

Máster Universitario en
Tecnologías de la Información y la Comunicación
en la Enseñanza y Tratamiento de Lenguas

TRABAJO FIN DE MÁSTER

Un enfoque computacional a la
preposición *in* en los adjuntos
periféricos de la Gramática del Papel y
la Referencia

Departamento de Filologías Extranjeras y sus
Lingüísticas

Facultad de Filología

UNED

Junio 2012

Directora: Dra. Rocío Jiménez Briones

Autora: Diana Hernández Pastor

1. INTRODUCCIÓN.....	4
2. ESTADO DE LA CUESTIÓN SOBRE EL TRATAMIENTO DE LAS PREPOSICIONES	4
2.1 En lingüística teórica y ciencia cognitiva.....	4
2.2 En procesamiento del lenguaje natural.....	9
2.2.1. La cuestión de la desambiguación de las preposiciones en PLN.....	11
3. MODELOS TEÓRICOS MÁS INFLUYENTES	12
3.1. La Gramática del Papel y la Referencia	14
3.1.1. El Lexicón en la GPR	16
3.1.2. La Estructura Relacional de la Cláusula	19
3.2. El Modelo Léxico Construccional	22
3.3. Introducción a FunGramKB.....	23
3.3.1. Estructura de FunGramKB	26
3.3.2. Lenguaje de Representación Conceptual (COREL)	30
3.3.3. FungramKB y el Lexicón.....	32
4. PROPUESTA DE TRABAJO.....	33
4.1. Herramientas de estudio	34
4.2. Metodología de investigación	35
4.3. Desarrollo de la investigación	36
4.4. La descripción semántica de la preposición <i>in</i>	38
5. CONCLUSIONES Y FUTURAS INVESTIGACIONES.....	46
6. BIBLIOGRAFÍA.....	¡Error! Marcador no definido.
ANEXOS	¡Error! Marcador no definido.

1. INTRODUCCIÓN

Este proyecto parte de la necesidad de tratar las preposiciones en la disciplina del Procesamiento del Lenguaje Natural (PLN). En este sentido, intentaremos aportar una aproximación al tratamiento computacional de las preposiciones en inglés en la base de conocimiento léxico-conceptual FunGramKB y su impacto en la teoría lingüística de la Gramática del Papel y la Referencia (GPR). Para llevar a cabo este proyecto realizaremos un análisis de la preposición *in* como ejemplo de nuestra propuesta a los modelos mencionados.

El objetivo de nuestra investigación se centra en la interpretación semántico-conceptual de los sintagmas preposicionales predicativos expresados por la preposición *in*, que funcionen como adjuntos en el marco de la Gramática del Papel y la Referencia, para su utilización en el PLN a través de FunGramKB.

A lo largo de esta investigación intentaremos mostrar las diferentes aproximaciones en torno a las preposiciones desde la perspectiva de la ciencia cognitiva, y más concretamente desde la lingüística computacional o procesamiento del lenguaje natural. Tras este análisis del estado de la cuestión, procederemos a explicar los principales modelos teóricos que influyen notablemente en nuestro trabajo, como son la Gramática del Papel y la Referencia y el Modelo Léxico Construccional. En el apartado final de este documento se realizará un análisis de la descripción semántica de la preposición *in* para su implementación en la base léxico computacional de FunGramKB, siguiendo los modelos teóricos descritos.

2. ESTADO DE LA CUESTIÓN SOBRE EL TRATAMIENTO DE LAS PREPOSICIONES

2.1 En lingüística teórica y ciencia cognitiva

La Lingüística Cognitiva surgió durante la década de 1980, con figuras como la de Lakoff y Langacker, los cuales representan piedras angulares en las investigaciones lingüísticas de hoy en día. La nueva ciencia nace del descontento por no encontrar fundamento a nuevas percepciones en las corrientes generativista y estructuralista. De acuerdo con la Lingüística Cognitiva, el lenguaje está

construido sobre papeles temáticos que los hablantes otorgamos al saber que obtenemos a partir de nuestras experiencias. Estos papeles temáticos conllevan la proporción de casos semánticos que expresan las funciones de los elementos del lenguaje.

(...) “human categorization is essentially a matter of both human experience and imagination -of perception, motor activity, and culture on the one hand, and of metaphor, metonymy, and mental imagery on the other.” (Lakoff, 1990:8)

De esta premisa, se puede deducir que las funciones semánticas son universales, pues son los conceptos que categorizados en nuestra mente los que determinan el lenguaje y no al contrario. Por esta razón, se ha intentado crear un inventario de papeles temáticos que representen nuestro pensamiento y en consecuencia nuestro lenguaje.

Las preposiciones constituyen una categoría léxica con una presencia frecuentemente alta en el lenguaje. A pesar de ello, constituyen una categoría gramatical notoriamente difícil de manejar. Las preposiciones han sido clasificadas como palabras relacionales debido a que su uso se fundamenta en la combinación con otras categorías gramaticales como verbos, nombres o adjetivos para establecer una relación funcional en un grupo u oración.

Muchas han sido las aproximaciones teóricas a las preposiciones, en especial, a las de significado espacial, desde una perspectiva de la Gramática Cognitiva en diferentes lenguas (Herskovits 1986; Linstromberg 1997; Tyler and Evans 2003). Estos estudios han proporcionado una visión de las limitaciones semánticas que caracterizan a esta categoría gramatical, como en el caso de las preposiciones combinadas con sintagmas nominales (SNs). Por ejemplo en el caso de la preposición *in* + SN donde el núcleo del SN puede ser un contenedor, *in the cup*; un lugar, *in England*; un modo, *in silence*, etc. El resultado de esas investigaciones se ha recogido en formato de listas como en el caso de Herskovits (1986) o redes semánticas como en el caso de Tyler and Evans (2003).

En el estudio de los papeles temáticos de las categorías gramaticales, las preposiciones han sido dejadas a un lado, bien por su complejidad, bien por haber sido consideradas no relevantes para las expresiones lingüísticas. Sin embargo, algunos estudios se han acercado a demostrar la importancia de esta categoría gramatical en un nivel semántico-sintáctico.

No obstante, en el momento de realizar un inventario de casos semánticos para las preposiciones, se debe concretar el tratamiento que se les ha asignado en las investigaciones de tipo sintáctico-semántico. Desde una perspectiva del significado, existe una distinción básica¹ de la semántica de las preposiciones (Downing y Locke, 2006:547):

- a) aquellas en las que el uso de la preposición viene determinado por el verbo, el nombre o el adjetivo que la precede: preposiciones gramaticalizadas
- b) aquellas que son elegidas por el emisor de manera voluntaria en función de sus intenciones: preposiciones lexicalizadas

La mayoría de las investigaciones que han abarcado el mundo de las preposiciones han girado en torno a las preposiciones gramaticalizadas y, en especial, a las preposiciones espaciales en las construcciones con verbos de movimiento. En este contexto, de entre los estudios con más influencia, cabe destacar la percepción cognitiva que Talmy (2000) ofrece sobre la construcción de los eventos en movimiento. Para su análisis, Talmy (1985) realizó un estudio de los conceptos básicos para describir los eventos de movimiento: Movimiento, Trayectoria, Manera y Figura. En su obra "Cognitive Semantics" (2000, vol. II. ch.3), Talmy defiende la existencia de un macroevento como unidad cognitiva, cuyos elementos conceptuales pueden ser considerados universales para una organización lingüística.

Supuso que las lenguas están clasificadas en dos grandes categorías tipológicas:

- las lenguas de marco verbal (*Verb Framed Languages*), en las que la ruta (Path) del movimiento viene determinada por el verbo:

¹ Existe un tercer uso de las preposiciones de carácter metafórico el cual puede incluirse dentro de las preposiciones léxicas, ya que el perfil definitorio del significado viene expresado por el concepto, pero queda fuera de la envergadura de nuestro trabajo.

The bottle exited from the cave, floating

- las lenguas de marco satelital (*Satellite-framed languages*), en las que la ruta viene determinada por las partículas que acompañan al verbo. En lenguas como el inglés, hablamos de las preposiciones:

The bottle floated out of the cave

La diferencia entre ellas radica en el modo en el que codifican los conceptos básicos de la estructura, de una manera sintáctica y lexicista, así como de su capacidad para expresar la dirección del movimiento, esto es, cuál es el marco en el que sucede la expresión lingüística.

Para nuestro proyecto nos centraremos en el estudio de las preposiciones léxicas o cuya semántica es composicional, ya que presumimos que la construcción de sintagmas preposicionales (SPs) puede clasificarse teniendo en cuenta los diferentes significados que aporta la preposición en la mencionada combinación.

Según Tyler y Evans (2003), la semántica de las preposiciones ha sido tratada de manera arbitraria a lo largo de las investigaciones, creyendo que los diferentes significados y usos de una misma preposición están organizados de manera accidental. Como propuesta resolutive a esta concepción, estas autoras establecen que los diversos significados asociados a las preposiciones en inglés pueden verse representados a través de una red polisémica en la que compartirían un significado central y abstracto.

En este sentido, las investigaciones de las preposiciones léxicas, desde la semántica profunda, han girado, fundamentalmente, sobre el eje de tres usos semánticos de las preposiciones:

- a. Uso temporal: determinan la localización en el tiempo ya sea al inicio, proceso o fin.
- b. Uso locativo: determinan la dimensión locativa en relación a un punto de referencia ya sea estático o en movimiento.
- c. Uso espacial: determinan la relación funcional entre el objeto y la interacción humana con los mismos.

En lo que a la sintaxis se refiere, las preposiciones han sido, por un lado, consideradas vitales para precisar la comprensión del lenguaje y, por otro lado, ignoradas en los ámbitos en los que se consideran sintácticamente promiscuas y vacías de significado.

Las dos consideraciones más importantes de las preposiciones a nivel sintáctico son:

1. La selección (Fillmore 1968; Bennett 1975; Kracht 2003), p.ej. la propiedad de una preposición a ser subcategorizada por el término regente (un verbo, un nombre o un adjetivo) como parte de su estructura. Las preposiciones seleccionadas (*selected prepositions*) carecen de significado, lo cual recibe el nombre de *bleached semantics* (Baldwin et al., 2009:120)

i.e.: The programme consists of three parts

Por el contrario, las preposiciones no seleccionadas (*unselected*), que no forman parte de ninguna estructura, sí tienen valor semántico:

i.e. : He is playing football with Tom

2. La valencia (Huddleston and Pullum 2002) de las preposiciones, la cual puede ser transitiva o intransitiva.

2.1. las preposiciones intransitivas no asumen ningún argumento. Suelen aparecer como:

a) componentes de amplias expresiones *multiword* (take it *off*)

b) predicados copulativos (Jane is *in*)

c) modificadores prenominales (an *off* day)

2.2. Las preposiciones transitivas seleccionan complementos para formar sintagmas preposicionales:

i.e.: Children play football in the garden

2.2 En procesamiento del lenguaje natural

Las investigaciones de esta categoría gramatical en el PLN no son muy numerosas. Sólo unos pocos trabajos, p.ej. The Preposition Project (Litkowski y Hargraves, 2005) o PrepNet (Saint-Dizier, 2005, 2008), se han ocupado de proporcionar una descripción de la semántica preposicional suficientemente robusta para las aplicaciones del PLN. De hecho, como nos recuerdan Sven, Helbig y Osswald (2006), las investigaciones lingüísticas sobre la semántica de las preposiciones no han tenido mucho impacto en las aplicaciones del PLN. No obstante, como reconoce Saint-Dizier (2008), se precisa tener un repositorio que organice y caracterice los comportamientos sintácticos y semánticos de las preposiciones, una tarea que resulta de vital importancia incluso en aplicaciones del PLN que requieran una forma simple de comprensión, p.ej. extracción de conocimiento o traducción automática. En este sentido, el principal problema que debemos afrontar es la resolución de la ambigüedad léxica en una categoría gramatical que resulta ser extremadamente polisémica, donde la interpretación semántica de la preposición depende estrechamente del contexto, por no mencionar la dificultad de identificar las regularidades interlingüísticas.

El adjunto de construcciones preposicionales ha resultado, junto con la desambiguación y el etiquetado morfosintáctico, una de las principales problemáticas en el desarrollo de aplicaciones de lingüística computacional. Tanto es así que en las últimas tres décadas se ha asistido a diversos cambios paradigmáticos en la investigación del adjunto de construcciones preposicionales, siendo objetivo de estudio de diferentes disciplinas como la psicolingüística, la IA o la sintaxis teórica entre otras.

En el siguiente apartado describimos, muy brevemente, algunos ejemplos de recursos léxico- semánticos desarrollados para el tratamiento de las preposiciones en el PLN:

1. *The English LCS Lexicon*: (Dorr, 1997) incluye 165 preposiciones con 122 sentidos intransitivos y 375 sentidos transitivos. Cada uno de los

sentidos de las preposiciones está descrito a través de una representación basada en la Estructura Lógica Conceptual de Jackendoff (1983, 1990).

2. *FrameNet*: (Baker, Fillmore y Lowe, 1998) es un proyecto orientado a construir una base de datos léxica del inglés que puede ser leída tanto por humanos como por máquinas, mediante la anotación de ejemplos del uso de las palabras en textos reales. Este proyecto tiene como fundamento la teoría del significado Semántica del Marco (Fillmore, 1982; 1985) que trata de describir el significado de la mayoría de las palabras a través de un marco semántico, esto es, a través de la descripción de eventos, relaciones o entidades junto con sus participantes.
3. *The Preposition Project*: (Litowski y Hargraves, 2005) se trata de un intento por desarrollar una base semántica comprehensiva para las preposiciones en lengua inglesa, orientada a aplicaciones del PLN. Para ello tomaron como fuente de información léxico-semántico el *New Oxford Dictionary of English*, cuyo contenido fue ajustado a través de una comparación cruzada del etiquetado funcional de las preposiciones en *FrameNet* y el tratamiento de las preposiciones por parte de una semántica descriptiva o tradicional del inglés. A fecha de 2009, el registro de preposiciones anotadas era de 334 con 673 sentidos.
4. *PrepNet*: (Saint-Dizier, 2005) tiene como propósito la explicación del tratamiento composicional de la semántica de las preposiciones que se interrelaciona con la semántica del predicado. Al igual que The English LCS Lexicon, PrepNet usa como lenguaje de descripción semántica la Estructura Lógica Conceptual de Jackendoff. Este proyecto destaca por intentar capturar las limitaciones de selección léxica, los sentidos metafóricos y los argumentos de mayor complejidad.

2.2.1. La cuestión de la desambiguación de las preposiciones en PLN

El proceso de ambigüedad se produce cuando un mismo elemento tiene, a su vez, diferentes sentidos o interpretaciones. En el caso de las preposiciones, la ambigüedad es un proceso latente, pues depende del carácter relacional de la preposición, esto es, de la combinación de la preposición con otras categorías gramaticales. Así en el siguiente ejemplo, una aplicación del PLN tendría muchos problemas para resolver la ambigüedad creada por la preposición “with”:

- (3) a. *Tom ate a yoghurt with a spoon.*
b. *Tom ate a yoghurt with cereals.*

En ambos casos la problemática surge con la función del SP. La tarea para la desambiguación radica en establecer cuál es el término regente del SP, si el verbo *ate* o el nombre *yoghurt*. En el caso (a) el SP depende del verbo, pues *Tom* usó una cuchara para comerse el yogur; en el caso (b) la preposición *with* introduce un SP dependiente del nombre, ya que el yogur es el que contiene los cereales.

No obstante, para afrontar plenamente el problema de la desambiguación, es necesario el desarrollo de una base de conocimiento que integre la semántica de las preposiciones desde una perspectiva profunda de la semántica. El principal problema dentro de la desambiguación de construcciones preposicionales es la búsqueda del término regente para un SP dado.

El interés en estudiar los casos de desambiguación semántica de las preposiciones surge de ser un hecho común de ocurrencia de errores en el análisis sintáctico de lenguas como el inglés. Las investigaciones, en este sentido, se han centrado en el desarrollo de modelos enfocados a estrategias de procesamiento humano basados en el análisis de árboles sintácticos independientes del contexto léxico o discursivo (Baldwin et al., 2009).

Sin embargo, a pesar del gran volumen de investigaciones hechas en el campo del Procesamiento del Lenguaje Natural en relación al tratamiento de la semántica de las preposiciones, aún no se ha desarrollado un modelo que permita

abordar los problemas de desambiguación semántica provocados por esta categoría gramatical.

Pocos son los estudios que realmente han aportado un paso hacia delante en la descripción sistemática de la semántica de las preposiciones para aplicaciones del PLN: Litowski (2002) con reglas manuales y O'Hara y Wiebe (2003) los cuales propusieron un enfoque estadístico basado en las colocaciones.

La selección de una preposición es importante para cualquier aplicación del PLN que opere a niveles sintáctico- semántico y traduzca restricciones formales a representaciones semánticas o viceversa. La selección forma un componente clave en el aprendizaje de subcategorización (Manning 1993; Briscoe and Carroll 1997).

3. MODELOS TEÓRICOS MÁS INFLUYENTES

Las perspectivas de estudio que aquí se plantean son de aspecto lingüístico, más concretamente, lingüístico-computacional, aunque ciertas referencias serán hechas a otras disciplinas como la Inteligencia Artificial y la Psicología Cognitiva. El objetivo común a estas disciplinas es conseguir la comprensión del lenguaje natural. La Psicología Cognitiva (PC) estudia todos los procesamientos de tipo cognitivo capaces de crear conocimiento; la Inteligencia Artificial (IA) es la ciencia que intenta crear, entre otros, sistemas de expertos; y la Lingüística Computacional (LC) es la encargada de estudiar el lenguaje a través de procesamientos computacionales. Es a partir de aquí, donde debemos clarificar en qué campo nos movemos, pues tras la consulta de diferentes definiciones de Lingüística Computacional, parece no estar clara cuál es su concreta ubicación:

“(…) desde el punto de vista de su vinculación a la lingüística, la lingüística computacional puede ser considerada una subdisciplina de la lingüística teórica, en tanto que uno de sus objetivos es la elaboración de modelos formales (e implementables informáticamente) del lenguaje humano”. (J. Gómez Guinovart, 1998:135)

“(…) Los lingüistas suelen hablar de la lingüística computacional como un área de conocimiento dentro de la lingüística aplicada. En cambio, debido a la posibilidad de desarrollar sistemas de computación que simulen algún aspecto de la capacidad lingüística del ser humano, los informáticos consideran la lingüística computacional como una rama de la inteligencia artificial, al igual que los sistemas expertos o la robótica, en cuyo caso prefieren hablar de procesamiento del lenguaje natural (PLN).” (Periñán-Pascual y Arcas- Túnez 2011: 1)

De este modo, es obvio decir que la LC está íntimamente relacionada con la rama de la Lingüística, por sus empeños en describir el lenguaje. Sin embargo, del mismo modo, no podemos ignorar su relación con la IA, y más específicamente con el PLN, cuando pasamos a contemplar las herramientas de las que hace uso la LC para llevar a cabo sus estudios del lenguaje.

Es posible encontrar, en algunos estudios, la alternancia del uso del término LC con el del PLN, pues ambos consideran idóneo el hecho de recurrir a los ordenadores para llevar a cabo sus investigaciones sobre los procesos de comprensión del lenguaje. Según J. Allen (1995:1)

“The goal of this research is to create computational models of language in enough detail that you could write computer programs to perform various tasks involving natural language. The ultimate goal is to be able to specify models that approach human performance in the linguistic tasks of reading, writing, hearing, and speaking.”

Por lo tanto, el PLN procede al diseño de técnicas de representación, recurriendo a un metalenguaje que permita la introducción de datos lingüísticos referidos a los distintos niveles lingüísticos.

Los principales fundamentos lingüísticos influyentes en el desarrollo que nos ocupa en nuestra investigación, se encuentran en la Gramática del Papel y la Referencia y el Modelo Léxico Construccional. Su influencia es relevante pues ambos contienen aspectos clave para cualquier base conceptual de procesamiento

del lenguaje. Por un lado comparten una visión funcional del lenguaje; por otro, contienen un carácter de adecuación tipológica, el cual les hace potentes para un estudio sistemático universal para las lenguas.

3.1. La Gramática del Papel y la Referencia

La Gramática del Papel y la Referencia es una teoría de corte funcional que posee una perspectiva comunicativa-cognitiva del lenguaje (Van Valin y La Polla, 1997). Esto se traduce en que el estudio de la lengua debe considerarse teniendo en cuenta las funciones semánticas y comunicativas del lenguaje.

En contraposición a las posturas tradicionales formalistas, las teorías funcionales centran su atención del estudio del lenguaje en las funciones, el significado y el uso real del hablante:

- las funciones, porque condicionan la forma gramatical,
- significado: porque la gramática es construida sobre contenidos conceptuales y semánticos, y
- uso: porque la función comunicativa del lenguaje es la que determina el uso que hace el hablante de ella.

De entre los modelos funcionales que han surgido en los últimos años, encontramos teorías funcionales conservadoras, moderadas y extremas. La GPR se considera de estilo moderado, pues consideran que la validez de la estructura gramatical viene determinada por las funciones semánticas y comunicativas.

De todo esto se concluye que se trata de una teoría que pretende analizar la interacción entre la estructura y la función en el lenguaje, de modo que los elementos determinantes del lenguaje son la semántica y la pragmática, dejando a la sintaxis como elemento dependiente de ambas.

Desde una perspectiva histórica, la GPR está íntimamente relacionada con la Gramática de Casos de Fillmore (1968), pues coinciden en el planteamiento de un paradigma lingüístico que cuente con un nivel de descripción semántica proyectado sobre una estructura sintáctica. Asimismo, ambas plantean la consideración de que

algunos elementos sintácticos, como el sujeto o el objeto directo, no tienen un carácter universal. La principal diferencia entre ellas es el modo en el que tratan la relación sintáctico-semántica.

El origen de la GPR tiene, como esencia, la búsqueda de una teoría lingüística capaz de dar explicación a los rasgos lingüísticos de todas las lenguas del mundo. En este sentido, la GPR pretende desarrollar un marco lingüístico que permita la descripción estructural de todas las lenguas ya sean lenguas configuracionales (lenguas cuyos elementos siguen un cierto orden) o no configuracionales (lenguas en las que el orden de los elementos no está predeterminado). (Mairal and Gil, 2003). Así, el orden por excelencia de las lenguas como el inglés o el español es el de SVO – *Peter is working at school*; mientras que lenguas como el lakota o el japonés mantienen una estructura SOV- *Peter at school work*. Otras lenguas como las amerindias siguen un patrón en el que el verbo precede al resto de elementos, VOS- *work school Peter*; y finalmente lenguas como el ruso tienen una estructura variable que no afecta al rasgo gramatical de la oración.

En relación a su estructura, la GPR se trata de una teoría monostratal, es decir, donde los elementos sintácticos y semánticos son trazados sin la intervención de abstractas representaciones sintácticas, sino que lo hace a través de un algoritmo de enlace que expresa la interfaz sintáctico-semántica.

En resumen, se trata de una teoría que busca la adecuación tipológica en el desarrollo de un marco lingüístico que sirva para describir los elementos que determinan la transmisión de significado en las lenguas. Más detalladamente, la estructura de GPR se compone de:

- a) un lexicón que almacena toda la información semántica de las unidades lingüísticas
- b) un inventario sintáctico que contiene plantillas sintácticas que se suceden en la gramática de una lengua concreta
- c) una representación de la estructura clausal del enunciado

Los dos primeros niveles están interrelacionados a través de un algoritmo de enlace que se caracteriza por ser bidireccional: de la sintaxis a la semántica y de la semántica a la sintaxis, de modo que da cuenta de aspectos formulados tanto por el emisor como por el receptor en una situación comunicativa.

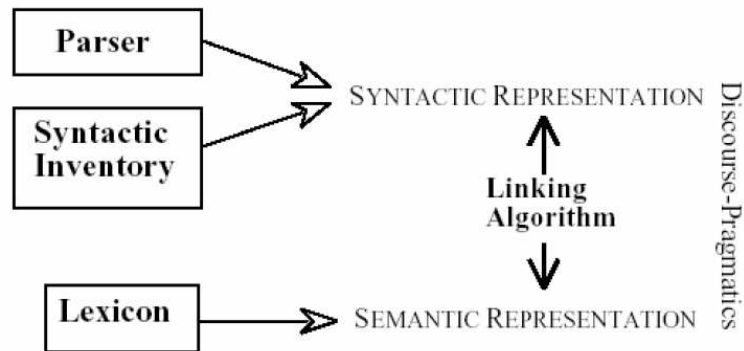


Figura 1. La arquitectura general de la GPR (Van Valin, 2005:131)

En el siguiente apartado pasaremos a explicar cómo la GPR procede a analizar el significado de las expresiones lingüísticas en su primer nivel, el lexicon.

3.1.1. El Lexicón en la GPR

Algunos estudios han demostrado que las estructuras morfosintácticas y gramaticales de una lengua vienen determinadas por el significado. Por esta razón, el lexicon, que se conoce como el lugar donde se almacena la información semántica, se ha convertido en una pieza fundamental en el análisis lingüístico de los últimos años:

(...) "It is now standardly assumed by most linguistic frameworks (both computational and theoretical) that much of the structural information of a sentence is best encoded from a lexicalized perspective."

(Pustejovsky, 1995:5)

En las teorías funcionalistas, las oraciones son tratadas como construcciones gramaticales para transmitir significado. En esta línea, las oraciones son producciones de los hablantes para describir situaciones, reales o no, y las

entidades envueltas en dichos contextos. Tales situaciones son llamadas Estado de Cosas (*State of Affairs*) en la GPR y se clasifican según los siguientes criterios:

a. Situaciones: son estáticas. Son situaciones típicas que describen alguna emoción interna o las condiciones de las entidades:

i.e.: Someone loving pasta

b. Acciones: al contrario que las situaciones, éstas son dinámicas:

p.ej: the earth rotating

c. Procesos: se oponen a las dos anteriores en que los procesos deben tener un final y suponen un cambio que se produce en un periodo de tiempo.

p.ej: a book falling on the floor

d. Eventos: son acciones instantáneas y por lo tanto contrarias a los procesos, pues no implica una duración en el tiempo.

Se ha demostrado que los hablantes hacen uso de estas diferencias semánticas y construyen oraciones según los tipos de Estado de Cosas descritos previamente. Ésta es la razón por la que la GPR comienza a construir la representación semántica de una oración a partir de su predicado y la descripción semántica de éste. Pero, ¿cómo se describe el significado de un predicado? La GPR clasifica los verbos en clases léxicas según su *Aktionsart* o sus propiedades internas temporales, siguiendo la clasificación de los verbos de Vendler (1967). A su vez cada tipo de *Aktionsart* puede tener una contraparte causativa, que establece que la acción ha sido provocada por un SoA diferente. A continuación mostramos algunos ejemplos de los *Aktionsart* con su opción causativa (González-Vergara, 2006:108):

- | | |
|-----------------------|-----------------------------|
| a. Estado: | El niño está asustado |
| a'. Estado causativo: | Los payasos asustan al niño |
| b. Actividad: | La pelota rodó por el campo |

- b'. Actividad causativa: El jugador rodó la pelota por el campo
- c. Logro: El jarrón se destrozó
- c'. Logro causativo: El gato destrozó el jarrón
- d. Realización: La ropa se secó
- d'. Realización causativa: El viento secó la ropa
- e. Realización activa: El perro paseó hasta el parque
- e'. Realización activa causativa: Mi hermana paseó al perro hasta el parque

En este sentido, en la GPR, cada clase léxica es representada formalmente como una Estructura Lógica (EL), la cual constituye el núcleo de la representación semántica de las unidades léxicas en el Lexicón. Esta EL necesita un sistema de descomposición semántica. La característica fundamental, que el metalenguaje para la construcción de estructuras lógicas de los verbos debe contener, es que sea válido para todas las lenguas. En la siguiente tabla se puede observar un inventario de clasificación de las EL usadas en la GPR:

Verb Class	Logical Structure
State	predicate' (x) or (x,y)
Activity	do' (x, [predicate' (x) or (x,y)])
Achievement	INGR predicate' (x) or (x,y), <i>or</i>
Accomplishment	INGR do' (x, [predicate' (x) or (x,y)]) BECOME predicate' (x) or (x,y), <i>or</i> BECOME do' (x, [predicate' (x) or (x,y)])
Semelfactive	SEML predicate' (x) or (x,y), <i>or</i> SEML do' (x, [predicate' (x) or (x,y)])
Active accomplishment	do' (x, [predicate'₁' (x, (y))] & INGR predicate'₂' (z,x) or (y))
Causative	α CAUSES β where α, β are LS of any type

Figura 2. Inventario de Estructuras Lógicas (Van Valin, 2005:45)

De esta manera, la GPR permite que cada Estado de Cosas sea representado a través de un metalenguaje semántico universal en formato de EL. Por ejemplo, la estructura lógica de la unidad léxica *querer* adopta la forma del estado (1).

(1) *want'* (*x*, *y*)



The boy(x) wanted some water (y)

3.1.2. La Estructura Relacional de la Cláusula

Desde una perspectiva sintáctica, existen dos aspectos básicos de la estructura de la cláusula que toda teoría lingüística debería considerar:

- La estructura relacional que se ocupa de las relaciones entre un predicado y sus argumentos de carácter semántico, sintáctico, pragmático o una combinación de ellos;
- La estructura no-relacional que se encarga de la organización jerárquica de oraciones, cláusulas y frases.

Más concretamente, acorde a la GPR, hay dos consideraciones generales que se deben reunir en una teoría sobre las estructuras de las cláusulas: la consideración de todos los rasgos universales de las cláusulas y la representación de un análisis comparativo de estructuras de diferentes lenguas.

Estas consideraciones son las que se contemplan en la concepción de la estructura clausal defendida en la GPR. En primer lugar, porque se ocupa de la combinación de la sintaxis, la semántica y la pragmática en sistemas gramaticales; en segundo lugar por su adecuación tipológica, es decir, la búsqueda de los aspectos comunes de la estructura clausal en todas las lenguas. Es en su búsqueda de la adecuación tipológica, cuando la GPR establece una estructura sintáctica en la que diferencia elementos universales de elementos específicos de una lengua. En los elementos universales encontramos elementos predicantes y no predicantes. La teoría propone tres unidades sintácticas que constituyen la estructura de la cláusula: (a) el núcleo (*nucleus*), que contiene el predicado, (b) el centro (*core*), que contiene el núcleo y los argumentos del predicado; y (c) la periferia (*periphery*), que incluye los elementos no argumentales del predicado. Estos elementos universales

son los que componen la noción de estructura clausal (no-relacional) en la GPR la cual recibe el nombre de *Estructura Jerárquica de la Cláusula*.

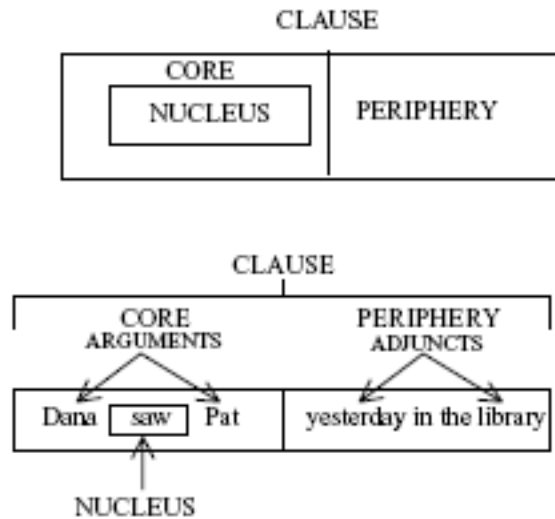


Figura 3. Componentes de la estructura jerárquica de la cláusula (Van Valin, 2005:4)

A modo de ejemplo, la figura 4 incluye un predicado verbal, el cual es el núcleo de la oración y tres argumentos nucleares, además de un elemento no argumental en la periferia:

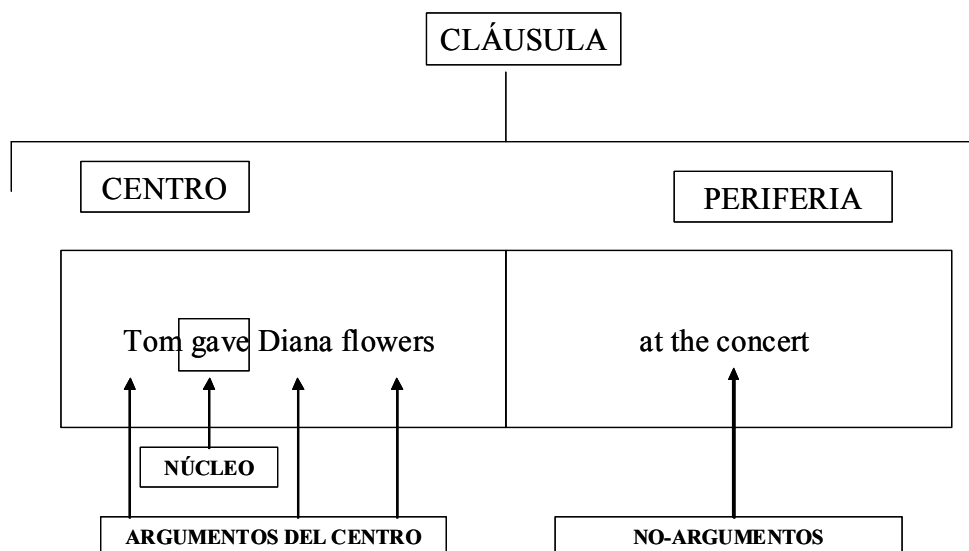


Figura 4. Estructura Jerárquica de la Cláusula

El núcleo de la oración es el verbo *give* que representa el predicado verbal y conlleva tres argumentos. Además encontramos en la periferia un elemento no argumental. De este modo, el centro está formado por el predicado y sus argumentos, mientras que la suma del *core* y la periferia son los que conforman la oración.

En los elementos específicos de la lengua, la GPR incorpora dos nuevas posiciones en la oración: *precore slot*, PCRs, el cual contiene elementos interrogantes; y *Right and Left Detached Positions*, RDP y LDP, que representan elementos adicionales a la oración a la derecha o izquierda de la enunciación y suelen estar separados del resto mediante una pausa. A modo de ilustración, la figura 5 contiene un adverbio temporal que modifica a toda la oración y está separado de la estructura jerárquica de la cláusula mediante una coma:

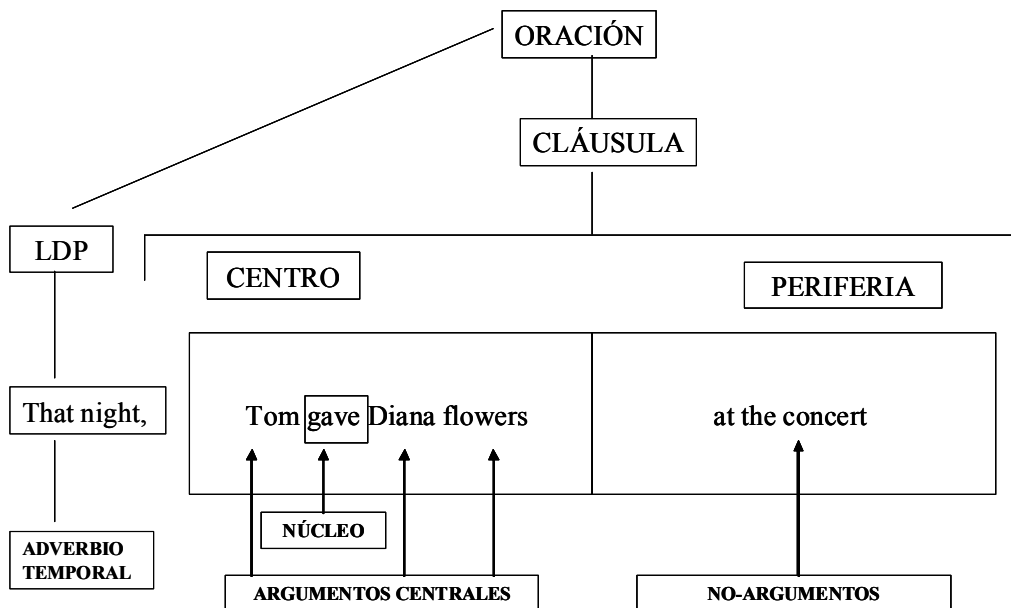


Figura 5. Estructura Jerárquica de la Cláusula con LDP

Como se ha comentado anteriormente, los componentes semánticos y sintácticos de esta teoría se proyectan por medio de un algoritmo de enlace que incluye un conjunto de reglas que da cuenta de la interfaz semántico-sintáctica. En

la siguiente tabla podemos observar las unidades semánticas que se esconden bajo las unidades sintácticas de la estructura jerárquica de la cláusula según la GPR:

Semantic element(s)	Syntactic unit
Predicate	Nucleus
Argument in semantic representation of predicate	Core argument
Non-arguments	Periphery
Predicate + Arguments	Core
Predicate + Arguments + Non-arguments	Clause (= Core + Periphery)

Figura 6. Elementos Semánticos tras las Unidades Sintácticas de la cláusula (Van Valin, 2005:5)

3.2. El Modelo Léxico Construccional

El Modelo Léxico Construccional (Ruiz de Mendoza y Mairal, 2008) es una teoría de construcción del significado que pretende explicar cómo todos los aspectos de la construcción del significado interaccionan entre sí. Para ello su arquitectura se compone de cuatro niveles que proporcionan los datos para la construcción del significado: el nivel argumental (core grammar), el nivel implicacional, el nivel ilocutivo y el nivel discursivo.

La interrelación entre los cuatro niveles se ve determinada por dos procesos cognitivos: el de subsunción y el de pautación. El primero (*subsumption*) da cuenta de la integración de estructuras semánticas al limitar bidireccionalmente la inserción de plantillas léxicas con construccionales: por un lado, las limitaciones internas determinan hasta qué punto el significado de un predicado léxico puede adaptarse al ser combinado con una plantilla construccional; y por otro lado, las limitaciones externas especifican de qué manera una plantilla construccional cambia el significado de un predicado.

El segundo proceso (*cueing*) proporciona la información de cómo damos sentido a las expresiones lingüísticas, al proporcionarle indicios de diferentes fuentes que pueden limitar la interpretación.

Debido a su intención de afinar la descripción de los rasgos de las unidades léxicas, el MLC supone un marco idóneo para el desarrollo de aplicaciones del PLN basadas en una semántica profunda, como es el caso de FunGramKB. Por ello, para la elaboración de nuestro proyecto nos centraremos en la información que el MLC proporciona en su primer nivel, argumental o Gramática Nuclear, donde encontramos las relaciones argumentales entre el léxico y la gramática.

3.3. Introducción a FunGramKB

FunGramKB (www.fungramkb.com) es una base de conocimiento léxico-conceptual que aporta un cambio de paradigma en los sistemas de comprensión del lenguaje natural. FunGramKB está siendo configurada por un grupo de investigadores para la construcción de sistemas para la comprensión del lenguaje fundamentados en la combinación de diferentes ciencias y disciplinas tales como la ciencia cognitiva, la lingüística y la inteligencia artificial. Su principal propósito es su aplicación en sistemas del PLN:

FunGramKB es de carácter multipropósito (Periñán-Pascual y Arcas- Túnez, 2010): por un lado es una base de conocimiento multifuncional pues se puede aplicar a diferentes tareas del PLN como la extracción de información, la traducción automática, la recuperación de documentos, etc.; por otro lado es multilingüe, pues su desarrollo se ejecuta se está ejercitando sobre diversas lenguas como el inglés, el búlgaro, el español o el francés, entre otras.

Su principal propósito es su aplicación en sistemas del PLN. Sus aportaciones en este campo producen un giro computacional proporcionando modificaciones en la orientación del modelo lingüístico, abandonando una base lexicista para adoptar una conceptualista u ontológica. (Mairal-Usón y Periñán Pascual, 2010).

Los niveles léxico y gramatical de FunGramKB se enmarcan en teorías lingüísticas que permiten al sistema captar las generalizaciones sintáctico-semánticas que proporcionan tanto explicaciones como predicciones sobre los fenómenos lingüísticos del fenómeno del lenguaje. En este sentido los fundamentos

lingüísticos de FunGramKB se sustentan sobre la Gramática del Papel y la Referencia (Van Valin y LaPolla, 1997; Van Valin 2005) y el Modelo Léxico Construccional (Ruiz de Mendoza y Mairal, 2008)

De este modo, tanto la GPR como el MLC comparten características fundamentales para cualquier modelo computacional del lenguaje: por un lado, su perspectiva funcional del lenguaje permite transmitir el matiz semántico de los fenómenos gramaticales; por otro lado, su adecuación tipológica permite la introducción de las distinciones universales como parte del marco lingüístico y, por lo tanto, resulta en un ingrediente esencial para los modelos multilingües.

Para el estudio de las influencias teóricas que encontramos en el modelo de FunGramKB, debemos remitirnos a la Ciencia cognitiva. El término “esquema”, creado en la psicología cognitiva, es equivalente a una representación mental de una entidad o evento en la mente del emisor y por lo tanto contiene un conocimiento basado en experiencias pasadas, lo cual simplifica el proceso de inferencia de la información a partir de nuestra conceptualización de la realidad. Los esquemas representan nuestro saber del mundo en el momento de la comunicación. De este modo, la comprensión del lenguaje se ve determinada por la competencia pragmática de los participantes. Por ejemplo, si una persona dice *Juan estuvo esperando en la consulta*, el conocimiento generalizado nos hará pensar que Juan es un paciente. Sin embargo, si seguimos escuchando y oímos *Juan atendió a más de treinta pacientes hoy*, podremos inferir que Juan es médico.

De este modo, el modelo de esquema se convierte en una pieza fundamental para la representación del conocimiento conceptual en FunGramKB. (Periñán-Pascual y Arcas- Túnez, 2010). A partir de los esquemas conceptuales se facilita la inferencia del conocimiento que tiene lugar en el proceso de comprensión del lenguaje.

FUNGRAMKB

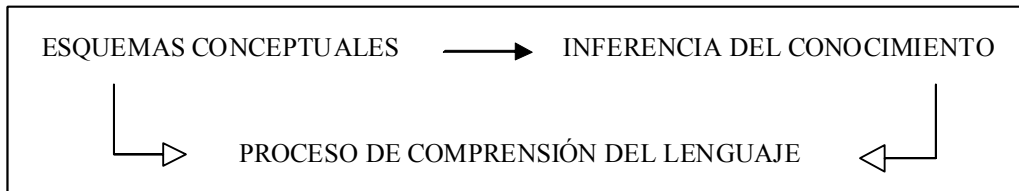


Figura 7. Proceso de comprensión del lenguaje en FunGramKB

Los esquemas conceptuales pueden ser clasificados siguiendo dos criterios:

- a) **Prototipicidad:** que da como resultado protoestructuras, es decir aquellos esquemas conceptuales que contienen un valor prototípico; y bioestructuras, que suponen descripciones de las instancias de eventos y entidades. A modo de ilustración, la unidad léxica *novela* implicaría la descripción del concepto prototípico al que va asociado; mientras que la palabra *Cervantes* implicaría la descripción de una entidad con nombre propio y por lo tanto el inventario de su bioestructura.
- b) **Temporalidad:** determina si el concepto necesita ser descrito dentro de un paradigma temporal (macroestructuras) o, si por el contrario, puede ser presentado de forma atemporal (microestructuras). Es decir, la descripción del esquema conceptual de la biografía de John Lennon necesita una macroestructura, mientras que una microestructura sería suficiente para la descripción de la profesión de cantante.

Como resultado de la unión de estos dos parámetros, obtenemos como resultado la tipología de esquemas conceptuales que muestra la figura 8:

		TEMPORALIDAD	
		-	+
PROTOTIPICIDAD	+	Protomicroestructura (postulado de significado)	Protomacroestructura (guión)
	-	Biomicroestructura (retrato)	Biomacroestructura (historia)

Figura 8. Tipología de los esquemas conceptuales en FunGramKB. (Periñán-Pascual y Arcas- Túnez, 2010:5)

3.3.1. Estructura de FunGramKB

La estructura de FunGramKB está compuesta por tres niveles de conocimiento, cada uno de los cuales está integrado por diferentes módulos independientes pero interrelacionados entre sí:

- a) Nivel Léxico
 - i. El Lexicón: está encargado de almacenar la información morfosintáctica de las unidades léxicas.

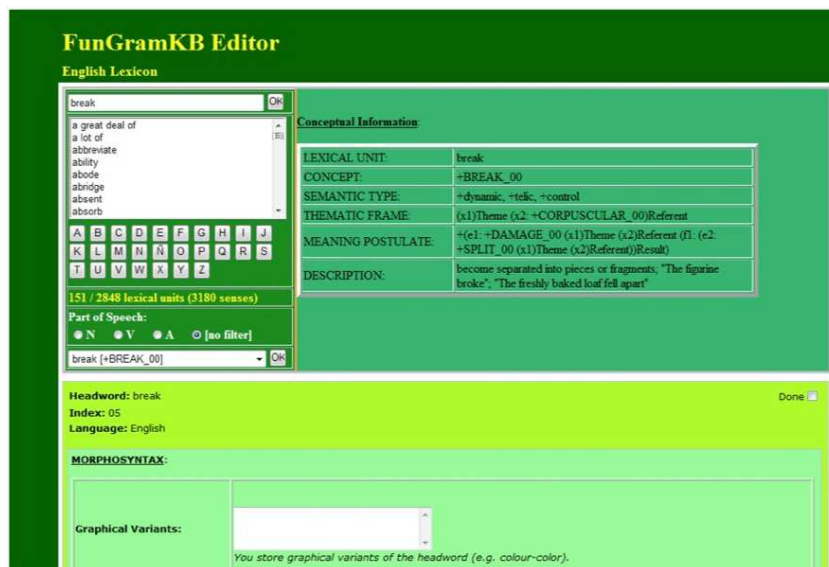


Figura 9 : Entrada léxica del verbo break en el Lexicón de FunGramKB.

- ii. El Morfición: ayuda al analizador y al generador con el tratamiento de los casos de morfología flexiva
- b) Nivel Gramatical
- i. El Gramaticón: almacena los esquemas constructivos que pueden ser utilizados por el algoritmo de enlace sintáctico-semántico de la Gramática del Papel y la Referencia.
- c) Nivel Conceptual
- i. La Ontología: Su estructura es de una jerarquía IS-A de unidades conceptuales. Dichas unidades conceptuales contienen el conocimiento semántico en forma de postulados de significado.

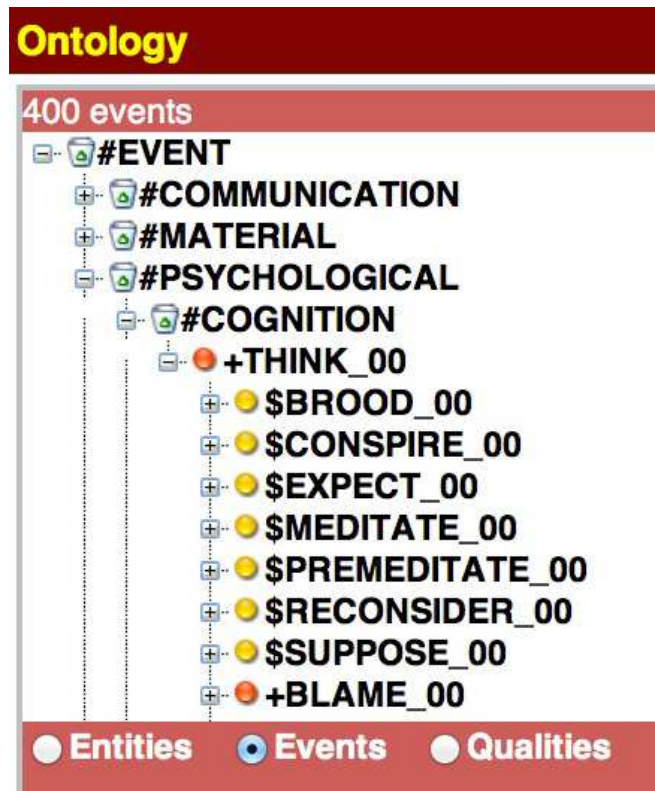


Figura 10: Estructura jerárquica en la Ontología de FunGramKB

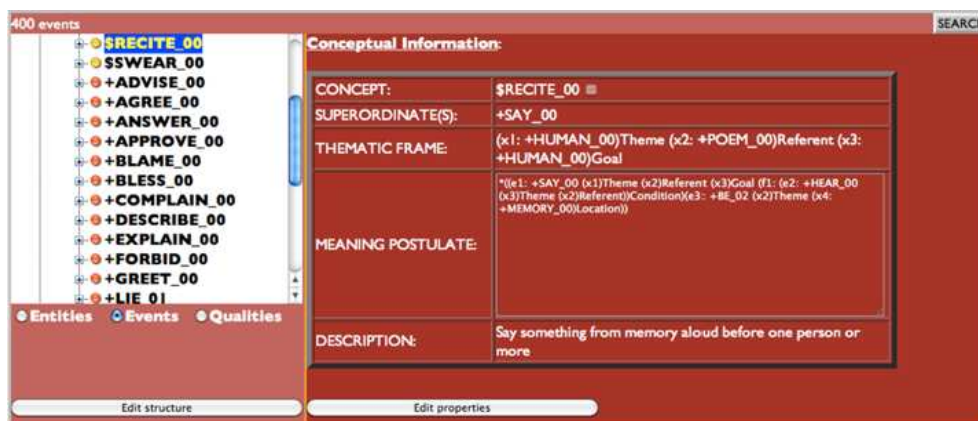


Figura 11: Entrada léxica del verbo recite en la Ontología de FunGramKB

- ii. El Cognición: almacena el conocimiento procedimental a través de guiones, esto es, una serie de eventos estereotípicos dentro de un marco temporal, más concretamente sobre la base del Modelo Temporal de Allen (1983).
- iii. El Onomasticón: es el encargado de almacenar el conocimiento sobre instancias de entidades y eventos. Este módulo almacena dos tipos diferentes de esquemas (retratos e historias) ya que las instancias pueden ser retratadas sincrónica o diacrónicamente.

De estos niveles el único que es compartido por todas las lenguas es el nivel conceptual, mientras que el nivel léxico y el nivel gramatical son particulares para cada una de las lenguas. De este modo se infiere que el lingüista que trabaje en FunGramKB debe crear una Ontología, un Cognición y un Onomasticón únicos para la creación de entradas de unidades léxicas- conceptuales a todas las lenguas.

La piedra angular de la base de conocimiento FunGramKB es la Ontología, ya que recordamos que estamos tratando con una base de conocimiento léxico-conceptual donde las unidades léxicas son descritas conceptualmente y no son tratadas tradicionalmente desde un enfoque lexicista. Esto se traduce en una reducción de la

redundancia y un incremento de la información en cada una de las entradas de unidades léxicas.

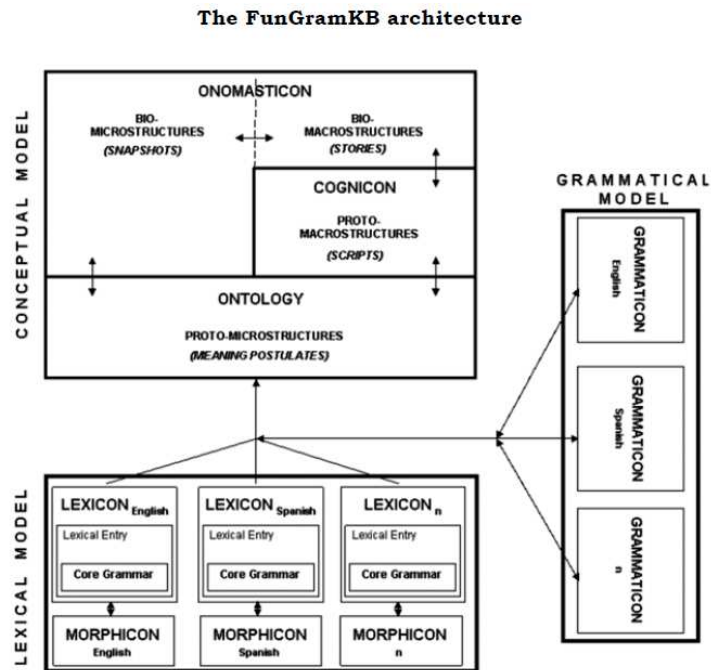


Figura 12. Arquitectura completa y la interrelación de los tres niveles en FunGramKB (Periñán- Pascual y Arcas Túnez, 2011:3)

Según Periñán-Pascual y Mairal Usón (2010), los tres módulos que integran el nivel conceptual, corresponden a los diferentes tipos de conocimiento que, acorde a las teorías de la psicología cognitiva, conforman el razonamiento humano:

- El conocimiento semántico que almacena la información sobre las palabras, coincide con la Ontología que representa esta información a través de proto-microestructuras.
- El conocimiento procedimental que recoge los datos sobre cómo las acciones se realizan en situaciones cotidianas, coincide con el Cognición que representa ese conocimiento en forma de macroestructuras cognitivas.

- El conocimiento episódico almacena información sobre eventos específicos o situaciones biográficas bajo un conjunto de bio-estructuras

Según Tulving (1985) la memoria a largo plazo se construye a partir de la combinación de estos tres conocimientos que nos ayudan a recuperar y almacenar la información. Del mismo modo es la interacción entre estos tres niveles (p.ej. Ontología, Cognición y Onomasticón) la que permite en FunGramKB una comprensión textual.

Para ello, FunGramKB ha creado COREL (*Conceptual Representation Language*-Periñán-Pascual y Mairal-Usón, 2010): un lenguaje de representación conceptual que está encargado de ser la herramienta para la definición de las unidades conceptuales que contiene FunGramKB:

“(...) una de las cuestiones claves para el desarrollo de un motor de razonamiento como parte de un sistema para el procesamiento del lenguaje natural consiste en el desarrollo de un único lenguaje de interfaz para la representación de cada uno de estos esquemas conceptuales, lo que permitirá compartir información más fácil y rápidamente entre cada uno de estos tres niveles.” (Periñán-Pascual y Mairal Usón, 2010: 4)

3.3.2. Lenguaje de Representación Conceptual (COREL)

El lenguaje COREL contiene dos niveles, sintáctico y semántico, que se encargan de describir las unidades conceptuales junto con los elementos que son relevantes en las construcciones, como son los argumentos y los satélites. Asimismo, COREL cuenta con un sistema de notación para otros elementos como los operadores.

Para Periñán-Pascual y Arcas-Túnez (2010) la ontología de FunGramKB contiene tres diferentes niveles conceptuales que contemplan, a su vez, conceptos de diferente tipo:

a) los metaconceptos: son el nivel más alto de esta clasificación taxonómica y están precedidos por el símbolo #. Conducen a un modelo metaconceptual cuyo diseño contribuye a la integración e intercambio de información con otras ontologías. El resultado final son 3 subontologías bajo las categorías de #ENTITY, #EVENT y #QUALITY que comprenden un total de 42 metaconceptos en total.

b) los conceptos básicos: están precedidos por el símbolo + y se usan en FunGramKB como unidades de definición que posibilitan la construcción de postulados del significado para conceptos básicos y terminales. Asimismo forman parte de las preferencias de selección en los marcos temáticos.

c) los conceptos terminales: van encabezados por el símbolo \$ y al contrario que los conceptos básicos no poseen poder definitorio al no tener una jerarquía.

Los conceptos básicos y terminales en FunGramKB son dotados de datos semánticos como el marco temático y el postulado del significado, los cuales construyen la descripción del significado a través de conceptos. Ambos constituyen esquemas conceptuales al ser contruidos a partir de conceptos y dar como resultado representaciones del conocimiento semántico formalizado en lenguaje COREL. De esta manera, a cada evento de la ontología se le otorga un marco temático, es decir, una construcción conceptual que establece el número y tipo de participantes implicados en una situación cognitiva descrita por el concepto. Como ilustración podemos observar el siguiente ejemplo que encontramos en Mairal-Usón y Periñán-Pascual (2009:223) sobre el marco temático del concepto +GIVE_00²:

(1) (x1: +HUMAN_00 ^ +ANIMAL_00)Agent (x2: +CORPUSCULAR_00)
Theme (x3)Origin (x4: +HUMAN_00 ^+ANIMAL_00)Goal

Los marcos temáticos también pueden abarcar aquellas preferencias de selección incluidas en la situación cognitiva, es decir, aquellas que ejercen algún tipo de poder predictivo sobre los participantes.

² El concepto +GIVE_00 está categorizado bajo el metaconcepto #TRANSFER y sus unidades léxicas pueden ser *give* o *provide*.

En este escenario es importante el hecho de la asignación de un único marco temático por concepto, mientras que las unidades léxicas pueden ser subcategorizadas bajo diferentes patrones de subcategorización. Asimismo, los conceptos descritos en la ontología de FunGramKB reciben un único postulado de significado, el cual es un conjunto de una o más predicaciones conectadas por las características del concepto.

3.3.3. FungramKB y el Lexicón

El lexicón está encargado de almacenar la información morfosintáctica de las unidades léxicas y de representar semánticamente los predicados de un idioma. La Gramática Nuclear del Lexicón contiene esos elementos cuyos valores permiten al sistema construir de manera automática la Estructura Lógica Conceptual (ELC) de las unidades léxicas.

El modelo de ELC de FunGramKB es un modelo mejorado de la Estructura Lógica presentada en la GPR. De este modo, cada variable en la ELC de una unidad léxica es referenciado a través de papeles temáticos a un participante del marco temático del concepto, al cual esa unidad léxica está unida, y, a su vez, cada participante del marco temático está referenciado a través de la coindización a un participante en el postulado del significado de ese concepto.

De este modo, la representación (1) hecha por la GPR en formato de Estructura Lógica es reemplazada por la ELC de FunGramKB en (2)

(1) Want (x, y)

(2) +WANT_00 (x-Theme, y-Referent)

Las entradas léxicas son representadas en forma de plantillas léxicas las cuales son una forma alternativa de representación léxica, basadas en una noción mejorada de las estructuras lógicas presentadas por la GPR. Por lo tanto, podemos concluir que la misión de la ELC es la de vincular en FunGramKB el nivel abstracto desarrollado en la Ontología con las características lingüísticas descritas para cada unidad léxica en el Lexicón. De este modo, según Periñán-Pascual y Mairal- Usón

(2009) “la ELC es usada como la interfaz entre la estructura semántica y la representación sintáctica de oraciones”.

4. PROPUESTA DE TRABAJO

Ya hemos visto que en la ciencia cognitiva, existe una extensa literatura en torno a las preposiciones, particularmente sobre las preposiciones espaciales y temporales (Langacker 1991; Talmy 2000; Tyler and Evans 2003) desde diferentes perspectivas.

Sin embargo, como se ha comentado previamente, las investigaciones de esta categoría gramatical en el PLN no son muy numerosas. Sólo unos pocos trabajos, p.ej. The Preposition Project (Litkowski y Hargraves, 2005) o PrepNet (Saint-Dizier, 2005, 2008), se han ocupado de proporcionar una descripción de la semántica preposicional suficientemente robusta para las aplicaciones del PLN. De hecho, como nos recuerdan Sven, Helbig y Osswald (2006), las investigaciones lingüísticas sobre la semántica de las preposiciones no han tenido mucho impacto en las aplicaciones del PLN. No obstante, como reconoce Saint-Dizier (2008), se precisa tener un repositorio que organice y caracterice los comportamientos sintácticos y semánticos de las preposiciones, una tarea que resulta de vital importancia incluso en aplicaciones del PLN que requieran una forma simple de comprensión, p.ej. extracción de conocimiento o traducción automática. En este sentido, el principal problema que debemos afrontar es la resolución de la ambigüedad léxica en una categoría gramatical que resulta ser extremadamente polisémica, donde la interpretación semántica de la preposición depende estrechamente del contexto, por no mencionar la dificultad de identificar las regularidades interlingüísticas.

Es en esta razón donde encontramos el origen de nuestra investigación. Nuestro propósito es la interpretación semántica de los sintagmas preposicionales predicativos en el marco de la GPR, más concretamente, aquellos sintagmas preposicionales que funcionan como elemento no argumental en la periferia de la cláusula.

Nuestro trabajo requiere el diseño y desarrollo de un componente conceptual que almacene la semántica de las preposiciones predicativas, las cuales “contribute semantic information to the clause in which they occur, both in terms of their own meaning and the meaning of the argument that they license” (Van Valin 2005: 21). En otras palabras, es preciso determinar (a) el componente semántico que afectará al esquema COREL de la oración, y (b) el papel temático que se asignará al adjunto en la ELC. Generalmente, los elementos que nos ayudan a desambiguar las preposiciones, y por tanto a etiquetarlas semánticamente, son principalmente dos: el término regente del sintagma y el complemento preposicional. En el marco de la GPR, nuestro interés se centra en el verbo como término regente del adjunto preposicional periférico de la cláusula. Además, nos ocupamos exclusivamente de los sentidos básicos de la preposición, donde los sentidos resultantes de extensiones metafóricas o metonímicas podrían ser derivados a partir de reglas de inferencia aplicadas a los sentidos básicos. A modo de ilustración, el siguiente ejemplo presenta una transposición metafórica del sentido espacial al dominio psicológico.

(1) She has a strange power over me. (Lakoff 1987: 435)

En definitiva, el resultado de esta propuesta será un repositorio sobre el conocimiento de la preposición *in* para su utilización en el PLN a través de FunGramKB. Como describimos a continuación, este repositorio se ubicaría principalmente dentro de la Ontología Nuclear y, en menor medida, en el Lexicón.

4.1. Herramientas de estudio

En este marco de trabajo, nuestra tarea comienza con el análisis manual de la preposición, a partir de los sentidos descritos en diversos diccionarios. Además, el análisis ha sido enriquecido con el estudio del contexto que suele acompañar a las diferentes construcciones de los sentidos mediante la exploración del *córpore*. De este modo, las herramientas que hemos utilizado para realizar nuestra investigación son:

- Longman Dictionary of Contemporary English Online: www.ldoceonline.com
- Collins British English Dictionary Online: www.collinsdictionary.com
- Merriam-Webster Dictionary Online: www.merriam-webster.com
- Cambridge Dictionaries Online: www.dictionary.cambridge.org
- Diccionario de la Real Academia Española (RAE): <http://rae.es/rae.html>
- BNC (*British National Corpus*): www.natcorp.ox.ac.uk/
- CREA (Corpus de Referencia del Español Actual): <http://corpus.rae.es/creanet.html>³

El uso del corpus sirvió para analizar las preferencias de selección de los sentidos clasificados para la preposición. En este sentido, se creó una lista de selección de la preposición en diferentes contextos ofrecidos por el corpus como pudieron ser: género teatral, periodístico, científico o miscelánea.

4.2. Metodología de investigación

Nuestra metodología es de naturaleza tanto cualitativa como cuantitativa pues hemos llevado a cabo un análisis comparativo de los sentidos ofrecidos por los diferentes diccionarios mencionados previamente. Dado que uno de nuestras opciones ha sido el interés por el verbo como término regente del adjunto preposicional, tras el análisis de la preposición, se ha utilizado el BNC para enriquecer el análisis con los datos encontrados a efectos de preferencias de selección que pueden aparecer en los marcos temáticos determinados por los sentidos semánticos. La clasificación semántica de los verbos que hemos utilizado para llevar a cabo este análisis ha sido la propuesta de interpretación semántica para los papeles temáticos de los argumentos en FunGramKB (anexo I).

Ya que la descripción semántica precisa del análisis de la preferencia de selección que sigue al SP, durante el estudio también se ha abarcado qué tipos de

³ Debido a la dimensión multilingüe de FunGramKB, es preciso consultar recursos de diversas lenguas.

unidades léxicas aparecían tras cada uno de los sentidos analizados. De este modo, se ha procedido a etiquetar la interpretación semántica con los papeles temáticos de los satélites proporcionada por FunGramKB (anexo II).

4.3. Desarrollo de la investigación

Al igual que los nombres, verbos y adjetivos corresponden cognitivamente a las entidades, los eventos y las cualidades respectivamente, las preposiciones se interpretarían conceptualmente por medio de las unidades ontológicas que denominaremos como “relaciones”. Por tanto, nuestra definición de relación sería la siguiente:

- (2) Una relación en FunGramKB es una unidad ontológica que establece una configuración relacional entre una o más entidades a través de un evento.

De esta forma, siguiendo con el sistema de organización y descripción utilizado en FunGramKB, las preposiciones almacenarían su interpretación semántica en una subontología de relaciones. En esta subontología, los metaconceptos vendrían determinados por los papeles temáticos que asignarían las preposiciones a los sintagmas que encabezarían, p.ej. #GOAL, #ORIGIN, #TIME, etc.⁴ Uno de los beneficios de este enfoque radicaría, por tanto, en poder ubicar los papeles temáticos en la configuración ontológica. En cambio, un tema que deberíamos estudiar con mayor profundidad sería la existencia o no de una jerarquía de relaciones basada exclusivamente en la relación de subsunción. En caso de existir, esta jerarquización sería probablemente poco profunda, pero deberíamos también analizar el impacto que tendrían los conceptos básicos sobre la definición de los terminales, con el fin de determinar el nivel ontológico intermedio de esta subontología. Igualmente, habría que averiguar si la herencia de conocimiento (i.e. postulado de significado) desde la relación superordinada hasta la subordinada tiene lugar de manera monotónica o no monotónica.

⁴ Aprovechamos para comentar que no descartamos la posibilidad de enriquecer el inventario de papeles temáticos de los satélites (Periñán-Pascual y Mairal-Usón 2010, anexo 2, p. 44) con el fin de proporcionar una interpretación semántica más precisa.

Microestructuralmente, las relaciones estarían provistas de un marco temático y un postulado de significado, al igual que ocurre con otras unidades ontológicas en FunGramKB. Siguiendo la definición (2), el marco temático estaría formado por una serie de participantes que denotarían las entidades prototípicas relacionadas. Por ejemplo, en el primer sentido descrito para la preposición *in* (p.ej. +INSIDE_00) el marco temático de la relación en una oración como “He lives in a card box” contaría con una variable:

(3) (x1: +CONTAINER_00)Location

donde x1 hace referencia a “a card box” (+BOX_00). En este caso, propondríamos la preferencia de selección +CONTAINER_00 para el argumento. Tras la creación del marco temático para el concepto, elaboraríamos el postulado de significado de +INSIDE_00, a través de la coindización de la variable. De este modo, el postulado de significado podría adoptar la siguiente forma:

(4) *(e2: +BE_02 (x1)Location (f1: +INSIDE_00)Position)

COREL es un lenguaje de representación conceptual suficientemente expresivo como para poder representar los principales matices en la semántica de las preposiciones, aunque necesitará ser mejorado en la descripción de los valores temporales y de cantidad de la preposición *in*.⁵

Ahora bien, siguiendo con la definición (2), también es preciso tener en cuenta el tipo semántico prototípico del evento centro de nuestra configuración relacional, el cual, aunque no sirva para exigir que un adjunto vaya encabezado por una determinada preposición, sí desempeña un papel importante en la compatibilidad semántica de determinados periféricos del evento.

De acuerdo con el enfoque conceptualista de FunGramKB, toda esta entrada conceptual estaría vinculada a una serie de preposiciones en diversas lenguas. Por

⁵ Ver nota al pie de página 6.

ejemplo, preposiciones como *in* o *inside* en inglés y *en* o *dentro de* en español estarían asignadas a la relación +INSIDE_00 descrita anteriormente. A modo de ilustración, en la oración “He put the milk back in the fridge” el concepto +INSIDE_00 sería expresado en:

- Polaco: Mleko **w** lodówke
- Francés: Il a mis le lait **dans** le frigo
- Alemán: Er stellte die Milch **in** den Kühlschrank
- Italiano: Ha messo il latte **in** frigorifero
- Español: Él puso la leche **en** el frigo

Por último, si fuera necesario especificar información más precisa sobre la realización lingüística de una preposición y esta información no fuera compartida por otras preposiciones vinculadas a una misma relación, entonces tendríamos la oportunidad de almacenar ese conocimiento en la entrada léxica del Lexicón.

4.4. La descripción semántica de la preposición *in*

Como se ha comentado previamente, nuestro proyecto consiste en proporcionar una subontología, #RELATION, a la base de conocimiento léxico-conceptual FunGramKB para abordar mediante una descripción semántica el papel de los argumentos introducidos por la preposición *in*.

En nuestro trabajo procedimos a la descripción semántica de las preposiciones desde una posición de abstracción con respecto al resto de componentes de la oración. La decisión vino determinada por el momento de coindizar las variables del marco temático del término regente en el postulado de significado de la relación. De este modo nos encontramos con la situación de decidir cuántas variables debíamos incluir en el marco temático de la relación.

En la primera fase del estudio, donde nos centramos en encontrar el componente semántico de la preposición *in* que afectará al esquema COREL de la

oración, observamos, a través del análisis de los diferentes sentidos ofrecidos por los diversos diccionarios para la preposición *in*, que dependiendo del término regente, evento o entidad, se proporcionaba una o más variables al marco temático. Éstas correspondían, normalmente, con los elementos que aparecían delante de la preposición en la oración. Sin embargo, tras el estudio de los ejemplos, consideramos que las variables ofrecidas por el término regente podían o no ocurrir, o incluso, podrían referirse a la variable (x1) del marco temático en el caso de los eventos. En este sentido, en la oración “He put the milk back in the fridge” el marco temático del concepto +INSIDE_00 quedaba con tres variables:

(x1)Agent (x2)Theme (x3) Location

de este modo, observamos cómo (x1) y (x2) eran determinadas por un evento.

En este punto, la problemática surgió en el momento de coindizar las variables del marco temático para integrarlas en el postulado de significado, pues el evento podría ser modificado y, por lo tanto, los participantes estimados en el marco temático de la relación podrían ser distintos. Esta conclusión nos hizo alejarnos de esta alternativa, pues el análisis del tipo semántico prototípico del evento centro de nuestra configuración relacional conllevaba mucho más de lo que este trabajo pretendía demostrar.

Por lo tanto, la opción que decidimos considerar fue la de crear un marco temático de la relación con tan sólo una variable, que correspondería al elemento que venía regido por la misma, es decir, al complemento preposicional. En este sentido, el componente semántico que afectaría al concepto +INSIDE_00, sería la variable (x1), la cual correspondería a la entidad que expresaría el lugar que indica la preposición. Así, el marco temático del concepto +INSIDE_00 en la oración “He put the milk back in the fridge” quedaría como sigue:

(x1: +CONTAINER_00)Location

el papel temático que se asignaría al adjunto en la ELC, sería el de *Location*, lo cual coincide con la interpretación semántica de uno de los papeles temáticos de los

satélites establecida por FunGramKB (anexo II). La preferencia de selección para este concepto admitiría los conceptos básicos +CONTAINER_00, +PLACE_00, +BUILDING_00 entre otros.

Esta última apreciación, nos haría considerar la posible creación de una jerarquía de relaciones basada exclusivamente en la relación de subsunción, pero deberíamos también analizar el impacto que tendrían los conceptos básicos sobre la definición de los terminales como se ha comentado previamente.

A continuación, se describen, uno a uno, los conceptos creados para la preposición *in*, siguiendo la descripción metodológica comentada previamente, y su aplicación computacional a través de lenguaje COREL.

1º Concepto: +INSIDE_00

El sentido asignado a este concepto es el de “dentro de un sitio” o, desde un enfoque cognitivo, “dentro de una entidad”. En consecuencia, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto correspondería a la que ha sido comentada en el apartado anterior.

2º Concepto: +INTO_00

El sentido asignado a este concepto es el de “hacia el interior de un sitio” o, desde un enfoque cognitivo, “hacia dentro de una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por un participante:

(x1)Goal

En un ejemplo como “They fell in the sea”, el marco temático del concepto +INTO_00 quedaría representado de la siguiente manera:

(x1: +PLACE_00)Goal

donde x1 hace referencia a “the sea” (+SEA_00). En este caso, propondríamos la preferencia de selección +PLACE_00 para el argumento.

El postulado de significado quedaría elaborado de la siguiente manera, tras la integración de la variable en el marco temático:

**(e1: +MOVE_00 (x2)Agent (x3)Theme (x4)Location (x5)Origin (x1)Goal (f1:
+INTO_00)Position)*

Este concepto se ubicaría dentro del metaconcepto #LOCATION creado en la nueva subontología.

3º Concepto: +INTO_01

El sentido asignado a este concepto difiere del anterior en que no expresa lugar, sino modo. Desde una perspectiva cognitiva, diríamos que la descripción semántica de la relación establecería “una entidad resultante”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por un participante:

(x1)Result

En un ejemplo como “The glass broke in pieces”, el marco temático del concepto +INTO_01 quedaría representado de la siguiente manera:

(x1: +PART_00)Result

donde x1 hace referencia a “pieces” (\$PIECE_00). En este caso, propondríamos la preferencia de selección +PART_00 para el argumento.

**(e1: +DO_00 (x2)Theme (x3)Referent (f1: (e2: +BECOME_00 (x4)Theme
(x1)Attribute))Result)*

Este concepto se ubicaría dentro del metaconcepto #RESULT creado en la nueva subontología.

4º Concepto: +IN_00

El sentido asignado a este concepto es el de “en un momento determinado” o, desde un enfoque cognitivo, “en el momento determinado por una entidad”. De este

modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Time

En esta línea, el marco temático de la relación en la oración “She came in January” podría describirse de la siguiente manera:

(x1: +MONTH_00)Time

donde x1 representaría a “January” (\$JANUARY_00). En este caso, propondríamos la preferencia de selección +MONTH_00 para el argumento.

El postulado de significado creado para el concepto +IN_00 quedaría representado de la siguiente manera:

**(e1: +DO_00 (x2)Theme (x3)Referent (f1: x1)Time)⁶*

La preferencia de selección en el marco temático para este concepto incluiría los conceptos básicos: +PERIOD_00 y +MOMENT_00 entre otros.

El sentido aquí descrito sería emplazado bajo el metaconcepto #TIME, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

5º Concepto: +DURING_00

El sentido asignado a este concepto es el de “durante un tiempo” o, desde un enfoque cognitivo, “durante el marco temporal determinado por una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Duration

⁶ Como se puede apreciar, los postulados de significado para los conceptos +IN_00 y +DURING_00 son iguales. Al fin de esta investigación no hemos encontrado el modo de crear un marco temático más específico para cada uno, pues resulta muy difícil la representación de ambos sentidos con COREL

En esta línea, el marco temático de la relación en la oración “They lived there in the winter” podría describirse de la siguiente manera:

(x1: +SEASON_00)Time

donde x1 representaría a “the winter” (+WINTER_00). En este caso, propondríamos la preferencia de selección +SEASON_00 para el argumento.

El postulado de significado creado para el concepto +DURING_00 quedaría representado de la siguiente manera:

**(e1: +DO_00 (x2)Theme (x3)Referent (f1: x1)Time)⁷*

La preferencia de selección en el marco temático para este concepto incluiría los conceptos básicos: +PERIOD_00 y +MOMENT_00 entre otros.

El sentido aquí descrito sería emplazado bajo el metaconcepto #TIME, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

6º Concepto: +WITHIN_00

El sentido asignado a este concepto es el de “en no más de un tiempo” o, desde un enfoque cognitivo, “en no más del marco temporal determinado por una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Time

En esta línea, el marco temático de la relación en la oración “We’ll send the letter in a minute” podría describirse de la siguiente manera:

(x1: +PERIOD_00)Time

⁷ Como se puede apreciar, los postulados de significado para los conceptos +IN_00 y +DURING_00 son iguales. Al fin de esta investigación no hemos encontrado el modo de crear un marco temático más específico para cada uno, pues resulta muy difícil la representación de ambos sentidos con COREL

donde x1 representaría a “a minute” (+MINUTE_00). En este caso, propondríamos la preferencia de selección +TIME_00 para el argumento.

Al igual que con los dos sentidos anteriores, el postulado de significado de este concepto resulta muy difícil de describir con el fin de diferenciarlo de los postulados precedentes.

La preferencia de selección en el marco temático para este concepto incluiría los conceptos básicos: +PERIOD_00 y +MOMENT_00

El sentido aquí descrito sería emplazado bajo el metaconcepto #TIME, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

7º Concepto: +IN_01

El sentido asignado a este concepto es el de “en cierto estado” o, desde un enfoque cognitivo, “el estado o condición expresado por una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Manner

En esta línea, el marco temático de la relación en la oración “Sarah was cooking in silence” podría describirse de la siguiente manera:

(x1: +STATE_00)Manner

donde x1 representaría a “silence” (+SILENCE_00). En este caso, propondríamos la preferencia de selección +STATE_00 para el argumento.

El postulado de significado creado para el concepto +IN_01 quedaría representado de la siguiente manera:

**(e1: +DO_00 (x2)Theme (x3)Referent (f1: x1)Manner)*

La preferencia de selección en el marco temático para este concepto incluiría los conceptos básicos: +STATE_00

El sentido aquí descrito sería emplazado bajo el metaconcepto #MANNER, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

8º Concepto: +IN_02

El sentido asignado a este concepto es el de “mediante un medio” o, desde un enfoque cognitivo, “mediante el medio o instrumento determinado por una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Means

En esta línea, el marco temático de la relación en la oración “They were built in bronze” podría describirse de la siguiente manera:

(x1: +METAL_00)Means

donde x1 representaría a “bronze” (+BRONZE_00). En este caso, propondríamos la preferencia de selección +METAL_00 para el argumento.

El postulado de significado creado para el concepto +IN_02 quedaría representado de la siguiente manera:

**(e1: +DO_00 (x2)Theme (x3)Referent (f1: x1)Means (f2)Instrument)*

La preferencia de selección en el marco temático para este concepto incluiría los conceptos básicos: +MATERIAL y +UNIT_OF_MEASURE_00 entre otros.

El sentido aquí descrito sería emplazado bajo el metaconcepto #INSTRUMENT, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

9º Concepto: +IN_03

El sentido asignado a este concepto es el de “en una cantidad” o, desde un enfoque cognitivo, “en la cantidad determinada por una entidad”. De este modo, la descripción de los participantes de las entidades prototípicas relacionadas en el marco temático de este concepto estaría constituida por una variable:

(x1)Quantity

En el análisis de este sentido encontramos la problemática de no hallar una entidad referida a numerales en FunGramKB. Los numerales se expresan con operadores de cuantificación con el inconveniente de que no pueden aparecer solos, sino que siempre van acompañados de una preferencia de selección en su argumento. Este análisis conllevaría una futura solución dentro de FunGramKB.

El sentido aquí descrito sería emplazado bajo el metaconcepto #QUANTITY, acorde a los papeles temáticos de los satélites descritos en FunGramKB dentro de la nueva subontología.

5. CONCLUSIONES Y FUTURAS INVESTIGACIONES

A lo largo de esta investigación se ha intentado aportar un enfoque innovador al PLN para la descripción semántica de las preposiciones. El estudio se ha hecho sobre los diferentes sentidos de la preposición *in*, más allá de los usos temporales y espaciales que tanta importancia han obtenido a lo largo de la historia de la lingüística cognitiva. Además se ha intentado la adaptación de la descripción semántica dentro de la base de conocimiento FunGramKB, a partir de un etiquetado elaborado para los papeles temáticos de los argumentos propio de la herramienta.

En relación a la categoría gramatical estudiada, i.e. la preposición, debemos incidir en el hecho de que, como se ha comentado en las investigaciones hechas al respecto en el mundo de la lingüística teórica y cognitiva, la problemática provocada por esta expresión lingüística se refleja tanto en su valor polisémico, como unidad léxica, como en su dependencia de un término regente determinado. Es por esto, que el estudio aquí desarrollado, refleja tan sólo una pequeña muestra de lo que sería el análisis de las construcciones introducidas por la preposición *in*. Mucho más profunda debería ser la consideración de las variables que participan en el marco temático de la preposición, a efectos de la creación de una interpretación semántica, independiente de su término regente.

Por otro lado, en el marco de la GPR, las preposiciones predicativas se consideran predicados de estado, los cuales son considerados básicos para la representación semántica del resto de predicados. Van Valin (2005:49) establece que las preposiciones predicativas, cuando funcionan como adjuntos modificadores de un verbo, asumen la estructura lógica de éste como uno de sus argumentos. En la presente investigación se ha demostrado que es posible la elaboración de un marco temático a nivel cognitivo para una preposición, sin atender a la regencia del evento. Sin embargo, un enfoque centrado en el evento podría resultar en una mayor cohesión cognitiva, al permitir establecer la prototipicidad del término regente de la relación y, por ende, el número de variables involucradas en el marco temático del concepto.

En definitiva, esta investigación aporta a la lingüística cognitiva un enfoque conceptual en la descripción semántica de las preposiciones, que permite la introducción de información semántica en una base de conocimiento léxico-conceptual para el desarrollo de una herramienta del PLN a través de un metalenguaje de representación conceptual como es COREL.

Como futuras investigaciones al respecto del tratamiento semántico de las preposiciones, se propone la ampliación de este estudio, p.ej. con la categorización de todas las preposiciones en lengua inglesa mediante un estudio comparativo con

una lengua románica como es el español para abarcar el objetivo multilingüe de FunGramKB.

6. BIBLIOGRAFÍA

- Allen, J.** (1995 [1987]): *Natural Language Understanding*. Redwood City, Ca.: Benjamin/Cummings, 2ª ed.
- Baker, C.F., Fillmore, C. J., y Lowe, J.B.** (1998). The Berkeley FrameNet Project. En *Proceedings of the COLING-ACL*. Montreal, Canada, pp. 86-90.
- Baldwin, T. Kordoni, V. y Villavicencio, A.** (2009). Prepositions in Applications: A Survey and Introduction to the Special Issue. *Computational Linguistics*, Vol. 35(2):119-147.
- Bennett, D.C.** (1975). *Spatial and Temporal uses of English Prepositions: An Essay in Stratificational Semantics*. London: Longman
- Briscoe, T. y Carroll, J.** (1997). Automatic extraction of subcategorization from corpora. En *Proceedings of the 5th Conference on Applied Natural Language Processing (ANLP)*, Washington DC, USA, pp. 356-363.
- Downing, A. y Locke, P.** (2006). *English Grammar: A University Course*. 2nd ed. Abingdon and New York: Routledge.
- Dorr, B.J.** (1997). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, Vo. 12(4): 271-322.
- Fillmore, C.J.** (1968). The case for case. En E. Bach and R. T. Harms (eds.) *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, pp. 1-88.
- Fillmore, C.J.** (1982). Frame semantics. En H. O. Hakhoe (ed.) *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., pp. 111-137.
- Fillmore, C.J.** (1985). Frames and the semantics of understanding. En *Quaderni di Semantica*, Vol. 6 (2): 222-254.
- Girju, R.** (2009). The Syntax and Semantics of Prepositions in the Task of Automatic Interpretation of Nominal Phrases and Compounds: a Crosslinguistic Study. En

Computational Linguistics Vol. 35(2): 185-228 - Special Issue on Prepositions in Application.

Gómez Guinovart, J. y Palomar, M. (coords.) (1998). *Lengua y Tecnologías de la Información*. Número monográfico de *Novática*, 133.

Gonzalez Vergara, C. (2006). La Gramática del Papel y la Referencia: una aproximación al modelo. *Onomazein*, Vol. 14: 101-140.

Herskovits, A. (1986). *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English*. Cambridge, England: Cambridge University Press.

Huddleston, R. y Pullum, G.K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, Mass: MIT Press.

Jackendoff, R. (1990). *Semantic Structures*. Cambridge, Mass: MIT Press

Kracht, M. (2003). Directionality selection. En *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, Toulouse, France, pp. 89-100.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal About the Mind*. Chicago: The University of Chicago Press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago / London: The University of Chicago Press.

Levin, B. y Tova, R. (1988). Lexical Subordination. En: Diane Brentari / Gary Laron / Lynn MacLeod (eds.): *Papers from the 24th annual regional meeting of the Chicago Linguistic Society*. Vol. 1: 275-289.

Linstromberg, S. (1997). *English Prepositions Explained*. Amsterdam/Philadelphia: John Benjamins Publishing Co.

- Litkowski, K. C.** (2002). Digraph analysis of dictionary preposition definitions. En *Proceedings of the ACL-02 Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, Philadelphia, USA, pp. 9-16.
- Litkowski, K. C. y Hargraves, O.** (2005). The preposition project. En *Proceedings of the 2nd ACL-SIGSEM Workshop on Prepositions*. Colchester, UK, pp. 171-179.
- Mairal Usón, R. y Gil, J.** (eds.) (2003). *En torno a los universales lingüísticos*. Cambridge: Cambridge University Press.
- Mairal Usón, R. y Perrián Pascual, C.** (2009). The anatomy of the lexicon component within the framework of a conceptual knowledge base. *Revista Española de Lingüística Aplicada* Vol. 22: 217-244.
- Mairal Usón, R. y Perrián Pascual, C.** (2010). Teoría lingüística y representación del conocimiento: una discusión preliminar. En Dolores García Padrón y María del Carmen Fumero Pérez (eds.) *Tendencias en lingüística general y aplicada*. Berlín: Peter Lang, pp. 155-168.
- Manning, C.D.** (1993). Automatic acquisition of a large subcategorization dictionary from corpora. En *Proceedings of the 31st Annual Meeting of the ACL*, Ohio, USA, pp. 235-42.
- Marcus, M. P., Santorini, B., y Marcinkiewicz, M.** (1993). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, Vol. 19(2):313-330.
- O'Hara, T. y Wiebe, J.** (2003). Preposition semantic classification via Treebank and FrameNet. En *Proceedings of the 7th Conference on Natural Language Learning*, Edmonton, Canada, pp. 79-86
- Perrián-Pascual, C. y Arcas-Túnez, F.** (2007). Deep semantics in an NLP knowledge base. En *12th Conference of the Spanish Association for Artificial Intelligence*, Universidad de Salamanca, Salamanca, España, pp. 279-288.

- Periñán-Pascual, C. y Arcas-Túnez, F.** (2010a). The architecture of FunGramKB. En *Proceedings of the Seventh International Conference on Language Resources and Evaluation, European Language Resources Association (ELRA)*. Valeta, Malta, pp. 2667-2674.
- Periñán-Pascual, C. y Arcas-Túnez, F.** (2010b). Ontological Commitments in FungramKB. *Procesamiento del Lenguaje Natural*, Vol.44: 27-34.
- Periñán-Pascual, C. y Arcas-Túnez, F.** (2011). Introducción a FunGramKB. En *AngloGermanica online: Revista electrónica periódica de filología alemana e inglesa*, nº 8: 1-14.
- Periñán-Pascual, C. y Mairal Usón, R.** (2009). Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural*, vol. 43: 265-273
- Periñán-Pascual, C. y Mairal-Usón, R.** (2010). La gramática de COREL: un lenguaje de representación conceptual. *Onomazein*, Vol. 21:11-45.
- Pustejovsky, J.** (1995). *The Generative Lexicon*. Cambridge, Mass.: MIT Press.
- Ruiz de Mendoza Ibáñez, F. J. y Mairal, R.** (2008). Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica* Vol. 42 (2): 355-400.
- Saint-Dizier, P.** (2005). PrepNet: A framework for describing prepositions: Preliminary investigation results. En *Proceedings of IWCS 6*, Tilburgo, Holanda.
- Saint-Dizier, P.** (2008). Syntactic and semantic frames in PrepNet. En *3rd International Joint Conference on Natural Language Processing*, Hyderabad, India, pp. 763-768.
- Sven, H., Helbig, H. y Osswald, R.** (2006). Semantic interpretation of prepositions for NLP applications. En *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*, Trento, Italy, pp. 29-36.

- Talmy, L.** (1985). Lexicalization patterns: semantic structure in lexical forms. En: Timothy Shopen (ed.): *Language Typology and Syntactic Description, Grammatical Categories and the Lexicon*. Vol. 3. Cambridge: Cambridge University Press, pp. 57-149.
- Talmy, L.** (2000). *Toward a Cognitive Semantics. Typology and Process in Concept Structuring*. Vol. 2. Cambridge: MIT Press.
- Tulving, E.** (1985). How many memory systems are there? *American Psychologist* Vol. 40: 385-398.
- Tyler, A. y Evans, V.** (2003) *The Semantics of English Prepositions. Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: Cambridge University Press.
- Van Valin, R.D. Jr.** (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- Van Valin, R.D. Jr. y LaPolla R. J.** (1997). *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Vendler, Z.** (1967). *Linguistics in philosophy*. Ithaca: Cornell University Press

Anexo I. Interpretación semántica de los papeles temáticos de los argumentos en FumGramKB

Metaconcepto	Papel	Definición
#COGNITION	[Agent]	Entity that makes another entity undergo a cognitive process.
	Theme	Entity that undergoes a cognitive process.
	Referent	Entity present in the consciousness of an entity that undergoes a cognitive process.
#COMMUNICATION	Theme	Entity that transmits a message.
	Referent	Message (i.e. set of propositions) that is transmitted.
	Goal	Entity that receives a message.
#CONSTITUTION	Theme	Entity that is made up of other entities.
	Referent	Entity that is part of another entity.
#CREATION	Theme	Entity that creates another entity.
	Referent	Entity that is created by another entity.
#EMOTION	Agent	Entity that makes another entity feel an emotion.
	Theme	Entity that feels an emotion.
	[Attribute]	Entity or quality that describes an attribute of an entity when feeling an emotion.
#EXISTENCE	Theme	Entity that exists.
#IDENTIFICATION	Theme	Entity that is identified by means of another entity.
	[Referent]	Entity that serves to define the identity of another entity.
	[Attribute]	Quality ascribed to an entity.

Metaconcepto	Papel	Definición
#INTENTION	Theme	Entity that pursues actively a determinate aim.
	Referent	Something which is actively pursued by an entity.
#LOCATION	Theme	Entity that stays in a location.
	Location	Location where an entity stays.
#MATERIAL	Theme	Entity that, volitionally or not, performs an event.
	[Referent]	Entity that is directly involved in the event caused by another entity.
#MOTION	Agent	Entity that makes another entity move.
	Theme	Entity that changes its place or position.
	[Location]	Location in which an entity moves.
	[Origin]	Location from which an entity moves.
	[Goal]	Location to which an entity moves.
#PERCEPTION	Theme	Entity that perceives another entity through any of the senses.
	Referent	Entity that is perceived through any of the senses.
#POSSESSION	Theme	Entity that owns another entity.
	Referent	Entity that is owned.
#TRANSFER	Agent	Entity that transfers another entity to a third entity.
	Theme	Entity that is transferred.
	Origin	Entity from which another entity is transferred.
	Goal	Entity to which another entity is transferred.
#TRANSFORMATION	Theme	Entity that transforms another entity.
	Referent	Entity that is transformed by another entity.

Anexo II. Interpretación semántica de los papeles temáticos de los satélites en FunGramKB

Papel	Definición
Beneficiary	Entity different from those of the arguments that derives benefit from the occurrence of the event.
Company	Entity that participates in a coordinated way with an entity of the arguments, usually Agent or Theme.
Comparison	Quality that is used as the basis of the comparison between two entities, usually Theme and Referent.
Condition	Predication that describes under which condition the event should occur.
Duration	Entity or quality that denotes the length of time from the beginning of the event to its end.
Frequency	Quality that describes how often the event occurs.
Instrument	Entity that is used to perform the event.
Manner	Entity or quality that describes the way in which the event occurs.
Means	Entity that, together with an Instrument, is used to perform the event.
Position	Quality that describes the position of Theme with respect to Location, Goal or Origin.
Purpose	Predication that describes the aim of the event.
Quantity	Entity or quality that describes the amount related to the occurrence of the event.
Reason	Predication that describes the cause of the event.
Result	Predication or entity that describes the consequence of the occurrence of the event.
Scene	Predication or entity that describes the situation in which the event occurs.
Speed	Quality that describes how fast the event is performed.
Time	Entity or quality that describes when the event is performed.

Anexo III. Ejemplos encontrados para los sentidos descritos de la preposición *in*:

+INSIDE_00:

- *They painted **in** the gallery*
- *They sent it **in** an envelope*
- *He found her **in** his bed*
- *My cat lives **in** a card box*
- *Mary saw John **in** the shop*
- *He hid it **in** the security box*

+INTO_00:

- *He came **in** the school*
- *They fell **in** the water*
- *She ran **in** the wood*
- *I threw it **in** the garbage*

+INTO_01:

- *He tore the paper **in** two parts*
- *The glass broke **in** pieces*
- *The book was translated **in** English*

+IN_00:

- *They broke up **in** January*
- *She came **in** April*
- *They met **in** 1996*
- *She first visited Russia **in** 1927*

+WITHIN_00:

- *We'll talk **in** a minute*
- *He'll come back **in** a month*
- *We'll make it up **in** a moment*
- *The judgement will be held **in** fifteen days*

+DURING_00:

- *They won the match **in** the second half*
- *They fell in love **in** the summer*
- *The Beatles got their fame **in** the early 60s*
- *We didn't have computers **in** those days*

+IN-01:

- *Can I have a word with you **in** private?*
- *It was therefore he, who became **in** a sense indispensable*
- *They gazed **in** admiration*
- *They got lost **in** the dark*
- *Sarah was cooking **in** silence*

+IN_02:

- *I sent it **in** Morse*
- *They were built **in** bronze*
- *The land was measured **in** metres*
- *The document was written **in** Word*

+IN_03:

- *Lamps of every size sprouted **in** pairs*
- *They came **in** thousands for the purpose*
- *Eggs are sold **in** a dozen*
- *Square metres would be expressed **in** thousands of bricks*