

ANÁLISIS DE LAS RELACIONES SEMÁNTICAS A TRAVÉS DE UNA TAREA DE LIBRE ASOCIACIÓN EN ESPAÑOL CON MAPAS AUTO-ORGANIZADOS *

ANALYSIS OF SEMANTIC RELATIONSHIPS FROM A
FREE-ASSOCIATION TASK IN SPANISH WITH
SELF-ORGANIZING MAPS

NATIVIDAD HERNÁNDEZ MUÑOZ
Universidad de Salamanca. Salamanca, España
natih@usal.es

MIGUEL LÓPEZ GARCÍA
Universidad Miguel Hernández. Alicante, España
m.lopezg@umh.es

RESUMEN

En Psicolingüística y Lingüística computacional las pruebas de libre asociación de palabras han servido tradicionalmente como fuente de datos para desarrollar redes conceptuales que, de diferente forma, simulan el espacio semántico mental de los hablantes de una lengua. La modelación de las redes léxicas se articula habitualmente a través de valores como el número de asociados de cada palabra, la fuerza asociativa y otra serie de factores cuantitativos de diversa índole. Menos abundante es, sin embargo, el número de estudios que cuantifican el entorno de una palabra a través del tipo de relaciones asociativas. Este trabajo realiza un análisis exploratorio de las relaciones semántico-asociativas de 150 palabras en español partiendo de una prueba de libre asociación discreta, con el objetivo de proponer una clasificación de las palabras por medio de las técnicas de los mapas auto-organizados (SOM), con el objetivo de que tanto las variables resultantes basadas en los tipos de asociados como la clasificación de palabras propuesta puedan utilizarse como variables independientes en estudios de psicolingüística experimental, para observar si características semánticas de corte asociativo influyen de algún modo en tareas de procesamiento lingüístico.

Palabras clave: Conocimiento semántico, asociados, redes semánticas, libre asociación, mapas auto-organizados.

* Este estudio se enmarca en el proyecto *El léxico fundamental del español*, financiado por el Ministerio de Ciencia e Innovación de España (FFI2009-08292).

ABSTRACT

In Psycholinguistics and Computational Linguistics free association tasks have been used commonly as general data collection to develop semantic networks that simulate semantic space of the speakers of a language. Semantic networks incorporate measures such as number of associates, associative strength and other groups of quantitative features of diverse entity. Nevertheless, the number of studies that considers a word from the point of view of the sort of association that it evokes is less abundant. This paper analyzes semantic-associative relationships of 150 Spanish words obtained from a discrete free-association task with the aim of proposing a word taxonomy based on the self-organizing maps (SOM). This procedure will generate both a set of semantic variables based on associates and word taxonomy to be used as independent variables in cognitive psychology experiments that focus on how associative semantic features of words predict linguistic performance.

Keywords: Semantic knowledge, word associations, semantic networks, free-association task, self-organizing maps.

Recibido: 20.03.2014. *Aceptado:* 24.07.2014.

1. INTRODUCCIÓN

En Psicolingüística el estudio sobre las características formales de las palabras, tales como su condición fonológica u ortográfica, ha recibido tradicionalmente una mayor atención que el estudio sobre las características semánticas y por esa razón poseemos menos conocimiento global sobre cómo funciona el nivel conceptual desde el punto de vista de los procesos cognitivos. Esto es debido parcialmente a la dificultad existente en encontrar medidas apropiadas que describan las diferentes dimensiones del amplio espectro que se conoce como *conocimiento semántico* (Balota & Coane, 2008), con la dificultad añadida que presentan las teorías que ponen en relación el componente semántico con el conocimiento general del mundo (Vigliocco & Vinson, 2007).

Sin embargo, en la última década ha habido un resurgimiento de las teorías que representan el componente semántico del lenguaje como una red compleja (*complex networks*) a través de los métodos e instrumentos integrados en la lingüística computacional (Borge-Holthoefer & Arenas, 2010; De Deyne, Navarro & Storms, en prensa). Estas investigaciones construyen redes léxicas basadas en las relaciones semánticas de las palabras en un sentido amplio, a partir de diferentes conjuntos de datos. En muchos casos estas redes muestran no sólo cómo está organizada la memoria semántica, sino que intentan predecir su dinámica, es decir, cómo aprenden nuevas palabras (Steyvers & Tenenbaum, 2005), cómo se realizan

las búsquedas en el lexicón mental y cómo se distribuye la activación entre los nodos (Steyvers, Shiffrin & Nelson, 2004; Borge-Holthoefer & Arenas, 2009; Kenett, Kenett, Ben-Jacob & Faust, 2011). Muchos de estos autores, además, proponen diferentes medidas que recogen características del entorno radial de las palabras (desde las más sencillas como la fuerza asociativa, medidas resultantes de la interacción dentro de las subredes o las más complejas derivadas de las características de la propia red como la centralidad: para una completa síntesis véase la descripción de De Deyne *et al.* (en prensa)), que son identificadas como predictores del procesamiento de tareas cognitivas lingüísticas sencillas como la decisión léxica (entre otros, Buchanan, Westbury & Burgess, 2001; Mirman & Magnuson 2008; De Deyne *et al.*, en prensa).

En este trabajo se propone realizar un análisis del componente semántico-asociativo de 150 palabras del español a partir de los resultados de una tarea libre asociación discreta, basándose en los modelos de autores que han hecho propuestas similares en otras lenguas. El carácter de este estudio es exploratorio y descriptivo, lejos de su alcance queda recoger datos para la elaboración de una red semántica compleja ni completa en español¹, sino que, basándose en principios semejantes, extrae medidas derivadas de la fuerza asociativa primaria y especialmente de los tipos de asociados para analizar su interrelación. De método y objeto serán, por tanto, las diferencias con estudios anteriormente citados, en primer lugar, planteamos una libre asociación discreta, basada en la primera respuesta que los hablantes dan ante una palabra concreta que funciona como estímulo. Así, a los participantes se les preguntó: para el término *x* ¿cuál es la primera palabra que se te viene a la mente? Las respuestas para la palabra *perro* podían variar entre términos como *gato*, *ladrar*, *animal*, etc. En segundo lugar, el alcance menor (150 palabras valoradas por 100 hablantes de español) está vinculado con el objetivo descriptivo y taxonómico de cada término y el conjunto, observar qué tipo de relación existe entre las diferentes variables derivadas de los tipos de asociados y su posible combinatoriedad a través de los mapas auto-organizados.

En esta línea, el estudio de las propiedades asociativas en español además de aportar datos cuantitativos al análisis semántico de las palabras, nos aporta gran información de carácter cualitativo, en relación con el tipo de relaciones que estructuran su entorno semántico inmediato (opuesto al entorno o vecindad del conjunto de toda la red). En este sentido intentaremos responder a lo largo de estas páginas a las siguientes preguntas de investigación: ¿cómo se organiza el espacio semántico inmediato en español a partir de las relaciones asociativas básicas?,

¹ Los estudios de normas de libre asociación en español son escasos y limitados en número de palabras (Fernández, Díez, Alonso & Beato, 2004). De hecho, no poseemos unas normas de libre asociación de carácter amplio, como las que recientemente se están recogiendo en lenguas como en holandés, con más de 70.000 participantes y redes resultantes de más de 12.000 nodos (De Deyne *et al.*, en prensa). Esperamos poder contar pronto con estudios ampliados en español.

¿qué tipo de relaciones básicas enlazan los nodos conceptuales?, ¿podemos establecer una taxonomía de tipos de asociados?, ¿cuáles son los más abundantes? y en una proyección futura ¿de qué forma podemos incorporar esta clasificación a la práctica experimental en psicolingüística? Para todo ello, antes de pasar a la parte metodológica, se realizará una breve descripción de las teorías sobre el enfoque cognitivo del significado que subyacen a la propuesta que aquí presentamos.

2. ANTECEDENTES TEÓRICOS

2.1. Dimensiones del conocimiento semántico

Dentro de la arquitectura cognitiva del procesamiento del lenguaje, la *memoria semántica* es el nivel donde los conceptos se almacenan. Posee un carácter independiente con respecto a las formas de las palabras o la información de carácter fonológico-ortográfico, pero está ligada al conocimiento sobre el mundo en un sentido amplio ya que en ella se encuentra codificada toda la información perceptiva o conceptual que refleja la concepción implícita que el sujeto tiene del mundo. En palabras de Tulving (1972: 386), “[Semantic memory] is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them and about rules, formulas and algorithms for the manipulation of these symbols, concepts and relations”.

En este contexto se entiende *conocimiento semántico* en un sentido amplio, con el resultado de que las fórmulas *concepto* y *significado de las palabras* funcionan como sinónimos en este planteamiento², asumiendo que la información de carácter conceptual general ligada a los términos, también está presente en el significado. En la bibliografía sobre la representación de la información semántica se alternan las teorías holísticas y las componenciales: las primeras consideran el concepto de las palabras como algo que no puede dividirse en unidades más pequeñas, mientras que las segundas consideran que los conceptos son un conjunto de rasgos procedentes del sistema conceptual con gran capacidad combinatoria (Katz & Fodor, 1963; Schreuder & Flores d’Arcais, 1989). Estos rasgos se rigen por los mismos principios de la conceptualización general y, por otro lado, en su consideración central son relativamente constantes en todas las culturas (Vigliocco & Vinson, 2007: 197). Todo ello hace que la distinción entre conceptualización general y conceptualización lingüística no sea necesaria (Jackendoff, 1983, 2003).

² Algunos autores establecen una diferencia funcional entre las representaciones semánticas de las palabras (el significado concreto de las palabras) y, por otro lado, la información conceptual (estructuras conceptuales generales no necesariamente lingüísticas), por tanto, podríamos hablar de dos niveles distintos de arquitectura y procesamiento cognitivo (véase para una ampliación Nuyts & Pederson, 2000).

Tal y como describen Balota y Coane (2008) la memoria semántica y, por tanto el conocimiento semántico, posee numerosas dimensiones: la palabra *perro* no solo tiene información sobre que sea un animal con cuatro patas y piel, que sea una mascota o que le guste perseguir gatos (o coches y ardillas), sino que también incorpora información sensorial sobre cómo se siente uno cuando lo toca, cómo es el sonido de su ladrido, el aspecto y cualidades de los diferentes tipos de perros, las respuestas emocionales ante las interacciones con ellos, etc. Igualmente, en los hablantes de español incorporaría elementos de *co-ocurrencia* contextual como el hecho de que suele aparecer en expresiones como “ser fiel como un perro”, “llevarse como perros y gatos” o “tratarle a uno como a un perro”.

La exploración de los tipos de información contenidos en el conocimiento semántico debe pasar por una clasificación de cada una de sus dimensiones o, al menos, por la delimitación de aquella que se está estudiando. De Deyne, Peirsman y Storms (2009) señalan, por ejemplo, cuatro tipos de fuentes de información semántica para definir el concepto de similitud (*semantic similarity*): los rasgos semánticos, la libre asociación, la *co-ocurrencia* léxica en textos y conocimiento experto en Thesaurus. Este mismo enfoque se observa en los estudios de psicolingüística experimental donde intentan separar los efectos de las diferentes fuentes de información semántica en las diferentes tareas, como, por ejemplo, en el *priming* o facilitación semántica (Lucas, 2000; Hutchison, 2003) o la falsa memoria (Hutchison & Balota, 2005), en especial, intentando explicar si son más destacados los efectos producidos por la superposición de rasgos semánticos o la información asociativa.

En la bibliografía sobre las teorías de representación y organización de la información semántica se reconocen cuatro grandes enfoques (véase Balota & Coane (2008)): las propuestas basadas en el análisis de los rasgos primitivos semánticos compartidos, aquellas basadas en el desarrollo de redes semánticas y, más recientemente, las teorías basadas en bases de datos o *corpora* de amplio alcance, o aquellas basadas en las teorías del *embodiment* que vinculan el conocimiento semántico con los sistemas motores y perceptivos generales del ser humano. Desde el punto de vista de los autores, poner en común los cuatro enfoques puede contribuir a entender la complejidad de la descripción del conocimiento semántico puesta de manifiesto. Sin embargo, para los objetivos de esta investigación, los autores se centrarán en las propuestas derivadas de la creación de redes de palabras que intentan recoger en sus versiones más ambiciosas la más completa representación del conocimiento léxico-semántico.

2.2. Redes semánticas

En el procesamiento lingüístico, tan importante como las características del con-

cepto mismo es la forma en la que está almacenado dicho concepto y la forma de acceso a dicho significado. De hecho, las características estructurales del sistema semántico no sólo pueden dar información sobre cómo son esas unidades sino que pueden predecir las actuaciones lingüísticas: “If a measure capture informative aspects of semantic representation, these values should predict responses on task assumed to rely on semantic activation” (Buchanan *et al.*, 2001).

En relación con la descripción de la estructura semántica en torno a redes, se puede reconocer dos líneas de desarrollo: por un lado, en la bibliografía clásica, a través de los modelos abstractos de procesamiento como las *semantic networks* (Collins & Quilian, 1969; Collins & Loftus, 1975) o, en un ámbito más reciente, a través de modelos computacionales basados en pruebas de psicología experimental, que dan como resultado las llamadas *complex networks*, a las que hacíamos referencia en la introducción (Steyvers & Tenenbaum, 2005; De Deyne & Storms, 2008; Borge-Holthoefer & Arenas, 2010; De Deyne *et al.*, en prensa).

Las redes semánticas suelen describirse en términos de proximidad y de semejanza semántica: “Networks are considered the natural representation of word associations, where nodes correspond to lexicalized concepts and links indicate semantic or lexical relationships between two nodes” (De Deyne *et al.*, en prensa: 1). En una red, una palabra está próxima a otra cuando comparte con ella algún rasgo de algún tipo. Cuanto más cercano es el camino, más alto es el grado de asociación. En este contexto, el componente compartido es entendido en un sentido amplio como cualquier relación asociativa, estrictamente semántica como *perro-gato*, o incluso derivada de la proximidad en el contexto, como resultado de la experiencia metalingüística del individuo como *tranvía-deseo* (dos palabras pueden estar relacionadas porque aparecen en contextos lingüísticos similares que necesariamente presuponen un sentido global compartido). Esta concepción amplia de las dimensiones del significado dentro de las redes, hace necesario que para su estudio se haga una identificación concreta de los tipos de relaciones asociativas que establecen las palabras entre sí, no sólo por el hecho de sistematizar la estructura de las redes, sino para poder independizar su capacidad de incidencia de los tipos de información semántica en los procesos lingüísticos.

2.3. Tipos de relaciones asociativas a partir de la libre asociación

En diversas ocasiones se han puesto de manifiesto las dificultades para acceder a muestras empíricas –especialmente en producción– de las que se puedan extraer conclusiones sobre cómo funciona el componente semántico. Existe, sin embargo, un conjunto de pruebas experimentales con una larga trayectoria de desarrollo experimental basadas en el paradigma de la libre asociación (Deese, 1962; Cramer, 1968), que son reconocidas como una valiosa fuente de datos para el acceso y la

representación del componente semántico: “We assume that a dynamic associative structure is created in memory that involves representations of the word themselves as well as connections to other words” (Nelson, McEvoy & Schreiber, 1998: 4).

De hecho, la libre asociación es uno de los paradigmas experimentales más importantes que se han utilizado para la elaboración de las redes semánticas complejas en diversas lenguas: para el inglés, Steyvers & Tenenbaum (2005); para el inglés y el español, Borge-Holthoefer & Arenas (2010); para el hebreo Kenett *et al.* (2011); o el holandés De Deyne & Storms (2008) y De Deyne *et al.* (en prensa). Así, es ampliamente asumido que la libre asociación es uno de los métodos más fiables de acceder al conocimiento semántico: “We hasten to add that Word association is the simplest possible form of linguistic processing that occurs during conceptual processing” (Barsalou, Santos, Simmons & Wilson, 2008). La aplicación actual de la asociación libre de palabras está centrada en la creación de espacios semánticos y de conjuntos de normas de proximidad léxica (Hirsh & Tree, 2001; Steyvers *et al.*, 2004), como las *Florida Free Association Norms* para el inglés (Nelson, McEvoy & Schreiber, 1998), el proyecto *Small World of Words* impulsado por Simon de Deyne y Gert Storms (2008), con importantes bases de datos en holandés, inglés y alemán³.

Junto a las clasificaciones clásicas de tipos de relaciones léxicas generales como, por ejemplo la de Clark (1970), basada en reglas paradigmáticas o sintagmáticas, o la de Aitchison (1994), a partir de una sistematización de las relaciones léxicas fuertes y las relaciones léxicas débiles, existen otras propuestas más específicas, que describen de forma particular los hallazgos obtenidos a través de la libre asociación. Como se ha visto, una palabra puede estar ligada semánticamente a otra de diversas maneras; dentro del conocimiento asociativo extraído de la libre asociación se manifiestan distintos tipos de relaciones que capturan diferentes dimensiones de información semántica (como relaciones categoriales, de causa y efecto, instrumentales, etc.). Tradicionalmente se han identificado dos grandes tipos de relaciones: las basadas en los rasgos semánticos compartidos (*featured-based*) y las basadas en asociaciones amplias (*association-based*)⁴. Además de estos dos tipos generales, algunos modelos recientes, que diseñan búsquedas lingüísticas más completas dentro de las redes asociativas, proponen otras distinciones más precisas que diferencian el carácter categorial (*animales-perro-gato-pastor alemán*) o de *co-ocurrencia* en contextos discursivos (*tranvía-deseo*) (De Deyne & Storms,

³ <http://smallworldofwords.com/blog/>.

⁴ Buchanan *et al.* (2001) consideran que las medidas semánticas extraídas de la libre asociación representan, en mayor medida, el segundo tipo de relaciones. En su investigación toman los resultados de una red basada en la libre asociación y una red basada en un conjunto de textos amplio (*global co-occurrences*). Ellos encuentran que el volumen de la vecindad semántica de una palabra predice la actuación en las tareas de *Word naming* y decisión léxica. De hecho, ellos proponen que las redes de asociados crean “free categories”.

2008; Borge-Holthoefer & Arenas, 2009).

Para este trabajo se ha establecido la siguiente clasificación basada en las propuestas de Buchanan *et al.* (2001) y De Deyne y Storms (2008), tal y como recoge la Figura 1.

Dentro de las asociaciones exclusivamente o propiamente semánticas se diferencia entre:

- Asociación categorial (taxonómicas): *perro-gato* (términos iguales), *animal-perro* (hiperónimo) o *pastor alemán-perro* (hipónimo).
- Asociación no categorial: relación funcional *coche-carretera*; causa efecto *fuego-humo*; partes de objetos *taza-asa*; etc.
- En cuanto a las relaciones de carácter amplio que reflejan rasgos de *co-ocurrencia* contextual y que reflejan relaciones semánticas más débiles encontramos:
- Asociación por *co-ocurrencia* o colocación exclusiva: *tranvía-deseo*⁵
- Asociación por *co-ocurrencia* o colocación y relación semántica: *abeja-picar*

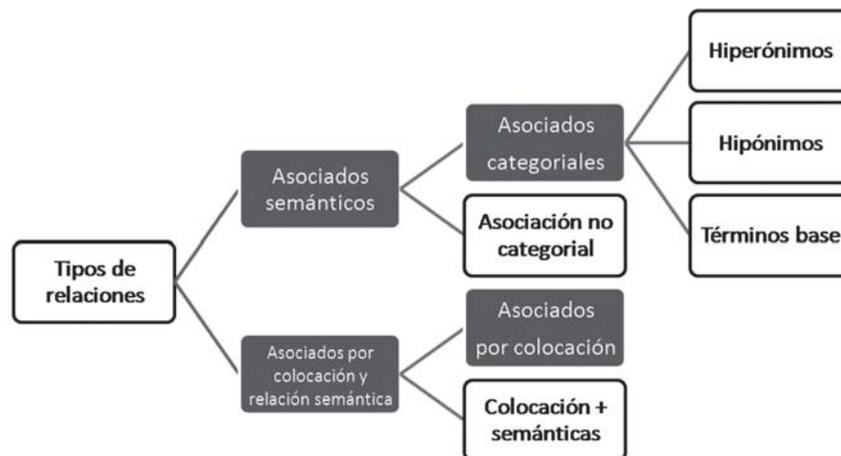


Figura 1. Tipos de relaciones asociativas consideradas en la investigación.

3. METODOLOGÍA

3.1. Prueba de libre asociación

En primer lugar se realizó una prueba escrita de libre asociación con 150 estímulos a 100 hablantes nativos de español donde se preguntaba por el primer asociado de

⁵ Basada en el título de la obra de Tennessee Williams.

cada palabra-estímulo. Para la recogida de los datos se utilizó tanto el método de papel y lápiz como la encuesta electrónica.

En segundo lugar, para cada una de las palabras se contabilizaron el número total de asociados, el número de asociados dados por dos o más personas, la frecuencia del primer asociado y el número de palabras idiosincrásicas (o asociados dados por un solo participante). También se obtuvo la conectividad de cada término: este índice valora tanto la fuerza de producción de un término como el número de ocasiones en las que cada palabra ha resultado un asociado, es decir, la fuerza asociativa directa e inversa de cada palabra, siguiendo la fórmula propuesta en Hernández e Izura (2011).

En tercer lugar, los asociados dados por más de dos informantes se clasificaron de acuerdo con la tipología descrita en el apartado 2.3, contabilizando exclusivamente los cuatro tipos marcados en oscuro en la Figura 1: asociados categoriales (aquellos que forman parte de la misma categoría semántica, llamados taxonómicos en la denominación de De Deyne y Storms (2008)), asociados semánticos (aquellos que mantienen una relación semántica asociativa de carácter general con la palabra estímulo); asociados por colocación (aquellos asociados que comparten contextos lingüísticos con la palabra estímulo) y asociados por colocación y relación semántica (aquellos que además de compartir contexto lingüístico tienen cierta relación semántica general). Cada asociado fue clasificado por dos investigadores y después se discutieron en conjunto los casos dudosos. Finalmente, se contabilizó para cada una de las palabras el número de asociados de cada uno de los cuatro tipos, así como la suma de las frecuencias. En algunos casos un mismo asociado podía pertenecer a dos categorías diferentes (siguiendo a De Deyne & Storms (2008)). El resultado final fue un conjunto de variables amplias que medían diferentes dimensiones de la vecindad asociativa de cada palabra producida a través de la libre asociación.

3. 2. Creación de la matriz de datos y del mapa auto-organizado SOM

Una vez clasificados todos los resultados, se obtuvo una matriz de datos en la que para cada palabra había trece valores diferentes, según el listado que aparece a continuación. Los datos del 1 al 4 son valores clásicos en la libre asociación, el 5 el único que refleja las relaciones bidireccionales entre estímulos y respuestas y, finalmente, los datos del 6 al 13 son datos basados en los tipos de relaciones asociativas descritas. La etiqueta *número de asociados/asociados* hace referencia al número de palabras diferentes obtenidas para cada palabra, mientras que *frecuencia* indica el número de participantes que dieron cada uno de los asociados correspondientes.

1. Número de asociados totales
2. Número de asociados dados por más de dos personas
3. Frecuencia asociativa del primer asociado
4. Asociados idiosincrásicos
5. Conectividad
6. Asociados categoriales
7. Frecuencia de los asociados categoriales
8. Asociados semánticos
9. Frecuencia de los asociados semánticos
10. Asociados por colocación
11. Frecuencia de los asociados por colocación
12. Asociados por colocación y relación semántica
13. Frecuencia de los asociados por colocación y relación semántica

A continuación se redujo la matriz a un vector de datos para cada palabra a través de la tecnología de creación de espacios multidimensionales que distribuye cada elemento (las palabras en nuestro caso) de forma regular con el fin descubrir la estructura subyacente de los datos introducidos. La técnica empleada para la representación gráfica del espacio multidimensional se denomina mapa auto-organizado (*self-organizing map*), más conocida por su acrónimo en inglés, SOM. Esta técnica fue desarrollada por Kohonen (1989) y se cuenta entre las redes neuronales de tipo no supervisado⁶. La principal característica de las redes SOM es su capacidad para representar vectores con gran número de componentes de información en espacios bidimensionales llamados *mapas*. En estos mapas, las celdas que están próximas entre sí contienen información similar, permitiendo de una manera gráfica clasificar vectores de datos complejos.

El proceso de creación de un mapa consiste en, a partir de un mapa aleatorio, aplicar un algoritmo de entrenamiento que permita aproximar el valor de las celdas del mapa a los valores de los vectores de entrenamiento. Este algoritmo se puede describir en 3 pasos:

- a) Seleccionar un dato de entrenamiento.
- b) Identificar la celda que contiene la información más similar al dato seleccionado.
- c) Modificar la información de la celda y la de las celdas vecinas para que se parezca aún más al vector seleccionado y volver al paso a).

⁶ La topología de una red SOM consiste en una capa de neuronas de entrada y otra de salida. La capa de entrada cuenta con tantas neuronas como componentes tiene el vector de datos. La capa de salida es una matriz de $n \times m$ neuronas con una relación de vecindad definida. El tamaño de la capa de salida se define en función del número de vectores disponibles para el entrenamiento.

Al cabo de las suficientes iteraciones, el espacio de datos de entrada es cubierto por el mapa y cada dato del espacio multidimensional de entrada tiene una proyección en el espacio bidimensional de salida, es decir, cada palabra puede ser ubicada en una celda del mapa, como puede observarse en la Figura 2. Además, celdas próximas entre sí contienen palabras con vectores de datos similares.

Finalmente, el aparato estadístico que acompaña a la técnica SOM permite identificar el rasgo de agrupación semántica que lidera la distanciaci3n entre diferentes grupos de unidades (palabras), creando tal y como se aprecia en la Figura 4 una taxonomía de cinco grupos dentro del mapa, que responden a las diferentes distribuciones del vector de datos.

4. RESULTADOS

4.1. Exploraci3n de los datos

A continuaci3n aparecen los resultados de la clasificaci3n general de los tipos de asociados producidos para las 150 palabras-estímulo recogidos en la Tabla I y en los Gráficos 1 y 2. Como puede observarse, más del 50% de los asociados pertenecen a los grupos de asociaci3n semántica, mientras que el número de asociados categoriales y asociados por colocaci3n más relaci3n semántica es semejante. En último lugar, aparecen los asociados exclusivamente debido a frecuencias contextuales. Si se considera no el total de asociados dados sino las frecuencias absolutas de cada uno (es decir, el número de participantes que dieron cada asociado), aunque parecen variar poco gráficamente los resultados muestran que los asociados por colocaci3n no son sólo menos numerosos, sino también menos productivos, en el sentido de que poseen frecuencias absolutas menores, mientras que los asociados categoriales fueron respuestas dadas por un mayor número de participantes.

Tabla I. Descriptores cuantitativos de tipos de asociados y frecuencias.

Típos de asociados	Número	% Núm.	Frecuencia	% Frec.
Categoriales	610	19,3	4735	22,9
Semánticos	1702	53,8	11071	53,6
Colocaci3n	217	6,9	936	4,5
Colocaci3n/ semánticos	637	20,1	3914	18,9
Total asociados	3166	100,0	20656	100,0

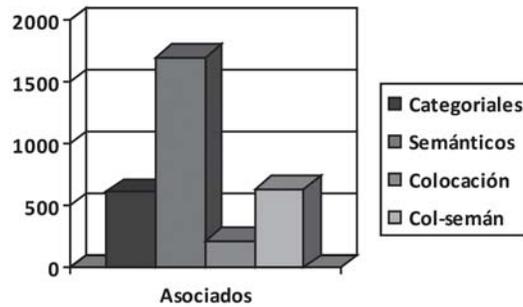


Gráfico 1. Tipos de asociados.

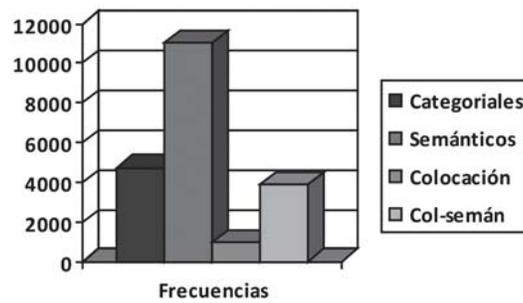


Gráfico 2. Frecuencias de los tipos de asociados.

4.2. Red auto-organizada

En primer lugar se hará una descripción general del mapa auto-organizado reflejado en la Figura 1. Para una correcta interpretación hay que destacar que las palabras que aparecen en la misma casilla o en casillas adjuntas poseen un vector de datos semejante, por lo que se puede deducir que las características mostradas en la libre asociación son semejantes en relación con las trece variables insertadas. Esta distribución gráfica en ningún caso valora características cualitativas del significado de las palabras, sino la matriz de datos cuantitativos obtenida para cada una de ellas. Por tanto, que dos palabras con significados semejantes estén juntas

en la misma celda o en celdas cercanas es fruto, no de los rasgos semánticos que aparentemente puedan compartir, sino de la semejanza del vector obtenido por cada una de ellas. Es decir, que, por ejemplo, el hecho de que *perro* y *gato* se encuentren en celdas contiguas (centro inferior del mapa) o *fuente* y *arroyo* en la misma celda (esquina inferior izquierda) no se debe a que los primeros sean animales y los segundos tengan relación con el agua, sino a que ambas parejas de palabras han obtenido un conjunto de variables numéricas semejantes.

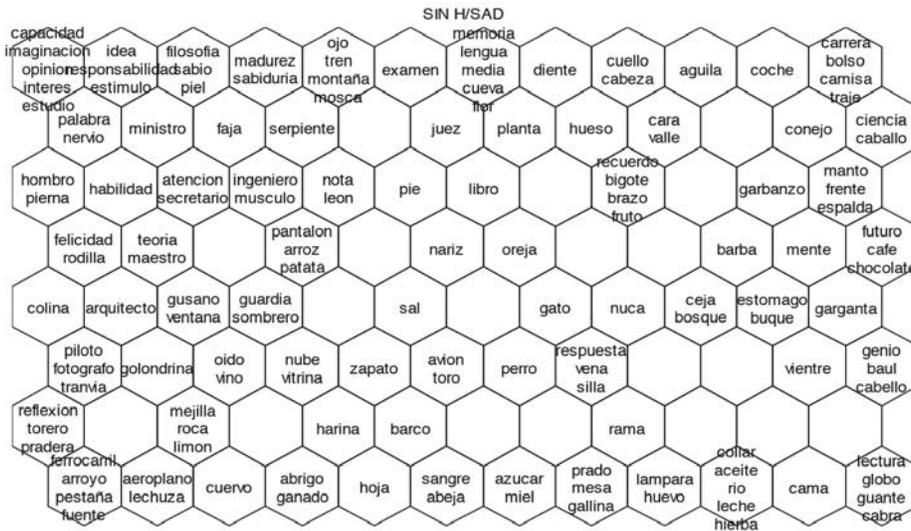


Figura 2. Red auto-organizada.

Junto al gráfico de las celdas, la red auto-organizada produce una plantilla donde se muestra la distribución de cada una de las variables en el mapa de forma independiente, como se ve en el ejemplo de la Figura 3, donde se aprecia en qué zonas del mapa aparecen los valores más altos o bajos de cada uno de los índices. Aunque la representación en blanco y negro no permita distinguir entre los valores extremos, en la imagen original en color se observa cómo los valores más altos, es decir las palabras con mayor número de asociados están en la esquina superior izquierda del mapa (como *capacidad* o *imaginación*), mientras que las que se encuentran en la zona central e inferior del mapa presentan palabras con pocos asociados (como *sangre* o *azúcar*).

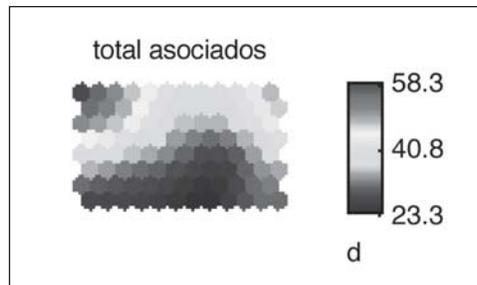


Figura 3. Distribución de los valores para la variable total de asociados⁷.

A continuación, antes de describir esta propuesta, en relación con los tipos de palabras según sus entornos asociativos basados en los resultados de la libre asociación, se comentarán algunas cuestiones generales que se desprenden de la observación de los datos. En primer lugar, parece que hay cierta similitud entre los vectores de datos de las palabras que pertenecen a una misma categoría semántica o campo nocional, a pesar de que no hay razón aparente para ello, más allá de que puedan formar una subred de palabras interconectadas entre sí con idénticas relaciones asociativas, pero no conectadas significativamente de forma individual con el resto de los términos. Así, las *Profesiones* tienden a situarse en la sección central izquierda, lo que representa unos valores bajos en cuanto al número y la frecuencia de asociados por colocación (sus medias son 0,3/0,6 respectivamente, mientras que el resto de las palabras obtienen unas medias de 1,5/6,7). Igualmente, las *Partes del cuerpo* tienden a aparecer en el centro-izquierda del mapa, lo que significa una muy alta frecuencia de los asociados exclusivamente categoriales (ellas obtienen una media de 44,9, mientras que el resto sólo 23).

Por otro lado, las palabras menos concretas e imaginables se sitúan en el extremo superior izquierdo debido principalmente a que son las que tienen mayor número de respuestas idiosincrásicas y, por tanto, mayor total de asociados (*capacidad, imaginación, estímulo, etc.*). Ambas cuestiones merecerán una reflexión posterior, puesto que, como hemos destacado, la proximidad en el mapa no responde a rasgos semánticos concretos (como ser un animal, o ser una profesión; o ser concreto o abstracto), por lo que se deduce que ciertos rasgos semánticos pueden hacer que las palabras tiendan a tener relaciones asociativas semejantes. Este aspecto deberá ser estudiado con mayor profundidad en el futuro.

⁷ Los valores numéricos representados en la escala de la derecha de la imagen no corresponden a los valores absolutos, sino al ajuste realizado después de reducir al vector de datos todos los valores insertados en el cálculo.

5. TIPOLOGÍA DE PALABRAS BASADA EN VECTORES ASOCIATIVOS

En cuanto a la clasificación de los términos en grupos de palabras con vecindades asociativas semejantes, la tecnología SOM ha permitido identificar cinco grupos de palabras con distribución de valores asociativos semejantes, según la Figura 4.

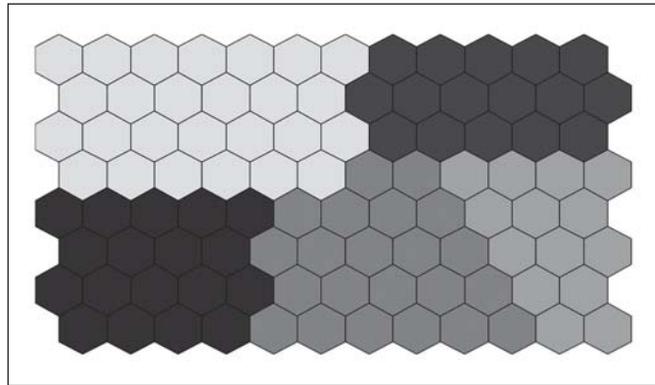


Figura 4. Regiones con similitudes vectoriales.

Grupo 1. Situado en la esquina superior izquierda. El rasgo más importante que lo define es que son palabras con un alto número de asociados donde predominan los asociados semánticos (ejemplo: *estímulo*). En este primer grupo se pueden identificar, además, dos claros subgrupos, aquéllos con un alto número de respuestas idiosincrásicas (*opinión*) o bajo número de idiosincrásicas (*tren*).

Grupo 2. Situado en la esquina superior derecha. Palabras con alto número de asociados, en los que predominan los de colocación (*bolso*).

Grupo 3. Situado en la parte inferior izquierda. Palabras con alta fuerza asociativa del primer asociado, alta frecuencia de relaciones semánticas y baja frecuencia de asociados por colocación (*fuelle*).

Grupo 4. Situado en la parte central inferior. Palabras con alta fuerza asociativa del primer asociado y alta conectividad (*miel*).

Grupo 5. Situado en la parte inferior izquierda. Palabras con alta y media fuerza asociativa del primer asociado y alta frecuencia de asociados de colocación (*baúl*).

Como puede observarse en la descripción de los grupos, los valores que han resultado más pertinentes para la clasificación han sido, en primer lugar, el número total de asociados (grupos 1 y 2), la frecuencia del primer asociado (grupos 3 y 5) y la conectividad (grupo 4). En segundo lugar, los asociados semánticos (grupo 1) y de colocación (grupo 2) y los valores de frecuencias de relaciones semánticas (grupo 3) y frecuencia de colocación (grupo 5). Por otro lado, llama la atención la menor cantidad de información dentro del mapa auto-organizado aportada por los asociados categoriales y las relaciones donde se unen la colocación y la semántica.

Por tanto, el hecho de que una palabra tenga muchos asociados y que éstos sean fundamentalmente de dos tipos, semánticos y estrictamente de colocación, así como que tenga fuerte asociación bidireccional (que sea estímulo y respuesta dentro de la red) parecen ser los valores más relevantes para la creación de grupos de palabras con vectores asociativos semejantes, por lo que podríamos predecir que éstas serán variables con mayor carga informativa a la hora de diferenciar tipos de palabras según el conocimiento asociativo que conllevan. De este modo, aunque habrá que contrastarlo en un futuro con estudios experimentales en español, estas medidas podrían mostrar cierta influencia en procesos lingüísticos donde el conocimiento asociativo estuviera implicado.

En el Anexo 1 puede encontrarse la clasificación completa de los 150 estímulos empleados en la prueba de libre asociación según los grupos creados por los vectores de datos basados en el número y tipo de asociados producidos. En el Anexo 2 aparece el listado completo de los asociados y las frecuencias de los ejemplos citados para cada uno de los grupos.

6. CONCLUSIONES

Se comenzó estas páginas relacionando las teorías sobre las diferentes dimensiones del conocimiento semántico con el paradigma metodológico de la creación de redes léxicas a través de la libre asociación para explorar parte del conocimiento asociativo de los hablantes de una lengua. A pesar de las limitaciones prácticas, de método y objeto, que presentaba este estudio en español con relación a los realizados en otras lenguas, los resultados de esta investigación destacan la importancia de hacer una correcta selección de las medidas que representan el conocimiento asociativo para fines experimentales y, en este sentido, realizan una nueva propuesta en esta dirección a partir de la exploración de los datos realizados con los mapas auto-organizados.

A lo largo de la presentación de los resultados se ha ido respondiendo a las preguntas que se hicieron en la introducción. Por un lado, se han analizado los

resultados de una prueba de libre asociación en español con el fin de analizar los tipos de asociados elicitados por cada uno de los estímulos: en ellos se observa que predominan las relaciones semánticas generales (instrumentales, causa y efecto, parte-todo, etc.), mientras que las relaciones jerárquicas categoriales y de colocación con contenido semántico suponen un segundo grupo menos numeroso, seguido de las relaciones estrictamente de colocación o dependencia contextual. En trabajos donde no se ha considerado la asociación discreta, sino que se han registrado varios asociados para cada estímulo, como el de De Deyne y Storms (2008: 228), se ha concluido que la información taxonómica/categorial está disponible más rápidamente que la información conceptual (como las propiedades del objeto). En una posible ampliación de este trabajo se podría dividir el grupo más numeroso, el de las relaciones semánticas generales, según las propuestas de estos autores, para llegar a un conocimiento más detallado de las posibilidades de vías de creación de conocimiento asociativo.

Desde el punto de vista metodológico se confirma el paradigma de la libre asociación como instrumento de observación del componente semántico, en línea con la amplia tradición recogida en la bibliografía. Asimismo, la tecnología de mapas auto-organizados (SOM) aparece como un método útil para reducir la matriz de datos cuantitativos complejos resultado de las respuestas dadas por los participantes a tareas con alta variabilidad como la libre asociación. A ello se suma la propuesta de un conjunto muy amplio de variables cuantitativas de descripción de los entornos asociativos, basadas en los tipos de relaciones semánticas y contextuales encontradas. Estas variables tal y como han propuesto otros investigadores, pueden incorporarse al estudio psicolingüístico experimental, paliando así la relativa falta de valores semánticos cuantificables, más allá de algunos tradicionales como la imaginabilidad o el número de significados, especificando, además, la dimensión del conocimiento semántico al que hacen referencia.

En relación con los entornos asociativos de las palabras, a través de nuestros resultados hemos podido establecer cinco grupos de términos según estén más o menos presentes en ellos ciertas cualidades asociativas, con el fin de clasificarlos según patrones más complejos que la simple fuerza asociativa primaria. Como destacábamos anteriormente, el siguiente paso es incluir la propuesta taxonómica como variable categórica independiente en pruebas de procesamiento lingüístico para observar su capacidad de predicción, tal y como diversos autores han hecho ante la propuesta de nuevas variables experimentales.

Para terminar, como otras propuestas de desarrollo futuro, habría que profundizar en el hecho de que la pertenencia a ciertas categorías semánticas parece predecir el número y tipo de asociados de las palabras, así como en que el concepto de abstracción parece estar ligado también a ciertas cualidades asociativas.

REFERENCIAS

- Aitchison, Jean. (1994). *Words in the mind: an introduction to the mental lexicon*. Oxford-Cambridge: Blackwell.
- Balota, David A. & Coane, Jennifer H. (2008). Semantic Memory. En J. H. Byrne, H. Eichenbaum, R. Menzel, H. L. Roediger III & D. Sweatt (eds.), *Handbook of learning and memory: A comprehensive reference* (pp. 512-531). Amsterdam: Elsevier.
- Barsalou, Lawrence; Santos, Ava; Simmons, W. Kyle & Wilson, Christine D. (2008). Language and simulation in conceptual processing. En M. de Vega, A. M. Glenberg & A. C. Graesse (eds.), *Symbols, embodiment and meaning* (pp. 246-283). Oxford: oxford University Press.
- Borge-Holthoef, Javier & Arenas, Alex. (2009). Navigating word association norms to extract semantic information. *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 2777-2782). Austin, TX: Cognitive Science Society.
- Borge-Holthoef, Javier & Arenas, Alex. (2010). Semantic Networks: Structure and dynamics. *Entropy*, 12, 1264-1302.
- Buchanan, Loti; Westbury, Chris & Burgess, Curt. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin and Review*, 8 (3), 531-544.
- Clark, Herbert H. (1970). Word associations and linguistic theory. En J. Lyons (ed.), *New Horizons in Linguistics* (pp. 271-286). Londres: Penguin.
- Collins, Allan M. & Quillian, M. Ross (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 240-247.
- Collins, Allan. M. & Loftus, Elisabeth F. (1975). A Spreading-activation theory of semantic processing, en *Psychological Review*, 82 (6), 407-428.
- Cramer, Phebe. (1968). *Word Association*. New York: Academic Press.
- De Deyne, Simon & Storms, Gert. (2008). Word Association: Network and semantic properties. *Behavior Research Methods*, 40 (1), 213-231.
- De Deyne, Simon; Navarro, Daniel J. & Storms, Gert. (en prensa). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods*. Disponible en http://smallworldofwords.com/blog/?page_id=2.
- De Deyne, Simon; Peirsman, Yves & Storms, Gert (2009). Sources of semantic proximity. En N.A. Taatgen y H. van Rijn (eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (pp.1834-1839). Austin, TX: Cognitive Science Society.
- Deese, James. (1962). On the structure of associative meaning. *Psychologica Review*, 69, 161-175.
- Fernández, Ángel; Díez, Emiliano; Alonso, María Ángeles & Beato, María Sole-

- dad. (2004). Free-association norms for the Spanish names of the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers*, 36 (3), 577-583.
- Hernández Muñoz, Natividad e Izura, Cristina. (2011). Dibujando el espacio semántico: La conectividad. *Actas del IX Congreso Internacional de Lingüística* (pp. 1168-1180). Valladolid: Universidad de Valladolid.
- Hirsh, Katherine W. & Tree, Jeremy T. (2001). Word association norms for two cohorts of British adults. *Journal of Neurolinguistics*, 14, 1-44.
- Hutchison, Keith. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin and Review*, 10 (4), 785-813.
- Hutchison, Keith & Balota, David. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language*, 52 (1), 1-28.
- Jackendoff, Ray. (1983). *Semantics and cognition*. Cambridge Mass.: MIT Press.
- Jackendoff, Ray. (2003). *Foundation of language: Brain, meaning, grammar*. Oxford: Oxford University Press.
- Katz, Jerrold J. & Fodor, Jerry A. (1963). The structure of a semantic theory. *Language*, 39, 170-210.
- Kenett, Yoed N.; Kenett, Dror Y.; Ben-Jacob, Eshel & Faust, Miriam. (2011). Global and local features of semantic Networks: Evidence from the Hebrew Mental Lexicon. *Plos One*, 6 (8), 1-14.
- Kohonen, Teuvo. (1989). *Self-organization and associative memory*. New York-Berlin: Springer-Verlag.
- Lucas, Margery. (2000). Semantic priming without association: a meta-analytic review. *Psychonomic Bulletin and Review*, 7 (4), 618-630.
- Mirman, Daniel & Magnuson, James S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbours and speeded by distant neighbours. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34 (1), 65-79.
- Nelson, Douglas L.; McEvoy, Cathy L. & Schreiber, Thomas A. (1998). *The University of South Florida Free Association Norms, Rhyme and word Fragment Norms*. Disponible en <http://w3.usf.edu/FreeAssociation/Intro.html>.
- Nuyts, Jan & Pederson, Eric. (2000). *Language y conceptualization*. Cambridge: Cambridge University Press.
- Schreuder, Robert & Flores d'Arcais, Giovanni B. (1989). Psycholinguistic issues in the lexical representation of meaning. En W. Marslen-Wilson (ed.), *Lexical representation and process* (pp. 409-436). Cambridge: MIT Press.
- Steyvers, Mark & Tenenbaum, Joshua B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41-78.

- Steyvers, Mark; Shiffrin, Richard M. & Nelson, Douglas L. (2004). Word Association spaces for predicting semantic similarity effects in episodic memory. En A. F. Healy (ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch and Thomas Landauer* (pp. 237-249). Washignton, DC: American Psychological Association.
- Tulving, Endel. (1972). Episodic and semantic memory. En Endel Tulvind & Wayne Donaldson (eds.), *Organization of memory* (pp. 381-492). Nueva York: Academic Press.
- Vigliocco, Gabriella & Vinson, David. (2007). Semantic representation. En M. G. Gaskell (ed.), *The Oxford Handbook of Psycholinguistics* (pp. 195-215). Oxford: Oxford University Press.

ANEXO 1

Tabla II. Lista de palabras por grupos derivados del análisis SOM.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
arroz	águila	abrigo	abeja	barba
atención	bigote	aeroplano	aceite	baúl
capacidad	bolso	arquitecto	avión	bosque
cueva	brazo	arroyo	toro	buque
estímulo	caballo	colina	azúcar	cabello
estudio	cabeza	cuervo	barco	cabra
examen	camisa	ferrocarril	collar	café
faja	cara	fotógrafo	gallina	cama
felicidad	carrera	fuelle	gato	ceja
filosofía	ciencia	ganado	hierba	chocolate
flor	coche	golondrina	hoja	estómago
habilidad	conejo	guardia	lámpara huevo	futuro
hombro	cuello	gusano	leche	garganta
idea	diente	harina	mesa	genio
imaginación	espalda	lechuga	miel	globo
ingeniero	frente	limón	nuca	guante
interés	fruto	mejilla	oreja	lectura
juez	garbanzo	nube	perro	mente
lengua	hueso	oído	prado	vientre
león	manto	pestaña	rama	
libro	planta	piloto	respuesta	
madurez	recuerdo	pradera	río	
maestro	traje	reflexión	sal	
media	valle	roca	sangre	
memoria		sombrero	silla	
ministro		torero	vena	
montaña		tranvía	zapato	
mosca		ventana		
músculo		vino		
nariz		vitrina		
nervio				
nota				
ojo				
opinión				
palabra				
pantalón				
patata				
pie				
piel				
pierna				
responsabilidad				
rodilla				
sabiduría				
sabio				
secretario				
serpiente				
teoría				
tren				

ANEXO 2

Tabla III. Ejemplo grupo 1: ESTÍMULO.

Asociados	Frec.	Semánticas gen.	Frec.	Concurren.	Frec.	Semán. + conc.	Frec.
respuesta	26	respuesta	26	visual	3	visual	3
sensación	7	sensación	7			respuesta	26
reacción	5	reacción	5			experimento	2
motivación	3	motivación	3				
visual	3	motivo	2				
motivo	2	nervio	2				
nervio	2	ánimo	2				
ánimo	2	empuje	2				
empuje	2	experimento	2				
experimento	2	feliz	2				
feliz	2	impacto	2				
impacto	2	impulso	2				
impulso	2	sentidos	2				
sentidos	2						

Tabla IV. Ejemplo del grupo 2: BOLSO.

Asociados	Frec.	Cat.	Fr.	Coloc.	Frec.
piel	10	bolsa	5	piel	10
grande	7	cartera	4	grande	7
cuero	6	complementos	3	cuero	6
bolsa	5	monedero	3	mujer	5
mujer	5	maleta	2	guardar	4
asa	4	prenda	2	cocodrilo	2
cartera	4	zapatos	2	colgar	2
guardar	4			marrón	2
complementos	3			rojo	2
mano	3			mano	3
monedero	3			lleno	2
bolsillo	2				
chica	2				

Continuación Tabla IV.

cocodrilo	2				
colgar	2				
compra	2				
dinero	2				
llaves	2				
lleno	2				
maleta	2				
marrón	2				
moda	2				
prenda	2				
rojo	2				
zapatos	2				

Tabla V. Ejemplos del grupo 3: FUENTE.

Asociados	Frecuencia
agua	87
fresca	2
río	2

Tabla VI. Ejemplos del grupo 4: MIEL (conectividad 1,75; valor entre 1 y 2).

Asociados	Frec.	Cat.	Frec.	Sem.	Frec.	Col.+Sem.	Frec.
abejas	42	azúcar	2	dulce	36	mosca	2
dulce	36			oso	3		
oso	3			azúcar	2		
azúcar	2			dulzor	2		
dulzor	2			líquido	2		
líquido	2						
mosca	2						

Tabla VII. Ejemplo del grupo 5: BAÚL.

Asociados	Frec.	Cat.	Frec.	Sem. G.	Frec.	Col.	Frec.	Col.+Sem.	Frec.
recuerdos	47	caja	4	guardar	11	recuerdos	47	recuerdos	47
guardar	11	cofre	4	ropa	6			guardar	11
ropa	6			madera	4			madera	4
caja	4			viejo	4			antiguo	2
cofre	4			tesoros	3			viejo	4
madera	4			antiguo	2				
viejo	4			cerrado	2				
tesoros	3			secretos	2				
antiguo	2			caja	4				
cerrado	2			cofre	4				
secretos	2								