

FrameNet Español. Una red semántica de marcos conceptuales¹

Carlos Subirats Rüggeberg
Universidad Autónoma de Barcelona
<carlos.subirats@uab.es>

1. Objetivos

*FrameNet Español*² es un proyecto de investigación de semántica cognitiva (Fillmore 1985), cuyo objetivo es el estudio de la organización conceptual de la red de clases semánticas que configura semánticamente el léxico de predicados de la lengua española. El propósito de FrameNet Español es triple:

1. identificar las clases semánticas (o marcos semánticos) que configuran conceptualmente el léxico de predicados del español;
2. determinar los argumentos semánticos (o *frame elements*) que determinan cada una de dichas clases o marcos;
3. anotar semántica y sintácticamente construcciones en las que aparecen predicados pertenecientes a dichas clases.

Estos tres objetivos se materializarán en una base de datos relacional en MySQL, que estará integrada por oraciones anotadas y que se podrá consultar y modificar con *FNDesktop*,³ un conjunto de aplicaciones que permite realizar básicamente dos operaciones:

1. anotar semántica y sintácticamente (cf. Fig. 2) oraciones que previamente han sido extraídas de forma automática de un corpus textual y, posteriormente, importadas en formato XML a la base de datos;
2. realizar consultas mediante una interfase web que permite acceder a la organización automática de la información resultante de la anotación semántica (cf. Fig. 3 y 4); esta interfase permite realizar consultas sobre:

¹ Quisiera darles las gracias a Rocío Donés, Jordi Duran, Mercedes García de Quesada, Paloma Garrido, Covadonga López Alonso, Lúdia Moya y Marc Ortega por su colaboración. Este proyecto de investigación está financiado por el Ministerio de Ciencia y Tecnología de España (TIC2002-01338) y se está desarrollando en colaboración con el proyecto FrameNet de Berkeley, California. Quisiera expresar mi agradecimiento a Collin Baker, Michael Ellsworth, Charles Fillmore y Miriam R.L. Petruck, sin cuya ayuda me hubiera sido imposible llevar a cabo este proyecto. Finalmente, quisiera dar las gracias al Ministerio de Ciencia y Tecnología por las ayudas que me ha concedido para poder realizar estancias de investigación en el International Computer Science Institute.

² <http://gemini.uab.es/SFN/>

³ FNDesktop es un software desarrollado por el proyecto FrameNet de Berkeley, California.

- la descripción de las clases semánticas y de sus respectivos argumentos conceptuales;
- la anotación semántica de un predicado perteneciente a una clase semántica específica; este tipo de consulta permite visualizar las distintas construcciones sintácticas en las que aparecen los distintos argumentos semánticos y, también, combinaciones de argumentos asociados a dicho predicado.

La base de datos de FrameNet se puede consultar asimismo con *Spanish FrameSQL*, una interfase web desarrollada por el Prof. Hiroaki Sato en la Universidad de Senshu (Japón). FrameSQL permite realizar consultas transversales, es decir, consultas sobre argumentos semánticos y/o construcciones sintácticas que se pueden realizar dentro de toda una clase semántica o, incluso, dentro de toda la base de datos.

Los resultados del proyecto FrameNet Español serán de dominio público a partir de enero de 2006 y consistirán, como ya hemos señalado anteriormente, en una base de datos relacional en MySQL, que contendrá:

1. oraciones etiquetadas en función de la aparición en ellas de predicados específicos, que pueden ser tanto verbales, nominales como adjetivos;
2. las descripciones semánticas de las clases semánticas a las que pertenecen los predicados analizados y
3. los subcórpora de cada entrada debidamente anotados.

Los resultados de FrameNet se podrán consultar vía web; asimismo, se podrá descargar de la red tanto la base de datos como el software para su consulta, con objeto de que los resultados del proyecto se puedan explotar también localmente.

La base de datos de FrameNet español tendrá fundamentalmente dos aplicaciones:

- Desde un punto de vista lingüístico, se podrá utilizar como un diccionario semántico *online*, ya que incluirá definiciones de los predicados estudiados, descripciones de las clases semánticas a las que pertenecen y tablas – organizadas automáticamente– que permitirán acceder a las oraciones anotadas, que ejemplificarán la realización sintáctica de los argumentos semánticos asociados a un predicado.
- Desde el punto de vista informático, se podrá usar como un corpus de entrenamiento para aplicaciones de etiquetación semántica automática (Gildea y Jurafsky 2002).

La interfase web de consulta de FrameNet Español se podrá enlazar con los resultados de otros proyectos de investigación de semántica léxica del español, como p. ej., el proyecto ADESSE (Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español)⁴ y, eventualmente, con otros bases de datos con información semántico-léxica del español, como p. ej., los resultados sobre el español del proyecto EuroWordNet⁵, en la medida en que la política de difusión y distribución de estos

⁴ <http://webs.uvigo.es/adesse/>

⁵ <http://www.illc.uva.nl/EuroWordNet/>

proyectos⁶ coincide con la de FrameNet Español, cuyos resultados, como hemos señalado anteriormente, serán de libre consulta y distribución.

2. El corpus textual del FrameNet Español

FrameNet Español utiliza un corpus textual de 350 millones de palabras⁷ como base textual para realizar las búsquedas de construcciones sintácticas y también para realizar extracciones automáticas de oraciones. Este corpus incluye tanto textos del español de América (60%) como del español europeo (40%)⁸. La consulta de las construcciones oracionales del corpus se realiza con XKWIC, una aplicación desarrollada por el Institut für Maschinelle Sprachverarbeitung de la Universidad de Stuttgart⁹, que se distribuye libremente con licencia.

La etiquetación, lematización y especificación de las propiedades morfológicas flexivas de las palabras del corpus –tanto formas simples como locuciones– se ha realizado con una aplicación de transducción léxica, que utiliza un diccionario electrónico de 600.000 formas¹⁰. Este diccionario electrónico expandido se genera automáticamente a partir de un diccionario que contiene 93.000 lemas, concretamente:

- 68.000 lemas simples –es decir, cadenas de caracteres entre dos espacios en blanco, independientemente de que se puedan descomponer morfológicamente o no–, como p. ej., *amaríamos*, *libertad*, *allí*, etc.;
- 25.000 lemas locutivos que no requieren información sintáctica para su identificación automática, como p. ej., *bombas atómicas*, *maestros de escuela*, *a troche y moche*, etc.

Los elementos léxicos que requieren información sintáctica para su identificación automática, como p. ej., las locuciones verbales *hacer caso omiso*, *levantar acta*, etc., y los grupos preposicionales predicativos *estar de moda*, *estar en condiciones*, *estar a disposición*, etc., se detectan automáticamente con ALIA (Algoritmo de Intersección de Automatas¹¹), un software de intersección de autómatas¹² que permite realizar análisis

⁶ EuroWordNet no permite la distribución libre de sus resultados.

⁷ Desgraciadamente, nos ha sido imposible utilizar el *Corpus de Referencia del Español Actual* (CREA) de la Real Academia Española (<http://corpus.rae.es/creanet.html>), ya que dicho corpus no está etiquetado e, incomprensiblemente, tampoco es de dominio público –ni siquiera parcialmente–, a pesar de haber sido desarrollado con dinero público.

⁸ Cf. http://gemini.uab.es/SFN/SFN_Corpus.html para un análisis detallado de la composición de este corpus.

⁹ <http://www.ims.uni-stuttgart.de/>

¹⁰ La creación de los léxicos electrónicos, las primeras versiones de los programas de (1) generación automática de léxicos expandidos y (2) etiquetación y lematización automáticas de formas simples y locuciones han sido desarrollados gracias a la financiación concedida por (1) el Ministerio de Educación de España CAICYT (PB85-371), CICYT (PB87-780 y PB92-0635), y (2) el Ministerio de Obras Públicas y Transportes (TIC90-403). Los diccionarios electrónicos se pueden consultar en la red en: http://gemini.uab.es:8080/servlets-examples/busca_dico.html.

¹¹ Cf. Ortega 2002.

¹² El desarrollo del software de intersección de autómatas y su aplicación al análisis sintáctico automático, así como la construcción de transductores con información léxica para la detección de unidades léxicas locutivas se han desarrollado en el marco de los proyectos del Ministerio de Educación de España CICYT (TIC96-0804 y TIC1999-0753).

sintácticos automáticos. ALIA utiliza la información de 2.300 transductores léxicos, que formalizan las propiedades léxicas y sintácticas que posibilitan el reconocimiento y la etiquetación automáticos de las locuciones verbales y los grupos preposicionales predicativos en español.

La extracción automática de las construcciones sintácticas que ejemplifican los argumentos semánticos de los predicados estudiados se realiza también con ALIA, que, para realizar dicha extracción, utiliza un conjunto de transductores específicos asociados a cada predicado, que se construyen semiautomáticamente con una aplicación gráfica, a partir de las características de las construcciones detectadas en el corpus.

3. Argumentos sintácticos y complementos adjuntos

La determinación sintáctica de los argumentos de un predicado se establece en función de determinadas propiedades formales que dicho predicado mantiene con sus argumentos:

1. la concordancia en número y persona entre el verbo predicativo y su primer argumento o el verbo de soporte y el primer argumento del predicado no verbal correspondiente;
2. la selección de clases disjuntas de pronombres clíticos de tercera persona en la pronominalización de segundos argumentos, concretamente, *lo, los, la, las* para el objeto directo y *le, les* para el objeto indirecto;
3. la selección de marcadores de argumentos como *a, de, en*, etc., que introducen los segundos argumentos (objetos preposicionales) en el caso de algunos predicados verbales y no verbales.

Sin embargo, además de los argumentos sintácticos, los predicados verbales y no verbales requieren argumentos con características conceptuales sistematizables, que están asociadas a clases de predicados. En muchas ocasiones, estos argumentos son complementos adjuntos y, por tanto, no responden a ninguna de las propiedades formales mencionadas anteriormente en 1., 2. y 3.

Analicemos un ejemplo. Desde el punto de vista sintáctico, *comentario* en la oración *Max le hizo un comentario a Eva* es un nombre predicativo de dos argumentos. No obstante, cuando se hace un *comentario*, se debe realizar necesariamente sobre un *tema*¹³, independientemente de que dicho *tema* sea explícito, como p. ej., *sobre estas irregularidades* en (1), o no sea explícito, como en (2):

- (1) *Max le hizo un **comentario** a Eva sobre estas irregularidades.*
- (2) *Max le hizo un **comentario** a Eva.*

Asimismo, un *comentario* se puede hacer en un *medio*, como p. ej. *en la prensa* en (3):

- (3) *Max hizo un **comentario** sobre estas irregularidades en la prensa.*

Desde un punto de vista estrictamente sintáctico, el *tema* y el *medio* de las anteriores oraciones (1) y (3) no constituyen argumentos sintácticos de *comentario*, al menos, tal

¹³ Los argumentos semánticos (o *frame elements*) aparecen en minúsculas y en cursiva.

como los hemos definido anteriormente. Asimismo, estos adjuntos sintácticos pueden ser formalmente muy distintos, como se puede observar en (4):

- (4) *Max hizo un **comentario** (sobre, a propósito de, en relación con, acerca de, etc.) estas irregularidades (en la prensa, por televisión, etc.).*

Los adjuntos sintácticos de la anterior oración (4), es decir, (*sobre, a propósito, en relación con, acerca de, etc.*) *estas irregularidades, y en la prensa, por televisión, etc.*, se podrían derivar sintácticamente mediante reducciones¹⁴. No obstante, una derivación sintáctica no podría dar cuenta del hecho de que, conceptualmente, *comentario* requiere complementos adjuntos que constituyen argumentos semánticos de dicho predicado y que tienen unas características conceptuales ligadas a la clase de predicados a la que pertenece, es decir, a la clase de predicados de comunicación. Por ello, la determinación de los argumentos de un predicado con criterios estrictamente sintácticos no coincide, en general, con la proyección sintáctica de sus argumentos conceptuales. Por este motivo, FrameNet parte de la caracterización conceptual de los predicados del léxico y estudia su proyección sintáctica como un reflejo de su estructura conceptual y, por tanto, se antepone la semántica a la sintaxis.

4. Clases de argumentos semánticos

Los argumentos semánticos de un predicado pueden ser básicamente de tres tipos: *externos, implícitos e incorporados*.

Los argumentos conceptuales *externos* se manifiestan fuera del grupo o constituyente encabezado por el predicado analizado. En general, se trata de argumentos que satisfacen el requerimiento conceptual de un predicado y tienen las siguientes características sintácticas:

1. Pueden ser primeros argumentos de:
 - verbos predicativos en forma personal, como p. ej. *encantar* en *A Juan le encanta [la paella]_{Externo}*;
 - nombres predicativos con verbo de soporte, como p. ej. *dar un ultimátum* en *[El presidente]_{Externo} les dio un ultimátum a los terroristas* o adjetivos predicativos, como p. ej., *ser rico* en *[Venezuela]_{Externo} es rica en tradiciones*;
 - infinitivos que han sufrido una reducción, que ha elidido su primer argumento, como p. ej. *bajar* en *[Los políticos]_{Externo} decidieron **bajar** los impuestos* o *firmar* en *[Le]_{Externo} obligaron a **firmar** el contrato*.

Los argumentos conceptuales *implícitos* son aquellos que se han elidido y no se manifiestan como argumentos o adjuntos de un predicado. Con objeto de agrupar automáticamente las oraciones con el mismo requerimiento conceptual, independientemente de que sus argumentos sean o no explícitos, la anotación semántica utiliza etiquetas específicas para dar cuenta de la elisión de argumentos conceptuales. Establecemos una distinción básica entre tres tipos distintos de elisión de argumentos conceptuales:

¹⁴ Cf. Subirats (2001:138-141).

1. elisión de argumentos definidos, como p. ej., la elisión de *Juan* como argumento de *trabajar* en *Juan estudia por las mañanas y trabaja por las tardes*;
2. elisión de argumentos indefinidos, como en la oración *Ya no se estila ser galante con las mujeres*, en la que *galante* ha sufrido una reducción de su primer argumento indefinido;
3. elisión de argumentos inducida por una construcción sintáctica, como p. ej. la elisión de la *causa* o el *agente* en una oración como *El barco se hundió por sí solo*, en la que la construcción de *se* medio bloquea la posibilidad de que una *causa* o un *agente* aparezcan como argumentos de *hundir*.

Llamamos argumentos *incorporados* a aquellos que se funden con un predicado. Así p. ej., en la oración *Les dispararon con una ametralladora*, el constituyente independiente *con una ametralladora* especifica el arma utilizada. Por el contrario, en la oración *Les ametrallaron*, el arma no es un constituyente independiente, como en nuestro anterior ejemplo, sino que se ha incorporado al propio predicado verbal *ametrallar*.

5. Clases conceptuales y argumentos semánticos

Las clases semánticas en las que se pueden integrar los predicados del léxico forman una red que se puede estructurar conceptualmente, de forma tal que se pueden establecer relaciones de herencia de argumentos entre dichas clases. Estas subredes de herencia de argumentos permiten explicar las características semánticas más generales del requerimiento conceptual de los predicados del léxico. Así p. ej., las características semánticas del verbo *empujar* están determinadas por sus argumentos conceptuales nucleares, es decir, un *agente* y un *tema* que realiza un desplazamiento que se puede determinar en función de un *origen*, una *trayectoria* y una *meta*. Pero *empujar* pertenece a una clase semántica más general, concretamente, a la clase de los predicados eventivos y, por ello, podemos considerar que *empujar* hereda de dicha clase los argumentos conceptuales no nucleares tanto temporales como locativos. Este hecho permite poner de manifiesto que las propiedades semánticas más generales, es decir, aquellas que afectan a múltiples clases semánticas, se pueden explicar por medio de las relaciones de herencia que se establecen en el marco de la red conceptual en la que se integran los predicados del léxico.

La identidad en la selección de argumentos conceptuales configuran las características del significado del léxico de predicados e inducen su partición en clases semánticas. Así p. ej., los predicados causativos de movimiento en español se caracterizan porque, en ellos, un *agente* es el causante de que un *tema* realice un desplazamiento, es decir, un *agente* mueve una figura en relación con una base. Sin embargo, dentro de este mismo esquema conceptual, existen distintas clases de predicados que ponen de relieve los diferentes argumentos conceptuales que integran dicho esquema.

Así p. ej., la clase conceptual de los predicados de colocación (*placing*), a la que pertenecería *poner* en la acepción que tiene en (5), evoca un esquema conceptual en el que un *agente* –*María*– desplaza un *tema* –*las flores*– en relación con una *meta* –*sobre la mesa*– y se enfatiza el *tema* que es desplazado por un *agente*:

(5) *María puso las flores sobre la mesa.*

Poner, por supuesto, no es el único predicado que tiene las características descritas, sino que existen otros, como *colocar*, *situar*, etc. Analicemos ahora la clase conceptual de separación (removing); esta clase incluye predicados como p. ej. *sacar* en (6), que se refieren a un esquema en el que un agente –Eva– retira un tema –los libros– de un origen –las cajas–:

(6) *Eva sacó los libros de las cajas.*

Los predicados que pertenecen a la clase de separación –al igual que los pertenecen a la clase de colocación– enfatizan el *tema*. Nótese que existen otros predicados que evocan el mismo esquema conceptual que *sacar*, como p. ej., *extraer*, *extracción*, *descargar*, etc.

A diferencia de lo que sucede con los predicados pertenecientes a las clase de colocación y separación, en las que, como hemos señalado anteriormente, se enfatiza el *tema*, los predicados de llenado (filling), como p. ej. *llenar* en (7), en los que un agente –Max– desplaza un tema –los discos– a una meta –la caja–, enfatizan la *meta*, que suelen ser contenedores, áreas, etc.:

(7) *Max llenó la caja de discos.*

Por el contrario, los predicados de la clase de vaciado (emptying), como p. ej. *vaciar* en (8), en los que un agente –Max– retira un tema –de trastos– de un origen –la casa– realzan el *origen*:

(8) *Max vació la casa de trastos.*

Obsérvese que *llenar* y *vaciar* comparten sus respectivos esquemas conceptuales con otros predicados, como p. ej., *inundar*, *untar*, etc., y *desinflar*, etc.

La integración de los predicados del léxico en clases conceptuales no implica su homogeneidad absoluta desde el punto de vista de los aspectos conceptuales que caracterizan sus argumentos semánticos. Analicemos, p. ej., las diferencias relacionadas con los argumentos semánticos de los predicados de movimiento que enfatizan la *trayectoria* que sigue un *tema* en su desplazamiento. Algunos predicados de esta clase indican la *trayectoria* en función de un alejamiento con respecto a la dirección de la trayectoria inmediatamente anterior al desplazamiento especificado por el predicado, p. ej., *girar*, *virar*, etc.:

(9) *Max giró a la derecha para tomar la carretera principal.*

Otros predicados indican la trayectoria en relación con un camino lineal implícito, como p.ej., *zigzaguear*; otros, como *desviarse*, denotan la trayectoria del *tema* como un alejamiento de su trayectoria anterior, sin que el predicado realce la dirección de dicho alejamiento:

(10) *El avión se desvió de su ruta por causas desconocidas.*

Otros predicados, como *cruzarse* en (11), realzan la *trayectoria* de un *tema* en relación con la *trayectoria* de otro *tema*:

(11) *Max se cruzó con Eva por la calle.*

En otros casos, el verbo realza la *trayectoria* en relación (1) con el *área* por la que discurre dicha *trayectoria*, como en el caso de *cruzar*, *atravesar*, etc., o (2) con el límite de dicha *área*, p. ej., *bordear*, *esquivar*, etc. Otros predicados de esta clase conceptual realzan la *trayectoria* del desplazamiento de un *tema* en relación con el tipo de curso que atraviesa dicho *tema*, como en el caso de *vadear*, que se utiliza únicamente para desplazamientos que atravesen corrientes de agua, como ríos, etc. Otros predicados de la clase conceptual que estamos analizando evocan desplazamientos con trayectorias no orientadas, p. ej., *vagar*, *merodear*, *vagabundear*, etc. Nótese que los predicados que especifican el curso o trayectoria en relación con los límites de un *área*, como *bordear* en (12):

(12) *Bordeamos el lago para llegar a la casa.*

pueden hacer referencia también a la forma física de un trayecto, como en (13), si bien en estos usos no se refieren al desplazamiento real de un *tema* y, por tanto, no existe movimiento propiamente dicho:

(13) *La antigua carretera bordeaba el lago.*

Es importante distinguir los predicados como *bordear* de los anteriores ejemplos (12), (13), de predicados como *bifurcarse*, que solo se pueden utilizar para describir la forma de una trayectoria física, como en (15), o de una trayectoria metafórica, como en (16), pero no para hacer referencia al desplazamiento de un *tema*, como podemos observar en (17):

(15) *La carretera se bifurca después del puente.*

(16) *A partir de aquel momento, nuestras vidas se bifurcaron para siempre.*

(17) **Max y Eva se bifurcaron a partir de aquel punto.*

6. Las cuatro tareas fundamentales en la construcción de una red de marcos conceptuales

La primera tarea que se tiene que llevar a cabo en el proceso de construcción de la red de marcos conceptuales del léxico del español consiste en la identificación de los esquemas semánticos que configuran dicha red y los elementos conceptuales que permiten delimitar dichos esquemas. Paralelamente, es necesario determinar los elementos léxicos que podrían incluirse inicialmente en esta primera aproximación a la definición de un esquema conceptual junto con sus argumentos, puesto que dichos esquemas tienen que ser un reflejo de las características conceptuales del léxico. Por ello, FrameNet no propone esquemas conceptuales abstractos desvinculados del léxico, puesto que estos no permiten construir un análisis conceptual del léxico ni de su organización en redes semánticas.

La segunda tarea en el proceso de creación de la red conceptual de predicados de FrameNet consiste en la identificación en el corpus de las distintas construcciones sintácticas que vehiculan el significado de los argumentos conceptuales asociados a cada uno de los predicados que integran una clase semántica determinada. Este proceso se realiza con la herramienta de consulta XKWIC (cf. Fig. 1) que permite acceder al corpus textual de FrameNet, que incluye 350 millones de palabras (cf. 2. en este artículo). A partir de la identificación de las construcciones sintácticas en las que aparece un predicado, se construyen semiautomáticamente gramáticas electrónicas con expresiones regulares (o autómatas), que permiten que el software ALIA (cf. 1. en este artículo) realice una selección automática de todas aquellas construcciones del corpus que tienen las características formales especificadas en el autómata. Posteriormente, el sistema de creación de subcórpora –asociado a un predicado– selecciona aleatoriamente 30 ejemplos, entre los que los lingüistas posteriormente eligen algunas oraciones para anotarlas semántica y sintácticamente.

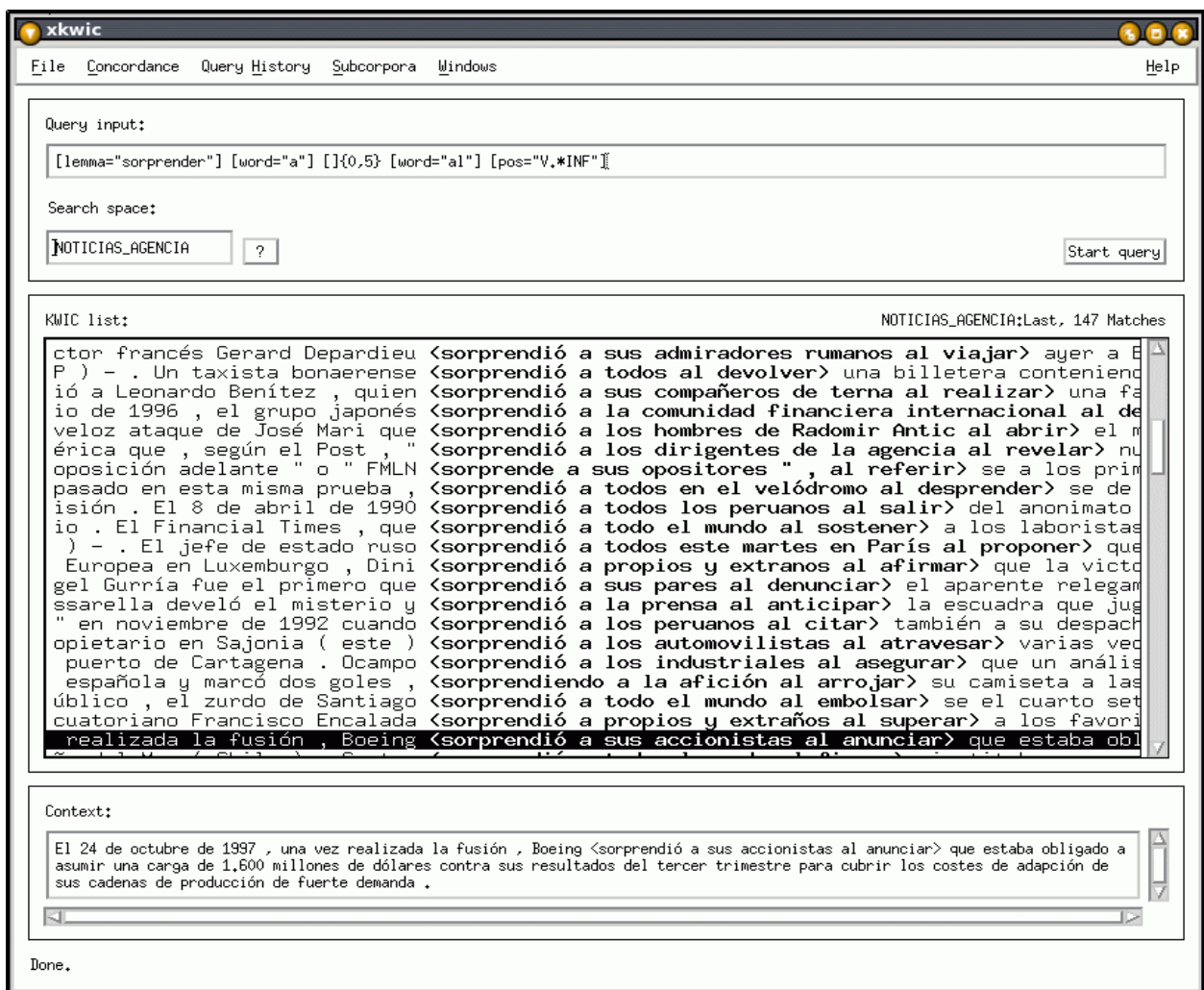


Fig. 1. Consulta de usos de *sorprender* en el corpus de FrameNet con la aplicación XKWIC

En suma, el objetivo de esta segunda tarea consiste en automatizar el proceso de extracción de aquellas oraciones que contienen determinadas proyecciones sintácticas

de los argumentos conceptuales de un predicado para organizar y facilitar su posterior anotación semántica.

La tercera tarea constituye el núcleo central del desarrollo del proyecto FrameNet: la anotación semántica y sintáctica semiautomática de las oraciones extraídas automáticamente del corpus en las que aparece un predicado en construcciones sintácticas previamente establecidas. Esta tercera tarea se realiza con la aplicación FNDesktop, que ha sido adaptada al español (cf. Fig. 2).

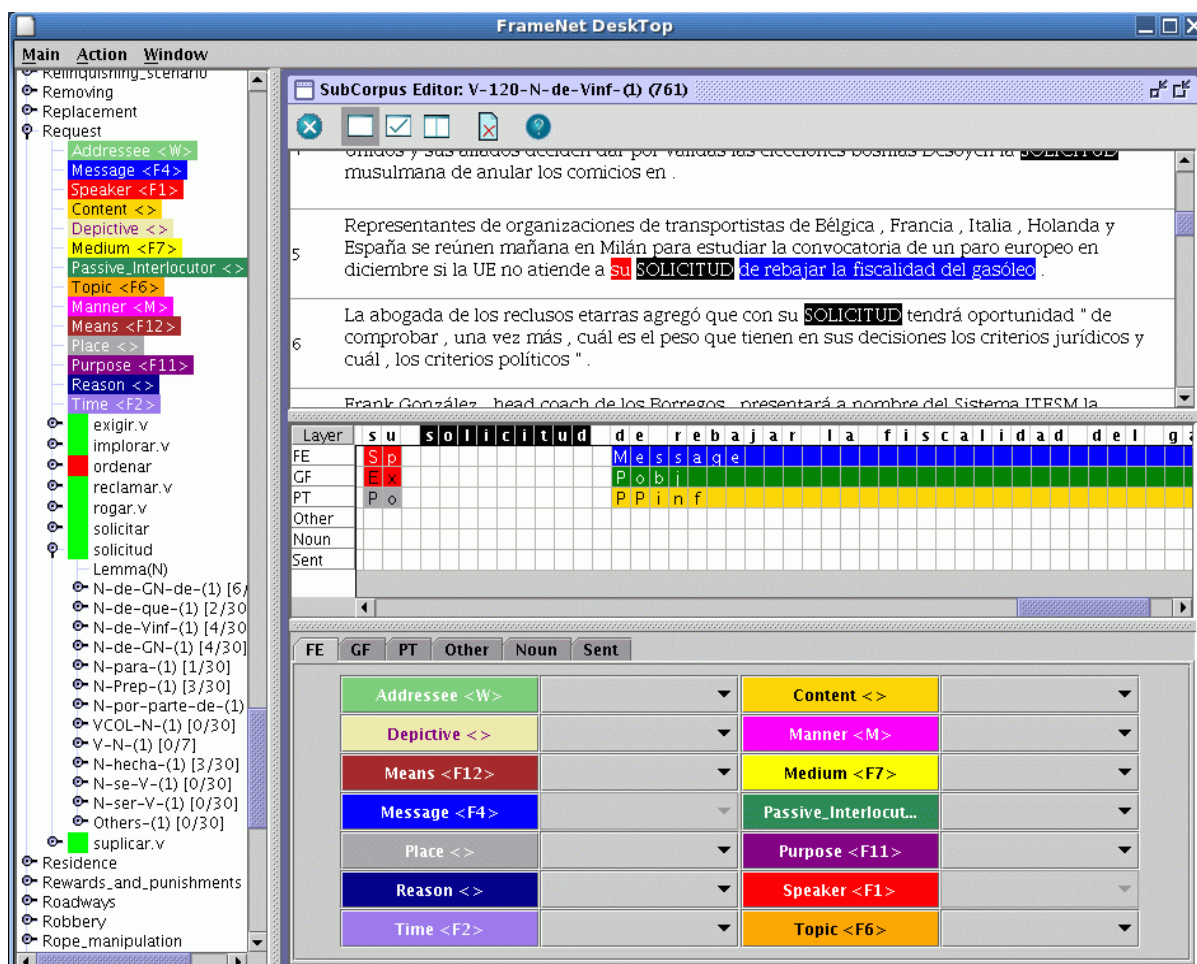


Fig. 2. Anotación del nombre predicativo *solicitud* perteneciente a la clase conceptual de predicados de petición (request) con la aplicación FNDesktop

Como se puede observar en la Fig. 2, la aplicación FNDesktop está dividida en un marco central y un menú. El menú es un marco de navegación, que da acceso directo a los principales objetos que integran la base de datos de FrameNet, es decir, las clases semánticas, sus argumentos conceptuales y las unidades léxicas que integran dichas clases. A su vez, el marco central alberga los objetos necesarios para llevar a cabo la etiquetación, es decir, las oraciones y las etiquetas semánticas. El menú incluye una lista de todas las clases conceptuales que integran la base de datos. Al pinchar sobre una clase conceptual, como p. ej., *Request*¹⁵ (petición) en la Fig. 2, se despliega la lista de

¹⁵ Los nombres de las clases semánticas, así como el de sus argumentos conceptuales, están en inglés en la base de datos.

sus argumentos semánticos y, debajo de ellos, aparece la lista de las unidades léxicas pertenecientes a la clase. Al pinchar sobre los elementos léxicos, como p. ej., *solicitud* en la Fig. 2, se despliega la lista de subcórpora asociados a dicho predicado que han sido extraídos automáticamente del corpus. El nombre de los subcórpora está en relación con las características de las construcciones sintácticas que los integran. Finalmente, al pinchar sobre los subcórpora, aparece, en el marco central, una lista de 30 oraciones, entre las que el lingüista puede elegir las que prefiere anotar semánticamente. Obsérvese que el marco central, que es donde se realiza la anotación, está dividido en tres secciones. En la sección superior, se encuentran las oraciones que se pueden seleccionar para su anotación. Pinchando una oración con el ratón, esta se despliega en la sección intermedia del marco central. Para llevar a cabo su anotación, se selecciona el constituyente que se quiere etiquetar y, en la sección inferior, se pincha el argumento semántico con el que se quiere anotar el constituyente seleccionado. Existen tres niveles de anotación, que permiten especificar (1) el argumento semántico, (2) su función sintáctica y (3) el tipo de constituyente. En la práctica, no es necesario especificar manualmente estos tres niveles de anotación, puesto que, al seleccionar la etiqueta correspondiente a un argumento semántico, como p. ej., *Speaker*, *Adressee*, etc., la aplicación FNDesktop asigna automáticamente –con un margen de error muy bajo– su función gramatical, como p. ej., *External*, *Direct Object*, etc., y el tipo de constituyente, p. ej., *NP*, *PP*, etc. Por ello, la anotación semántica y sintáctica constituye un proceso semiautomático, puesto que, al asignar una etiqueta semántica a un constituyente, FNDesktop añade automáticamente la especificación de su función sintáctica y el tipo de constituyente.

La cuarta y última tarea en el desarrollo del proyecto FrameNet consiste en la consulta vía web de la organización automática de los resultados de la anotación, tanto para verificar dicha anotación, como para examinar la caracterización semántica del predicado analizado. La consulta vía web se puede realizar con FNDesktop (cf. Fig. 3) y, también, con la aplicación Spanish FrameSQL¹⁶ (cf. Fig. 4). Una de las consultas más básicas que se puede efectuar con FNDesktop consiste en acceder a los resultados de la anotación de una unidad léxica, a partir de los argumentos conceptuales que han sido utilizados en su anotación, como p. ej. en la Fig. 3:

- en la columna de la izquierda, aparece el nombre de los argumentos conceptuales que han sido utilizados en la anotación del predicado *solicitud*;
- en la columna del centro, encontramos la especificación del número de oraciones anotadas en las que aparece el correspondiente argumento conceptual con enlaces que permiten acceder a las correspondientes oraciones anotadas;
- en la columna del derecha, se especifica el tipo de constituyente y la función sintáctica del argumento o argumentos semánticos especificados en la columna de la izquierda, con enlaces que permiten acceder a los ejemplos correspondientes.

La aplicación Spanish FrameSQL permite realizar consultas más complejas sobre los resultados de la anotación de un predicado. Así p. ej. en la Fig. 4, podemos observar el resultado de la consulta de la combinatoria de todos los argumentos conceptuales de la anotación semántica del nombre predicativo *solicitud*. Obsérvese que en la Fig. 4, FrameSQL especifica la posición que ocupa *solicitud* en relación con sus argumentos conceptuales e indica los verbos de soporte, junto con los auxiliares que

¹⁶ Esta aplicación ha sido desarrollada por el Prof. Hiroaki Sato.

intervienen en la formación de los tiempos verbales compuestos; asimismo, en la columna de la izquierda, aparece el número de predicados en los que aparece la combinación de argumentos especificada, con un enlace que permite visualizar el ejemplo correspondiente. En el caso de un nombre predicativo como *solicitud*, la detallada especificación combinatoria que ofrece FrameSQL nos permite distinguir fácilmente los usos con verbo de soporte, que se corresponden con las combinaciones especificadas en las 7 primeras filas de la Fig. 4, de las construcciones con reducción del verbo de soporte¹⁷, que aparecen en las filas 8-15, en las que podemos apreciar claramente cómo los argumentos semánticos aparecen siempre a la derecha del predicado.

Frame Element	Number Annotated	Realizations(s)
Addressee	6 exx	PP.Comp 2 exx PP.Pobj 4 exx
Manner	1 exx	AVP.Mod 1 exx
Medium	1 exx	PP.Comp 1 exx
Message	14 exx	PP.Pobj 1 exx PPinf.Pobj 8 exx PqueSsubj.Pobj 5 exx
Purpose	5 exx	PPinf.Adct 4 exx PqueSsubj.Adct 1 exx
Speaker	20 exx	AJP.Ext 1 exx NP.Ext 1 exx PP.Adct 16 exx Poss.Ext 2 exx
Time	1 exx	NP.Mod 1 exx

Fig. 3. Consulta de la realización sintáctica de los argumentos conceptuales de *solicitud*.

7. Conclusiones

El proyecto FrameNet proporcionará un mapa de la organización conceptual de léxico del español y ofrecerá, además, descripciones detalladas de la proyección de este mapa conceptual en el léxico de predicados del español. Asimismo, la anotación semántica y sintáctica de un número significativo de oraciones en las que aparecerán predicados del léxico del español pertenecientes a clases conceptuales muy diversas, como p. ej., emoción, percepción, cognición, movimiento, comunicación, etc., permitirá poner de manifiesto las características específicas de la manifestación léxica de dicho mapa conceptual.

La visualización vía web de la reorganización automática de la información de la base de datos de FrameNet mediante FNDesktop y FrameSQL, en función de las clases semánticas y sus argumentos, la combinatoria de argumentos, etc., proporcionará un diccionario semántico *online*, que abrirá nuevas perspectivas para el análisis cognitivo de las características semánticas de los predicados del léxico español. Asimismo, la

¹⁷ Cf. Subirats (2001:141-143).

posibilidad que ofrece FrameSQL de realizar consultas cruzadas y simultáneas sobre FrameNet en español y en inglés permitirá que nuestra base de datos se pueda utilizar como un diccionario semántico bilingüe *online* inglés-español y español-inglés, el cual, además de tener aplicaciones para la consulta humana, tendrá sin duda repercusiones en el desarrollo de sistemas de traducción automática basados en el análisis cognitivo del léxico.

Finalmente, la creación de una base de datos con un número significativo de oraciones anotadas con información semántica, sintáctica, categorial y léxica permitirá desarrollar nuevas aplicaciones en el ámbito del tratamiento automático del español, que no solo posibilitarán el desarrollo de nuevas tecnologías para el procesamiento semántico automático, sino también para las nuevas formas de tratamiento de la información textual que va a requerir el futuro desarrollo de la web semántica en español.

Num	FE/LUset (sort = FE; Request, solicitud, N,)
01	<i>haber_hacer</i> + solicitud.N + Addressee + Purpose
01	<i>hacer</i> + solicitud.N + Addressee + Purpose
01	Speaker + <i>hacer</i> + solicitud.N + Addressee
02	Speaker + solicitud.N + Message
01	solicitud.N + <i>hacer</i> + Time + Speaker + Purpose
01	solicitud.N + <i>irse_hacer</i> + Addressee + Speaker
01	solicitud.N + <i>irse_hacer</i> + Medium + Addressee + Speaker
03	solicitud.N + Message
01	solicitud.N + Message + Speaker
01	solicitud.N + Addressee
01	solicitud.N + Manner + Speaker
01	solicitud.N + Purpose
03	solicitud.N + Speaker
08	solicitud.N + Speaker + Message
01	solicitud.N + Speaker + Purpose
27	

Fig. 4. Combinaciones de argumentos conceptuales de las oraciones en las que aparece el nombre predicativo *solicitud*.

Referencias bibliográficas

- BÁEZ, Valerio. *Desde el hablar a la lengua. Prolegómenos a una teoría de la sintaxis y la semántica textual y oracional*. Málaga; Ágora.
- BAKER, Collin F.; FILLMORE, Charles; CRONIN, Beau. 2003. The Structure of the FrameNet Database. *International Journal of Lexicography* 16.3:281-296 (http://www3.oup.co.uk/lexico/hdb/Volume_16/Issue_03/pdf/160281.pdf).
- CASAS, Miguel. 1999. *Las relaciones léxicas*. Tübingen: Max Niemeyer.
- CUARTERO, Juan. 2003. *Cosas que se hacen. Esquemas sintáctico-semánticos agentivos del español*. Frankfurt a.M.: Peter Lang.
- CIFUENTES, José Luis. 1999. *Sintaxis y semántica del movimiento*. Alicante: Instituto de Cultura Juan Gil-Albert.
- FILLMORE, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semántica* 6.2:222-254.
- FILLMORE, Charles J.; BAKER, Collin F.; SATO, Hiroaki. 2002. The FrameNet Database and Software Tools. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*. Las Palmas, pp. 1157-1160 (<http://framenet.icsi.berkeley.edu/~framenet/papers/demo4.pdf>).
- FILLMORE, Charles J.; BAKER, Collin F.; SATO, Hiroaki. 2002. Seeing Arguments through Transparent Structures. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*. Las Palmas, pp. 787-791 (<http://framenet.icsi.berkeley.edu/~framenet/papers/LREC12.pdf>).
- GARCÍA-MIGUEL, J.M.; COMESAÑA, Susana. (en prensa). Verbs of Cognition in Spanish: Constructional Schemas and Reference-points. *Language, Culture and Cognition. An International Conference on Cognitive Linguistics*. Braga (<http://webs.uvigo.es/adesse/textos/LCC2003%20texto%20final.pdf>).
- GILDEA, Daniel; JURAFSKY, Daniel. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28.3:245-288.
- IÑESTA, Eva M.; PAMIES, Antonio. 2002. *Fraseología y metáfora: aspectos tipológicos y cognitivos*. Granada: Método Ediciones.
- KOIKE, Kazumi. 2001. *Colocaciones léxicas en el español actual: estudio formal y léxico-semántico*. Madrid: Universidad de Alcalá y Takushoku University.
- LENZ, Rodolfo. 1920 [1925, 2ª edición]. *La oración y sus partes. Estudios de gramática general y castellana*. Madrid: Publicaciones de la Revista de Filología Española,
- LUQUE DURÁN, J. de D. 2001. *Aspectos universales y particulares del léxico de las lenguas del mundo*. Granada: Método.
- MORIMOTO, Yuko. 2001. *Los verbos de movimiento*. Madrid: Visor Libros.
- ORTEGA, Marc. 2002. *Transductores en el análisis léxico y sintáctico de un texto*. Tesis de licenciatura, Universidad Politécnica de Cataluña.
- PALÉS, Marisol. 2003. *Diccionario de sinónimos y antónimos*. Madrid: Espasa Calpe.
- PENADÉS, Inmaculada. 2000. *La hiponimia en las unidades fraseológicas*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.
- SANTOS, Luis A.; ESPINOSA, Rosa M. 1996. *Manual de semántica histórica*. Madrid: Síntesis.
- SUBIRATS, Carlos. 2001. *Introducción a la sintaxis léxica del español*. Madrid/Frankfurt: Iberoamericana/Vervuert.
- SUBIRATS, Carlos; SATO, Hiroaki. 2004. Spanish FrameNet and FrameSQL. *4th International Conference on Language Resources and Evaluation (LREC)*,

Workshop on Building Lexical Resources from Semantically Annotated Corpora, May 30, 2004. Lisboa (http://seneca.uab.es/csubirats/Subirats-Sato_LREC-2004.doc).

SUBIRATS, Carlos; PETRUCK, Miriam R. L. 2003. Surprise: Spanish FrameNet! *Workshop on Frame Semantics, International Congress of Linguists. July 29, 2003.* Praga (<http://framenet.icsi.berkeley.edu/~framenet/papers/SFNsurprise.pdf>).

TALMY, Leonard. 2000. *Toward a Cognitive Semantics*, 2. vols. Cambridge, MA: The MIT Press.

WOTJAK, Gerd. 2000. ¿Qué pasa a los significados al sustantivarse los verbos? En G. Wotjak, ed. *En torno al sustantivo y adjetivo en el español actual. Aspectos cognitivos, semánticos, (morfo)sintácticos y léxicogenéticos.* Madrid/Frankfurt: Iberoamericana/Vervuert, pp. 153-172.