

LA CONSTRUCCIÓN AUTOMÁTICA DE LOS MARCOS PREDICATIVOS DE LOS VERBOS DE MOVIMIENTO A PARTIR DEL DICCIONARIO DEL ESPAÑOL ACTUAL

Periñán Pascual, José Carlos

*Profesor Adjunto
Unidad Central de Idiomas
Universidad Católica San Antonio de Murcia
Campus de los Jerónimos s/n
30107 Guadalupe (Murcia)
Tlfn: 968278800
Fax: 968307066
E-mail: jcperinan@pdi.ucam.edu*

(Recibido Abril 2003; aceptado Junio 2003)

BIBLID [1133-682X (2002-2003) 10-11;225-240]

Resumen

Functional Grammar Workbench (FGW) se concibe como un laboratorio virtual de ingeniería lingüística dentro del marco de la Gramática Funcional de Simon C. Dik. Uno de los métodos más utilizados para la construcción de lexicones computacionales extensos es el análisis automático de diccionarios legibles por la máquina. La presente versión de FGW permite a este respecto la construcción automática de los marcos predicativos de los verbos de movimiento en español a partir de las entradas léxicas del Diccionario del Español Actual (1999) de Manuel Seco, Olimpia Andrés y Gabino Ramos. Una de las razones por la que hemos elegido este diccionario se encuentra en el hecho de que se procura seguir con mayor rigor el modelo de definición sinonimica: la verdadera definición se deslinda del entorno sintagmático del término titular.

Palabras clave: lexicografía computacional, diccionario, lexicón, marco predicativo, Gramática Funcional.

Abstract

Functional Grammar Workbench (FGW) can be conceived as a virtual linguistic engineering laboratory within the framework of Simon C. Dik's Functional Grammar. One of the best methods for the construction of large computational lexicons is the automatic analysis of machine-readable dictionaries. In that respect, the present version of FGW is able to build the predicate frames of Spanish motion verbs from the lexical entries in Diccionario del Español Actual (1999) by Manuel Seco, Olimpia Andrés and Gabino Ramos. One of the reasons for the selection of this dictionary is found in its lexicographical method: the definition text is presented with much more precision than in other Spanish dictionaries, which usually omit the information concerning selectional restrictions or present them in a rather restricted way.

Key words: computational lexicography, dictionary, lexicon, predicate frame, Functional Grammar.

Résumé

Functional Grammar Workbench (FGW) est conçu comme un laboratoire virtuel de génie linguistique, dans le cadre de la Grammaire Fonctionnelle de Simon C. Dik. Une des méthodes les plus utilisées pour la construction de lexicons de computation étendus est l'analyse automatique de dictionnaires lisibles par la machine. Cette version de FGW permet la construction automatique des cadres prédicatifs des verbes de mouvement en espagnol à partir des entrées lexique du Diccionario del Español Actual (1999) de Manuel Seco, Olimpia Andrés et Gabino Ramos.

C'est parce qu'il essaie d'établir avec la rigueur la plus grande le modèle de définition synonymique : la définition véritable se sépare de l'environnement syntagmatique du terme titulaire, que nous avons choisi ce dictionnaire.

Mots clés: Lexicographie computationnelle, dictionnaire, lexicons, cadre prédicatif, Grammaire Fonctionnelle.

Sumario

1. Introducción. 2. El diccionario como fuente de conocimiento. 3. Functional Grammar Workbench 2.1 y la Gramática Funcional de Dik. 4. La tipología semántica del verbo de movimiento. 5. La construcción del marco predicativo. 5.1 La identificación de los argumentos y la asignación de las funciones sintácticas y las preferencias selectivas. 5.2. La asignación de las funciones semánticas. 6. Las futuras líneas de investigación.

1. Introducción

La lingüística computacional, o lingüística informática, está considerada como una de las ramas de la inteligencia artificial, y en general de la ingeniería, de ahí que aparezca también con el nombre de *ingeniería lingüística*¹. Todos los campos de la inteligencia artificial, ya sea la robótica, la construcción de sistemas expertos o el procesamiento del lenguaje natural, se ocupan de la investigación y sistematización de una capacidad cognitiva², que en el caso de la lingüística computacional resulta ser la capacidad lingüística. Uno de los objetivos más ambiciosos de la lingüística computacional es dotar a la máquina de competencia lingüística con el objeto de automatizar ciertas actividades humanas en lenguaje natural (Cassen et alii (1991: 411)), siendo la traducción automática uno de los campos más investigados del procesamiento del lenguaje natural.

Dentro de este contexto, la lingüística computacional proporciona al lingüista las herramientas necesarias para realizar una descripción homogénea y completa de una lengua. Se trata de instrumentos que permiten procesar datos y extraer información que permita establecer generalizaciones o determinar tendencias estadísticas. A este respecto, *Functional Grammar Workbench 2.1* (de aquí en adelante, FGW) puede ser considerado como un laboratorio virtual de ingeniería lingüística con fines investigadores y docentes dentro del marco de la Gramática Funcional de S. C. Dik (1978) (1980) (1989a) (1997). Actualmente, y desde la perspectiva de la lexicografía computacional, nuestra principal preocupación es la construcción automática³ de un lexicón computacional a gran escala a partir de un diccionario legible por la máquina para su aplicación en el procesamiento del lenguaje natural. Con este propósito, la presente versión de FGW ha conseguido construir de forma automática los

¹ Moure y Llisterrí (1996) distinguen entre la lingüística computacional (o procesamiento del lenguaje natural) y la ingeniería lingüística (o industrias de la lengua). Desde su punto de vista, mientras que la primera desarrolla las técnicas, la segunda construye las aplicaciones comerciales. No obstante, reconocen que a veces la frontera que delimita estos conceptos es muy difusa.

² Haugeland (1998: 30) llega a definir la inteligencia artificial como una especie de esencia destilada de la ciencia cognitiva.

³ Aunque FGW requiere en una fase inicial la introducción por parte del lexicógrafo de la información que contiene el DEA, el proceso de construcción de los marcos predicativos a partir de esta información es realizado por la máquina sin ningún tipo de intervención humana. Podemos imaginar con relativa facilidad que si tuviéramos el DEA en un formato legible por la máquina, o incluso utilizáramos un programa convencional de OCR y escaneáramos el diccionario para obtenerlo en formato ASCII, nuestro programa podría implementarse con muy pocas modificaciones para llevar a cabo los mismos procesos constructivos de forma totalmente automática.

marcos predicativos correspondientes a los verbos de movimiento del español⁴ a partir de sus entradas léxicas en el *Diccionario del Español Actual* (1999) de M. Seco, O. Andrés y G. Ramos (de aquí en adelante, DEA).

2. El diccionario como fuente de conocimiento

El proceso de construcción de un lexicón computacional, o en nuestro caso los marcos predicativos de sus predicados, puede realizarse siguiendo tres enfoques metodológicos diferentes. Una primera posibilidad es construirlo manualmente a partir de nuestros conocimientos como hablantes. Se trata de un método que consume bastante tiempo, ya que es preciso determinar qué entradas vamos a incluir en el lexicón, además de todas sus propiedades sintácticas y semánticas.⁵

En segundo lugar, el lingüista puede construir el lexicón manualmente pero en esta ocasión utilizando la información que le aportan las obras lexicográficas. Esta es la línea que sigue actualmente el Modelo Lexemático-funcional,⁶ el cual utiliza los diccionarios como textos que dan forma a nuestro conocimiento compartido sobre la lengua y el mundo:

The purpose of dictionaries is to provide definitions of senses of words and, in so doing, supply knowledge about not just language, but the world. (Wilks et alii (1993: 341))

El uso de diccionarios convencionales es un método mucho más seguro para generar información léxica que depender de las intuiciones del ingeniero lingüista acerca del comportamiento sintáctico o la representación semántica de las palabras. Dentro de los diccionarios se encuentra implícita una riqueza de información que puede ser extraída para su utilización por un sistema del procesamiento del lenguaje natural (PLN).

En realidad, se pretende obtener un diccionario tratable por la máquina a través del proceso de transducción de un diccionario legible por la máquina, dando como resultado una base de datos léxica almacenada en un formato estructurado a la que la máquina pueda tener acceso algorítmicamente. De todas las formas de texto, es en los diccionarios donde la estructura semántica de la lengua es más explícita, y por tanto más accesible para la representación del conocimiento. Esta característica se acentúa todavía más en el caso de los diccionarios legibles por la máquina a los que los ingenieros lingüistas pueden acceder; la mayoría de ellos están dirigidos a los estudiantes de inglés, por lo que su nivel de formalización interna es mucho mayor en cuanto a la información sintáctica, morfológica y semántica.

⁴ La presente versión incluye 295 verbos de movimiento con un total de 521 sentidos, limitándose a cubrir verbos primarios del grupo A en términos de la clasificación de Dixon (1991: 88), según la cual estos verbos sólo pueden admitir sintagmas nominales como realizaciones de las funciones de sujeto y objeto directo.

⁵ A pesar de todo, la mayoría de los modelos computacionales del lenguaje natural basados en la teoría de la Gramática Funcional de S. C. Dik utilizan lexicones que se crean a través de este primer método de una forma bastante *ad hoc* (Voogt-van Zutphen (1989: 152)).

⁶ El Modelo Lexemático-funcional, desarrollado inicialmente por Martín Mingorance durante los años 80 y 90, integra la Gramática Funcional de S. C. Dik y la Lexemática de E. Coseriu. Mientras el primer modelo permite construir principalmente el marco predicativo de cada lexema (eje sintagmático), el segundo modelo estructura el lexicon 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.

Una tercera estrategia que se puede adoptar es la construcción automática de un lexicón computacional a través de una serie de algoritmos que permitan leer las entradas léxicas de un diccionario electrónico ya existente. Si pasamos la aplicación informática por todo el diccionario, el resultado final implicará la elaboración de un lexicón muy extenso. Es deseable aplicar métodos computacionales para la construcción de un lexicón tratable por la máquina, principalmente porque implica una minimización del esfuerzo humano y una reducción del tiempo de trabajo. Además, el intento de construir lexicones para sistemas computacionales tomando como fuente de información los diccionarios electrónicos disponibles conlleva la utilización de métodos más empíricos.

FGW se ubica dentro de esta última línea de investigación. A pesar de que no trabajamos con una versión electrónica del DEA, la adaptación de nuestro sistema a ese tipo de fuente léxica implicaría un coste mínimo de procesamiento adicional debido a la fidelidad de la información extraída de este diccionario. Ciertos lingüistas, como Faber y Mairal (1999: 89-90), muestran serias dudas sobre la utilidad de la extracción automática de información semántica a partir únicamente de un diccionario; la propia limitación que tiene el lexicógrafo de describir en poco más de una línea el sentido de un lexema puede convertir a la obra lexicográfica en una fuente de conocimiento insuficiente para el procesamiento del lenguaje natural. El uso de un diccionario legible por la máquina como fuente para la adquisición de conocimiento léxico no ha dado grandes frutos después de muchos años de investigación y proyectos, debido principalmente a una serie de inconvenientes que diversos investigadores han señalado (Ide y Véronis (1994), Wilks et alii (1996), Sierra y McNaught (2000)):

- * Los diccionarios presentan una información incompleta. Esta falta de información se debe principalmente a dos razones: por una parte, los diccionarios se han diseñado para ser utilizados por humanos, los cuales, a diferencia de las máquinas, poseen un conocimiento lingüístico general; por otra parte, los diccionarios están limitados por unas restricciones de espacio.
- * Los diccionarios presentan a menudo inconsistencias en la información que contienen, siendo la circularidad un claro ejemplo. Dos posibles motivos se alegan ante tal deficiencia: el diccionario ha sido compilado por varios lexicógrafos, o bien el lexicógrafo no ha sido capaz de reconocer ciertas similitudes entre diversas entradas léxicas. Precisamente, por la falta de homogeneidad y consistencia en las definiciones, la extracción automática de la información implícita resulta más compleja de lo que parece.
- * Los diccionarios legibles por la máquina contienen errores tipográficos.
- * Las entradas lexicográficas carecen de generalizaciones, las cuales sirven para reducir la redundancia de información.
- * Los diccionarios son objetos estáticos cuya información puede llegar a ser incluso obsoleta.

A pesar de todo, muchos otros 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 parece ser el enfoque más recomendable. Los beneficios que aporta la computerización lexicográfica, principalmente en lo que concierne a la minimización del esfuerzo humano y a la reducción del tiempo de trabajo, compensan notablemente sus posibles deficiencias. No obstante, aunque la estrategia a seguir es la automatización del proceso, es inevitable realizar una revisión manual en la fase de posesición para 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)).

El propósito de nuestra investigación no es tanto el estudio de la lengua, sino más bien la idoneidad de la transducción de un diccionario convencional en un lexicón computacional. No intentamos hacer una gramática del corpus, ya que en ese caso analizaríamos un inventario de textos reales escritos en esa lengua, y no exclusivamente entradas léxicas de un determinado diccionario. Investigamos la estructuración y uso de la lengua desde el prisma de la lexicografía. No obstante, no sólo es necesario recopilar una serie de resultados; es imprescindible igualmente aplicar una interpretación correcta. De esta forma, se integran la objetividad y la subjetividad características de todo método científico (Candalija Reina (1998)).

3. Functional Grammar Workbench 2.1 y la Gramática Funcional de Dik

Hasta ahora, el lector ha ido leyendo términos como *funcional* o *marco predicativo* asumiendo por nuestra parte que no requerían explicación alguna. No obstante, dedicamos este apartado a describir sucintamente el lexicón de la Gramática Funcional de S. C. Dik, modelo teórico a partir del cual se ha implementado nuestro programa.⁷ Si un sistema computacional del procesamiento del lenguaje natural no se basa en unos sólidos fundamentos lingüísticos, diseñaremos programas similares a aquellos que aparecieron en los primeros tiempos de la Inteligencia Artificial, los cuales se caracterizaban por seguir el principio erróneo de “mi teoría se basa en lo que mi programa hace” (Kwee Tjoe Liong (1994: 387)).

Hoy en día el número de teorías funcionales es muy grande, debido en parte a la polisemia del término *funcional*. Por ejemplo, Gutiérrez (1997: 468-469) ubica bajo el paraguas del *Funcionalismo* toda perspectiva científica de acercamiento al lenguaje que utilice el concepto de *función*. En el panorama actual de la lingüística funcional podemos destacar el funcionalismo formal de Alarcos Llorach, la Gramática Funcional de Dik, la Gramática Léxico-funcional de Kaplan y Bresnan, la Gramática Sistemática de M.A.K. Halliday y la Gramática Tagnémica de Pike, entre otros muchos modelos. Ante tanto *mare magnum*, nuestra aplicación adopta el enfoque funcional del modelo de S. C. Dik, el cual comprende básicamente

⁷ La propuesta más ambiciosa del diseño de un modelo computacional basado en la teoría de la Gramática Funcional la presenta el propio Dik en el proyecto *Functional Grammar Computational Model of the Natural Language User* (1989b).

tres componentes: el lexicón, la estructura de predicación y las reglas de expresión. Debido al objeto de investigación de este artículo, nos limitamos a describir el lexicón, y más concretamente la construcción de los marcos predicativos.

En la Gramática Funcional de S. C. Dik todas las unidades léxicas de sentido pleno son tratadas como predicados cuyos marcos predicativos se almacenan en el lexicón, componente central de la descripción lingüística funcional. A través de la información proporcionada por el marco predicativo, se puede predecir el comportamiento sintáctico de un predicado; de hecho, este modelo considera que la semántica está supeditada a la pragmática, mientras que la sintaxis lo está a la semántica (Dik (1980: 2)). Muchas teorías lingüísticas actuales adoptan este enfoque lexicista:

... la sintaxis ha pasado a considerarse un epifenómeno en beneficio de un panlexicismo que ha desembocado en la concepción del componente lexicón como una fuente de predictibilidad de las propiedades sintácticas de los predicados. (Mairal (1999: 42))

En el marco predicativo se halla la información fundamental que permite a la máquina construir la predicación subyacente de una oración de entrada. Los marcos predicativos verbales que construye nuestra aplicación de forma totalmente automática son bastante fieles a los propuestos por Dik. Tomemos como ejemplo el marco predicativo de *andar 1*:

ANDAR 1

Recorrer [un lugar, esp un camino, o una distancia].

ANDAR [V, mov-des] (X₁)Entidad (X₂: lugar, camino, distancia)Locativo

Podemos decir que se trata de un esquema estructural que especifica la forma del predicado (*ANDAR*), su categoría sintáctica (*V*), el tipo semántico al que pertenece (*mov-des*), las valencias cuantitativa (dos argumentos subcategorizados) y cualitativa (las funciones semánticas *Entidad* y *Locativo*),⁸ y las preferencias de selección⁹ de cada argumento (*lugar*, *camino* y *distancia* para el segundo argumento). El orden de presentación de los argumentos no corresponde necesariamente con el orden lineal de los constituyentes en una expresión lingüística, siendo este último establecido por las reglas de expresión de la lengua en cuestión.

En cuanto a la naturaleza del marco predicativo, la Gramática Funcional adopta un enfoque relacional en la descripción del significado, donde no se recurre a un metalenguaje, como la utilización de primitivos semánticos o predicados abstractos, sino a las unidades léxicas de la propia lengua (Dik (1989b: 14) (1997: 83)).

⁸ Véase el apartado 5.2 acerca del inventario y naturaleza de nuestras funciones semánticas.

⁹ Dik utiliza el término *restricciones de selección*, pero pensamos que el concepto de *preferencia de selección* refleja más fielmente su utilidad, es decir, no sirven para determinar la aceptabilidad de una expresión lingüística, sino más bien la prototipicidad de la expresión que formalice el argumento. También nos gustaría subrayar el hecho de que 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, etc).

Los marcos predicativos que FGW construye siempre incluyen las funciones semánticas centrales, aunque en bastantes ocasiones también se especifican algunas funciones semánticas periféricas (o, según la terminología de Dik, *satélites*). En realidad, existe cierta problemática a la hora de diferenciar los argumentos y los satélites (Faber y Mairal (1999: 51)), llegando a la conclusión de que es realmente el tipo de predicado el que determina si una función es central o periférica (Rojo (1983: 83)).

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, PROFGLLOT (Dik (1992) (1994)), implementación multilingüística en Prolog de la Gramática Funcional, maneja unos lexicones muy poco extensos, precisamente porque se elaboraron manualmente. A continuación presentamos un ejemplo de los marcos predicativos que se almacenan en este componente léxico (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 que el segundo argumento posee la función semántica objetivo/meta (*go*) y podrá formalizarse en un término concreto (*concr*). La *t* indica la ranura donde se puede insertar un término, i.e. una expresión que se utiliza para hacer referencia a una entidad.

Por otra parte, 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. Por ejemplo, el marco predicativo de *mismanage* tiene básicamente la siguiente forma (Voogt-van Zutphen (1989: 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*).

4. La tipología semántica del verbo de movimiento

FGW construye los marcos predicativos de los verbos de movimiento del español a partir de la información lexicográfica del DEA, a saber: la palabra titular, su categoría gramatical, el texto definitorio y su número de acepción. Toda esa información es introducida por el usuario a través de la propia aplicación, con el fin de ir aumentando el número de registros de la base de datos. No obstante, nuestras entradas léxicas poseen además el campo del subtipo semántico al que pertenece el verbo. Este campo tiene que ser rellenado igualmen-

te por el usuario, pero en esta ocasión utilizando su conocimiento como hablante y ajustándose a la clasificación que proponemos a continuación:¹⁰

- a- El verbo causativo (*mov-caus*) denota que alguien provoca que algo o alguien se mueva (p.ej. agitar, expulsar, lanzar).
- b- El verbo de dirección (*mov-dir*) expresa un movimiento extrínseco, el cual está centrado en el espacio en el que el desplazamiento se efectúa. Este tipo de verbos se integra en una geometría espacial en la que diferenciamos tres ejes: interior (hacia dentro/fuera: p.ej. entrar, salir), vertical (hacia arriba/abajo: p.ej. subir, caer) y horizontal (desde/hacia un punto: p.ej. venir).
- c- El verbo de desplazamiento (*mov-des*) expresa un movimiento intrínseco, el cual está centrado en el sujeto que se desplaza, conceptualizando el movimiento en sí además de la manera, el medio o la ruta del desplazamiento (p.ej. gatear, patinar, bucear).
- d- Finalmente, el verbo corporal (*mov-corp*) no indica un desplazamiento desde un lugar anterior a un lugar posterior, sino el movimiento de un cuerpo (p.ej. arrodillarse, auparse, temblar).

Por consiguiente, hemos clasificado los verbos de movimiento desde una perspectiva cognitiva, prescindiendo totalmente de las características formales que se reflejan en el nivel lingüístico (p.ej. el aspecto verbal, la reflexivización o el tipo de complementos que le acompañan). Según nuestra metodología de clasificación, no se tienen en cuenta criterios como, por ejemplo, que los verbos de dirección suelen ir acompañados de un complemento direccional introducido por *a*, *hacia* o *hasta*.

5. La construcción del marco predicativo

5.1 La identificación de los argumentos y la asignación de las funciones sintácticas y las preferencias selectivas

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, Andrés y Ramos (1999: xxii)). De esta manera, se presenta una fórmula definitoria más precisa que en otros diccionarios, ya que la verdadera definición se deslinda del entorno sintagmático del término titular. La máquina debe buscar los diferentes argumentos del predicado verbal precisamente en ese contorno de la palabra titular, el cual siempre se presenta entre corchetes:

¹⁰ Aunque esta clasificación fue inicialmente inspirada por el trabajo de Lamiroy (1991), el cual diseña una tipología según las características formales del verbo, nuestra postura se enclava en el plano cognitivo, más concretamente en la línea de Cifuentes (1998) en lo que respecta a los verbos de desplazamiento y dirección.

DESCARRILAR 1

Salirse del carril [un tren u otro vehículo que va sobre carriles].

DESCARRILAR [V, mov-dir] (X₁: tren, vehículo) Entidad

No obstante, no todos los argumentos de un marco predicativo están explícitos en el texto definitorio. Por ejemplo, en el caso de los verbos causativos de movimiento, casi un 95% de las entradas léxicas del DEA no explicita preferencia selectiva alguna para la función semántica *Causativo (remolinar 2)*, mientras que un 8% de los casos no presenta preferencias selectivas para ninguno de los argumentos del predicado titular (*alejarse 2*):¹¹

REMOLINAR 2

Formar remolinos [en algo (cd)].

REMOLINAR [V, mov-caus] (X₁) Causativo (X₂: cosa) Entidad

ALEJAR 2

Ahuyentar o hacer huir. Tb fig.

ALEJAR [V, mov-caus] (X₁) Causativo (X₂) Entidad

Cuando el contorno de la palabra titular encierra entre sus corchetes varios argumentos, es preciso diseñar un algoritmo que los identifique. Esos argumentos pueden ser fácilmente delimitados si el DEA explicita su función sintáctica (*expeler 1*):¹²

EXPELER 1

Despedir o expulsar [alguien o algo (suj) algo (cd)] de su interior. Tb abs.

EXPELER [V, mov-caus] (X₁: persona, cosa) Causativo (X₂: cosa) Entidad

En cambio, la tarea no es tan fácil cuando la definición lexicográfica no especifica las funciones sintácticas de los argumentos (*arribar 1*):

ARRIBAR 1

Llegar [una embarcación o sus ocupantes a un lugar].

ARRIBAR [V, mov-dir] (X₁: embarcación, ocupante) Entidad (X₂: lugar) Objetivo

En estos casos, nuestra aplicación averigua estas funciones sintácticas a partir de dos criterios: la categoría gramatical del verbo titular y la realización morfológica del argumen-

¹¹ De hecho, tan sólo un 30% de las entradas léxicas de nuestro lexicón especifica las preferencias selectivas del argumento que se presenta como sujeto, cuya función semántica es *Causativo* o *Entidad* para los verbos causativos y no causativos respectivamente. Aunque en la mayoría de las ocasiones el sujeto suele referirse a un ser animado, hemos preferido construir un marco predicativo a partir únicamente de la información aportada por el diccionario.

¹² El DEA sólo explicita las funciones sintácticas cuando éstas no se deducen del propio texto definitorio. Por ejemplo, en un 2,5% de las entradas léxicas de los verbos de movimiento aparece explicita la función de sujeto, frente al casi 60% de las entradas de los verbos transitivos de movimiento donde se especifica la función de complemento directo.

to en cuestión. En caso de que se exprese formalmente a través de un sintagma preposicional, se considera además la preposición que lo introduce. Por ejemplo, supongamos que un texto definitorio sólo especifica un argumento cuya función sintáctica no se explicita. Si el verbo titular es intransitivo o pronominal, entonces la aplicación asigna la función sintáctica de sujeto (*fluir 1*), siempre que no se trate de un sintagma preposicional, en cuyo caso asigna la función de complemento (*transitar 1*). En cambio, con los verbos transitivos, este argumento sin función sintáctica explícita se identifica como un complemento directo (*izar 1*):

FLUIR 1

Correr o deslizarse [un fluido].

FLUIR [V, mov-des] (X₁: fluido) Entidad

TRANSITAR 1

Ir o pasar [por una calle o camino]. Tb sin compl.

TRANSITAR [V, mov-des] (X₁) Entidad (X₂: calle, camino) Locativo

IZAR 1

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

IZAR [V, mov-caus] (X₁) Causativo (X₂: cosa, bandera) Entidad

Las preferencias de selección son identificadas a través de la presencia de un nombre o los pronombres indefinidos *alguien*, *algo* u *otro* dentro del sintagma que representa el argumento. Mientras en el caso del nombre FGW activa un lematizador, con los pronombres indefinidos *alguien* o *algo* las preferencias selectivas son *persona* y *cosa* respectivamente (*expeler 1*), en aras a una mayor coherencia en la representación del marco predicativo. En el caso del pronombre indefinido *otro* y sus alomorfos, se toma como preferencia selectiva el referente del pronombre (*echar 40*):

ECHAR 40

Salir de un lugar [a otro], esp de manera rápida y decidida.

ECHARSE [V, mov-dir] (X₁) Entidad (X₂: lugar) Objetivo

Somos conscientes de que nuestro marco predicativo podría verse enriquecido considerablemente al identificar otros argumentos compatibles con determinadas clases semánticas de verbos o, también a partir de esta información, aumentar las preferencias selectivas. No obstante, insistimos en nuestra postura de trabajar únicamente con la información aportada por el DEA. Bajo esta premisa, sí ha sido posible, en cambio, enriquecer el marco predicativo aprovechando parte de la explicación complementaria al texto definitorio que aporta esta obra lexicográfica. De las indicaciones sintáctico-semánticas que nos ofrecen estas explicaciones, la máquina sólo considera la información sobre la introducción de algún argumento opcional, teniendo la mayoría de ellos la función sintáctica de complemento (*arrastrar 2*):

ARRASTRAR 2

Llevar [una pers o cosa otra] tras de sí, tirando de ella. Tb fig. A veces con un compl de dirección.

ARRASTRAR [V, mov-caus] (X₁: persona, cosa) Causativo (X₂: cosa, persona) Entidad (X₃) Objetivo

5.2 La asignación de las funciones semánticas

Con todo lo explicado hasta el momento, la máquina puede construir un marco predicativo que presente la siguiente forma:

TRANSITAR [V, mov-des] (X₁) Suj (X₂: calle, camino) Compl

Hasta aquí, nuestra aplicación es capaz de establecer las dependencias sintácticas que contraen los argumentos con su predicado nuclear. No obstante, aunque se opte por mantener los esquemas sintácticos de dependencias, es imprescindible expresar estas dependencias como relaciones semánticas (Bosque (1990)). A este respecto, FGW dispone de un aparato conceptual que permite designar las funciones semánticas de los argumentos:

Las relaciones sintácticas tienen, en último término, una justificación en las relaciones cognitivas que establecen. (Horno (1998: 631))

FGW utiliza cinco funciones semánticas en la descripción del marco predicativo de los verbos de movimiento: *Entidad* (entidad que se mueve), *Causativo* (entidad que hace que algo o alguien entre en movimiento), *Locativo* (espacio o tiempo donde se ubica una entidad), *Origen* (entidad desde donde algo se mueve) y *Objetivo* (entidad hacia donde algo se mueve). Nuestras funciones *Entidad* y *Causativo* corresponden perfectamente a los papeles *Móvil* y *Causativo* respectivamente dentro de la gramática semántica de Dixon (1991: 94), quien describe a los verbos del tipo semántico *Movimiento* como básicamente intransitivos, cuyo sujeto tiene el papel temático *Móvil* en caso de no aparecer en una construcción causativa. También existe una evidente correspondencia con los argumentos semánticos básicos del Modelo Localista de Rasmussen (1994): i.e. *Entidad*, *Instigador*, *Lugar*, *Origen* y *Destino*.

Igualmente, como vemos en la tabla 1, podemos comprobar que varias funciones semánticas del inventario propuesto por Dik se amalgaman en unos pocos argumentos semánticos. Por una parte, *Agente*, *Posicionario* y *Fuerza* se fusionan en *Causativo*, porque todos ellos pueden concebirse como *agentes causativos* en un estado de cosas no dinámico:

Since Positioner and Force react in the same way to the passive test, they appear as Agent variants, to judge from their syntactic behaviour. This amounts to saying that on a more basic level these SFs correspond to the same semantic notion. (Rasmussen (1994: 57-58))

Por otra parte, fusionamos los papeles semánticos *Dirección* y *Recipiente* en *Objetivo* respaldándonos en una fundamentación sintáctico-semántica. Desde una perspectiva semántica, podemos decir que el *Recipiente* puede identificarse como un punto de destino animado (Rasmussen (1994: 57)). Además, seguimos la propuesta sintáctica de Herslund y

Sorensen (1994) de considerar el sujeto, objeto y adjunto como las relaciones fundamentales y universales de la estructura de la cláusula, por lo que la relación gramatical de adjunto abarca todos aquellos casos que la gramática tradicional etiqueta como complemento de sujeto, complemento de objeto, complemento preposicional, complemento, objeto indirecto y complemento circunstancial de lugar. Así, y con referencia a las entradas léxicas de los verbos de movimiento en el DEA, podemos hablar de un adjunto dativo y un adjunto locativo. Debemos recordar que los complementos locales y dativos no pueden aparecer juntos en una misma oración, por lo que el verbo se ve obligado a elegir entre las dos manifestaciones de la relación de adjunto. Son numerosas las ventajas que presenta esta postura teórica en la implementación de nuestro programa, como vemos en el siguiente ejemplo:

ENVIAR 1

Hacer que [alguien (cd)] vaya [a un lugar o a una pers distante], esp con una misión.
Tb sin el 2º compl. Frec en part sustantivado.

ENVIAR [V, mov-caus] (X₁) Causativo (X₂: persona) Entidad (X₃: lugar, persona)

Objetivo

Finalmente, y como apunta Rasmussen (1994: 58), siempre que se denote un estado de cosas dinámico, el *Agente* y el *Procesado* pueden integrarse en el papel *Entidad*.

En los marcos predicativos construidos por FGW, los papeles semánticos funcionan más como papeles conceptuales, los cuales determinan la función o posición de la entidad en el conocimiento del mundo que posee el hablante. La información semántica almacenada en el marco predicativo está íntimamente relacionada con el tipo de conocimiento cognitivo que se va a conceptualizar. El hecho de que, por ejemplo, un verbo tenga como argumentos una entidad móvil y un espacio donde se ubica esa entidad se debe a su propio significado, de ahí que las funciones semánticas de los argumentos de un predicado sean fácilmente identificables a partir de su texto definitorio:

Given that the purpose of dictionaries is to provide definitions of words and their senses, it might well be expected that, of all forms of text, it would be in dictionaries that the semantic structure of language would be the most explicit and hence most accessible for examination and comparison with the semantic structure of knowledge representations. (Wilks et alii (1996: 106))

En el proceso de asignación de los papeles semánticos, el programa tiene en cuenta diversos parámetros: el tipo semántico del verbo titular, la función sintáctica del argumento en cuestión y, finalmente, si éste se representa formalmente por un sintagma preposicional, y en caso afirmativo la preposición que lo introduce. La tabla 2 describe las diversas condiciones que tienen que cumplirse para que la aplicación asigne una determinada función semántica a un argumento. Por ejemplo, si un predicado que pertenece a los verbos de desplazamiento, dirección o movimiento corporal posee un argumento con la función de sujeto, a dicho argumento se le asigna la función semántica *Entidad*. En cambio, si se trata de un verbo causativo, se aplica entonces la función *Causativo*.

6. Las futuras líneas de investigación

El valor de este tipo de implementaciones informáticas es muy diverso:

The lexicon can be used in parsing and generating systems; it can be used, together with a monolingual dictionary of another language, to create a bilingual dictionary which can be used in translation; it can be used as an object of research in its own right; etc. (Voogt-van Zutphen (1989: 170))

Nuestro propósito es utilizar FGW como un laboratorio de ingeniería lingüística, cuyos resultados utilizaremos en el proceso de diseño e implementación de un sistema de traducción mecánica. En segundo lugar, podemos considerar nuestro lexicón funcional como un corpus informatizado, el cual se convierte en objeto de investigación filológica con la ayuda de una serie de instrumentos o paquetes de gestión de corpora.

Actualmente FGW utiliza también las entradas léxicas del *Collins COBUILD English Dictionary* (1995) para construir de forma similar los marcos predicativos de los verbos de movimiento en inglés. Además, estamos pensando en ampliar en un futuro inmediato el alcance de nuestro programa; no sólo construiremos marcos predicativos de otros tipos semánticos de verbos, sino también trabajaremos con otras partes del discurso, preferentemente nombres y adjetivos. En plazo muy breve, tendremos la posibilidad de estudiar los verbos de movimiento del español y el inglés en su plano paradigmático, asociando así a cada predicado un marco predicativo y un postulado de significado, el cual también se construye de forma totalmente automática a partir del texto definitorio. Ya que nuestra intención es comprobar la validez interlingüística e intercultural de nuestro lexicón funcional en los planos sintagmático y paradigmático, resulta imprescindible utilizar un mismo entorno computacional para llevar a cabo nuestra investigación: FGW nos proporciona ese escenario.

Tabla 1

	S. C. DIK	FGW
VERBOS CAUSATIVOS DE MOVIMIENTO	Meta	Entidad
	Agente, Posicionario, Fuerza	Causativo
	Lugar	Locativo
	Origen	Origen
	Dirección, Recipiente	Objetivo
VERBOS NO CAUSATIVOS DE MOVIMIENTO	Agente, Procesado	Entidad
	Lugar	Locativo
	Origen	Origen
	Dirección	Objetivo

Tabla 2

	Condiciones		
	Función Sintáctica	Tipo Semántico	Preposición
ENTIDAD	SUJ CD	mov-des mov-dir mov-corp mov-caus	
CAUSATIVO	SUJ	mov-caus	
LOCATIVO	CD CD COMPL	mov-des mov-dir	excepto <i>a, hacia, hasta</i> excepto <i>a, hacia, hasta, de, del</i>
ORIGEN	COMPL		<i>de, del</i>
OBJETIVO	CI CD COMPL	mov-dir	<i>a, hacia, hasta</i> <i>a, hacia, hasta</i>

Referencias

BOSQUE MUÑOZ, I. (1990), *Las Categorías Gramaticales. Relaciones y Diferencias*, Madrid, Síntesis.

CANDALIJA REINA, J. A. (1998), "Sobre la científicidad de la gramática: el uso de corpora informatizados como método de análisis lingüístico", en Cifuentes Honrubia, J. L. (ed.), *Estudios de Lingüística Cognitiva*, 2 vols, Alicante, Universidad de Alicante, 295-307.

CASSEN, B., J. F. DÉGREMONT, H. SCHNELLE y A. ZAMPOLLI (1991), "Formación del personal investigador y estudios de doctorado en lingüística computacional", en Vidal Beneyto, J. (ed.), *Las Industrias de la Lengua*, Madrid, Pirámide, 411-415.

CHEN, S. y L. XU (1995), "A full and efficient machine tractable dictionary for natural language processing: a revised version of the CUVOALD", *Computers and the Humanities* 28-3, 141-152.

CIFUENTES HONRUBIA, J. L. (1998), "Verbos con incorporación conceptual direccional", en Cifuentes Honrubia, J. L. (ed.), *Estudios de Lingüística Cognitiva*, 2 vols, Alicante, Universidad de Alicante, 479-505.

- Collins COBUILD English Language Dictionary* ((1987) 1995) Londres, Collins.
- DIK, S. C. (1978), *Functional Grammar*, Dordrecht, Foris.
- DIK, S. C. (1980), *Studies in Functional Grammar*, Londres y Nueva York, Academic Press.
- DIK, S. C. (1989a), *The Theory of Functional Grammar. Part I: The Structure of the Clause*, Dordrecht, Foris.
- DIK, S. C. (1989b), "FG*C*M*NLU: Functional Grammar Computational Model of the Natural Language User", en Connolly, J. H. y S. C. Dik (eds.), *Functional Grammar and the Computer*, Dordrecht, Foris, 1-28.
- DIK, S. C. (1992), *Functional Grammar in Prolog. An Integrated Implementation for English, French and Dutch*, Berlín y Nueva York, Mouton de Gruyter.
- DIK, S. C. (1994), "The lexicon in a computational functional grammar", en Atkins, B. T. S. y A. Zampolli (eds.), *Computational Approaches to the Lexicon*, Oxford, Oxford University Press, 279-294.
- DIK, S. C. (1997), *The Theory of Functional Grammar. Part I: The Structure of the Clause*, Berlín y Nueva York, Mouton de Gruyter.
- DIXON, R. M. W. (1991), *A New Approach to English Grammar on Semantic Principles*, Oxford, Clarendon Press.
- FABER, P. y R. MAIRAL (1999), *Constructing a Lexicon of English Verbs*, Berlín, Mouton de Gruyter.
- GUTIÉRREZ ORDÓÑEZ, S. (1997), *Principios de Sintaxis Funcional*, Madrid, Arco.
- HANDKE, J. (1995), *The Structure of the Lexicon: Human versus Machine*, Berlín y Nueva York, Mouton de Gruyter.
- HAUGLAND, J. (1998), "Semantic engines: an introduction to mind design", en Clark, A. y J. Toribio (eds.), *Machine Intelligence: Perspectives on the Computational Model*, Nueva York, Garland, 29-62.
- HERSLUND, M. y F. SORENSEN (1994), "A valence based theory of grammatical relations", en Engberg-Pedersen, E., L. F. Jakobsen y L. S. Rasmussen (eds.), *Function and Expression in Functional Grammar*, Berlín y Nueva York, Mouton de Gruyter, 81-95.
- HORNO CHÉLIZ, M. C. (1998), "Conceptualización y categorización lingüística de las relaciones espaciales en verbos locativos", en Cifuentes Honrubia, J. L. (ed.), *Estudios de Lingüística Cognitiva*, 2 vols, Alicante, Universidad de Alicante, 629-637.
- IDE, N. y J. VÉRONIS (1994), "Machine readable dictionaries: what have we learned, where do we go?", en Calzolari, N. y C. Guo (eds.), *Proceedings of the Post-Coling 94 International Workshop on Directions of Lexical Research*, Beijing, 137-146.
- KWEE TJOE LIONG (1994), "Prolog for a functional grammar of Hungarian: A programmer's look at grammar writing", en Engberg-Pedersen, E., L. F. Jakobsen y L. S. Rasmussen (eds.), *Function and Expression in Functional Grammar*, Berlín y Nueva York, Mouton de Gruyter, 385-401.

- LAMIROY, B. (1991), *Léxico y Gramática del Español: Estructuras Verbales de Espacio y de Tiempo*, Barcelona, Anthropos.
- MAIRAL, R. (1999), “El componente lexicón en la Gramática Funcional”, en Butler, C., R. Mairal, J. Martín Arista y F. J. Ruiz de Mendoza (eds.), *Nuevas Perspectivas en Gramática Funcional*, Barcelona, Ariel, 41-98.
- MOURE, T. y J. LLISTERRI (1996), “Lenguaje y nuevas tecnologías: el campo de la lingüística computacional”, en Fernández Pérez, M. (ed.), *Avances en Lingüística Aplicada* 4, Santiago de Compostela, Universidad de Santiago de Compostela, 147-227.
- RASMUSSEN, L. S. (1994), “Semantic functions in perspective – reconsidering meaning definitions”, en Engberg-Pedersen, E., L. F. Jakobsen y L. S. Rasmussen (eds.), *Function and Expression in Functional Grammar*, Berlín y Nueva York, Mouton de Gruyter, 41-63.
- ROJO, G. (1983), *Aspectos Básicos de Sintaxis Funcional*, Málaga, Ágora.
- SECO, M., O. ANDRÉS y G. RAMOS (1999), *Diccionario del Español Actual*, 2 vols., Madrid, Aguilar.
- SIERRA, G. y J. MCNAUGHT (2000), “Extracting semantic clusters from MRDs for an onomasiological search dictionary”, *International Journal of Lexicography* 13-4, 264-286.
- VOOGT-VAN ZUTPHEN, H. (1989), “Towards a lexicon of Functional Grammar”. En Connolly, J. H. y S. C. Dik (eds.), *Functional Grammar and the Computer*, Dordrecht, Foris, 151-176.
- WILKS, Y., D. FASS, C. GUO, J. E. MCDONALD, T. PLATE y B. M. SLATOR (1993), “Providing machine tractable dictionary tools”, en Pustejovsky, J. (ed.), *Semantics and the Lexicon*, Dordrecht, Kluwer Academic Publishers, 341-401.
- WILKS, Y., B. M. SLATOR y L. M. GUTHRIE (eds.) (1996), *Electric Words. Dictionaries, Computers and Meanings*, Cambridge: Massachusetts, MIT Press.