente irregulares, como el patrón supletivo de los verbos fuertes ingleses (por ejemplo, *break, broke, broken*), está sustentada por la alta frecuencia de uso de tales unidades, de ahí que en el caso de las formas infrecuentes se haya dado una clara tendencia a la regularización motivada porque estas no reciben suficiente refuerzo a través del uso⁵.

La fuerza léxica desempeña también un papel muy importante en el establecimiento de lo que Bybee (1988: 132) denomina *the basic/derived relation* (la relación básico/derivado). Esta se refiere a que los hablantes construyen conexiones unidireccionales entre formas relacionadas (como en el caso de *white* → *whiteness* o *weep* → *wept*), de modo que lo que determina qué unidad se constituye como básica no es sólo la mayor simplicidad formal y semántica, sino que la fuerza léxica juega un papel crucial. Así, en los casos de regularización la forma de la raíz que permanece y que por tanto se adopta como básica, suele ser la más fuerte; ese es el caso de pasados regularizados como *weaped* y *leaped* (Bybee 1995: 236), donde es la forma de presente (la de mayor fuerza y simplicidad semántica) la que sobrevive en la raíz⁶. En algunos estudios la fuerza se ha revelado como un factor más decisivo en la regularización que la simplicidad semántica⁷. Además de la regularización, también el aprendizaje léxico está condicionado por la relación básico/derivado: según la autora (1988: 133), el aprendizaje funciona en el campo léxico como en otros ámbitos de conocimiento, es decir, las palabras se adquieren mediante un proceso de integración con otras unidades ya almacenadas; pues bien, en este proceso las palabras más frecuentes, y por tanto más fuertes, se adquieren antes que las débiles, las cuales se aprenden y almacenan en términos de las primeras.

En conclusión, la frecuencia, cuya representación en el LM se ha expresado a través de metáforas como fuerza léxica, latencia, o nivel de activación, se ha revelado

como un factor de gran peso en el desarrollo del procesamiento y en la explicación de determinados hechos lingüísticos relativos a la organización y evolución léxica. Su incidencia es tal que determina que ciertas unidades ganen un nivel de autonomía y activación que las convierte en líderes del procesamiento, la organización y la evolución.

3. MODELOS DE PROCESAMIENTO LÉXICO Y HECHOS DE FRECUENCIA

Los hechos descritos arriba apoyan la construcción de modelos de procesamiento léxico que incorporen en una arquitectura explícita el valor del factor frecuencia, de modo que la ubicación o el reflejo de los efectos de frecuencia resulte del propio procesamiento. Eso es lo que ocurre, en definitiva, en el léxico mental (LM) de los humanos, y los modelos de procesamiento deben, tan explícitamente como sea posible, simular el funcionamiento de aquel.

3.1. *Teorías de ruta única y teorías de doble ruta*

Hace años se planteaba la cuestión del procesamiento de las palabras complejas en términos globales y dicotómicos: unos autores defendían el procesamiento por la ruta de la palabra, como si se tratara de palabras simples, y otros se inclinaban por la ruta del morfema (lo que podemos denominar procesamiento amorfemático y morfemático, respectivamente)⁸.

En la actualidad las teorías de ruta única han sido sustituidas por modelos de doble ruta que admiten una combinación de procedimientos. Existen, no obstante, dos maneras de entender esa combinación. Para unos (Pinker 1991) el acceso a cada palabra compleja se ejecuta a través de una única vía, dependiendo de si la palabra es irregular y está por tanto almacenada en la memoria (vía amorfemática), o regular (Pinker asume que las formas regulares no están almacenadas y por tanto sólo se accede a ellas por la vía de los morfemas, que sí lo están). Pero la mayoría (Clahsen 1997; Burani *et al.* 1997; Schreuder y Baayen 1995; Chialant y Caramazza 1995, entre otros) adopta un enfoque más interactivo en el que las dos modalidades de acceso actúan en paralelo y contribuyen conjuntamente al procesamiento de la mayoría de las palabras complejas. Defenderemos más adelante que un modelo de este tipo, en concreto el de Schreuder y Baayen 1995, es capaz de acomodar los hechos de frecuencia, pero antes vamos a referirnos brevemente a la propuesta de Bybee 1988 y 1995 (*viz.* apartado 2), cuyas interesantes intuiciones no pueden ser ignoradas.

3.2. *El modelo conexionista de Bybee*

El de Bybee no es en realidad un modelo de procesamiento, sino de organización del LM, pero merece la pena detenerse en él porque, según veremos en breve,

organización y procesamiento mantienen una relación de recíproco condicionamiento. Para esta autora el LM es un almacén de representaciones (palabras simples y complejas) que no se relacionan mediante operaciones de computación o reglas: lo único que hay son patrones o esquemas estáticos que emergen de las relaciones entre las palabras. Se trata de un modelo tipo red donde cada nodo representa una palabra (asociada con un determinado nivel de fuerza léxica, -viz. más arriba) y todas sus propiedades, es decir, no hay niveles distintos para forma y significado. Las múltiples relaciones entre las palabras se expresan mediante líneas que conectan los nodos, con independencia de si se trata de relaciones formales, semánticas o de ambos tipos. No existen nodos independientes para los morfemas, cuya realidad se limita al entramado de relaciones fonológico-semánticas. Se trata, en fin, de un modelo que reconoce dos mecanismos en el LM: “the ability to form networks among stored elements of Knowledge, and the ability to register the frequency of individual items or patterns” (Bybee 1988: 125).

3.3. Schreuder y Baayen 1995: un modelo explícito de procesamiento

Mientras que en Bybee conexionismo y fuerza léxica son dos propiedades que en gran medida se le suponen al LM, lo que intentaremos mostrar a partir de aquí es que lo interesante del tratamiento de Schreuder y Baayen (S&B) es que permite, en el terreno de las palabras complejas, derivar ambas propiedades del procesamiento y llegar así a una explicación funcional de las mismas (en el sentido de que ambas propiedades resultan de la manera en que funciona el procesamiento). Esto se debe a que proponen un modelo explícito de procesamiento que refleja, a nuestro entender, el condicionamiento mutuo que existe entre organización (entendida como el conjunto de relaciones y unidades que existen en el LM) y procesamiento (el conjunto de operaciones conducentes a la comprensión/producción de las formas complejas): así, lo que existe en el LM determina la manera en que discurre el procesamiento, y a la inversa, este contribuye a reforzar determinadas relaciones y unidades en detrimento de otras.

S&B diseñan un modelo para la comprensión oral que capta el desarrollo temporal del proceso de comprensión: este se inicia con el acceso a un nivel de representación inicial y culmina con la traslación de la forma “interpretada” a los sistemas de procesamiento supraléxicos. La hipótesis de partida es que las unidades del LM están asociadas con distintos niveles de representación⁹ que se activan

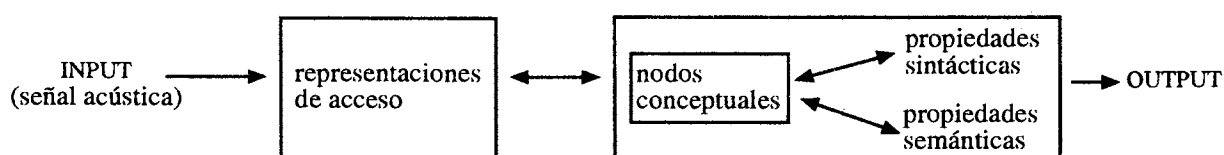


Fig. 1. Modelo de procesamiento de Schreuder y Baayen 1995. Versión abreviada.

sucesivamente¹⁰, de modo que el procesamiento morfológico consiste en computaciones que operan sobre esas representaciones y cuya función primordial es el cálculo del significado. La figura 1 es una versión simplificada de este modelo:

En el LM, ante el estímulo de la señal acústica, se activan representaciones de acceso (representaciones A), que son representaciones formales de carácter fonológico correspondientes tanto a palabras como a morfemas, que a su vez activan nodos conceptuales (nodos C) conectados con diversas propiedades sintáctico-semánticas (propiedades S). Los nodos C se conciben como “estaciones de enlace” que median entre las representaciones A y sus propiedades S y reflejan la categorización conceptual de un subconjunto de rasgos sintáctico-semánticos efectuada por las palabras y los morfemas. Una propiedad fundamental de este modelo (típica, por lo demás, de los modelos de activación sucesiva) es que, además de la corriente de activación que parte de la señal acústica y avanza de izquierda a derecha en la Fig. 1, existe una segunda corriente de retroactivación (lo que S&B (1995: 140) llaman *activation feedback*) que avanza en el sentido contrario -las flechas de doble punta en la Fig. 1 representan el doble sentido de la activación.

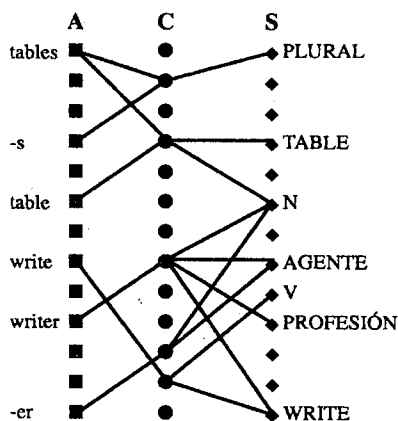


Fig. 2. *Procesamiento de tables y writer.*
A = representaciones A; C = nodos C; S = propiedades S.

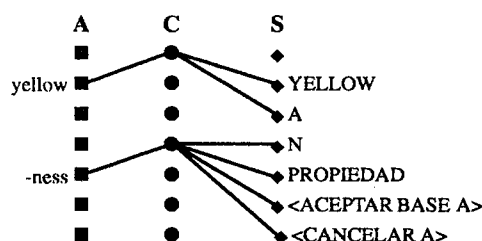


Fig. 3. *Procesamiento de yellowness. Estadio 1.*

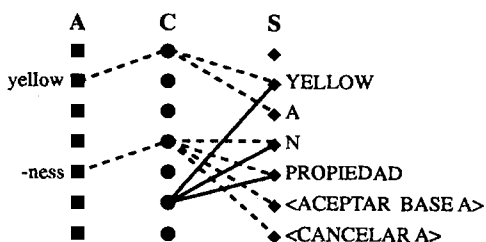


Fig. 4. *Procesamiento de yellowness. Estadio 2.*

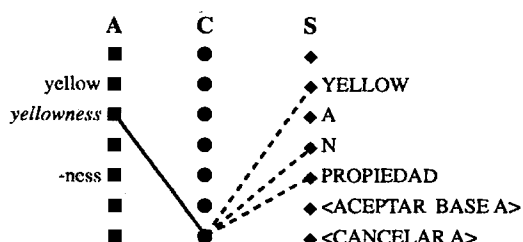


Fig. 5. *Procesamiento de yellowness. Estadio 3.*

Antes de considerar la cuestión de la frecuencia, vamos a ver cómo se ejecuta en la práctica el procesamiento de algunas palabras complejas: *tables*, *writer* y *yellowness*¹¹.

La Fig. 2 expresa el procesamiento de *tables* y *writer* en un LM que dispone de representaciones A para ambas formas y también para los morfemas que las constituyen (*table*, *-s*, *write* y *-er*) (viz. el caso de *yellowness* en relación con el establecimiento de una representación A para un neologismo). Esto implica que puede haber contribución de las dos rutas, la de la palabra y la del morfema, tal como se asume en (2), dependiendo, la contribución de cada ruta, de factores que trataremos en 3.4. Una vez que la señal acústica desencadena la activación de representaciones formales en A, se inicia una corriente de activación que conduce de A a C y de C a S. Cuando para una palabra dada se han activado varios nodos en S correspondientes a dos nodos conceptuales, la computación de la forma compleja se reparte, según S&B (1995: 134, 139), en dos operaciones: una de legitimación (*licensing*), mediante la cual el LM comprueba la compatibilidad de rasgos categoriales (rasgos S) entre base y afijo, y otra de composición (*composition*) que calcula el significado del todo a partir de los rasgos S que expresan el significado de las partes. El *output* de estas operaciones (una forma compleja perfectamente integrada e interpretada) pasa a los niveles supraléxicos de procesamiento (es decir, procesamiento frástico, oracional, etc.). En el modelo de S&B (1995: 137), la consecuencia de tales operaciones para el propio LM es distinta según se trate de flexión o de derivación: tras la composición, no se establecen nodos C para las palabras con flexión (lo cual capta la intuición de que estas no se corresponden con categorías conceptuales, -viz. la ausencia de nodo C para *tables* en (2)), pero sí se establecen para los derivados (con independencia de que estén lexicalizados -el caso de *writer-*, o sean transparentes -viz. explicación para *yellowness*)¹². Esto implica que para las primeras sólo puede haber una representación unitaria en el nivel de acceso, a partir de ahí el procesamiento sigue la vía del morfema, mientras que para los derivados también hay representación unitaria en el nivel conceptual, de modo que todo el procesamiento puede seguir la vía de la palabra, con independencia de que la otra ruta pueda contribuir en mayor o menor medida (viz. más abajo). En el caso de los derivados lexicalizados la participación de la ruta de la palabra es obligada dado que sólo ella puede conducir a los rasgos idiosincrásicos de S, que no se derivan de los morfemas constituyentes (viz. en (2) el rasgo “profesión” asociado con *writer*).

En contraste con *tables* y *writer*, las Figs. 3, 4 y 5 dan cuenta del procesamiento de un neologismo: en este caso un derivado que no existe *a priori* en el LM. Naturalmente ese procesamiento es posible sólo a través de la vía del morfema, lo que supone que ha de haber en el LM representaciones A, C y S para *yellow* y *-ness*. La Fig. 3 refleja la activación sucesiva de las mismas en lo que constituye la fase inicial del procesamiento, tras la cual se producen la legitimación y la composición (viz. más arriba) de la nueva palabra. A su vez estas dan lugar a un nuevo estadio (Fig. 4) que consiste en la creación de un nodo C para la nueva forma con la consiguiente

difuminación de la corriente originaria de activación que circula de izquierda a derecha (expresada ahora con líneas discontinuas). Obsérvese que esta operación de conceptualización, expresada con trazos gruesos en (4), se ejecuta a través del mecanismo de retroactivación (*viz.* más arriba), esto es, de derecha a izquierda. También se sigue este camino de retorno en la Fig. 5, que refleja el establecimiento de una representación A para el derivado: “the newly created concept node is also linked with the memory trace of its full form in the access system” (S&B: 140). Nótese pues que, si bien el acceso originario es a través de la base y el afijo, se asume que la forma compleja ha dejado su huella en la memoria y ello posibilita el vínculo entre el concepto y la representación formal. La pervivencia de la nueva representación en el LM está en función del refuerzo producido por nuevas exposiciones a la misma: de no producirse este, las representaciones asociadas con *yellowness* irán borrándose progresivamente.

Un rasgo importante del mecanismo de S&B es que la retroactivación no sólo se aplica a la creación de nodos C y su vinculación con representaciones A, sino que es un mecanismo interno de autoimplementación que también actúa cada vez que se procesa una forma ya presente en el LM, de modo que ha de entenderse que las líneas que unen los distintos nodos en las figs. de arriba representan una corriente que circula en los dos sentidos. De este modo, el nivel de fuerza léxica de una palabra o morfema está en función de la frecuencia con que el LM resulta expuesto a ese elemento (a nivel externo) y también del *feedback* interno que procura el sistema.

Tras esta breve descripción del mecanismo de S&B, no resulta difícil percibir que el conexionismo, que Bybee expresa mediante una metáfora espacial estática (el listado de familias morfológicas cuyos miembros se vinculan mediante conexiones estáticas), se deriva en este modelo de la información fonológica y semántico-sintáctica compartida por las palabras complejas que se activa en el LM cada vez que se procesan palabras relacionadas (así, la representación A y las propiedades S asociadas con *-ness* se activan cada vez que se procesa una palabra con ese sufijo, y lo mismo ocurre respecto de una raíz como *write*, etc.), de modo tal que el conexionismo resulta del procesamiento; es este, de hecho, el que permite establecer las conexiones pertinentes. Una consecuencia de este enfoque es que las propiedades S compartidas no se repiten para las distintas formas, lo que contribuye a dotar al LM de un deseable nivel de parsimonia.

3.4. Los hechos de frecuencia en el modelo de Schreuder y Baayen

Veamos a continuación cómo el modelo de S&B da cuenta de los hechos de frecuencia descritos en el apartado 2. El menor nivel de error y de demora en el procesamiento de determinadas formas flexivas, la persistencia de formas irregulares, y la relación básico-derivado se captan en este modelo a través del alto nivel de activación o fuerza léxica de las representaciones A correspondientes (recuérdese que esos tres fenómenos se refieren a las palabras con flexión, y estas no tienen nodos C

independientes en S&B), que se acumula debido a la alta frecuencia de uso de esas formas, y en consecuencia, a la frecuencia con que son procesadas. En relación con los errores y la demora, hay que recordar que los resultados recogidos en Stemberger y MacWhinney 1988 se refieren a la producción, y no a la comprensión, de formas flexivas, mientras que el mecanismo de S&B está diseñado para la comprensión; sin embargo, dada su versatilidad, creemos que este mecanismo se puede concebir también como un modelo de producción, en cuyo caso el proceso va, no del sonido al significado, sino del significado al sonido, de modo que las representaciones fonológicas en lugar de ser de acceso son de salida, y los escasos errores y la mayor rapidez en la producción de las formas más frecuentes son fruto de que para estas la codificación fonológica es más directa y simple que para las de baja frecuencia porque existe en el LM una representación A precomputada y muy reforzada¹³.

La persistencia de formas irregulares y la relación básico/derivado no son fenómenos de procesamiento sino de organización y contenidos del LM. No obstante, dado que organización y contenidos están en función de lo que se procesa, estos fenómenos se explican igual que el anterior: debido al continuo refuerzo que supone el procesamiento frecuente, las representaciones A de determinadas formas irregulares (como los pasados y participios de ciertos verbos fuertes ingleses) acumulan un nivel de fuerza léxica en el LM que garantiza su supervivencia, mientras que las de otras formas (en el caso de la relación básico/derivado) ganan suficiente autonomía en relación con otras palabras morfológicamente emparentadas con ellas como para convertirse en básicas en procesos de regularización y adquisición (aunque a veces no sean básicas formal y/o semánticamente, como vimos en el apartado 2).

Otro de los hechos de frecuencia descritos en 2 es el que se refiere al valor cuantitativo de los sufijos derivativos y su influencia en el procesamiento. Recordemos que en Burani *et al.* (1997) ese valor es la suma de la frecuencia de uso y la frecuencia afijal (*viz.* comienzo del apartado 2). Pues bien, aunque estos autores asumen que a mayor valor, mayor *saliency* (es decir, relevancia o activación) del afijo en el LM, no desarrollan un mecanismo explícito donde tal activación resulte del procesamiento, mientras que el modelo de S&B permite derivar el grado de activación en un momento dado del procesamiento, y en concreto, de la manera en que este se ve condicionado por el valor cuantitativo. Veamos cómo.

El asentamiento de un afijo en el LM se produce a través de lo que S&B (1995: 144) denominan *affix discovery procedure* (procedimiento de descubrimiento afijal), mediante el cual se establece un nodo C para aquellos afijos que presentan un nivel suficiente de frecuencia y de transparencia afijal: tras procesar un determinado número de palabras donde se repite el mismo afijo (este procesamiento, antes del descubrimiento afijal, es como el de las palabras simples, es decir, sigue la ruta de la palabra), el LM llega a descubrir la coactivación de un conjunto de propiedades S y una determinada secuencia fónica, de modo que se crea un nodo C por la vía del retorno o retroactivación (partiendo de las propiedades S)¹⁴ que se enlaza posteriormente con una representación A en el nivel de acceso.

Pues bien, los mismos factores que contribuyen al descubrimiento afijal contribuyen al posterior reforzamiento de las incipientes representaciones descritas arriba; en palabras de S&B (1995: 142): “the activation level of the concept node of an affix is a function of the number of semantically transparent formations in which that affix occurs and of the frequencies of those formations”. Nótese que la transparencia semántica es un factor muy relevante pues garantiza la activación del mismo subconjunto de rasgos S para las palabras que comparten un afijo dado y es en gran medida la repetición de esta activación lo que dota de fuerza a un afijo en el LM. En este sentido, tanto S&B como Burani *et al.* asumen que el grado de activación del afijo se beneficia tanto de su frecuencia afijal como de la frecuencia de uso de las palabras que lo llevan. Hemos visto, a propósito del estudio de Stemberger y MacWhinney, que las altas frecuencias de uso tienden a favorecer el desarrollo de representaciones completas para las palabras complejas en el LM y a su vez estas representaciones favorecen el procesamiento de esas formas por la ruta de la palabra. La pregunta que surge entonces es si el procesamiento de las palabras complejas muy frecuentes puede incidir en alguna medida en el fortalecimiento de sus afijos. La respuesta en el marco de S&B es positiva: aun en el supuesto de que el procesamiento se hiciera únicamente por la ruta de la palabra, al llegar al nivel S se activarían, entre otros, un subconjunto de rasgos semánticos que, además de estar vinculados al nodo C de la forma compleja, están también vinculados al nodo C de un afijo (*viz.* Figs. 2-5), de modo que, gracias a la corriente de retroactivación, el nodo C y la representación A del afijo obtendrían algún tipo de refuerzo. De todos modos, dado que S&B defienden la actuación paralela de las dos rutas (la de la palabra y la del morfema) siempre y cuando haya representaciones para el todo y para las partes en el LM, ni siquiera en el caso de las palabras muy frecuentes hay razón para descartar la participación de la vía del morfema, y si esta participa las representaciones del afijo se refuerzan directamente en el viaje de ida del procesamiento (de izquierda a derecha), a la vez que siempre reciben una porción extra de refuerzo de esa corriente de retorno que S&B asumen que actúa de manera automática (bajo la condición de la transparencia semántica). En cuanto a la frecuencia afijal, ni que decir tiene que el procesamiento de muchas palabras que comparten el mismo afijo refuerza ineludiblemente las representaciones para el mismo en el LM a través de la continua activación de las mismas. En definitiva, pues, el mecanismo de S&B da cuenta de manera explícita, de cómo el nivel de activación de un afijo es mayor cuanto más altas sean su frecuencia de uso y su frecuencia afijal (es decir, cuanto mayor sea su valor cuantitativo).

En cuanto a los resultados de los experimentos de Burani *et al.* (1997), la interpretación más plausible parece la siguiente: la mayor demora en la tarea de decisión léxica (la de descarte de las pseudo-palabras) con los sufijos de mayor valor cuantitativo se debe a que el alto grado de activación de que gozan en el LM hace que este emplee más tiempo en asegurarse de la incompatibilidad entre base y sufijo antes de rechazar la combinación; con los de escaso valor, en cambio, la decisión sobre el estatus de no palabra es más rápida porque no se activa ninguna unidad en el léxico mental (o no se activa con suficiente fuerza) y por eso mismo tampoco hay necesidad de comprobar

compatibilidades. Por contra, resulta lógico que la tarea de lectura sea más rápida con los sufijos muy frecuentes porque para estos, y no para los otros, existe en el léxico mental una representación fónica precomputada que acelera la lectura¹⁵. En suma, parece claro que el valor cuantitativo no es sólo un parámetro externo sino que tiene su reflejo en el léxico mental y en consecuencia incide sobre el procesamiento.

Volviendo a la hipótesis de la doble ruta, hay que señalar que en el modelo de S&B es posible predecir la contribución relativa de cada ruta para una palabra compleja dada a partir del análisis de su frecuencia y su transparencia: así, la ruta morfológica tendrá mayor protagonismo cuanto mayor sea el grado de opacidad y de frecuencia de uso de la palabra, en tanto que la vía morfológica contribuirá en mayor grado que la otra en el procesamiento de aquellas formas que presenten un nivel alto de transparencia y una baja frecuencia de uso. Además, una ventaja importante frente a otros modelos, donde las dos rutas se desarrollan como dos caminos separados, es que en el de S&B (1995: 151) “the two access mechanisms interactively converge on the desired meaning representations”; esta propiedad, que se puede observar en la Fig. 2, enlaza con la no repetición de idénticas propiedades S que vimos más arriba, y demuestra el espíritu interactivo que caracteriza el modelo de S&B.

4. CONCLUSIÓN

Hemos defendido que es posible dar una explicación fundamentada a los fenómenos de frecuencia relacionados con las palabras complejas si se adopta un modelo de procesamiento morfológico donde la frecuencia de palabras y morfemas queda registrada en el léxico mental como consecuencia del propio procesamiento: cuanto más frecuente sea una unidad en la lengua más posibilidades hay de que un léxico mental concreto se vea repetidamente expuesto a ella, y de que por tanto llegue a adquirir a través del repetido procesamiento un nivel de activación (o fuerza léxica, como diría Bybee) más alto que las unidades de menor frecuencia. Cuanto más alto sea ese nivel, la producción/comprensión de esa unidad será más rápida y eficiente; tendrá más posibilidades de sobrevivir aunque sea irregular; se convertirá en básica (respecto a otras unidades relacionadas) en procesos de adquisición o regularización; el léxico mental tardará más en rechazar combinaciones ilegítimas donde participe esa unidad y tardará menos en leer combinaciones de ese tipo.

El modelo diseñado por Schreuder y Baayen (1995) se puede definir como un modelo de estrategias combinadas cuyo alto nivel de implementación interna lo hace adecuado para dar cabida a la variedad de posibilidades que presentan las palabras complejas. Se trata de un modelo que combina dos rutas paralelas en el acceso a las palabras complejas (la de la palabra y la del morfema) a la vez que distingue varios niveles de representación para las unidades del léxico mental; además combina la activación sucesiva con la retroactivación, dotando al sistema de un mecanismo interno de autoaprovechamiento; acomoda igualmente el recíproco condicionamiento que se da entre organización y procesamiento. Y todo ello con una considerable economía de medios.

NOTAS

1. Hall (1992: 26 y 27) defiende modelos de explicación en lingüística que no descansen únicamente en los niveles de generalización y predicción de los principios chomskyanos, sino que tomen en cuenta también factores funcionales, entre otros los derivados de los mecanismos de aprendizaje y procesamiento del lenguaje. Por su parte, Escribano (1991: 205) argumenta que “una teoría que no permita la construcción de modelos eficientes de procesamiento no es una buena teoría gramatical”.
2. Para un uso más general (es decir, no limitado al ámbito de las palabras complejas) de estas dos nociones de frecuencia, véase Lyons (1977: 13 y ss.).
3. Para establecer esta correlación Stemberger y MacWhinney utilizan los índices de frecuencia (calculados por millón de palabras de texto impreso) que se ofrecen en Francis y Kucera 1982.
4. Burani *et al.* (1997: 61) usan combinaciones de raíz verbal más sufijo que requiere raíz nominal o adjetival, de modo que las pseudo-palabras vienen dadas por la violación de las restricciones de selección entre base y sufijo. Sus ejemplos se reparten en dos grupos que representan dos extremos de frecuencia (1997: 70): en uno se incluyen sufijos de alta frecuencia (con ejemplos como *vendario, rendario, dividario...*; *coprezza, cedezza, prendezza...*, etc.), y en el otro sufijos de baja frecuencia (*ubbidardo, spegnardo, stupardo...*; *rodigia, sentigia, splendigia...*, etc.). La frecuencia de las raíces verbales se mantiene constante en los dos grupos, de modo que este factor no puede provocar diferencias en el procesamiento de ambos grupos.
5. Bybee (1988: 132) señala que la frecuencia media de las formas de pasado de aquellos verbos fuertes del Inglés Antiguo que han permanecido fuertes supera en 20 veces la de aquellas que se han regularizado.
6. Como ya hemos explicado, en el enfoque de Bybee es la escasa fuerza léxica de algunas formas irregulares (en este caso *wept* y *leapt*) el factor que de hecho desencadena la regularización. En contraposición, Bybee apunta que la alta frecuencia (y por lo tanto fuerza) de los pasados de verbos como *sleep* o *keep* previene la regularización de los mismos (**sleaped, *keeped*).
7. En este sentido, Tiersma (1982) presenta una serie de ejemplos de nombres del frisio en los que la forma semánticamente marcada, la del plural, es la que sirve de base a la regularización, de manera que los nuevos plurales regulares están compuestos por la antigua forma irregular del plural más la terminación regular del singular. Lo relevante es que se trata de nombres en los que el plural es, por razones pragmáticas, más frecuente que el singular (designan objetos que suelen aparecer en pares o grupos: ojos, dientes, medias...), y ha desarrollado por ello un mayor nivel de fuerza léxica.
8. Katamba (1994: 229 y ss.) ofrece un resumen de este debate, en el que la vía de la palabra fue defendida entre otros por Butterworth 1983 a través de su *Full Listing Hypothesis* (hipótesis del almacenamiento pleno). Taft (1991) aboga por el procesamiento morfemático, apoyado por el propio Katamba, quien argumenta que la otra ruta requiere el almacenamiento de todas las formas complejas, algo improbable en lenguas aglutinantes con una morfología tan rica como por ejemplo el turco.
9. Chialant y Caramazza (1995), cuyo modelo es, como el que nos ocupa, de doble ruta, distinguen dos niveles en el procesamiento de las formas complejas: unidades de acceso y representación léxica. Sin embargo, la vaguedad de su propuesta no permite comprender la diferencia entre ambos, porque ¿qué pueden ser las unidades de acceso sino algún tipo de representación léxica? Por lo demás, su modelo está muy poco implementado a nivel de ejecución: no presenta desarrollo temporal, a pesar de que el procesamiento es un evento que transcurre en el tiempo.
10. Nótese pues que el modelo de S&B es un *spreading activation model*, o sea, un modelo de activación sucesiva. Véase, en este sentido, Katamba (1994: 255 y ss.) y Aitchison (1994: 206 y ss.).
11. Nótese que en las figs. 2, 3, 4 y 5 las propiedades S se refieren a categoría -N(ombre), A(djetivo) y V(erbo); subcategorización –por ejemplo, <aceptar base A>; y también al conjunto de rasgos puramente semánticos asociados con cada nodo C. En general, estos últimos no los hemos representado en detalle, sino sólo a través de un nodo S seguido de la palabra en mayúsculas. S&B no elaboran tampoco este aspecto.

12. Schreuder y Baayen (1995: 137) entienden que el proceso de composición en el caso de la flexión es trivial, consistiendo simplemente en la unión de dos conjuntos de rasgos activados en S. En cambio, en la derivación la computación es más compleja que una simple unión y eso conlleva la creación de nodos C, con independencia de que estos sean totalmente necesarios para dar cuenta de las propiedades irregulares que se asocian con tantos derivados.
13. Como hemos señalado, es posible interpretar el modelo de S&B como un modelo de producción, de modo que para las formas flexivas esta comenzaría con la activación de los nodos C de la base y el afijo, que llevaría a la activación de las propiedades S correspondientes, tras lo cual se produciría la integración semántica (consistente, como hemos visto, en la simple unión de las propiedades S de los dos nodos C). Seguidamente, el procesamiento se encaminaría a la codificación fonológica, con lo cual el sentido a seguir es ahora de derecha a izquierda (esto no supone obstáculo alguno, porque como hemos explicado, este modelo permite que la activación discorra en ambos sentidos). Recuérdese que las formas flexivas no dan lugar a nodos C, pero sí pueden estar asociadas con representaciones fonológicas, de modo que ese camino hacia la izquierda ha de ser por la ruta de los morfemas hasta llegar al nivel fonológico, donde las formas más frecuentes dejan su huella a través de una representación unitaria. Esto significa que en esos casos no es necesaria una operación de integración fonológica, puesto que dos nodos C dados llevan a una representación ya precomputada; de ahí la rapidez y el menor riesgo de error.
14. Nótese que es natural que ese nodo C se establezca por la vía del retorno puesto que el descubrimiento afijal se produce al final del procesamiento: tras la repetida activación en S de las propiedades de un afijo, se produce la categorización de las mismas en una única unidad y eso se representa mediante una corriente de retroactivación que parte de las propiedades S pertinentes, que resultan conectadas a un nodo C. *viz.* S&B (1995: 144-145) en relación con la creación de un nodo C para el morfema de plural de los nombres holandeses en un LM joven.
15. Nótese que la rápida lectura de las palabras con sufijos muy frecuentes en Burani *et al.* es equiparable a la producción rápida y sin errores de las formas flexivas muy frecuentes en el estudio de Stemberger y MacWhinney: el elevado nivel de activación de las respectivas representaciones favorece un procesamiento rápido y libre de error.

BIBLIOGRAFÍA

- Aitchison, J. 1994 (1987). *Words in the Mind*. Oxford: Blackwell.
- Burani, C., F. Dovetto, A. Thornton y A. Laudanna. 1997. "Accessing and Naming Suffixed Pseudo-words". *Yearbook of Morphology 1996*. Eds. G. Booij y J. Van Marle. Dordrecht: Kluwer. 55-72.
- Butterworth, B. L. 1983. "Lexical Representation". *Language Production, Vol.2: Development, Writing and Other Language Processes*. Ed. B. L. Butterworth. Londres: Academic Press. 257-294.
- Bybee, J. 1988. "Morphology as Lexical Organization". *Theoretical Morphology*. Eds. M. Hammond y M. Noonan. San Diego: Academic Press. 119-141.
- Bybee, J. 1995. "Diachronic and Typological Properties of Morphology and their Implications for Representation". *Morphological Aspects of Language Processing*. Ed. L. B. Feldman. Hillsdale (NJ): Lawrence Erlbaum. 225-246.

- Chialant, D. y A. Caramazza. 1995. "Where is Morphology and How is it Processed? The Case of Written Word Recognition". *Morphological Aspects of Language Processing*. Ed. L. B. Feldman. Hillsdale (NJ): Lawrence Erlbaum. 55-76.
- Chomsky, N. 1986. *Knowledge of Language*. Nueva York: Praeger.
- Clahsen, H. 1997. "The Representation of Participles in the German Mental Lexicon. Evidence for the Dual Mechanism Model". *Yearbook of Morphology 1996*. Eds. G. Booij y J. Van Marle. Dordrecht: Kluwer. 73-95.
- Escribano, J. L. G. 1991. *Una teoría de la oración*. Oviedo: Universidad de Oviedo.
- Francis, W. N. y H. Kucera 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Hall, J. H. 1992. *Morphology and Mind*. Londres: Routledge.
- Katamba, F. 1994. *English Words*. Londres: Routledge.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Pinker, S. 1991. "Rules of Language". *Science* 253: 530-535.
- Schreuder, R. y R. H. Baayen, 1995. "Modeling Morphological Processing". *Morphological Aspects of Language Processing*. Ed. L. B. Feldman. Hillsdale (NJ): Lawrence Erlbaum. 131-154.
- Stemberger, J. P. y B. MacWhinney. 1988. "Are Inflected Forms Stored in the Lexicon?". *Theoretical Morphology*. Eds. M. Hammond y M. Noonan. San Diego: Academic Press. 101-116.
- Taft, M. 1991. *Reading and the Mental Lexicon*. Hillsdale (NJ): Lawrence Erlbaum.
- Tiersma, P. 1982. "Local and General Markedness". *Language* 58: 832-849.