

LAS LISTAS DE DENSIDAD SEMÁNTICA Y LA LEXICOGRAFÍA FUNCIONAL COMPUTACIONAL¹

José Carlos Perrián Pascual

*Unidad Central de Idiomas
Universidad Católica de San Antonio*

RESUMEN: La mayoría de los investigadores en lingüística computacional coinciden en afirmar que el mejor método para la construcción de un lexicon computacional es el análisis automático de un diccionario legible por la máquina. En la entrada léxica de un diccionario podemos encontrar toda la información que necesita el lexicon en un sistema del procesamiento del lenguaje natural. No obstante, debemos descodificar la información presentada por el lexicógrafo humano con el fin de hacerla accesible computacionalmente. En el proceso de transducción de un texto lexicográfico en un formato característico de la Gramática Funcional proponemos un modelo enriquecido del componente nuclear de la teoría de S.C. Dik. El postulado de significado asociado a cada marco predicativo se almacena como una lista de densidad semántica, la cual permite que el sistema construya una red conceptual que conecte los predicados del lexicon. Esta estructuración cognitiva del lexicon, abogada igualmente por el Modelo Lexemático-funcional, se fundamenta en el enfoque de la Descomposición Léxica Gradual. En este artículo describimos precisamente la construcción automática de estas listas de densidad semántica para los predicados verbales en español.

ABSTRACT: Most researches conclude that the best method for building computational lexicons is the automatic analysis of machine-readable dictionaries. All the information needed for NLP lexicons can be found in a dictionary text. We have to decode the information presented by the human lexicographer in order to make it accessible in a computational way. In the transduction of a dictionary text into a Functional Grammar format we propose an enriched model of the basic component of S.C. Dik's theory, the Fund. The meaning postulate associated to each predicative frame will take the form of a semantic density list, which enables the system to create a conceptually-motivated network connecting the predicates of the lexicon. This cognitive structuring of lexicon, also advocated by the Functional-Lexematic Model, is grounded on the Stepwise Lexical Decomposition approach. In this paper we therefore describe the automatic construction of semantic density lists for verbal predicates in Spanish.

1. El presente artículo está basado en la ponencia "Towards a Cognitive-Functional Transduction of a Machine-Readable Dictionary of Spanish" presentada en el 9th International Functional Grammar Conference (20-23 Septiembre 2000, Madrid).

1. Functional Grammar Workbench 2.2

Functional Grammar Workbench (de aquí en adelante, FGW) se concibió originariamente como un laboratorio virtual de ingeniería lingüística con fines investigadores y docentes dentro del marco de la gramática funcional. Actualmente, y desde la perspectiva de la lexicografía computacional, nuestra principal preocupación es la construcción automática de un lexicón computacional a partir de un diccionario legible por la máquina. Con este propósito, la versión 2.2 de FGW permite elaborar de forma automática los marcos predicativos, los postulados de significado y las listas de densidad semántica de los verbos de movimiento del español a partir de las entradas léxicas del *Diccionario del Español Actual* (1999) de Manuel Seco, Olimpia Andrés y Gabino Ramos (de aquí en adelante, DEA).²

FGW utiliza en esta versión un procedimiento asistido en la fase de inserción de las nuevas entradas léxicas; es decir, la introducción de las nuevas entradas se lleva a cabo a través de una interfaz agradable al usuario no experto en lingüística que le permite “copiar” la información que contienen las entradas léxicas del DEA. Por ejemplo, la entrada léxica de *trepar 1b* se introduce en el lexicón de FGW a través de la interfaz que se muestra en la ilustración 1. Podemos imaginar con facilidad que si tuviéramos el DEA en un CD-ROM con formato legible por la máquina, o incluso utilizáramos un programa convencional de OCR y escaneáramos el diccionario para obtenerlo en formato ASCII, el programa llevaría a cabo los mismos procesos constructivos de forma totalmente automática, implicando muy pocas modificaciones en sus procedimientos.

Ilustración 1

The screenshot shows a software window with a dark background and white text. At the top right, there are two input fields: 'Entry #' with the value '1' and 'Subentry' with the value 'B'. Below these are four rows of data, each with a label on the left and a text field on the right:

HEADWORD	trepar
Grammatical Category	V-INTR
Semantic Type	MOV-DIR
Definition	Subir [una planta] agarrándose a otra o a un objeto.

At the bottom of the window, there are two buttons: 'OK' on the left and 'Cancel' on the right.

2. La presente versión incluye 295 verbos de movimiento con un total de 521 sentidos, limitándonos a cubrir aquellos verbos cuyos argumentos son realizados únicamente por sintagmas nominales.

Tomando la Gramática Funcional de S.C. Dik (1978, 1980, 1989, 1997) como modelo para una descripción formalizada de la lengua, en este artículo presentamos una estructura más rica del componente léxico. La Gramática Funcional de S.C. Dik consiste básicamente en tres componentes: el lexicón, la estructura de predicación y las reglas de expresión. El lexicón es el componente nuclear de la descripción lingüística, el cual almacena todos los predicados de la lengua (o unidades léxicas de sentido pleno) y sus correspondientes marcos predicativos. El marco predicativo es un esquema estructural que define las propiedades sintáctico-semánticas más importantes del predicado. Combinando los predicados y los términos podemos construir las predicaciones, las cuales definen las propiedades estructurales y funcionales de las expresiones lingüísticas.

Nosotros proponemos un lexicón funcional que no sólo sea un depósito de las propiedades sintáctico-semánticas de los predicados, sino que además explicita las relaciones léxicas que se establecen en la lengua a partir de las funciones semánticas de sus componentes de significado. Como postula el Modelo Lexemático-funcional (Martín Mingorance, 1990, 1995; Faber y Mairal, 1994, 1999; Mairal, 1999), el significado de un predicado viene constituido por la intersección de los niveles sintagmático y paradigmático. El Modelo Lexemático-funcional integra la Gramática Funcional de S.C. Dik y la Semántica Estructural de E. Coseriu. Mientras el primer modelo permite construir principalmente el marco predicativo de cada lexema (eje sintagmático), el segundo modelo estructura el lexicón en campos semánticos (eje paradigmático). Estos dos ejes se conectan por medio de los “esquemas conceptuales cognitivos”, los cuales se interrelacionan con otros esquemas formando una macrorred semántica:

The FLM [Functional Lexematic Model]³ lexicon is a network of lexemes with both a micro- and a macro-structure. Microstructurally, lexemes are characterised as complexes of phonological, syntactic, semantic, pragmatic and phrase structural information. On the level of macrostructure, lexical entries are characterised as interconnected by cohesive, associative and encyclopaedic functions. In this sense, it is conceived of as a grammar, capable of accounting for a speaker's lexical competence. (Faber y Mairal, 1998: 4)

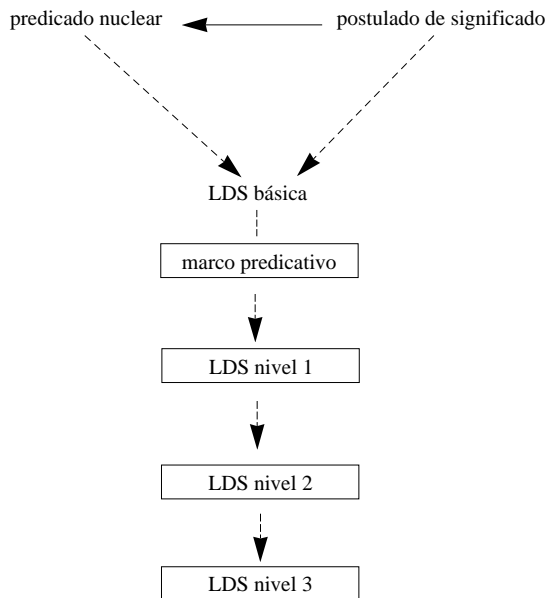
Por consiguiente, esta organización onomasiológica del lexicón posibilita la identificación de las diferentes relaciones de significado que se establecen entre los lexemas.

En FGW, el componente léxico se enriquece considerablemente al asociar a cada predicado no sólo un marco predicativo y un postulado de significado, sino además una lista de densidad semántica (de aquí en adelante, LDS), la cual sirve para establecer la macroestructura del lexicón propuesta por el Modelo Lexemático-funcional. El propósito de este artículo es precisamente describir con detalle cómo se construye

3. El paréntesis es nuestro.

automáticamente la LDS asociada a un determinado predicado a partir de su correspondiente texto definitorio en el DEA. Como observamos en la ilustración 2, el marco predicativo y el postulado de significado de un predicado configuran la materia prima a partir de la cual se elabora la LDS. Por este motivo, será preciso describir brevemente los marcos predicativos y los postulados de significado que construye nuestra aplicación, para posteriormente especificar su influencia en los procesos de creación y expansión de la LDS.

Ilustración 2



2. El Diccionario como fuente de conocimiento de un sistema de procesamiento del lenguaje natural.

A la hora de elaborar el lexicón de un sistema computacional, debemos tener en cuenta que podemos realizar esta tarea de forma manual o automática. La construcción manual de un lexicón computacional es una tarea muy laboriosa y consume mucho tiempo, ya que, por ejemplo, debemos decidir qué especificaciones sintácticas y semánticas son necesarias incluir en cada entrada. Además, el uso de diccionarios convencionales como fuente de información es mucho más seguro para generar información léxica que depender de nuestras propias intuiciones sobre las palabras. Por todo ello, resulta una idea bastante atractiva utilizar diccionarios legibles por la máquina en la construcción de un lexicón computacional. De hecho, el interés compu-

tacional en los diccionarios surgió en el mismo momento en que algunos editores, particularmente en Gran Bretaña, construían formatos lexicográficos más formalizados en cuanto a la sintaxis y el vocabulario utilizados para definir los sentidos de las palabras. Diccionarios como *Longman Dictionary of Contemporary English* (Procter, 1978) o *Collins' COBUILD Dictionary of the English Language* (1987) parecen casi haber sido diseñados para la investigación lingüística computacional, a pesar del hecho de que sus formatos radicalmente nuevos fueron diseñados para los estudiantes de inglés (Guthrie et alii, 1996: 64).

La construcción de lexicones a gran escala para el procesamiento del lenguaje natural a partir de recursos legibles por la máquina (i.e. diccionarios y corpórea) implica la extracción automática, o semi-automática, de la información léxica que contienen estos recursos a través de procedimientos de ingeniería lingüística que permiten el análisis lingüístico de los contenidos. El proceso de análisis está pensado para extraer la información más importante a partir de las definiciones sin tener que producir necesariamente un análisis completo de todo el texto de una definición. Cuando la información léxica que se ha obtenido a partir de los diccionarios convencionales se presenta en un formato estructurado accesible algorítmicamente, p.ej. una base de datos relacional, hablaremos de “base de datos léxica”. En otras palabras, transformar un diccionario legible por la máquina en un diccionario tratable por la máquina implica la transducción del texto del diccionario en un formato estructurado al que pueda tener acceso un programa del procesamiento del lenguaje natural.

El principal inconveniente que se suele encontrar en estas bases de datos léxicas radica en el tipo de información que pueden proporcionar los diccionarios convencionales para su utilización en un sistema del procesamiento del lenguaje natural. No debemos olvidar que los diccionarios impresos se diseñan originariamente para su uso humano, y no para ser explotados por una máquina:

An MRD [machine-readable dictionary]⁴ relies heavily on the user's background linguistic and common-sense knowledge to retrieve and comprehend the information it contains, so a human is not overly concerned when, say, circularity of definition occurs (Ooi, 1998: 31).

Algunos investigadores (Boguraev y Levin, 1993) afirman que existe una línea divisoria entre el conocimiento léxico y el conocimiento del mundo, aunque a veces sea bastante ténue y difusa, y que resulta difícil extraer dicho conocimiento del mundo a partir únicamente del diccionario. En cambio, otros investigadores (Wilks et alii, 1993, 1996) piensan que los diccionarios proporcionan conocimiento no sólo sobre la lengua, sino también sobre el mundo, e incluso otros llegan a afirmar que ambos conocimientos son totalmente inseparables (Cuenca y Hilferty, 1999: 94). No obstante, la mayoría coincide en que los diccionarios proporcionan información detallada a

4. El paréntesis es nuestro.

nivel léxico, pero la información sintáctica no está lo suficientemente especificada (Ooi, 1998: 31) y, además, las entradas léxicas carecen a menudo de la información pragmática que interviene en la determinación de un sentido (Ide y Véronis, 1998: 11). Debido precisamente a que, por limitaciones evidentes de espacio, la definición lexicográfica es muy sucinta, algunos lingüistas piensan que el conocimiento que la máquina extrae de un diccionario no permite una construcción exacta de los conceptos de las palabras. Por ejemplo, a pesar de reconocer la valía de los recursos lexicográficos, Faber y Mairal (1999: 90) prefieren la construcción manual del lexicón por construir entradas léxicas más ricas en detalles y por proporcionar un mayor grado de coherencia en la información presentada.

Precisamente otro de los inconvenientes que se suele señalar en el modelo lexicográfico de la conversión automática es que los diccionarios convencionales presentan numerosas inconsistencias en la información léxica que poseen. En realidad, no podemos esperar que ningún algoritmo sea capaz de recuperar toda la información que contienen los diccionarios elaborados manualmente debido a las inconsistencias, los errores de omisión y los errores de comisión (Fontenelle, 1990: 91).

A pesar de todo, la mayoría de los investigadores afirman que la elaboración manual de un lexicón computacional extenso es poco viable, por lo que la aplicación de un procedimiento de transducción de un diccionario legible por la máquina a un lexicón computacional parece ser el enfoque más recomendable. No obstante, aunque la estrategia a seguir es la automatización del proceso, es inevitable realizar una revisión manual en la fase de posesición para resolver ambigüedades y corregir falsos análisis (Handke, 1995: 234). Introduciendo una serie de correcciones y extensiones podemos mejorar el acceso a la información del diccionario, aumentar su cobertura sintáctica y asegurar el comportamiento adecuado del sistema (Chen y Xu, 1995: 146).

3. El lexicón computacional y la gramática funcional

La mayoría de los modelos computacionales del lenguaje natural basados en la teoría de la Gramática Funcional utilizan un lexicón de tamaño y contenido limitados, generado a mano de una forma bastante *ad hoc*, razón por la cual se reduce enormemente la potencia del sistema y se restringe el alcance de la aplicación. Por ejemplo, PROFLOT (Dik, 1992, 1994), implementación multilingüística en Prolog de la Gramática Funcional, maneja unos lexicones muy poco extensos, precisamente porque se elaboraron manualmente. Estos lexicones almacenan junto a cada predicado un marco predicativo básico y un postulado de significado. A continuación presentamos un ejemplo de los marcos predicativos que se almacenan en estos componentes léxicos (Dik, 1994: 283):

bpredv ([[kick, [act], [[[anim], t, [ag]], [[concr], t, [go]]]])

Esta estructura nos informa de que *kick* es un predicado verbal básico (bpredv) que denota acción (act) cuyo primer argumento posee la función semántica agente (ag) y será realizado por un término animado (anim), mientras el segundo argumento posee la función semántica objetivo/meta (go) y podrá formalizarse en un término concreto (concr).⁵

De forma similar, las definiciones de significado de PROFGLOT presentan la siguiente apariencia (*ibídem*, 290):

mean ([*kick*, T, [[[anim], X, [ag]], [[concr], Y, [go]]]],
 [_, [hit, T, [[anim], X, [ag]], [[concr], Y, [go]]]],
 [_, [[idiom], [the, foot], [instr]], _, _]])

Este postulado de significado expresa formalmente que “X kicks Y means X hits Y with the foot”.

Observamos que la estructura formal del *definiens* y el *definiendum* en las definiciones de significado es la misma: ambos adoptan el carácter de marcos predicativos. Esto implica que si en una estructura subyacente nos encontramos el predicado *kick*, entonces éste puede ser sustituido por la paráfrasis *hit ... with the foot*. En FGW los marcos predicativos y los postulados de significado adoptan igualmente estructuras similares, pero en nuestro caso ambos se construyen de forma totalmente automática.

En cambio, el proyecto ASCOT (Voogt-van Zutphen, 1989) contiene un lexicón a gran escala construido a partir de las entradas léxicas del *Longman Dictionary of Contemporary English*. La limitación se encuentra en este caso en la empobrecida naturaleza del lexicón, ya que sólo almacena marcos predicativos y estructuras de términos. Por ejemplo, el marco predicativo de *mismanage* tiene básicamente la siguiente forma (*ibídem*, 153 y 157):

mismanage-V [T1] (x1: <H>) (x2: <T>)

Es decir, el predicado titular *mismanage* es un verbo (V) monotransitivo (T1) que coge típicamente un sujeto (x1) humano (H) y un objeto (x2) abstracto (T).

Finalmente, y en relación con nuestro concepto de LDS, podemos establecer un cierto paralelismo a nivel paradigmático entre FGW y el proyecto LINKS (Meijs, 1989, 1992; Vossen, 1989, 1994; Meijs y Vossen, 1992), el cual establece automáticamente la interrelación semántica de las palabras (principalmente, nombres) a través de sus definiciones en el *Longman Dictionary of Contemporary English*. El objetivo de este proyecto es construir una base de datos semántica con el propósito de cubrir las necesidades de los sistemas de análisis textual. LINKS traza conexiones entre las palabras con el fin de crear una enorme red multidimensional, donde el significado de una palabra es una función de la posición que ocupa en esta red y de sus conexiones con

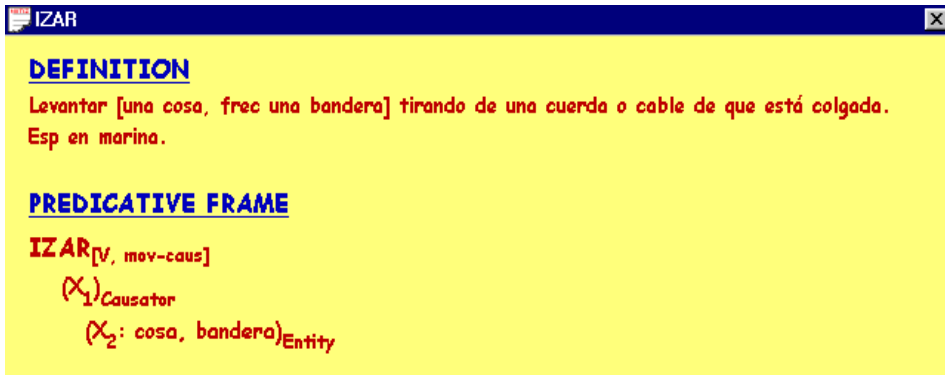
5. La t indica la ranura donde se puede insertar un término, i.e. una expresión que se utiliza para hacer referencia a una entidad.

todas las palabras de la red con las que se encuentra más estrechamente relacionada. En nuestro caso, FGW crea una red relacional del lexicón a partir de las definiciones del DEA.

4. El marco predicativo en FGW

Los marcos predicativos verbales que construye FGW de forma totalmente automática son bastante fieles a los propuestos por Dik. Tomando como ejemplo el marco predicativo de la ilustración 3, vemos que se trata de un esquema estructural que especifica la forma del predicado (*izar*), su categoría sintáctica (verbo), el tipo semántico al que pertenece (movimiento causativo), la valencia cuantitativa (dos argumentos subcategorizados), la valencia cualitativa (las funciones semánticas de *Causativo* y *Entidad*) y las preferencias de selección de cada argumento (*cosa* y *bandera* para el segundo argumento).⁶

Ilustración 3



The screenshot shows a window titled "IZAR" with a yellow background. It contains the following text:

DEFINITION
Levantar [una cosa, frec una bandera] tirando de una cuerda o cable de que está colgada.
Esp en marina.

PREDICATIVE FRAME
IZAR_[V, mov-caus]
(X₁)_{Causator}
(X₂: cosa, bandera)_{Entity}

Si analizamos con detalle los marcos predicativos resultantes en FGW, observamos algunas diferencias con los propuestos por el modelo de Dik. En primer lugar, nuestras preferencias de selección siempre se expresan por medio de predicados de la lengua, a diferencia del modelo funcional que puede además utilizar rasgos de selección de tipo general: p.ej. *animado*, *humano*, *concreto*.

En segundo lugar, nuestros marcos predicativos siempre incluyen las funciones semánticas centrales, aunque en ciertas ocasiones también se especifican algunas fun-

6. Desde un punto de vista computacional, una de las inconsistencias del DEA es precisamente que no todas las definiciones ofrecen información sobre las preferencias de selección, como muestra el ejemplo de la ilustración 3. Ya que esto suele ocurrir principalmente cuando el argumento se formaliza como un sujeto animado, existe la posibilidad de que sea la propia máquina la que corrija estas deficiencias.

ciones semánticas periféricas o “satélites”. De todas formas, no es suficiente decir que los argumentos son los constituyentes nucleares, y por tanto obligatorios para formar una predicación nuclear completa, mientras que los satélites son los constituyentes periféricos, los cuales proporcionan la información opcional de la cláusula. A veces resulta difícil ver si un determinado término aparece en condición de satélite o argumento:

This is especially true of the borderline between arguments and predicate satellites ... since both types of terms tell us something about the internal properties of the SoA [state of affairs]⁷ designated by the core predication (Dik, 1997: 87).

Otra diferencia gira en torno a la tipología semántica utilizada para la clasificación del verbo de movimiento: verbo de dirección, desplazamiento, corporal o causativo. El uso de una subclasificación del verbo de movimiento nos permite asignar con mayor precisión determinados papeles semánticos a los argumentos identificados. Por otra parte, utilizar una subclasificación de carácter cognitivo (Cifuentes, 1998) permite a los investigadores establecer, entre otros paralelismos, los diversos patrones sintácticos en que se puede formalizar un determinado concepto.

Finalmente, se observan algunas diferencias en las funciones semánticas utilizadas tanto en la descripción del marco predicativo como en el postulado de significado. Por razones semánticas y sintácticas, e inspirados por el modelo localista de Rasmussen (1994), hemos preferido aglutinar ciertos papeles semánticos de la Gramática Funcional, como muestra la siguiente tabla:

<u>S.C. Dik</u>	<u>FGW</u>
Agente, Posicionador, Fuerza	Causativo
Agente, Procesado, Meta, Cero	Entidad
Dirección, Recipiente	Objetivo

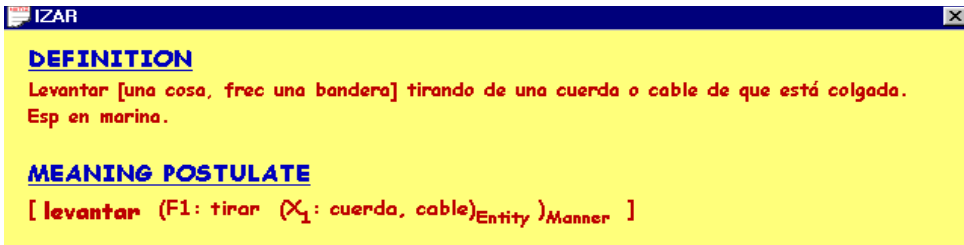
En FGW, y más concretamente en el caso de los verbos de movimiento, el papel *Causativo* hace referencia a la entidad que hace que algo o alguien entre en movimiento, *Entidad* se refiere a la entidad que se mueve y *Objetivo* es la entidad hacia donde algo o alguien se mueve. El resto de las funciones semánticas se utilizan de la misma forma que en la Gramática Funcional. En cuanto a su naturaleza, sólo decir que las funciones semánticas de la Gramática Funcional se ubican entre los papeles conceptuales y las categorías de casos (Faber y Mairal, 1999: 52), mientras que los papeles semánticos del FGW funcionan más bien como papeles conceptuales, los cuales determinan la función o posición de la entidad en el conocimiento del mundo que posee el hablante.

7. El paréntesis es nuestro.

5. El postulado de significado en FGW

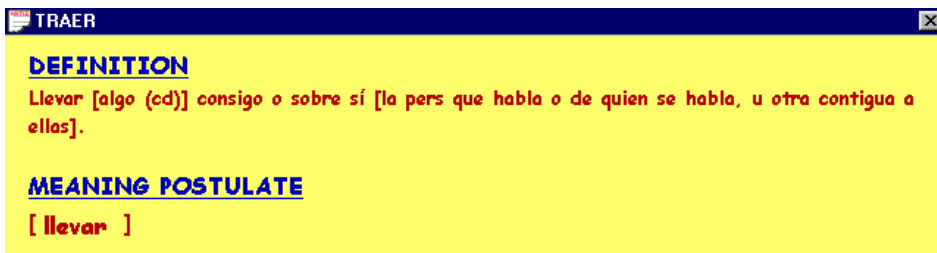
Como el objetivo de los diccionarios es proporcionar las definiciones de las palabras, podemos esperar que sea en el texto definitorio donde la estructura semántica de la lengua resulte más explícita y por tanto más accesible para el estudio de la estructura semántica de las representaciones del conocimiento (Wilks et alii, 1996: 106). En FGW, los postulados de significado se describen con estructuras similares a las predicaciones subyacentes (ilustración 4).

Ilustración 4



Preferimos utilizar el término “postulado de significado” en lugar de “definición de significado”, el cual podría parecer a primera vista más apropiado, debido a que este último implica una relación de sinonimia total entre el predicado y su Paráfrasis, y no todas las representaciones de significado construidas por FGW pueden ajustarse a este principio, ya que a veces nuestros postulados proporcionan una especificación parcial del significado del predicado (ilustración 5).

Ilustración 5



Existen dos nociones fundamentales que se aplican a lo largo de todos los niveles de la descripción semántica de la cláusula: los satélites y los operadores. Mientras los satélites indican rasgos léxicos en una posición periférica, los operadores por su parte indican rasgos gramaticales. Los satélites utilizados en FGW coinciden con los propuestos por la Gramática Funcional para la descripción de las predicaciones central y extendida, indicando su naturaleza con las variables F1 y F2 respectivamente. Por otra

parte, la presente versión también comparte con la Gramática Funcional los operadores de conjunción y disyunción y los operadores de aspecto (ingresivo, prospectivo, etc) y negación, además de incorporar el operador de causatividad (ilustraciones 6 y 7).

Ilustración 6

TUMBAR

DEFINITION
 Derribar o hacer caer [algo o a alguien]. Tb fig.

MEANING POSTULATE
 [**derribar** _{v caus} **caer**]

Ilustración 7

ARRANCAR

DEFINITION
 (Taur) Echar a correr [el toro] para acometer. Frec pr.

MEANING POSTULATE
 [**ingr correr** (F2: acometer)_{purpose}]

A través de las evidencias que nos proporcionan la aparición de estos operadores y satélites, podemos llegar a afirmar que nuestros postulados de significado adoptan la forma de predicaciones extendidas, las cuales sitúan la predicación central respecto a parámetros espacio-temporales y cognitivos.

Llegado a este punto podemos ahora decir que una de las razones por la que hemos elegido el DEA es que en este diccionario se procura seguir con mayor rigor que en otros diccionarios de la lengua española el modelo de definición sinonímica. Para ello, el DEA utiliza un método lexicográfico que permite distinguir claramente los “elementos habituales del contorno” del definiendum y los “elementos sustanciales del contenido” de esa palabra (Seco et alii, 1999: xxii). De esta forma se presenta la fórmula definitoria con mayor precisión, ya que la verdadera definición se deslinda del entorno sintagmático del término titular. Por tanto, la información necesaria para la construcción del marco predicativo se encuentra en el contorno de la palabra titular, presentada siempre entre corchetes, mientras que el texto restante de la definición se convierte en el aducto del proceso de construcción del postulado de significado.⁸

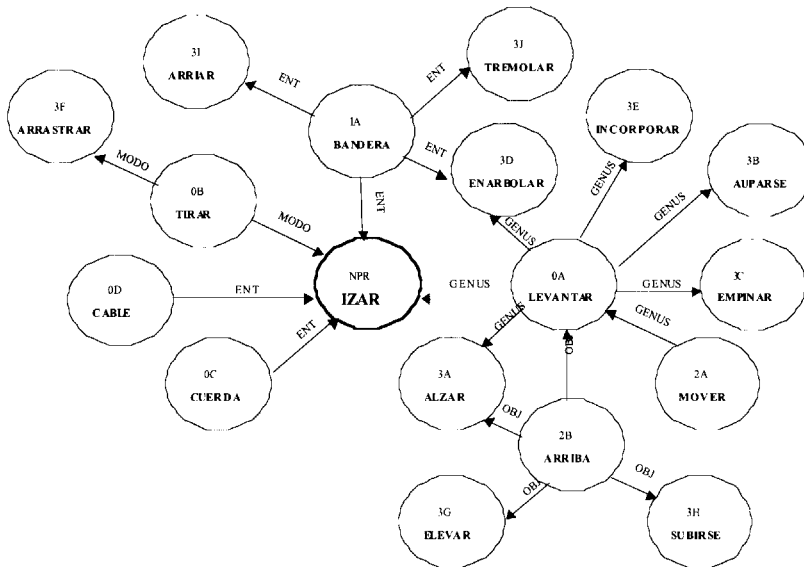
8. Precisamente una de las mejoras de este diccionario recae en el tratamiento que reciben las preferencias selectivas frente al de otros diccionarios de la lengua española, en los que o bien se omite dicha información, o bien aparece codificada sólo en algunas categorías gramaticales mediante la fórmula *Referido a* (Ballesteros y Jiménez, 1998: 241).

6. La lista de densidad semántica

Al igual que el Modelo Lexemático-funcional, nuestro objetivo es expandir y enriquecer el componente léxico de la Gramática Funcional, elaborando una estructuración semántica del lexicon de la lengua, en nuestro caso por medio de las LDS, i.e. una lista de los predicados asociados conceptualmente a un determinado predicado titular. Tiene lugar una factorización de los componentes de significado utilizando los diccionarios como textos que dan forma a nuestro conocimiento general sobre las palabras. También compartimos con el Modelo Lexemático-funcional, a diferencia del tesoro, el modo ascendente de construir la estructuración del lexicon: a partir de los datos que se extraen de los textos definitorios podemos explicitar las asociaciones o relaciones semánticas que se crean en el léxico (Faber y Mairal, 1999: 60).

En el lexicon del FGW, los predicados forman una compleja red en la que cada uno de ellos se conecta con otros a través de diferentes relaciones. Esta red conceptual se representa a través de un grafo dirigido y etiquetado, cuyos nodos denotan los predicados⁹ y las aristas indican las relaciones conceptuales o asociativas que se establecen entre los nodos. Dichas aristas van etiquetadas con una función semántica o con la relación *genus* (ilustración 8).

Ilustración 8

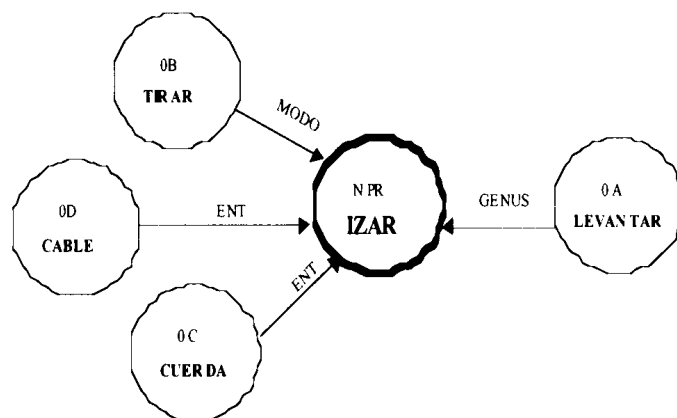


9. Para que se refleje mejor la idea de que los miembros que configuran una LDS son en realidad conceptos, y no simples unidades ortográficas de la lengua, en futuras versiones los nodos se referenciarán por una palabra titular de una entrada léxica y un número de acepción. Ya que el lexicon de la presente versión de FGW sólo incluye las entradas léxicas de los verbos de movimiento, la mayoría de los casos de ambigüedad léxica que tienen lugar durante el proceso de elaboración de las LDS se resuelven simplemente asignando el sentido más usual.

Observamos que, a diferencia del Modelo Lexemático-funcional, este tipo de redes conceptuales no presenta una arquitectura jerárquica. A pesar de todo, nuestros grafos poseen la misma eficiencia en la organización de la información: es decir, presentan la mayor cantidad de información posible de la forma más económica, permitiendo la inferencia léxica al navegar a través de la red conceptual.

El desarrollo de una red conceptual se concibe como una extensión de la LDS básica, la cual se forma a partir del postulado de significado. Supongamos que deseamos construir la LDS básica de *izar 1*, cuyo postulado de significado se presentaba en la ilustración 4. La LDS básica la constituyen todos aquellos predicados que aparecen en el postulado de significado y que contribuyen significativamente a la descripción del sentido de la palabra titular, a la cual denominamos predicado nuclear. Por ello, no se tienen en cuenta los predicados que pertenecen a la *stop list* (Wilks et alii, 1996: 191) de nuestra aplicación, que en nuestro caso contiene aquellos predicados de sentido genérico que aparecen con bastante frecuencia en los textos definitorios del DEA (p.ej. *cosa, lugar, persona, etc.*). La ilustración 9 muestra en forma de grafo la LDS básica asociada a *izar 1*.

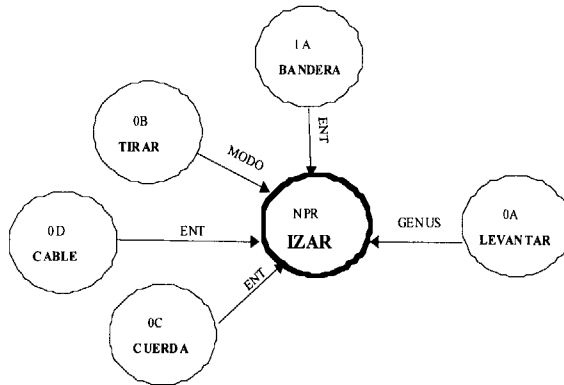
Ilustración 9



A partir de ahora, se produce un proceso de expansión en el que se distinguen diversos niveles.¹⁰ En un primer nivel, la microrred se amplía con las preferencias semánticas del marco predicativo del predicado nuclear (ilustración 10).

10. En los grafos de las ilustraciones, los diferentes niveles de expansión de la LDS se reflejan en el primer dígito del código de los nodos, siendo la LDS básica el nivel cero.

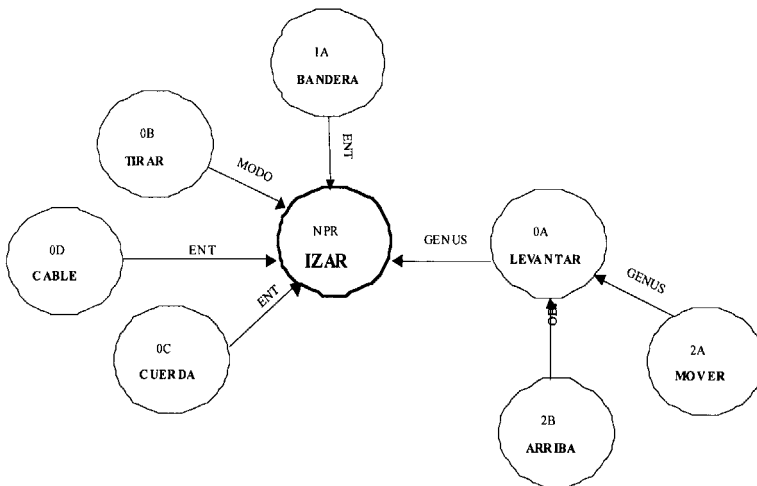
Ilustración 10



Las preferencias de selección se incluyen como nodos o conceptos de la red porque son parte de la semántica léxica orientada a la ontología; es decir, permiten designar un artículo léxico a un determinado concepto siempre y cuando sus argumentos sean realizados por términos de determinados tipos semánticos (Nirenburg y Levin, 1992: 11).¹¹

En un segundo nivel de expansión, la estructura conceptual se amplía con la LDS básica asociada a cada uno de los *genus* identificados en el postulado de significado del predicado nuclear (ilustración 11).

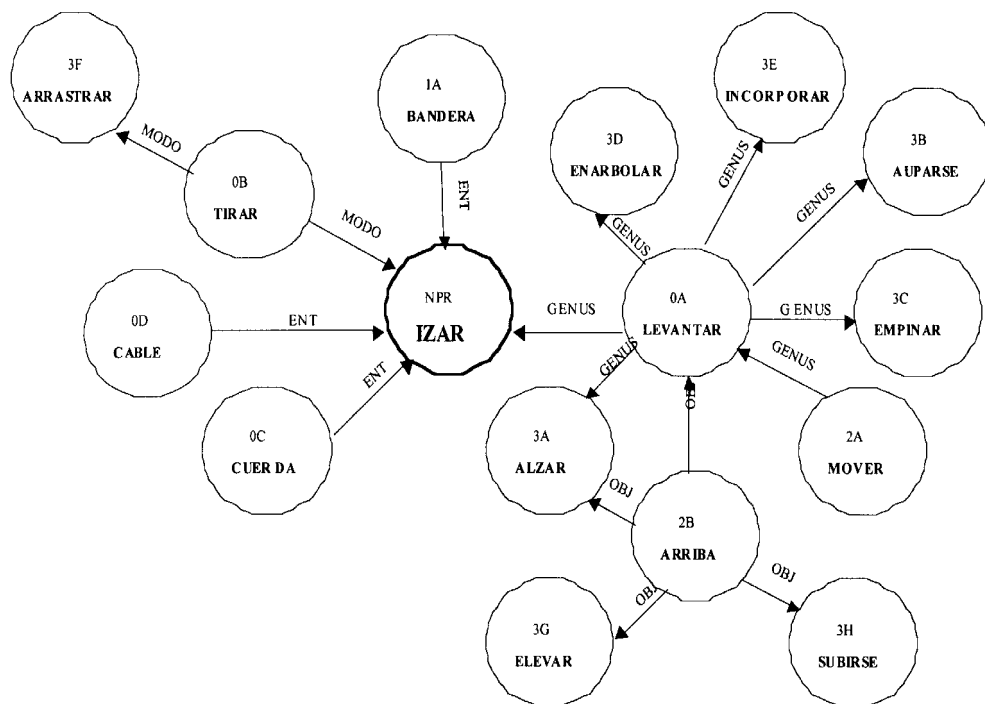
Ilustración 11



11. La integración del postulado de significado y el marco predicativo en la LDS está en la línea de las últimas tendencias funcionales en el análisis léxico. Recordemos que en el modelo funcional de S.C. Dik tanto el *definiendum* como el *definiens* de una definición de significado adoptan las formas de marcos predicativos.

Finalmente, la red conceptual se extiende con miembros periféricos identificados a través de dos fases. En una primera fase, se buscan en todas las LDS básicas del lexicon, independientemente de la parte del discurso a la que pertenezca el término titular, los miembros que coincidan con alguno de los nodos del grafo, además de con la correspondiente función semántica de una de sus aristas (ilustración 12).¹²

Ilustración 12



En una segunda fase, la red conceptual se amplía con aquellos predicados del lexicon que posean en su marco predicativo una preferencia semántica que coincida con un nodo del primer nivel, además de compartir su correspondiente función semántica, y siempre y cuando el tipo semántico de la palabra titular sea el mismo que el del predicado nuclear. Todo ello da como resultado final el grafo de la ilustración 8.

FGW permite que el usuario conozca esta microestructura conceptual a través de la siguiente tabla:¹³

12. En esta fase, se exceptúan los nodos del grafo con la función de *genus* identificados en el segundo nivel de expansión por resultar muy genéricos conceptualmente.

13. La última columna de esta tabla identifica además la dirección de la arista, ya que primero se indica el nodo origen y luego el o los nodos destinos, donde *NPr* referencia el predicado nuclear.

CÓDIGO	NODO	ETIQUETA	ARISTAS
0A	levantar	GENUS	0A-NPr/3A/3B/3C/3D/3E
0B	tirar	MODO	0B-NPr/3F
0C	cuerda	ENTIDAD	0C-NPr
0D	cable	ENTIDAD	OD-NPr
1A	bandera	ENTIDAD	1A-NPr/3I/3D/3J
2A	mover	GENUS	2A-0A
2B	arriba	OBJETIVO	2B-0A/3A/3G/3H
3A	alzar		
3B	auparse		
3C	empinar		
3D	enarbolar		
3E	incorporar		
3F	arrastrar		
3G	elevar		
3H	subirse		
3I	arriar		
3J	tremolar		

La distinción explícita de estos tres niveles de expansión de la LDS básica nos permite el estudio del concepto de distancia semántica,¹⁴ además de su influencia en la ubicación de un predicado en un determinado paradigma cognitivo.

Con respecto al tipo de términos utilizados en la estructuración conceptual del léxico, el Modelo Lexemático-funcional utiliza básicamente los *genus* de las definiciones para elaborar jerarquías léxicas a través del método de la “descomposición léxica gradual”. Este principio establece un sistema de interrelación de las entradas léxicas en el cual el predicado del *definiens* de un postulado de significado puede convertirse en el *definiendum* de otro postulado de significado (Dik, 1997: 100-101). De esta forma, su organización conceptual, al igual que la de WordNet (Miller, 1995; Fellbaum, 1998), estructura los predicados verbales basándose principalmente en la relación de troponimia y sinonimia. En cambio, nuestro modelo utiliza tanto el *genus* como los componentes del *differentiae* para interrelacionar los predicados del léxico, por lo que se establecen otras muchas relaciones semántico-conceptuales. En este sentido, FGW se parece bastante a EuroWordNet (Vossen, 1998; Alonge et alii, 1998). Por un lado, ambos modelos son capaces de establecer relaciones entre predicados con diferentes partes del discurso. Por otra parte, se introducen relaciones nuevas que permiten codificar datos sobre argumentos y adjuntos que están fuertemente implicados

14. Algunos experimentos psicológicos en torno a las redes semánticas como modelo de la memoria humana han descubierto que la información directamente asociada a un nodo es más fácilmente accesible que la información que tiene que ser transferida a otros nodos por medio de la herencia de las propiedades (Handke, 1995: 95).

en el significado de un verbo, ya que la asociación que se establece entre ellos se refleja claramente en las definiciones lexicográficas (Alonge et alii, 1998: 101-102).

7. Conclusión

Nuestro proyecto propone un modelo de lexicón bipartito, en el que el marco predicativo expresa las propiedades formales, mientras que el postulado de significado, y por extensión la LDS, expresa las propiedades semánticas del predicado. A diferencia de lo que ocurre en la Gramática Funcional de S.C. Dik, FGW asigna a los postulados de significado un papel tan primordial en el lexicón como el que tienen los marcos predicativos. La utilización de las LDS en el proceso de interpretación del aducto servirá para identificar las relaciones conceptuales que se establezcan entre los predicados de un texto. En definitiva, las LDS son un medio de capturar el conocimiento extralingüístico necesario que nos permita establecer las relaciones paradigmáticas entre las palabras del aducto.

La aportación del lexicón de FGW en el marco de la lexicografía computacional se podría definir como la construcción de un componente semántico fuerte implementado en una base de datos léxica enriquecida. Disponemos de un componente léxico más potente que una simple base de datos léxica, al permitir la construcción de una red conceptual que interconecte los distintos sentidos del lexicón por medio de las LDS, pero además nuestro diseño es menos complejo que el de una base de conocimiento, ya que no es necesario diseñar un modelo ontológico ni definir un grupo restringido de primitivos semánticos que funcionen como conceptos ontológicos. En definitiva, nuestro modelo de lexicón intenta combinar las ventajas de ambos enfoques.

Al permitir que la información extraída de las entradas léxicas del DEA se almacene en una base de datos, podemos tener acceso sistemático a este conocimiento. Por consiguiente, un lexicón de estas características permite diversos tipos de aplicaciones. Por ejemplo, puede utilizarse en los sistemas de análisis y generación del lenguaje natural. Además, con la ayuda de un lexicón similar en otra lengua, podemos construir un diccionario bilingüe, el cual podrá ser utilizado en la traducción tanto asistida como automática. Apoyándonos en ciertas herramientas auxiliares, nos ayudaría igualmente en tareas de investigación lingüística, donde destacaríamos dos vertientes: desde la perspectiva de la lingüística teórica, podríamos establecer más puntos de contacto entre la estructura gramatical (marco predicativo) y la definición léxica (postulado de significado y LDS) de un mismo predicado, o bien, con la ayuda de un diccionario bilingüe y/o un tesoro, podríamos establecer similitudes y diferencias entre marcos predicativos o LDS de predicados equivalentes en dos lenguas; desde la perspectiva del procesamiento del lenguaje natural, y más concretamente en el marco de la traducción automática y la recuperación de información, podríamos comprobar la eficacia de la complementariedad del marco predicativo y la LDS en la resolución de la ambigüedad léxica.

Bibliografía

- ALONGE, Antonietta, CALZOLARI, Nicoletta, VOSSEN, Piek, BLOKSMA, Laura, CASTELLÓN, Irene, MARTÍ, María Antonia y PETERS, Wim (1998) "The linguistic design of the EuroWordNet database". En *Computers and the Humanities* 32, nº 2-3, pp. 91-115.
- BALLESTEROS VALLADARES, Laura y JIMÉNEZ BRIONES, Rocío (1998) "The 'tolerance' of Spanish dictionaries regarding selection restrictions". En José Luis Cifuentes Honrubia (ed.), *Estudios de Lingüística Cognitiva*, 2 vols, Alicante, Universidad de Alicante, pp. 235-245.
- BOGURAEV, Branimir y LEVIN, Beth (1993) "Models for lexical knowledge bases". En James Pustejovsky (ed.), *Semantics and the Lexicon*, Dordrecht, Kluwer Academic Publishers, pp. 325-340.
- CHEN, Si-Qing y XU, Luomai (1995) "A full and efficient machine tractable dictionary for natural language processing: a revised version of the CUVOALD". En *Computers and the Humanities* 28, nº 3, pp. 141-152.
- CIFUENTES HONRUBIA, José Luis (1998) "Verbos con incorporación conceptual direccional". En José Luis Cifuentes Honrubia (ed.), *Estudios de Lingüística Cognitiva*, 2 vols, Alicante, Universidad de Alicante, pp. 479-505.
- Collins COBUILD English Language Dictionary* (1987) Londres, Collins.
- CUENCA, Maria Josep y HILFERTY, Joseph (1999) *Introducción a la Lingüística Cognitiva*, Barcelona, Ariel.
- DIK, Simon C. (1978) *Functional Grammar*, Dordrecht, Foris.
- _____ (1980) *Studies in Functional Grammar*, Londres-Nueva York, Academic Press.
- _____ (1989) *The Theory of Functional Grammar. Part I: The Structure of the Clause*, Dordrecht, Foris.
- _____ (1992) *Functional Grammar in Prolog. An Integrated Implementation for English, French and Dutch*, Berlín-Nueva York, Mouton de Gruyter.
- _____ (1994) "The lexicon in a computational functional grammar". En B.T. Sue Atkins y Antonio Zampolli (eds.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, pp. 279-294.
- _____ (1997) *The Theory of Functional Grammar. Part I: The Structure of the Clause*, Berlín-Nueva York, Mouton de Gruyter.
- FABER, Pamela y MAIRAL, Ricardo (1994) "Methodological underpinnings for the construction of a functional lexicological model". En *Miscelánea: A Journal of English and American Studies* 15, pp. 193-217.
- _____ (1998) "Methodological criteria for the elaboration of a functional lexicon-based grammar of the semantic domain of cognitive verbs". En Hella Olbertz, Kees Hengeveld y Jesús Sánchez García (eds.), *The Structure of the Lexicon in Functional Grammar*, Amsterdam-Philadelphia, John Benjamins, pp. 3-24.
- _____ (1999) *Constructing a Lexicon of English Verbs*, Berlín-Nueva York, Mouton de Gruyter.

- FELLBAUM, Christiane (1998) "A semantic network of English: the mother of all wordnets". En *Computers and the Humanities* 32, nº 2-3, pp. 209-220.
- FONTENELLE, Thierry (1992) "Automatic extraction of lexical-semantic relations from dictionary definitions". En *EURALEX '90 Proceedings. Fourth International Congress on Lexicography*, Barcelona, Bibliograf, pp. 89-103.
- GUTHRIE, Louise, PUSTEJOVSKY, James, WILKS, Yorick y SLATOR, Brian M. (1996) "The role of lexicons in natural language processing". En *Communications of the ACM*, vol. 39, nº 1, pp. 63-72.
- HANDKE, Jürgen (1995) *The Structure of the Lexicon: Human versus Machine*, Berlín-Nueva York, Mouton de Gruyter.
- IDE, Nancy y VÉRONIS, Jean (1998) "Introduction to the special issue on word sense disambiguation: the state of the art". En *Computational Linguistics* 24, nº 1, pp. 1-40.
- MAIRAL, Ricardo (1999) "El componente lexicón en la Gramática Funcional". En Christopher Butler et alii (eds.), *Nuevas Perspectivas en Gramática Funcional*, Barcelona, Ariel, pp. 41-98.
- MARTÍN MINGORANCE, Leocadio (1990) "Functional grammar and lexematics in lexicography". En Jerzy Tomaszczyk y Barbara Lewandowska-Tomaszczyk (eds.), *Meaning and Lexicography*, Amsterdam-Philadelphia, John Benjamins, pp. 227-253.
- _____ (1995) "Lexical logic and structural semantics. Methodological underpinnings in the structuring of a lexical database for natural language processing". En Amalia Marín Rubiales (ed.) (1998) *El Modelo Lexemático Funcional. El Legado Lingüístico de Leocadio Martín Mingorance*, Granada, Universidad de Granada, pp. 177-197.
- MEIJS, Willem (1989) "Spreading the word: knowledge-activation in a functional perspective". En John H. Connolly y SIMON C. Dik (eds.), *Functional Grammar and the Computer*, Dordrecht, Foris, pp. 201-215.
- _____ (1992) "The expanding lexical universe: extracting taxonomies from machine-readable dictionaries". En *EURALEX '90 Proceedings. Fourth International Congress on Lexicography*, Barcelona, Bibliograf, pp. 119-128.
- MEIJS, Willem y VOSSEN, Piek (1992) "In so many words: knowledge as a lexical phenomenon". En James Pustejovsky y Sabine Bergler (eds.), *Lexical Semantics and Knowledge Representation: First SIGLEX Workshop*, Berlín, Springer-Verlag, pp. 137-153.
- MILLER, George A. (1995) "WordNet: a lexical database for English". En *Communications of the ACM*, vol. 38, nº 11, pp. 39-41.
- NIRENBURG, Sergei y LEVIN, Lori (1992) "Syntax-driven and ontology-driven lexical semantics". En James Pustejovsky y Sabine Bergler (eds.), *Lexical Semantics and Knowledge Representation: First SIGLEX Workshop*, Berlín, Springer-Verlag, pp. 5-20.

- OOI, Vincent B. Y. (1998) *Computer Corpus Lexicography*, Edinburgh, Edinburgh University Press.
- PROCTER, Paul (1978) *Longman Dictionary of Contemporary English*, Harlow, Longman.
- RASMUSSEN, Lone Schack (1994) "Semantic functions in perspective – reconsidering meaning definitions". En Elisabeth Engberg-Pedersen et alii (eds.), *Function and Expression in Functional Grammar*, Berlín-Nueva York, Mouton de Gruyter, pp. 41-63.
- SECO, Manuel, ANDRÉS, Olimpia y RAMOS, Gabino (1999) *Diccionario del Español Actual*, 2 vols., Madrid, Aguilar.
- VOOGT-VAN ZUTPHEN, Hetty (1989) "Towards a lexicon of Functional Grammar". En John H. Connolly y Simon C. Dik (eds.), *Functional Grammar and the Computer*, Dordrecht, Foris, pp. 151-176.
- VOSSEN, Piek (1989) "The structure of lexical knowledge as envisaged in the LINKS-project". En John H. Connolly y Simon C. Dik (eds.), *Functional Grammar and the Computer*, Dordrecht, Foris, pp. 177-199.
- ____ (1994) "The end of the chain: Where does decomposition of lexical knowledge lead us eventually?". En Elisabeth Engberg-Pedersen et alii (eds.), *Function and Expression in Functional Grammar*, Berlín-Nueva York, Mouton de Gruyter, pp. 11-39.
- ____ (1998) "Introduction to EuroWordNet". En *Computers and the Humanities* 32, nº 2-3, pp. 73-89.
- WILKS, Yorick, FASS, Dan, GUO, Cheng-ming, MCDONALD, James E., PLATE, Tony y SLATOR, Brian M. (1993) "Providing machine tractable dictionary tools". En James Pustejovsky (ed.), *Semantics and the Lexicon*, Dordrecht, Kluwer Academic Publishers, pp. 341-401.
- WILKS, Yorick A., SLATOR, Brian M. y GUTHRIE, Louise M. (eds.) (1996) *Electric Words. Dictionaries, Computers and Meanings*, Cambridge: Massachusetts, MIT Press.