

**TESIS DOCTORAL**



**2016**

La desambiguación semántica de los sintagmas prepositivos como adjuntos periféricos en el marco de la Gramática del Papel y la Referencia: un enfoque desde la lingüística computacional y la ingeniería del conocimiento

Diana Hernández Pastor

Licenciada en Filología Inglesa  
Máster en Tecnologías de la Información y la  
Comunicación en la Enseñanza y Tratamiento de las  
Lenguas

Departamento de Filologías Extranjeras y sus Lingüísticas  
Facultad de Filología

Director: Dr. D. Carlos Perriñán Pascual

Tutor: Dr. D. Ricardo Mairal Usón



**TESIS DOCTORAL**



**2016**

La desambiguación semántica de los sintagmas prepositivos como adjuntos periféricos en el marco de la Gramática del Papel y la Referencia: un enfoque desde la lingüística computacional y la ingeniería del conocimiento

Diana Hernández Pastor

Licenciada en Filología Inglesa  
Máster en Tecnologías de la Información y la  
Comunicación en la Enseñanza y Tratamiento de las  
Lenguas

Departamento de Filologías Extranjeras y sus Lingüísticas  
Facultad de Filología

Director: Dr. D. Carlos Perriñán Pascual

Tutor: Dr. D. Ricardo Mairal Usón



Departamento de Filologías Extranjeras y sus Lingüísticas

Facultad de Filología

La desambiguación semántica de los sintagmas prepositivos como adjuntos periféricos en el marco de la Gramática del Papel y la Referencia: un enfoque desde la lingüística computacional y la ingeniería del conocimiento

Diana Hernández Pastor

Licenciada en Filología Inglesa  
Máster en Tecnologías de la Información y la Comunicación en la Enseñanza y Tratamiento de las Lenguas

Director: Dr. D. Carlos Periñán Pascual

Tutor: Dr. D. Ricardo Mairal Usón



*A mi madre, por todo.*

*A mis "Tomis", por ser lo mejor de mi vida.*



## **Agradecimientos**

Esta tesis ha sido fruto de muchas colaboraciones de personas a las que quiero y admiro y sin las cuales no habría sido posible su elaboración.

En primer lugar, quiero agradecer a mi director, D. Carlos Perriñán, su ayuda en todo momento; por descubrirme el complejo mundo del PLN y saber guiarme en el confuso camino en el que me he encontrado en varias ocasiones. Por atenderme en cualquier instante en el que me encontraba bloqueada con un guiño de simpatía. Sin su ayuda y su conocimiento, no habría sido posible llegar hasta el final. Muchas gracias por animarme y confiar en mí.

Asimismo, quiero agradecer a D. Ricardo Mairal por aceptar gentilmente ser mi tutor. Es un honor para mí que haya representado el enlace institucional para esta tesis.

Quiero transmitir mi sincera gratitud al Dr. Ken Litkowski y al Dr. Patrick Saint-Dizier por compartir conmigo sus estudios. Gracias, igualmente, a todos los investigadores del grupo FunGramKB, Lexicom y Seconds por acogerme con tanta amabilidad en cada congreso y compartir conmigo su conocimiento. En especial, quiero agradecer a la Dra. Rocío Jiménez Briones y al Dr. Avelino Corral su luz en el camino de la RRG.

Gracias a todos los profesores que he tenido a lo largo de mi carrera. Todos en mayor o menor medida han aportado algo a mi conocimiento. Gracias a D. Manuel, por enseñarme la importancia de los libros y el saber. Gracias por inculcarme un espíritu curioso por querer aprender y seguir pendiente de mí tantos años después.

Inmensa gratitud va dirigida a D. Thomas H. Schmidt, por ser mi gran postor y haberme apoyado desde el primer día. Gracias por confiar en mí y ofrecerme todas las oportunidades que he podido disfrutar. A este agradecimiento quiero sumar un profundo reconocimiento al cariño y apoyo que todos los miembros del departamento de idiomas de la UCAM me han mostrado a lo largo de este camino. Su compañerismo, amistad y comprensión me han ayudado a no perder el ritmo. Gracias, especialmente a Imelda, Jorge, Raquel, M<sup>a</sup> del Mar y Alba por escucharme, animarme y esclarecer los agobios

que me asaltaban casi constantemente. Asimismo, quiero transmitir un intenso agradecimiento a Ana Rosa y José Andrés, a quienes no les ha importado introducirse en el complejo mundo de esta tesis para ayudarme a “construir el armario”; a Sara por ofrecerme sus ojos para ver los pequeños detalles; y, muy en especial, a Gloria y a Jonás por todos sus ánimos y su tiempo con la “segunda tesis”. Gracias a todos de corazón.

Gracias a todos mis amigos, en especial a aquellos que han añorado mi presencia en todas esas ocasiones en las que no he podido estar. Gracias a todos por preocuparos, escucharme y hacerme reír cuando a veces no podía casi respirar. Gracias, en especial, a Cris y Ele, por compartir mis sueños y estar a mi lado siempre que las he necesitado.

Gracias a mis suegros, Tomás y Elvira, por cuidar de mi pequeño Tomi para que pudiera acabar este trabajo. Gracias por recibirme con tanto cariño y tratarme como a una hija.

Un agradecimiento muy especial está dedicado a mi madre. Gracias por saber guiarme y hacerme reflexionar siempre desde una perspectiva crítica. Gracias por invertir en mí todo el sacrificio que sé que ha supuesto. Gracias por hacer de mí lo que soy hoy. Sin sus consejos, amor y confianza no podría haber llegado hasta aquí. Gracias por todo, mamá.

Gracias a mi hijo, Tomi, porque ha representado la ilusión de querer ser algo mejor. Aún no lo sabe, pero es, sin duda, el mejor motivo de mi existencia.

Y por último, pero no menos importante, gracias a mi marido, Tomás. Por su paciencia, cariño, buen humor y todas esas cualidades por las que le elegí. Gracias por apoyarme y comprender mi ausencia en muchos momentos juntos con nuestro pequeño. Y gracias por quererme y demostrármelo cada día.

A todos, muchas gracias.

# ÍNDICE

<b>ÍNDICE</b> .....	<b>i</b>
<b>LISTA DE FIGURAS</b> .....	<b>iii</b>
<b>LISTA DE TABLAS</b> .....	<b>v</b>
<b>LISTA DE ANEXOS</b> .....	<b>vii</b>
<b>CAPÍTULO 1. INTRODUCCIÓN</b> .....	<b>1</b>
<b>1.1 Motivación del estudio</b> .....	<b>1</b>
<b>1.2 Estructura de la tesis</b> .....	<b>9</b>
<b>CAPÍTULO 2. ESTADO DE LA CUESTIÓN</b> .....	<b>11</b>
<b>2.1 Introducción</b> .....	<b>11</b>
<b>2.2 Estado de la cuestión de las preposiciones</b> .....	<b>12</b>
2.2.1 Las preposiciones en gramática .....	12
2.2.2 Las preposiciones en lingüística cognitiva .....	20
2.2.3 Las preposiciones en el procesamiento del lenguaje natural .....	25
<b>CAPÍTULO 3. MARCO DE LA LINGÜÍSTICA TEÓRICA</b> .....	<b>39</b>
<b>3.1 La Gramática del Papel y la Referencia</b> .....	<b>39</b>
3.1.1 La estructura de la cláusula.....	42
3.1.2 La representación semántica en la RRG.....	47
3.1.3 Las preposiciones en la RRG.....	50
<b>CAPÍTULO 4. MARCO DE LA INGENIERÍA DEL CONOCIMIENTO Y EL PROCESAMIENTO DEL LENGUAJE NATURAL</b> .....	<b>55</b>
<b>4.1 FunGramKB</b> .....	<b>55</b>
4.1.1 Introducción.....	55

4.1.2 Paradigmas influyentes en FunGramKB .....	58
4.1.2.1 Teorías cognitivistas .....	58
4.1.2.2 Teorías constructivistas .....	63
4.1.3 La arquitectura de FunGramKB .....	66
4.1.3.1 El Lexicón.....	70
4.1.3.2 La Ontología .....	72
<b>4.2 ARTEMIS .....</b>	<b>79</b>
<b>CAPÍTULO 5. MODELIZACIÓN CONCEPTUAL DE LAS RELACIONES .....</b>	<b>87</b>
<b>5.1 Introducción .....</b>	<b>87</b>
<b>5.2 Metodología del estudio .....</b>	<b>88</b>
<b>5.3 Principios metodológicos para el diseño de la subontología de las relaciones.....</b>	<b>93</b>
<b>5.4 Modelización conceptual de la subontología de las relaciones.....</b>	<b>98</b>
5.4.1 La dimensión cognitiva temporal: #TIME .....	98
5.4.2 La dimensión cognitiva espacial: #LOCUS .....	108
5.4.3 La dimensión cognitiva circunstancial: #CIRCUMSTANCE.....	124
<b>5.5 Entrada conceptual de las relaciones .....</b>	<b>138</b>
<b>CAPÍTULO 6. MODELIZACIÓN LINGÜÍSTICA DE LAS PREPOSICIONES .....</b>	<b>143</b>
<b>6.1 Motivación teórica.....</b>	<b>143</b>
<b>6.2 Análisis de los rasgos lingüísticos para los sentidos de las preposiciones .....</b>	<b>152</b>
<b>CAPÍTULO 7. CONCLUSIONES Y FUTURAS INVESTIGACIONES .....</b>	<b>161</b>
<b>BIBLIOGRAFÍA.....</b>	<b>167</b>
<b>ANEXOS .....</b>	<b>191</b>

## LISTA DE FIGURAS

Figura 1. Ejemplo de entrada para un sentido de la preposición in según PDEP.....	35
Figura 2. Ejemplo del sentido <i>EndState</i> .....	37
Figura 3. Arquitectura general de la RRG.....	41
Figura 4. Componentes de la estructura jerárquica de la cláusula.....	43
Figura 5. Contrastes universales en la estructura jerárquica de la cláusula.....	43
Figura 6. Estructura jerárquica de la cláusula y sus componentes ilustrados.....	44
Figura 7. Interrelación sintáctico-semántica de los elementos que componen la estructura jerárquica de la cláusula.....	44
Figura 8. Estructura jerárquica de la cláusula con LDP. ....	45
Figura 9. Operadores de la estructura jerárquica de la cláusula. ....	46
Figura 10. Representación formal de los operadores en una oración en inglés.....	47
Figura 11. Estructura del sintagma preposicional de una preposición predicativa en la periferia de la RRG.....	52
Figura 12. Operadores en la estructura jerárquica de unsintagma nominal.....	52
Figura 13. Proceso de comprensión del lenguaje en FunGramKB. ....	59
Figura 14. Tipología de los esquemas conceptuales en FunGramKB.....	60
Figura 15. La interrelación de los tres niveles conceptuales de la Ontología según el tipo de conocimiento.....	62
Figura 16. Arquitectura del MLC. ....	64
Figura 17. Arquitectura modular de FunGramKB.....	67
Figura 18. El papel de la ELC como interlingua. ....	72
Figura 19. Visión parcial de la ontología de los eventos en FunGramKB.....	74
Figura 20. Interfaz de edición de ARTEMIS. ....	80
Figura 21. Estructura estratificada de la cláusula en la RRG.....	84
Figura 22. Estructura estratificada de la cláusula en ARTEMIS.....	84
Figura 23. Análisis sintáctico-semántico de <i>The juice froze black in the refrigerator.</i> .	85
Figura 24. Taxonomía jerarquizada del metaconcepto #DURATION_F1.....	101
Figura 25. Taxonomía jerarquizada del metaconcepto #DURATION_F2.....	102
Figura 26. Taxonomía jerarquizada del metaconcepto #INTERVAL_F1. ....	104
Figura 27. Taxonomía jerarquizada del metaconcepto #INTERVAL_F2. ....	106
Figura 28. Taxonomía de relaciones #TIME.....	107

Figura 29. Taxonomía metaconceptual de la dimensión cognitiva #PLACE.....	113
Figura 30. Distribución metaconceptual de la dimensión cognitiva #PATH.....	118
Figura 31. Distribución metaconceptual de la taxonomía #LOCUS .....	121
Figura 32. Distribución de los niveles conceptuales en la dimensión cognitiva #PLACE. .....	122
Figura 33. Distribución de los niveles conceptuales en la dimensión cognitiva #PATH .....	123
Figura 34. Lista de relaciones de Srikumar y Roth. ....	125
Figura 35. Distribución de los niveles metaconceptuales de la taxonomía #CIRCUMSTANCE.....	127
Figura 36. Estructura del nivel metaconceptual #MANNER.....	129
Figura 37. Distribución de los niveles conceptuales para la taxonomía #CIRCUMSTANCE.....	137
Figura 38. Revisión hecha por Piñón de la división entre trayectorias delimitadas y no delimitadas de Jackendoff. ....	149

## LISTA DE TABLAS

Tabla 1. Estructuras lógicas en la RRG para los diferentes tipos de <i>Aktionsart</i> . ....	49
Tabla 2. Rasgos para la caracterización de las distintas clases de <i>Aktionsart</i> . ....	49
Tabla 3. Las diez preposiciones más frecuentes en inglés hablado y escrito según el BNC. ....	88
Tabla 4. Ejemplo del número de los sentidos para las ocho preposiciones seleccionadas proporcionados por los cuatro diccionarios consultados. ....	89
Tabla 5. Número final de sentidos de las ocho preposiciones. ....	91
Tabla 6. Lista de elementos primitivos de PV para algunas preposiciones inglesas según Zhang. ....	111
Tabla 7. Entradas conceptuales para las relaciones creadas en la taxonomía #TIME..	139
Tabla 8. Entradas conceptuales para las relaciones creadas en la taxonomía #LOCUS .....	139
Tabla 9. Entradas conceptuales para las relaciones creadas en la taxonomía #CIRCUMSTANCE.....	140
Tabla 10. Rasgos aspectuales en las capas de la cláusula según la RRG. ....	145



## LISTA DE ANEXOS

Anexo I. La interpretación semántica de los papeles temáticos de los argumentos en FunGramKB. ....	193
Anexo II. La interpretación semántica de los papeles temáticos de los satélites en FunGramKB. ....	197
Anexo III. Sentidos ofrecidos por PDEP para la preposición <i>in</i> . ....	199
Anexo IV. Sentidos ofrecidos por <i>Supersenses</i> para la preposición <i>in</i> . ....	201
Anexo V. Dimensión cognitiva #TIME ilustrada con unidades léxicas.....	203
Anexo VI. Dimensión cognitiva #PLACE ilustrada con unidades léxicas .....	205
Anexo VII. Dimensión cognitiva #PATH ilustrada con unidades léxicas .....	207
Anexo VIII. Dimensión cognitiva #CIRCUMSTANCE ilustrada con unidades léxicas .....	209
Anexo IX. Relación de preposiciones con sus conceptos básicos.....	213



*Preposition is the 1st word in:*  
*Genesis (“In the beginning”)*  
*the Quran (“In the name of the merciful and compassionate God”),*  
*Paradise Lost (“Of Mans First Disobedience”),*  
*Oliver Twist (“Among other public buildings in a certain town”),*  
*Don Quixote (“En un lugar de la Mancha”/“In some village in la*  
*Mancha”), and The Hobbit (“In a hole in the ground there lived a hobbit”)*  
Schneider (2014, p. 78)



# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1 Motivación del estudio

La cuestión de la ambigüedad siempre ha representado una tarea ardua en la lingüística, la ciencia cognitiva y el procesamiento del lenguaje natural (PLN).<sup>1</sup>

---

<sup>1</sup> Es importante destacar que el tratamiento computacional del lenguaje ha originado numerosas etiquetas para referirse a esa disciplina, como lingüística computacional, procesamiento del lenguaje natural, tecnologías lingüísticas, ingeniería lingüística, industrias de la lengua o lingüística informática. Dependiendo del aspecto que se desee enfatizar, los investigadores prefieren una etiqueta sobre las demás. Por ejemplo:

[...] mientras la lingüística computacional se centra más en la modelización del conocimiento lingüístico para posibilitar la construcción de sistemas computacionales que analicen y/o generen textos en lenguaje natural, el PLN hace un mayor énfasis en la búsqueda de soluciones a los problemas que plantea la lingüística computacional, pero en el marco de aplicaciones concretas: p.ej. recuperación y extracción de información, resúmenes automáticos, traducción mecánica, etc. (Periñán-Pascual, 2012b, p. 16)

En este estudio, no obstante, utilizo los términos lingüística computacional y PLN de manera indistinta, refiriéndome en cualquier caso al procesamiento informatizado de cualquier tarea que implique el análisis o la generación de la lengua.

Tradicionalmente, la ambigüedad ha sido analizada o bien como ambigüedad léxica donde una palabra puede tener más de una posible interpretación debido a su carácter polisémico, o como ambigüedad estructural (o también conocida como ambigüedad sintáctica), donde una cláusula puede tener más de una posible interpretación debido a su estructura gramatical. En ambos casos, el problema se traduce en una ambigüedad semántica, ya que inevitablemente se produce un error en la interpretación. Con el fin de solventar los diferentes tipos de ambigüedades, se aplican los análisis morfológico, sintáctico y semántico sobre la realización lingüística.

Por otra parte, en el marco del PLN, las preposiciones han provocado un obstáculo notorio que sigue siendo objeto de investigación. Debido a su alta frecuencia en el lenguaje y su carácter relacional con otras categorías léxicas, como los verbos, nombres o adjetivos, constituyen una clase gramatical manifiestamente difícil de procesar. Sorprendentemente, como comentaron Hartrumpf, Helbig y Osswald (2006), las investigaciones lingüísticas sobre la semántica de las preposiciones no han tenido mucho impacto en las aplicaciones del PLN. De hecho, siempre ha sido una tendencia generalizada que los ingenieros informáticos no recurran a las investigaciones en lingüística teórica a la hora de diseñar un sistema del PLN, debido en la mayoría de las ocasiones al hermetismo, la poca practicidad y la falta de aplicabilidad inmediata de las propias investigaciones lingüísticas, tal y como señalaron Wintner (2009) y Periñán (2012b). No obstante, EAGLES (1999) reconoció el valor de la información proporcionada por las preposiciones como uno de los temas más discutidos en torno a la estructura argumental del marco predicativo en las entradas léxicas de los lexicones computacionales. Además, según indicó Saint-Dizier (2008), se precisa tener un repositorio que organice y caracterice los comportamientos sintácticos y semánticos de las preposiciones, lo cual resulta en una tarea de vital importancia para el avance científico en PLN, donde Baldwin, Kordoni y Villavicencio (2009) destacaron algunos de los campos en que los investigadores han mostrado un mayor interés por esta categoría gramatical:

- (i) Las aplicaciones que incorporan la generación del lenguaje natural.
- (ii) La extracción de información, especialmente en el papel que desempeña la preposición al conectar elementos del texto.

- (iii) Las aplicaciones que requieren la representación explícita de un espacio 3D, p.ej. la robótica, los agentes animados y la realidad virtual, donde se interpreta la información espacial de las preposiciones.
- (iv) La traducción automática basada en el enfoque interlingüístico.
- (v) La corrección automática de errores en el uso de preposiciones en textos escritos por hablantes no nativos.

Paradójicamente, las preposiciones han sido generalmente ignoradas en las aplicaciones del PLN, limitándose a tratar a un subconjunto específico de preposiciones en una lengua determinada, principalmente el inglés. Pero, ¿qué problema plantea la preposición a la lingüística computacional? A modo de ilustración, una aplicación del PLN encontraría dificultades para solucionar la ambigüedad expresada por el sentido de la preposición *with* en las siguientes oraciones:

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

El inconveniente surge con la función sintáctica del sintagma preposicional y a su vez con la interpretación semántica del mismo. La adjunción del sintagma preposicional (*PP attachment*) es el área de la sintaxis que intenta dar una explicación a este tipo de ambigüedad sintáctica, teniendo como principal objetivo el poder determinar cuál es el término regente de la preposición. De este modo, para las oraciones en (1) el objetivo sería determinar qué término es el que rige a la preposición (i.e. si el verbo *ate* o el nombre *yoghurt*). En (1a) el sintagma preposicional depende del verbo, ya que *with* introduce a *spoon* como instrumento para la acción verbal. Sin embargo, en (1b) la preposición encabeza el sintagma *with cereals*, el cual proporciona una característica a la entidad referenciada por el nombre *yoghurt*, funcionando así como complemento adjetival. En este sentido, el principal problema al que se debe hacer frente es la desambiguación de una categoría léxica que resulta ser extremadamente polisémica, donde la interpretación semántica de la preposición está íntimamente relacionada con el contexto.

No obstante, este tipo de problema no siempre puede ser solventado fácilmente tras examinar únicamente el contexto inmediato, es decir, cuando el analizador se apoya en comprobaciones locales para conseguir la información que necesita para la desambiguación. La resolución de ambigüedades sintácticas a menudo requiere un contexto más amplio que la simple cláusula, tal y como se ejemplifica en la ambigüedad ocasionada por la preposición que se encuentra en las siguientes oraciones (Lytinen, 1987, p. 169):

- 2) a. *The cleaners dry-cleaned the coat that Mary found at the rummage sale for \$10.*
- b. *The cleaners dry-cleaned the coat that Mary found in the garbage for \$10.*

Sólo con el procesamiento sintáctico, estas dos oraciones seguirían siendo ambiguas para la máquina. En ambas oraciones, el sintagma preposicional *for \$10* puede vincularse a cualquiera de los dos verbos (i.e. *dry-cleaned* o *found*). La estructura sintáctica de las oraciones en (2) no proporciona pista alguna para la adjunción del sintagma preposicional, ni tampoco las preferencias de selección son de gran ayuda, ya que en ambas oraciones aparecen los mismos verbos.

En estos casos, los analizadores suelen aplicar el análisis sintáctico en primer lugar, dejando algunas ambigüedades sintácticas sin resolver hasta la etapa del análisis semántico. No obstante, si las ambigüedades sintácticas no se resuelven inmediatamente, hay un coste computacional que pagar, debido a que la ambigüedad irresoluta puede afectar al resto del análisis sintáctico, resultando en ocasiones en una explosión combinatoria en el número de ambigüedades sintácticas que tendrán que ser tratadas posteriormente durante el proceso de análisis semántico. En definitiva, los procesamientos sintáctico y semántico deben estar integrados en el analizador con el fin de construir una representación semántica del texto. Ambos tipos de procesamiento deben ocurrir al mismo tiempo, apoyándose uno al otro para proporcionar la información necesaria para resolver las ambigüedades sintácticas y semánticas de forma paralela. Por este motivo, el propósito de esta tesis es el desarrollo de un modelo teórico que permita la descripción léxico-conceptual los diversos sentidos de las preposiciones,

el cual permita la creación de una base de conocimiento que almacene la información sintáctico-semántica que se requiere para solventar el problema de la ambigüedad de los sintagmas preposicionales durante el proceso de comprensión textual realizado por un sistema del PLN. Aunque el modelo que propongo en esta tesis tiene una aplicabilidad multilingüe, me centré en treinta y tres preposiciones del inglés que permitieran ilustrar la configuración ontológica de los conceptos correspondientes a los sentidos expresados por las preposiciones en las dimensiones cognitivas de tiempo, lugar y circunstancia, además de en ocho preposiciones (i.e. *at, by, for, in, of, on, to* y *with*) elegidas de entre las quince preposiciones más frecuentes en la lengua inglesa según Leech, Rayson & Wilson (2001), con el fin de capturar su comportamiento lingüístico dentro de la oración. Con el fin de facilitar la integración de los procesamientos sintáctico y semántico, el modelo de descripción preposicional presentado en este estudio se origina a partir del paradigma funcionalista de la teoría lingüística de la Gramática del Papel y la Referencia y a partir de la base de conocimiento léxico-conceptual FunGramKB, la cual integra a su vez el modelo lingüístico anterior con los esquemas cognitivos idealizados propios de la teoría de Lakoff (1987) y con un enfoque simbólico propio de la ingeniería del conocimiento y la inteligencia artificial. Por una parte, la Gramática del Papel y la Referencia (*Role and Reference Grammar*, en adelante RRG correspondiente a sus siglas en inglés) (Van Valin 1993, 2005, 2008; Van Valin & LaPolla, 1997) es actualmente una de las teorías funcionales más relevantes del panorama lingüístico. En contraposición a los enfoques tradicionales formalistas, las teorías funcionales centran su atención en las funciones del lenguaje, el significado y el uso real del hablante. La RRG se considera de estilo moderado (Nichols, 1984), pues reconoce que la validez de la estructura gramatical viene determinada por las funciones semánticas y comunicativas. Por tanto, 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. Más concretamente, la RRG presenta un modelo gramatical de carácter monoestrático, donde los componentes sintáctico y semántico se vinculan a través de un algoritmo de enlace bidireccional. Además, la RRG intenta desarrollar un aparato lingüístico que permita la descripción estructural de todas las lenguas, tratando así de encontrar la tipicidad lingüística universal. Como apuntó Perinián-Pascual

(2012b), la RRG posee determinadas características que hacen que esta teoría gramatical resulte bastante atractiva para un modelo computacional del lenguaje. En primer lugar, un enfoque funcional del lenguaje permite capturar las generalizaciones sintáctico-semánticas que sirven para explicar la motivación semántica de los fenómenos gramaticales. En segundo lugar, la adecuación psicológica es relevante para los modelos cuya finalidad es la comprensión del lenguaje. Finalmente, la adecuación tipológica es un requisito imprescindible para los sistemas multilingües.

Dentro de la RRG, esta tesis se centra en los sintagmas preposicionales que actúan como adjuntos periféricos en la cláusula, los cuales, como se explicará más adelante, son aquellos que contribuyen semánticamente a toda la oración y no están regidos directamente por ningún elemento concreto. De este modo, mi estudio pretende dar respuesta al comportamiento sintáctico de esos sintagmas en el procesamiento de la cláusula.

Por otro lado, FunGramKB (Periñán-Pascual & Arcas-Túnez, 2004, 2007, 2008, 2010a, 2010b, 2011; Mairal-Usón & Periñán-Pascual, 2009; Periñán-Pascual & Mairal-Usón, 2010b) es una base de conocimiento léxico-conceptual multipropósito diseñada principalmente para su aplicación en sistemas del PLN que requieran la comprensión del lenguaje. Esta base de conocimiento es multipropósito en el sentido de que es tanto multifuncional como multilingüe. En otras palabras, FunGramKB ha sido diseñada con el fin de ser potencialmente reutilizada en diversas tareas del PLN y con diversas lenguas, siendo los lexicones del inglés y del español los más desarrollados actualmente. Además, la arquitectura de FunGramKB comprende tres niveles principales de conocimiento, cada uno de los cuales está constituido por diversos módulos independientes aunque claramente interrelacionados: (i) nivel léxico, formado por los módulos de Lexicón y Morficón, (ii) nivel gramatical, configurado por el Gramaticón, y (iii) nivel conceptual, formado por los módulos de Ontología, Cognición y Onomasticón. En esta base de conocimiento, mientras cada lengua tiene sus propios módulos léxico y gramatical, cada módulo conceptual es compartido por todas las lenguas. La mayoría de teorías lingüísticas contemporáneas siguen mostrando un enfoque lexicista, donde las unidades léxicas no son consideradas como átomos de la gramática sino como objetos complejos generalmente representados por estructuras de

rasgos (Gibbon, 2000). En cambio, FunGramKB adopta un enfoque conceptualista, donde la Ontología se convierte en el pivote de toda la arquitectura de la base de conocimiento, gracias al cual, además de compartir las ventajas del modelo lexicista, se proporciona un tratamiento más adecuado al multilingüismo dentro del marco del PLN.<sup>2</sup> Una de las consecuencias de este enfoque conceptualista en FunGramKB se encuentra en el hecho de que la modelación de los conceptos por parte de los ingenieros del conocimiento es siempre previa al desarrollo de las entradas léxicas por parte de los lingüistas. En otras palabras, sólo una vez que se crea una unidad conceptual (p.ej. +GIVE\_00) en la Ontología junto con su correspondiente entrada conceptual, la cual contiene la representación del conocimiento semántico, los lingüistas pueden crear las entradas de todas las unidades léxicas que estén vinculadas al concepto recién creado, las cuales almacenarán la información morfosintáctica relevante en cada lengua. Esta dinámica de desarrollo de la base conocimiento es la que se ha realizado con las entidades, los eventos y las cualidades, los cuales corresponden lingüísticamente a los nombres, los verbos y los adjetivos respectivamente, y fue precisamente el mismo procedimiento que se aplicó en el caso de las relaciones, las cuales se instancian en preposiciones según el modelo propuesto

Dentro de FunGramKB, esta tesis se centra en la construcción de una nueva subontología de relaciones y en el modelo para la creación de las entradas léxicas de las preposiciones que serán almacenadas en el Lexicón. Debo señalar que, aunque las preposiciones trabajadas en esta tesis pertenecen al inglés, tanto el modelado cognitivo de las relaciones que se describen como el modelo de integración de la información sintáctico-semántica pueden ser reutilizados en otras lenguas, especialmente indoeuropeas.

En definitiva, el presente estudio se aproxima al campo de la desambiguación preposicional desde el prisma de la lingüística computacional y la ingeniería del conocimiento, creando así un modelo que permita almacenar de forma integrada la información sintáctico-semántica necesaria para una interpretación adecuada de los

---

<sup>2</sup>Léanse Montiel-Ponsada, Aguado-de-Cea, Gómez-Pérez y Peters (2008, 2011) y Montiel-Ponsada (2011) para una descripción pormenorizada del tratamiento del multilingüismo en los proyectos de ingeniería del conocimiento.

sintagmas preposicionales predicativos que funcionen como adjuntos periféricos en el marco lingüístico de la RRG. Por tanto, las investigaciones resultantes de esta tesis contribuyen a la implementación de aplicaciones más robustas de PLN, p.ej. en áreas como la extracción automática de conocimiento o las interfaces de diálogo persona-máquina, y en particular al desarrollo de ARTEMIS (Periñán-Pascual, 2013b; Periñán-Pascual & Arcas-Túnez, 2014; Díaz-Galán & Fumero-Pérez, 2016), un sistema basado en FunGramKB que implementa computacionalmente el modelo teórico de la RRG con el fin de extraer la representación semántica de una oración en términos de su estructura lógica conceptual. En este sentido, y siguiendo la estela del equipo de investigación de FunGramKB, considero que los desafíos que afrontemos en sintaxis, morfología, fonología, semántica o pragmática nos permitirán avanzar significativamente en el diseño de la próxima generación de sistemas de PLN, tal y como también concluyen otros investigadores:

As it has become increasingly pervasive in our lives, HLT [Human Language Technology] has become an increasingly central component of computer science too. Central to future research and developments in this area is the convergence of knowledge of computer science and linguistics and other related disciplines. Crucially, the key ingredient is an understanding of linguistics and how languages work to meet the levels of adequacy needed to be successful and accepted and to guide computer science in application development and in the engineering of solutions of value to us.

(Nolan, 2016, p. 247)

In the rapidly growing field of NLP, where natural language data are processed, there is a growing demand for computer systems to learn and adapt to the user's needs, and to interact with the user in a very natural manner. This is a great opportunity for linguists and an inspiring challenge for linguistic theories. As NLP systems are no longer just statistic and rule-based, the broad range of linguistic topics and modern theories can be usefully applied to NLP software and the language models they require.

(Diedrichsen, 2016, p. 249)

## 1.2 Estructura de la tesis

El resto de apartados de esta tesis se estructura en dos partes, las cuales corresponden, por un lado, al análisis del estado de la cuestión y la descripción del marco de actuación y, por otro, al estudio realizado para la modelización conceptual de las relaciones y la modelización lingüística de las preposiciones.

La primera parte está formada por los capítulos 2, 3 y 4. En el capítulo 2, se lleva a cabo un análisis del estado de la cuestión en lo que respecta al tratamiento recibido por las preposiciones en diferentes áreas lingüísticas, como la gramática, la lingüística cognitiva y la lingüística computacional. En el capítulo 3, se presenta la RRG como la teoría lingüística que sirve de marco de actuación para la descripción de la interfaz sintáctico-semántica entre los niveles léxico y conceptual de FunGramKB. Precisamente, en el capítulo 4, se profundiza en la descripción de FunGramKB, especialmente en los módulos del Lexicón y la Ontología. Asimismo, se ha dedicado un apartado para la presentación de ARTEMIS con el fin de dar una mejor explicación de los fenómenos sintácticos en los que contribuiría mi aportación investigadora en el marco de la lingüística computacional.

La segunda parte está constituida por los capítulos 5 y 6. En el capítulo 5, muestro la descripción conceptual de las relaciones y la modelización de tres taxonomías correspondientes a las dimensiones cognitivas de tiempo, lugar y circunstancia que serán incluidas en la Ontología de FunGramKB. En el capítulo 6,

presento el estudio de la información sintáctica de los adjuntos periféricos en la RRG y proporciono dicha información en un formato legible por la máquina de modo que pueda ser tratada computacionalmente desde el Lexicón.

Finalmente, en el capítulo 7 presento las conclusiones de esta investigación, donde se plasman igualmente las futuras investigaciones que se pueden derivar de la misma.

## CAPÍTULO 2

### ESTADO DE LA CUESTIÓN

#### 2.1 Introducción

A lo largo de la tradición lingüística, las preposiciones han sido consideradas como una clase gramatical carente de significado (Baldwin, 2006a), de manera que no se les ha prestado suficiente atención en su papel como elemento relacional de la cláusula. En este capítulo, presento una descripción de algunos de los principales estudios que tratan la preposición desde diferentes perspectivas:

- desde la gramática, atendiendo más concretamente a las descripciones realizadas por Quirk, Greenbaum, Leech y Svartvik (1985) y Huddleston y Pullum (2002), Downing y Locke (2006), Ávila (2007) y Girju (2009). También se mostrarán determinados comportamientos sintácticos según Baldwin, Kordoni y Villavicencio (2009) y Suppes (1979, 2005);
- desde la lingüística cognitiva, presentando los enfoques de investigadores como Bennett (1975), Lakoff y Johnson (1980), Brugman (1981), Herskovits (1986, 1997), Lakoff (1987), Brugman y Lakoff (1988), Zelinski-Wibbelt (1993), Boroditsky (2000), Talmy (1985, 1991, 2000), Pratt y Francez (2001), Brée y Pratt-Hartmann (2002), Von Stechow (2002), Tyler y Evans (2003), Svenonius (2008) y Lindstromberg (2010); y

- desde la lingüística computacional, centrándome en los estudios más relevantes acerca del tratamiento de esta categoría gramatical en el ámbito del PLN (Hindle & Rooth, 1993; Brill & Resnik, 1994; Harabagiu, 1996; Dorr, 1997, 2001; Baker, Fillmore & Lowe, 1998; Fillmore, Wooters & Baker, 2001; Agirre, Lersundi & Martínez, 2002; Lersundi & Agirre, 2003; Mokhov & Rudzic, 2003; O’Hara & Wiebe, 2003, 2009; Alam, 2004; Boonthum, Toida & Levinstein, 2005; Litkowski & Hargraves, 2005,2007; Saint-Dizier, 2005, 2008; Hartrumpf, et al., 2006; Ye & Baldwin, 2007; Yuret, 2007; Popescu, Tonelli & Pianta, 2007; Tratz & Hovy, 2009, 2011; Hovy, Tratz & Hovy, 2010; Hovy, Vaswani, Tratz, Chiang, & Hovy, 2011; Srikumar & Roth, 2011; 2013a; 2013b; Litkowski, 2013a, 2013b, 2014; Schneider, Srikumar, Hwang & Palmer, 2015).

## 2.2 Estado de la cuestión de las preposiciones

### 2.2.1 Las preposiciones en gramática

Quirk et al. (1985) definieron las preposiciones como “(...) a relation between two entities, one being that represented by the prepositional complement, the other by another part of the sentence” (1985, p.657). Según estos autores, una de las características de esta clase léxica es que su complemento suele ser un sintagma nominal, una cláusula interrogativa indirecta o una cláusula de gerundio:

- 3) a. *He was surprised at her attitude.*
- b. *He was surprised at what he saw.*
- c. *He was surprised at meeting you.*

Así, en la lengua inglesa, las preposiciones son aquellas palabras que no admiten (a) una cláusula introducida por *that*<sup>3</sup>, (b) una cláusula de infinitivo, o (c) un pronombre en caso de sujeto como complemento de la preposición (Quirk et al. 1985, p. 659):

- 4) a. *He was surprised \*at that she noticed him.*  
b. *He was surprised \*at to see her.*  
c. *He was surprised \*at she.*

De este modo, cuando la preposición esté regida por un verbo o un adjetivo, la preposición puede ser omitida cuando vaya seguida de una cláusula introducida por *that* o una cláusula de infinitivo:

- 5) a. *He was surprised that she noticed him.*  
b. *He was surprised to see her.*

Asimismo, estos autores distinguieron tres tipos de sintagmas preposicionales acorde a su función sintáctica, clasificándolos como postmodificadores (p.ej. *the people on the bus*), adverbiales (p.ej. *the people were singing on the bus*) y complementos (p.ej. *we were looking at his awful paintings*) (Quirk et al., 1985, p.657). Cabe mencionar que, de entre las funciones sintácticas correspondientes a los sintagmas preposicionales adverbiales, debo destacar aquellos sintagmas que funcionan como adjuntos de la oración, ya que son el motivo investigador de este estudio.

Según estos autores, los adjuntos están divididos en dos tipos. Por un lado, existen los adjuntos predicativos (*predication adjunct*), que afectan sólo al predicado de la oración y suelen tener una posición fija al final de la oración<sup>4</sup>; y por otro lado, están los denominados adjuntos oracionales o no predicativos (*sentence adjunct*), que aportan carga semántica a toda la cláusula.

---

<sup>3</sup>Las oraciones introducidas por *that* pueden aparecer como complementos preposicionales siempre y cuando vayan precedidas del nombre *fact*, como en *They convinced him of the fact that they needed more troops* (Quirk et al., 1985, p. 659).

<sup>4</sup> Si el complemento del verbo es demasiado largo, el adjunto puede aparecer justo detrás del verbo en posición intermedia entre el verbo y su complemento.

Los adjuntos predicativos se clasifican en obligatorios u opcionales, estando los primeros regidos para el completo sentido del verbo (1985, p. 505):

- 6) a. *He lived.\**  
b. *He lived in Chicago.*

Sin embargo, los adjuntos predicativos opcionales son aquellos que no alteran la relación entre el verbo y su complemento, pero que enriquecen la semántica del predicado:

- 7) *The rabbit disappeared behind the bush.*

(Quirk et al., 1985, p. 510)

Por su lado, los adjuntos no predicativos disfrutan de la libertad de aparecer o no en la oración tanto en posición inicial como final, mostrando así su contribución semántica a la cláusula como un todo, y no solamente al verbo y sus complementos. Por ejemplo:

- 8) *She kissed her mother on the platform.*

Por consiguiente, la diferencia que estos autores establecieron entre los adjuntos predicativos opcionales y los adjuntos no predicativos radica en la posición que puede ocupar el adjunto en relación al núcleo de la cláusula. Defienden que cuanto más periférico es el adjunto, más coherente es considerarlo como adjunto no predicativo. A modo de ilustración, compárense las siguientes oraciones mostradas en Quirk et al. (1985, p. 511):

- 9) a. *She kissed her mother on the platform.*  
b. *She kissed her mother on the cheek.*

En el ejemplo (9b) la posición del adjunto *on the cheek* es relativamente más cercana al núcleo que en el ejemplo (9a) *on the platform*. Este hecho es sencillo de observar si se cambia la posición del adjunto, tal y como se indica en Quirk (1985, p.512):

10) a. *On the platform, she kissed her mother.*

b. *On the cheek, she kissed her mother.\**

Así, el ejemplo (10b) resulta incorrecto al mover el adjunto a posición inicial. Además, los dos ejemplos pueden fusionarse como en (11).

11) *She kissed her mother on the cheek on the platform.*

En este caso, es fácil apreciar cómo la oración cobra más coherencia cuando el adjunto *on the cheek* aparece en una posición más cercana al verbo y sus complementos, mientras que *on the platform* aparece un poco más alejado del núcleo de la oración. En consonancia con este hecho, existe la razón por la que un adjunto no predicativo puede aparecer separado del núcleo de la oración por una coma:

12) *On the platform, she kissed her mother on the cheek.*

Este ejemplo evidencia la capacidad que tienen los adjuntos no predicativos de contribuir semánticamente a la oración completa y no sólo al predicado verbal. Esta concepción de los sintagmas preposicionales como adjuntos no predicativos y, por lo tanto, periféricos conlleva un análisis más profundo de su autonomía como elementos de la cláusula que se trata en esta tesis.

Desde un punto de vista semántico, Dixon (1991) identificó tres casos en el uso de las preposiciones en inglés:

- (i) La preposición introduce un sintagma nominal periférico, proporcionando especificación locativa o temporal (p.ej. *in the house, at three o'clock*) o marcando un instrumento (p.ej. *with a stone*), un beneficiario (p.ej. *for Mary*), a recipient (p.ej. *to John*), etc.
- (ii) La preposición aparece como un elemento inherente al verbo. Por tanto, este verbo no suele aparecer sin la preposición, p.ej. *refer to, rely on*, etc.
- (iii) La preposición aparece dentro de una locución idiomática (p.ej. *phrasal verb*).

Según este autor, existen dos formas de tratar los casos de tipo (ii): o bien se trata de un lexema formado por dos palabras que se comporta como un verbo transitivo, como ocurre en el análisis (13), o bien se trata de un verbo intransitivo, el cual debe ir obligatoriamente seguido por un sintagma preposicional, como ocurre en (14).

13) *She referred to my book.*

V      SN

14) *She referred to my book.*

V      SP

La primera alternativa presenta la ventaja de concebir al sintagma nominal que sigue al verbo como un objeto transitivo que puede convertirse a un sujeto pasivo, como se muestra en (15).

15) *Jespersen was referred to by everyone attending the symposium.*

La segunda alternativa presenta además el inconveniente de que los verbos intransitivos pueden ir seguidos de diversos tipos de preposiciones (p.ej. *travel to/toward/around/in Japan*), mientras que los supuestos verbos intransitivos del tipo (ii) están confinados a una sola preposición. Por consiguiente, la mejor forma de tratar a un verbo con una preposición inherente es como un verbo transitivo formado por dos palabras. Según Huddleston y Pullum (2002), existen muchas palabras que pueden ser clasificadas como preposiciones de acuerdo con tres criterios: (a) la posibilidad de admitir como complementos otros elementos que no necesariamente tienen que pertenecer a la categoría nominal (p.ej. *I didn't know about it until recently*), (b) su función como adjunto no predicativo o como complemento de una oración (p.ej. *She had slept in the attic* o *She put the key in her bag*) y (c) la admisión de modificadores (p.ej. *She died two years after their divorce*). Sin embargo, estos mismos autores reconocieron que los sintagmas preposicionales prototípicos suelen estar compuestos por una preposición y un sintagma nominal, proponiendo además la siguiente definición para las preposiciones: “a relatively closed grammatically distinct class of words whose most central members characteristically express spatial relations or serve to mark various syntactic functions and semantic roles” (Huddleston & Pullum 2002, p. 603).

Otro aspecto que estos autores estudiaron fue el hecho de que las preposiciones disponen de capacidad para seleccionar sus propios complementos de manera similar a la que lo hacen los verbos, los adverbios, los nombres y los adjetivos y que, por este motivo, la perspectiva de considerar a las preposiciones como núcleos es razonable. De hecho, desde un punto de vista semántico, la relación entre la preposición y su complemento configura el sentido de la preposición. Como se verá más adelante, esta información ha sido relevante para la creación de uno de los criterios elaborados para la descripción conceptual de los sentidos de las preposiciones que se presenta en el capítulo 5.

Por su lado, Downing y Locke (2006) defendieron que, en la relación existente dentro de un sintagma preposicional entre la preposición y su complemento, ambos elementos disfrutan del mismo estatus gramatical. En otras palabras, afirmaron que no es posible la construcción de un sintagma preposicional si falta alguno de esos dos elementos. Por este motivo, estos autores consideraban a la preposición como un elemento que puede desempeñar diversas funciones sintácticas dentro de la cláusula dependiendo de su semántica. Así, mantuvieron que existe una distinción básica de la semántica de las preposiciones entre aquellas en las que el uso de la preposición viene determinado por el verbo, el nombre o el adjetivo que la precede, denominadas preposiciones gramaticalizadas, y aquellas que son elegidas por el emisor de manera voluntaria en función de sus intenciones, llamadas preposiciones lexicalizadas. De este modo, la mayoría de preposiciones disponen de usos gramaticalizados, tal y como se ilustra en (16) y (17) con el uso de *in* y *with* respectivamente (Downing & Locke, 2006, p. 547):

16) *I have confidence in you.*

17) *I agree with you.*

Cuando una preposición es seleccionada por otra categoría gramatical (p. ej., un verbo o un adjetivo), se establece un uso gramaticalizado de la misma que conlleva que su significado esté influido por el tipo semántico de su término regente y, en consecuencia, provoque una pérdida semántica de la preposición con respecto a sus significados más habituales. En estos casos, las preposiciones no admiten modificadores

y sus complementos suelen verse restringidos a complementos nominales. La mayoría de las investigaciones que han abarcado el estudio de la semántica de las preposiciones han girado en torno a este uso gramaticalizado de la preposición (Baldwin, 2006b) y, en especial, a las construcciones con verbos de movimiento. A este respecto, en la sección 2.2.2, se prestará especial atención al estudio realizado por Talmy (1985,1991, 2000).

En el caso de las preposiciones lexicalizadas, es la combinación entre la preposición, su término regente y el complemento de la preposición la que determina el tipo semántico del sintagma preposicional y su contribución a la cláusula (Ávila, 2007; Girju, 2009). En la mayoría de ocasiones, los adjuntos no predicativos están formados por preposiciones de uso lexicalizado.

Desde un punto de vista semántico, Downing y Locke (2006) comentaron que los sentidos más frecuentes de las preposiciones son los que se refieren al tiempo o al espacio. De hecho, es habitual encontrar usos metafóricos derivados de esos sentidos locativos y temporales, como sucede en las expresiones *in love* o *in anger*, donde las emociones son conceptualizadas como contenedores, por lo que el sentido de la preposición es el mismo que el utilizado en el sintagma preposicional *in Spain*. Esta afirmación está en consonancia con la atención prestada a las dimensiones de tiempo y espacio sobre las que la mayoría de estudios de corte cognitivo han mostrado interés, como se muestra en la sección 2.2.2.

Con respecto a la sintaxis, paradójicamente 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. Según Baldwin et al. (2009), las dos consideraciones más importantes de las preposiciones a nivel sintáctico son la selección y la valencia, tal y como describo a continuación:

- a) La selección se refiere al rasgo de las preposiciones que determina si una preposición particular es seleccionada o no por un término regente. Por ejemplo, las preposiciones seleccionadas son aquellas que pueden estar regidas por un verbo o adjetivo como en la oración *everything depends on you*, y, consecuentemente, no conllevan una carga semántica, ya que la

presencia de la preposición depende unilateralmente del elemento regente. Por otro lado, las preposiciones no seleccionadas no dependen de ninguna otra categoría léxica, sino más bien contribuyen semánticamente a la cláusula en la que se encuentran y, por lo tanto, al significado global de la oración, como en el caso de *Jane and Tom talked about the situation in the workshop*. El sintagma preposicional *in the workshop* no está seleccionado por ningún elemento de la oración y, consiguientemente, se comporta como un elemento relativamente independiente que ayuda a ubicar la acción del verbo principal en el esquema situacional descrito por la oración.

- b) La valencia hace alusión al rasgo de la transitividad de la preposición. Se consideran preposiciones transitivas aquellas que tienen la capacidad de seleccionar un argumento (normalmente un sintagma nominal) con el fin de construir un sintagma preposicional, p.ej. *many people come to my farm in the afternoon*. Por el contrario, aquellas preposiciones que no pueden asumir argumentos son consideradas intransitivas y suelen ocurrir como “(a) components of larger multiword expressions (e.g., verb particles constructions such as pick it up), (b) copular predicates (e.g., the doctor is in), or (c) prenominal modifiers (e.g., an off day)” (Baldwin et al., 2009, p. 120).

Desde una perspectiva sintáctico-semántica, Suppes (2005) otorga a las preposiciones la responsabilidad de llevar consigo la carga de expresar ideas cognitivas sobre espacio, tiempo y muchas otras relaciones entre los humanos y objetos de alrededor. En su análisis sobre la oración *John is pouring water in the pitcher* (Suppes, 1979, p. 390) afirma que el sintagma preposicional no hace referencia a la ubicación de *John*, sino a la acción expresada por el verbo, es decir, al evento de la oración. De este modo, concluye en la imposibilidad de analizar la oración a través de una semántica natural independiente del contexto, pues la combinación sintáctico-semántica parece exigir la presencia de elementos oracionales que permitan la correcta interpretación del sintagma preposicional.

Como conclusión de este apartado, es relevante recapitular que las preposiciones analizadas en esta tesis son aquellas consideradas como no seleccionadas y transitivas

según Baldwin et al. (2009), o lexicalizadas según Downing y Locke (2006), que funcionan como núcleo del sintagma preposicional seleccionando su propio complemento, según Huddleston y Pullum (2002), y que permiten introducir adjuntos no predicativos que contribuyen semánticamente a la oración en su totalidad, según Quirk et al. (1985).

### 2.2.2 Las preposiciones en lingüística cognitiva

La Semántica Cognitiva nació en los años 80 como reacción a los estudios formales y estructurales sobre la lengua que se habían desarrollado hasta entonces. La perspectiva de la Semántica Cognitiva (Talmy, 1985, 1991, 2000; Jackendoff, 1983, 1990; Langacker, 1986; Lakoff, 1987) marca una línea de investigación donde toda conceptualización mental procede de la experiencia física que los hablantes obtienen del entorno que les rodea. En otras palabras, pretende explicar cómo las intenciones del hablante limitan su conceptualización y, por lo tanto, su selección léxica. Por tanto, este enfoque establece que la estructura conceptual que los hablantes crean es la encargada de determinar la manifestación semántica de los enunciados lingüísticos que expresamos. Como Talmy (2000, p.4) afirmó, “research on cognitive semantics is research on conceptual content and its organization in language”.

En consonancia con esta afirmación, Evans y Green (2006) defendieron que existen cuatro principios fundamentales en la Semántica Cognitiva. El primero de ellos es el principio de cognición corpórea (*embodied cognition thesis*), el cual mantiene que la organización conceptual procede de las experiencias que los humanos tenemos con el mundo. El segundo principio implica que la estructura conceptual es la encargada de determinar la estructura semántica de nuestro lenguaje, es decir, las decisiones léxicas vienen determinadas por conceptos que han sido creados en la mente del hablante. El tercer principio muestra cómo estos esquemas mentales son, de algún modo, definidos en nuestra mente a partir de rasgos característicos que adquirimos conforme a nuestra experiencia. En otras palabras, establece que la definición conceptual que almacenamos en nuestra mente no depende de un amplio número de definiciones colegiadas sino de características comunes a elementos del mundo junto con un conocimiento cultural y enciclopédico que nos permite crear excepciones, incluso cuando un elemento reúne las características necesarias para ser conceptualizado de algún modo. Por último, el cuarto

principio trata de la creación de significado en el proceso de conceptualización. Esta idea sostiene que el proceso de creación de significado no sólo está basado en la descripción conceptual del conocimiento enciclopédico, sino que además se necesitan estrategias de inferencia que permitan la conexión de diversos esquemas mentales que, a su vez, puedan proporcionar un nuevo significado. Como explicaré más detalladamente en el Capítulo 4, estos principios teóricos de la Semántica Cognitiva son los que se han tenido en cuenta en el momento de diseñar y desarrollar la base de conocimiento FunGramKB.

Es importante destacar que existen diversos estudios que afirman que la dimensión espacial es la base semántica a partir de la cual la mente construye el resto de significados (Landau & Jackendoff, 1993; Mandler, 2004, Spivey, 2007). Más concretamente, con respecto al significado de las preposiciones, existe una literatura relativamente extensa, y más en particular, sobre las preposiciones espaciales y temporales en diferentes lenguas (Bennett, 1975; Lakoff & Johnson, 1980; Brugman, 1981; Herskovits, 1986, 1997; Lakoff, 1987; Brugman & Lakoff, 1988; Zelinski-Wibbelt, 1993; Boroditsky, 2000; Talmy, 1985, 1991, 2000; Pratt & Francez, 2001; Brée & Pratt- Hartmann, 2002; Von Stechow, 2002; Tyler & Evans, 2003; Svenonius, 2008; Lindstromberg, 2010).

Por un lado, estos estudios de índole cognitivo han intentado dar respuesta al alto carácter polisémico de las mismas. Bennett (1975) afirmaba que las diferencias entre los sentidos de las preposiciones inglesas *in*, *on* y *at* residen en el contexto de la preposición más que en las preposiciones en sí mismas. Defendía que estas preposiciones disfrutaban de un único sentido locativo que abarcaba tanto los usos temporales como espaciales. En esta línea, Herskovits (1986) describió tres niveles de representación para las preposiciones espaciales. En el primer nivel, definió un significado ideal que reflejaba la definición de la preposición y mostraba la representación léxica de la preposición; en el segundo nivel, determinó los tipos de uso derivados del primer nivel, a los que se vinculaban elementos pragmáticos que configuraban la situación en la que el ítem léxico podía aparecer; y por último, en el tercer nivel, definió un conjunto de principios pragmáticos que en combinación permitían un uso más riguroso del segundo nivel.

Por otro lado, Lakoff (1987) sugirió que las estructuras lingüísticas de las que los humanos hacemos uso proceden de nuestras experiencias corpóreas que en definitiva son las encargadas de modelar nuestro sistema de conceptualización. Para él, la dimensión espacial es la base de la estructura conceptual. Lakoff usó la expresión “categorías radiales” para transmitir la idea de que a partir de un sentido prototípico principal se organizan de modo convencional otros sentidos relacionados con el anterior. Siguiendo esta idea, Brugman y Lakoff (1988) realizaron un estudio sobre la preposición *over* desde el enfoque de la semántica léxica cognitiva. Sostenían que las palabras son en sí mismas categorías conceptuales que abarcan significados que, aunque muestren pequeñas diferencias, disfrutan de cierto grado de tipicidad. En consecuencia, el análisis de *over* es considerado un prototipo de los análisis del lenguaje espacial desde una perspectiva de la lingüística cognitiva. Sin embargo, el modelo establecido por Brugman y Lakoff resultó en un análisis de sentidos asociados individualmente a unidades léxicas, lo que dio lugar a un filtro muy desgranado que dificultaba en gran medida la tarea de la conceptualización.

No obstante, Tyler y Evans (2003) afirmaron que la semántica de las preposiciones había sido manejada de manera arbitraria a lo largo de las investigaciones, creyendo que los diferentes significados y usos de una misma preposición estaban organizados de manera accidental. De hecho, estas autoras dirigieron la atención a una distinción similar a la presentada por Velardi, Pazienza y Fasolo (1991) cuando éstos diferenciaban entre “semántica superficial” y “semántica profunda” en el marco del PLN.<sup>5</sup> De este modo, Tyler y Evans (2003) propusieron una

---

Según Velardi, Pazienza y Fasolo (1991), los procesadores del lenguaje natural suelen adoptar uno de los siguientes dos modos de representación del significado léxico:

- (a) Se opta por expresar el significado relacional, i.e. las asociaciones de significado entre las palabras o las clases de palabras por medio de relaciones de significado; el significado relacional es “superficial” en el sentido de que no busca el verdadero sentido de una palabra, sino más bien describe sus usos en el lenguaje cotidiano, o en una sublengua.
- (b) Se opta por expresar el significado conceptual, i.e. el contenido cognitivo de las palabras por medio de rasgos o primitivos; el significado conceptual es “profundo” en el sentido de que expresa fenómenos que están profundamente incorporados en la lengua.

alternativa al modelo planteado por Brugman y Lakoff (1988), estableciendo que los diversos significados asociados a las preposiciones en inglés podían verse representados a través de una red polisémica en la que compartirían un significado central y abstracto. Esta organización sistemática a partir de un significado principal pretendía simular la organización conceptual del lexicón en la mente humana. Consecuentemente, la creación de significado consistiría en un proceso conceptual en el que a través de la integración de información pragmática y enciclopédica obtenida de la percepción y experiencia humana, la mente construiría significados atribuidos a un mismo elemento léxico.

En esta línea, cabe destacar la presencia de estudios que aportan distintas perspectivas para la descripción de dos elementos fundamentales que participan en el marco de la dimensión espacial de la preposición: por un lado, el objeto cuya localización se trata de describir y, por otro, el escenario que sirve como punto de referencia para la localización. Las denominaciones para referirse a estos dos conceptos varían según el autor: *theme* y *referent object* (Jackendoff 1983), *primary object* y *secondary object* (Talmy 1985), *figure* y *ground* (Langacker 1986), y *trajector* y *landmark* (Lakoff 1987) respectivamente.

Asimismo, otra descripción semántica y conceptual de las preposiciones espaciales es la distinción entre preposiciones locativas y direccionales (Clark, 1973; Bennet, 1975; Jackendoff, 1983, 1990; Zwarts, 2005). Considerando este contexto, es relevante comentar la percepción cognitiva que Talmy (1985, 1991, 2000) ofreció sobre la expresión de los eventos en movimiento. Para su análisis, realizó un estudio sobre los conceptos que intervienen en la acción de este tipo de eventos y estableció cuatro componentes semánticos: (a) el *Movimiento*, i.e. la acción de movimiento expresada por el evento en sí mismo, (b) la *Trayectoria*, i.e. la dirección que se propone, (c) la *Figura* (*primary object*), i.e. la entidad que realiza o padece el movimiento, (d) el *Fondo* (*secondary object*), i.e. el marco de referencia para el movimiento de la *Figura*. De este

---

Aunque representar los sentidos de las palabras con una semántica superficial es útil para muchas aplicaciones del PLN, estos investigadores reconocen que no es completamente suficiente.

<sup>6</sup> Talmy también usa los términos derivados de la psicología Gestalt, i.e. *Figure* y *Ground*.

modo, defendió la existencia de un macroevento como unidad cognitiva, cuyos elementos conceptuales pueden ser considerados universales para una organización lingüística. Asimismo, Talmy afirmó que existen dos posibles elementos externos al evento que pueden influir en la descripción del movimiento: (a) la *Manera*, i.e. el modo en el que se realiza el movimiento, y (b) la *Causa*, i.e. la entidad que origina el movimiento de la *Figura*.

En este sentido, Talmy (1985, 1991, 2000) propuso que las lenguas están clasificadas en dos grandes categorías tipológicas dependiendo del modo en el que codifican los componentes semánticos de la estructura (i.e. léxica o sintácticamente), así como en su capacidad para expresar la dirección del movimiento, esto es, cuál es el marco en el que sucede la expresión lingüística. Así, su clasificación resulta en:

- las lenguas de marco verbal, en las que la *Trayectoria* del movimiento viene determinada por el propio verbo. Por ejemplo, éste es el caso del español:

18) La botella salió de la cueva.

- las lenguas de marco satélite, en las que la *Trayectoria* del movimiento viene determinada por los satélites que acompañan al verbo. Este tipo de lenguas incluye a todas las lenguas indoeuropeas, salvo las lenguas romance. En lengua inglesa<sup>7</sup>, concretamente, estos satélites suelen adoptar la forma de sintagmas preposicionales:

19) *The bottle floated out of the cave.*

Las lenguas de marco satélite suelen expresar el *Movimiento* con la ayuda de elementos periféricos o modificadores, gramaticalmente relacionados con el verbo, que conllevan componentes semánticos externos al evento como es la *Manera* o la *Causa*. Por el contrario, las lenguas de marco verbal pueden expresar la *Trayectoria* de una manera inherente al evento, i.e. este componente puede ser expresado lingüísticamente en forma de amalgama verbal en el proceso de lexicalización, mientras que los

---

<sup>7</sup>Aunque en inglés también existen verbos que expresan implícitamente la *Trayectoria*, como es el caso de *enter*.

componentes semánticos externos son omitidos o expresados a través de una cláusula (Asbury, Gehrke, Van Riemsdijk, & Zwarts, 2008). Por ejemplo, Cifuentes (1998) recordó que algunos verbos del español, como *bordear*, *atravesar* o *sobrevolar* entre otros, han amalgamado junto con la acción verbal no sólo la *Trayectoria* sino también la relación que expresa la preposición, por lo que se construyen con sintagmas nominales en función de complemento directo, en vez de hacerlo con sintagmas preposicionales.

No obstante, a pesar de los diversos intentos por tratar de describir la semántica de las preposiciones desde una perspectiva cognitiva, estos estudios tienden a proponer una visión de conjunto de las limitaciones semánticas características de esta clase gramatical sin tener en cuenta su función sintáctica en la oración. De acuerdo con Bennett (1975), que defendía el papel que desempeñaba el contexto para la correcta definición del significado de la preposición, considero necesario el estudio de la sintaxis y los elementos presentes en la oración en la que aparece la preposición para poder desarrollar un repositorio sintáctico-semántico que permita la creación de una subontología de relaciones para su aplicación en el marco del PLN.

### 2.2.3 Las preposiciones en el procesamiento del lenguaje natural

Según Baldwin, Kordoni y Villavicencio (2009), podemos hablar de tres posturas opuestas que adoptan los sistemas del PLN con respecto al tratamiento de la semántica de las preposiciones:

- (a) Las preposiciones están semánticamente vacías y no vale la pena su representación semántica. Esta postura suele adoptarse en investigaciones sobre recuperación de información o clasificación documental, donde los textos suelen procesarse como “bolsas de palabras”<sup>8</sup> donde unidades léxicas como artículos, conjunciones o preposiciones son totalmente ignoradas.

---

<sup>8</sup> Este modelo de representación vectorial sólo tiene en cuenta la ocurrencia de las palabras (i.e. su presencia o no en el documento), donde incluso el orden de aparición de las palabras en el texto es completamente irrelevante.

- (b) La semántica de la preposición es una función sobre las palabras que seleccionan a la preposición y sobre las palabras que la preposición a su vez selecciona, de tal forma que es muy complicado diseñar de forma independiente una caracterización semántica de las preposiciones.
- (c) La semántica preposicional es compleja pero puede ser capturada en un recurso léxico.

En este apartado, describo las investigaciones más relevantes dentro del marco de la lingüística computacional que han contribuido a mejorar el tratamiento de las preposiciones en aplicaciones del PLN. Aunque en esta área no son numerosos los estudios sobre las preposiciones, existen algunos proyectos de notable relevancia en este ámbito, los cuales adoptan la postura (c) arriba descrita, como son *The English LCS Lexicon* (Dorr, 1993, 1994, 1997, 2001), *The Preposition Project* (Litkowski & Hargraves, 2005, 2007) o *PrepNet* (Saint-Dizier, 2005, 2008). Aunque estos proyectos han proporcionado una descripción relativamente detallada de la semántica de las preposiciones para aplicaciones del PLN, aún existe la necesidad de construir un repositorio que sea capaz de almacenar una formalización sintáctico-semántica que permita lidiar con los problemas derivados de la ambigüedad semántica de las preposiciones, como subrayó Saint-Dizier (2008).

El intento por construir modelos que permitan a sistemas del PLN inferir la interpretación semántica de las preposiciones ha sido de diversa naturaleza (O'Hara & Wiebe, 2003, 2009; Litkowski & Hargraves, 2005, 2007; Tratz & Hovy, 2009, 2011; Srikumar & Roth, 2011; 2013a; 2013b; Schneider et al., 2015). El estudio de la desambiguación de los sintagmas preposicionales, junto con el análisis de las construcciones preposicionales y el etiquetado morfosintáctico, ha resultado uno de los principales problemas en el desarrollo de sistemas que reflejasen claramente un avance en la lingüística computacional. Tanto es así que en las últimas tres décadas se ha asistido a diversos cambios paradigmáticos en la investigación de las preposiciones, siendo objeto de estudio de diferentes disciplinas, como la psicolingüística, la inteligencia artificial o la sintaxis teórica, entre otras, donde la mayoría de estos estudios se han centrado en la estructura “verbo + nombre + sintagma preposicional”.

Diversas investigaciones enfocadas a la desambiguación de los sintagmas preposicionales (cf. Jensen & Binot, 1987; Brill & Resnik, 1994; Agirre et al., 2002; Lersundi & Agirre, 2003; Boonthum et al., 2005; Hovy et al., 2010; Hovy et al., 2011) se centraron en el desarrollo de modelos que pretendían simular estrategias de razonamiento humano, basándose en el análisis de árboles sintácticos independientes del contexto léxico o discursivo en el que aparecía el sintagma preposicional. Sin embargo, la dificultad para formalizar el conocimiento semántico de las preposiciones para una aplicación del PLN es una tarea compleja. A continuación, se muestra una breve revisión de los estudios más significativos que se han aproximado a describir la semántica de las preposiciones, bien sea a través de recursos léxico-semánticos desarrollados para el PLN o bien mediante la evaluación de sistemas que utilicen los recursos ya creados con el fin de conseguir la desambiguación de los sintagmas preposicionales.

Hindle y Rooth (1993) propusieron el uso de información léxica extraída de la relación entre el sintagma preposicional y su término regente (i.e. sustantivo o verbo), considerando su distribución en un corpus. De este modo, obtuvieron resultados con respecto a la frecuencia con la que un determinado verbo o nombre aparecía con una preposición concreta. Sin embargo, su método contaba con limitaciones con respecto al tipo de relaciones a ser analizadas y el alto número de combinaciones resultantes. En esta línea, Brill y Resnik (1994) trataron de mejorar este enfoque aportando un método basado en un conjunto de reglas para intentar predecir ciertos sintagmas analizados en un corpus a partir de marcadores lingüísticos contextuales y poder así resolver el problema de la ambigüedad en los sintagmas preposicionales, es decir, aportaron una nueva perspectiva al análisis de las construcciones preposicionales, enfatizando la importancia del contexto más allá del término regente de la preposición.

Harabagiu (1996) trató la desambiguación de los sintagmas prepositivos, concretamente aquellos introducidos por la preposición *of*, mediante un corpus creado a partir de *The Penn Treebank* (Marcus, Marcinkiewicz & Santorini, 1993) y haciendo uso de la base de conocimiento léxico *WordNet* (Miller & Teibel, 1991). Su objetivo consistía en categorizar los argumentos de la preposición con conceptos ya etiquetados en *WordNet* para nombres y verbos, estableciendo conexiones semánticas mediante la

aplicación de tres reglas de heurística inferencial que permitían la lectura de los diferentes tipos de estructura preposicional.

Al igual que ha sucedido con otras categorías gramaticales, como los nombres, los verbos y los adjetivos, ha habido intentos por almacenar el contenido léxico de las preposiciones mediante la construcción de bases de conocimiento. Por ejemplo, *The English LCS Lexicon* (Dorr, 1993, 1994, 1997, 2001), componente fundamental para la construcción de la representación interlingüe del aducto en el sistema de traducción automática UNITRAN, fue uno de los primeros proyectos en almacenar información sintáctico-semántica sobre las preposiciones, a pesar de que su objetivo principal consistió en proporcionar un tratamiento más robusto de los verbos siguiendo la clasificación de Levin (1993). *The English LCS Lexicon* incluye 165 preposiciones con un total de 497 sentidos que forman parte de las construcciones verbales, siendo cada uno de ellos descrito a través de una representación semántica llamada Estructura Léxico-Conceptual<sup>9</sup> la cual es una versión modificada de la Estructura Conceptual de Jackendoff (1983, 1990). La Estructura Conceptual consiste en una representación semántica que se origina de la descomposición del significado del verbo por medio de un conjunto de tipos (p.ej. Evento, Estado, Tiempo, Modo, etc.) y primitivos (p.ej. GO, STAY, BE, AT, TODAY, etc.). Para Jackendoff, no hay distinción entre la representación semántica y la representación conceptual subyacente de una oración. Por tanto, sus representaciones semánticas son "estructuras conceptuales" concebidas como objetos mentales. No obstante, como apunta Lin (2004), las teorías sobre la estructura del argumento verbal que asumen una representación semántica léxica independiente de la estructura sintáctica (p.ej. el enfoque semántico de Jackendoff) son incompletas sin un componente independiente que realice proyecciones explícitamente entre los argumentos en la representación del evento y los argumentos en la estructura sintáctica, tal y como ocurre entre el esquema COREL y la estructura lógica conceptual en ARTEMIS. De lo contrario, se originaría el problema de intentar alinear las representaciones de la semántica léxica con los tipos de eventos definidos aspectualmente, ya que estos últimos no vienen determinados únicamente por el verbo

---

<sup>9</sup>Es importante que el lector no confunda la Estructura Léxico-Conceptual de Dorr con la Estructura Lógica-Conceptual que se origina con ARTEMIS a través de FunGramKB, donde esta última está basada en la estructura lógica de la RRG.

aislado, sino también por otros componentes de la cláusula. A modo de ilustración, el ejemplo (20) muestra la Estructura Léxico-Conceptual propuesta por Dorr (1993, p. 237) para la oración *John went happily to school*:

20) [Event GOLoc  
 ([Thing JOHN],  
 [Path TOLoc ([Position ATLoc ([Thing JOHN], [Location SCHOOL])])])]  
 [Manner HAPPILY])]

Otro recurso léxico relevante en el campo del PLN, y que ha sido usado de forma recurrente en el estudio de la descripción semántica y la desambiguación de las preposiciones, es *FrameNet* (Baker et al., 1998; Fillmore et al., 2001). Se trata de una base de datos léxica para verbos, nombres y adjetivos del inglés con una anotación manual del uso de las palabras en diferentes corpus. Este proyecto tiene como fundamento la teoría lingüística de la *Semántica del Marco* (Fillmore, 1982; 1985), la cual trata de describir el significado de los diversos sentidos de las palabras a través de un marco en el que se señalan los elementos necesarios para construir un esquema conceptual de la situación. En otras palabras, cada marco representa una estructura conceptual que contiene propiedades semánticas del evento, las entidades y las relaciones que aparecen en el contexto más cercano del concepto a analizar. Además, contiene información sobre las unidades léxicas asociadas a estas descripciones a través de ejemplos. En relación a los sintagmas preposicionales, se observa que, en numerosas ocasiones, éstos adoptan la posición de argumentos verbales, de modo que las etiquetas utilizadas para la interpretación de los elementos del marco podrían ser utilizadas para la desambiguación de la función sintáctica de la preposición. Sin embargo, *FrameNet* no fue creado con el objetivo de la desambiguación de esta clase gramatical, por lo que los análisis sólo son válidos cuando se realizan aplicando la metodología usada en *FrameNet*.

Mokhov y Rudzic (2003) plantearon un enfoque que buscaba la categorización automática de los sintagmas prepositivos a través de *WordNet*. A partir de la

información semántica de las unidades léxicas, estos autores propusieron un método de clasificación automática que consistía en cuatro niveles dependientes. El primer nivel trataba de detectar sintagmas preposicionales que acompañaran a nombres o verbos. El segundo nivel etiquetaba los sintagmas detectados, haciendo uso de siete etiquetas obtenidas de *The Penn Treebank*, a través del contenido semántico almacenado en *WordNet*. El tercer nivel determinaba el término regente del sintagma preposicional, (p. ej. sustantivo o verbo). Finalmente, el cuarto nivel, que no llegaron a desarrollar, suponía el refinamiento de la semántica de la relación expresada por el sintagma preposicional. No obstante, aunque los resultados fueron prometedores, se requerían herramientas más robustas que permitieran un análisis más detallado del comportamiento de los sintagmas preposicionales.

O'Hara y Wiebe (2003) realizaron una clasificación de las etiquetas asignadas a los sintagmas preposicionales a partir de los inventarios *The Penn Treebank* y *FrameNet*, para lo cual se utilizaron dos clasificadores semánticos: uno basado en las colocaciones de las palabras de manera estándar y otro basado en las colocaciones de las clases léxicas en forma de hiperónimos almacenados en *WordNet*. El experimento dio como resultado una mayor eficacia en la clasificación al usar de manera conjunta ambos clasificadores.

Alam (2004) realizó un análisis manual de la preposición *over* en relación a dos factores que estaban representados por categorías ontológicas en forma de hiperónimos: por un lado, el significado determinado por el objeto de la preposición y, por otro, el significado determinado por el término regente de la preposición. En consecuencia, diseñó dos árboles de decisiones, es decir, un modelo de predicción para cada uno de los factores, que pretendían definir los sentidos de *over*. Su metodología consistía, en primer lugar, en comprobar el árbol diseñado para el objeto de la preposición y, en el caso de que éste no proporcionara el sentido idóneo, se comprobaría el segundo árbol para poder así desambiguar los diferentes sentidos de la preposición.

Boonthum et al. (2005) trataron de ampliar la metodología usada por Alam con respecto a los dos posibles factores determinantes para el significado de la preposición, añadiendo la relación de meronimia utilizada por *WordNet*. De este modo, afirmaron que la relación que la preposición establece con su complemento y su término regente

sirve como técnica para la desambiguación, aunque estimaron la importancia del contexto y de cierto conocimiento enciclopédico para mejorar el proceso.

En esta misma línea, Hartrumpf et al. (2006) insistieron en que, a pesar del incremento de estudios lingüísticos sobre la semántica de las preposiciones, aún era necesario mejorar el protocolo de desambiguación orientado a las aplicaciones del PLN.

En esta última década, uno de los repositorios léxico-semánticos más populares en el área de la desambiguación de las preposiciones fue *PrepNet* (Saint-Dizier, 2005, 2008). Saint-Dizier (2005) presentó *PrepNet* como un estudio preliminar de la categorización de las preposiciones en francés en la línea que se hizo para nombres y verbos en *FrameNet*, *VerbNet* y *WordNet*. A partir de la consideración de los sentidos de las preposiciones como nociones abstractas, creó tres niveles de clasificación para la descripción de los sentidos. En su estudio propuso una descripción de las preposiciones a partir de una versión simplificada de la Estructura Léxico Conceptual de Dorr, con el fin de construir un repositorio de los comportamientos sintácticos y semánticos de las preposiciones. De este modo, presentó una categorización conceptual de las preposiciones a través de un inventario de papeles temáticos junto con una caracterización de diferentes niveles para los diversos usos analizados. Así, el primer nivel, compuesto a su vez por tres subniveles, caracterizaba, en primer lugar, la familia semántica del sentido de la preposición en forma de papeles temáticos (p. ej., localización, modo, compañía, etc.); en segundo lugar, categorizaba el rasgo (*facets*) del primer subnivel (p. ej., origen, destino, modo y actitudes, etc.); y finalmente especificaba las modalidades de los rasgos del segundo subnivel (*modalities of a facet*) si era necesario (p. ej. modo de comparación, modo con punto de referencia, etc.). Tras la categorización de los sentidos, en el segundo nivel se realizaba la entrada conceptual que consistía en etiquetar cada sentido con un número, un nombre y una glosa; asignarle un marco lingüístico; y realizar una representación conceptual que consistía en una descripción de la semántica composicional del sentido de la preposición. A modo de ilustración de la metodología usada en la elaboración de *PrepNet*, el ejemplo (21)

muestra la descripción de la entrada léxica y la construcción de la entrada conceptual para un sentido de la preposición *from* donde se observan los niveles comentados<sup>10</sup>:

21) *from* = *un origine*

- *famille* : *localisation*
- *facette* : *source*
- *Domaine* : *temp, loc, comp*
- *gloss*: *exprime un origine*

<FRAME>X Pred [80] Y</FRAME>

<FRAMESTRATA/>

<FRAMECONTRAINTE>Y:place/thing</FRAMECONTRAINTE>

<LCS>X: *from* (D,Y)</LCS>

<LCSPRIM>*from*</LCSPRIM>

<LCSDOM>*temp, loc, comp*</LCSDOM>

<LCSCHAMP>*path*</LCSCHAMP>

<COMMENTAIRES>*on peut coupler la primitive from avec les primitives suivantes: back, front, under, above, inside, outside, east, west, north, south*

Saint-Dizier (2008) presentó un análisis más profundo de los sentidos incluidos en cada categorización conceptual y trasladó a otras lenguas las realizaciones léxicas de las descripciones presentadas. En esta ocasión, el proyecto tuvo como propósito, además de la categorización previamente comentada, la explicación del tratamiento composicional de la semántica de las preposiciones que se interrelaciona con la semántica del predicado. También se intentó capturar las limitaciones de las selecciones

---

<sup>10</sup>(Saint-Dizier, comunicación personal, 20 de noviembre de 2014).

léxicas, los sentidos metafóricos y los argumentos de mayor complejidad. Sin embargo, Saint-Dizier tuvo que abandonar el proyecto por falta de financiación según comunicación personal.

Otro proyecto de gran envergadura en este campo es *The Preposition Project* (TPP) (Litkowski & Hargraves 2005, 2007; Litkowski, 2013a, 2013b, 2014), el cual supuso un intento mejorado para desarrollar una base de datos semántica para las preposiciones en lengua inglesa orientada a aplicaciones del PLN. Para ello, extrajeron de *FrameNet* un corpus de aproximadamente 27.000 oraciones con ejemplos de marcos de colocación de las preposiciones. Para la descripción, tomaron como fuente de información léxico-semántica el *New Oxford Dictionary of English* (1997) y el *Oxford Dictionary of English* (2003), cuyo contenido fue seleccionado a través de una comparación cruzada del etiquetado funcional de las preposiciones en *FrameNet* y el tratamiento de las preposiciones realizado desde la perspectiva de la semántica descriptiva de Quirk et al. (1985). De este modo, este proyecto supuso el análisis de 334 preposiciones en inglés asociadas a 673 sentidos, a las que se les otorgó un papel temático y se les dotó de una plantilla con información sintáctico-semántica con respecto a su complemento y posición en la oración. Asimismo, cada sentido fue provisto de una definición, de otras preposiciones con el mismo papel temático y de un análisis del sentido dentro de un marco, siguiendo la metodología de *FrameNet*.

*SemEval-2007*,<sup>11</sup> el cuarto seminario internacional sobre evaluaciones semánticas, dio origen a diversos estudios dirigidos a la desambiguación de las preposiciones con el propósito de fomentar el desarrollo de sistemas que permitieran lidiar con esta problemática en el PLN. Por ejemplo, Litkowski y Hargraves (2007) presentaron un estudio para tratar de averiguar hasta qué punto las preposiciones podían ser desambiguadas y estudiar las limitaciones lingüísticas de esta clase léxica. En consecuencia, se desarrolló un conjunto de 25.000 ejemplos, que abarcaban 34 de las preposiciones más frecuentes en inglés, con el objetivo de determinar si el comportamiento general de las preposiciones de manera aislada se sustentaba con un conjunto de ejemplos extraídos de *FrameNet*. Los resultados preliminares

---

<sup>11</sup>*SemEval* (Semantic Evaluation) son seminarios organizados por ACL-SIGLEX donde se realizan diferentes proyectos para evaluar sistemas de análisis semántico.

proporcionaron un avance en el campo del PLN, manifestado en el trabajo de diversos investigadores que hicieron uso de técnicas estadísticas para identificar rasgos fiables asociados a la desambiguación de las preposiciones (cf. Ye & Baldwin, 2007; Yuret, 2007; Popescu et al., 2007).

Más adelante, como tentativa para enriquecer el análisis de las preposiciones, se desarrollaron tres corpus lingüísticos (Litkowski, 2013a):<sup>12</sup>

- *SemEval Corpus*, que fue construido a partir de *FrameNet*,
- *Oxford English Corpus*, el cual fue cimentado a partir de los ejemplos de los sentidos en el *Oxford Dictionary of English*, y
- *CPA Corpus of Prepositions*, que fue creado sobre un conjunto de oraciones obtenidas del *British National Corpus*, adoptando la metodología del *Corpus Pattern Analysis project* (CPA) para los verbos (Hanks, 2004).

No obstante, Litkowski (2013b) destacó la necesidad de cambiar el curso de las investigaciones dirigidas a la desambiguación de los sentidos de las preposiciones motivado por la importancia de esta clase gramatical en la clasificación semántica de los papeles temáticos en relación con otras unidades léxicas, donde la atención reside en las relaciones semánticas más que en las preposiciones en sí mismas. Por consiguiente, Litkowski (2014) presentó un nuevo recurso para el análisis de las preposiciones, *Pattern Dictionary of English Prepositions* (PDEP)<sup>13</sup>, en el que 304 preposiciones están vinculadas a 81.509 oraciones analizadas a través de datos almacenados en TPP, resultando en un inventario de sentidos extraídos de los corpus mencionados previamente. En realidad, este proyecto sirvió tanto para la descripción semántica como para el almacenamiento de la información sintagmática prototípica de las estructuras a las que las preposiciones se asocian. Sin embargo, el alto número de etiquetas creadas en TPP dificulta en gran medida el proceso de desambiguación para un analizador sintáctico. La Figura 1 presenta un ejemplo de entrada para un sentido de la preposición *in* almacenado en TPP y mostrado en PDEP:

---

<sup>12</sup>Estos corpus están disponibles en <http://www.clres.com>

<sup>13</sup>La base de datos de este proyecto puede ser consultada en <http://www.clres.com/db/TPPEditor.html>

Figura 1. Ejemplo de entrada para un sentido de la preposición in según PDEP

En esta línea, O’Hara y Wiebe (2009) también contribuyeron a la descripción semántica de las preposiciones, creando un repositorio de relaciones a partir de diversos recursos semánticos como *The Penn Treebank* y *FrameNet*.

Como se ha comentado, *SemEval-2007* sirvió para la presentación de diferentes estudios que mostraron enfoques y técnicas con el fin de desambiguar las preposiciones. De entre los más relevantes, cabe destacar la aportación de Tratz y Hovy (2009). Siguiendo los estudios de O’Hara y Wiebe (2003) y Alam (2004) en lo referente a la relación sintáctica que existe entre el término regente, el complemento de la preposición y la preposición en sí misma, Tratz y Hovy (2009) diseñaron una clasificación para los sentidos de las preposiciones teniendo en cuenta la estructura completa de la oración en la que aparecía la preposición. Examinaron el comportamiento de las 34 preposiciones usadas en *SemEval-2007*, etiquetadas vía TPP, consiguiendo mostrar una mejor precisión en la desambiguación que los sistemas presentados por Ye y Baldwin (2007), Yuret (2007) y Popescu et al. (2007). Posteriormente, Hovy et al. (2010) consideraron la importancia de establecer unos parámetros para conseguir resultados significativos en el proceso de desambiguación. De este modo, afirmaron que, para obtener una alta precisión en el proceso de desambiguación de los sentidos de las preposiciones, “one needs to consider context, features, and granularity” (Hovy et al., 2010, p. 460). De hecho, Tratz y Hovy (2011) presentaron un analizador sintáctico formado por cuatro módulos, estando uno de ellos destinado a la desambiguación de los sentidos de las preposiciones, el cual alcanzó una precisión de 85,7% a través de *The Penn Treebank*.

Srikumar y Roth (2011, 2013b) presentaron una metodología enfocada a predecir el comportamiento semántico de las preposiciones en un análisis que considera al término regente de la preposición, a su complemento y al papel semántico de los

mismos como piezas clave para el correcto etiquetado de las relaciones expresadas por la preposición. El punto de partida para su propuesta estuvo marcado, igualmente, por los resultados obtenidos en *SemEval* 2007. Considerando estos datos y las anotaciones hechas a los sentidos de las preposiciones en TPP, consiguieron reducir las 332 etiquetas anotadas en TPP a únicamente 32 relaciones (Srikumar & Roth, 2013b), aglutinando los sentidos de las diferentes preposiciones más relacionados semánticamente. Para su estudio, hicieron uso de dos modelos: el primer modelo usaba una máquina de soporte vectorial con el fin de predecir la etiqueta de la relación sin atender a los argumentos, mientras que el segundo hacía uso de anotaciones semánticas parciales para la predicción de todos los elementos que configuraban la relación. Sus resultados fueron de un 89,43% de precisión en la predicción de las relaciones.

Más recientemente, Schneider et al. (2015) propusieron una taxonomía basada en el inventario de relaciones de Srikumar y Roth (2013a) a la que llamaron *Supersenses*. Schneider y su equipo realizaron un análisis contrastivo del inventario de relaciones de Srikumar y Roth con los papeles temáticos de *VerbNet* (Bonial, Corvey, Palmer, Petukhova & Bunt, 2011) y AMR (Banarescu et al., 2013). Este recurso resultante, el cual está disponible públicamente<sup>14</sup>, proporciona 73 sentidos para las preposiciones, ampliando no sólo el inventario de relaciones de Srikumar y Roth, sino también proporcionando información acerca de ejemplos, definiciones, sinónimos y su relación con otros recursos semánticos. La Figura 2 presenta un ejemplo para el sentido *EndState* ofrecido por Srikumar y Roth (2013a, p. 4)

---

<sup>14</sup><http://demo.ark.cs.cmu.edu/PrepWiki>

## 2.9 EndState

**Definition:** The object of the preposition indicates the resultant state of an entity that is undergoing change. The governor of the preposition is usually a verb or noun denoting transformation.

**Examples:**

- protest turned into a violent confrontation
- forced the club into a special meeting
- profits dropped to \$75 million
- she was moved to tears
- smashed to smithereens
- return towards traditional Keynesianism

Prep.	Senses
into	6(6) 7(7)
to	3(1b)* 5(2)* 6(2a)
towards	2(1a)*

**Comments:**

1. This relation requires some change of state, where the object of the preposition indicates the final state at the end of the transformation.
2. This relation includes cases like *level fell to 3 feet*, where the object of the preposition is a numeric quantity. See note for `Numeric`.

Figura 2. Ejemplo del sentido *EndState*



## CAPÍTULO 3

### MARCO DE LA LINGÜÍSTICA TEÓRICA

#### 3.1 La Gramática del Papel y la Referencia

La RRG (Foley & Van Valin, 1984, Van Valin 1993, 2005, 2008; Van Valin & LaPolla 1997) es una teoría gramatical de corte funcional que posee una perspectiva comunicativa-cognitiva del lenguaje. Esto se traduce en que el estudio de la lengua debe considerarse teniendo en cuenta las funciones semánticas y comunicativas del lenguaje. Desde sus orígenes, la RRG intenta dar respuesta a dos principales preguntas:

*firstly ,what would a linguistic theory be like if it were based on the analysis of typologically diverse languages rather than on English; and secondly, what is the best way of accounting for the interaction of syntactic, semantic and pragmatic phenomena?*

(Van Valin, 2005, p. 1)

Las teorías gramaticales están clasificadas en tres diferentes tipos: estructurales, formales y funcionales. Las gramáticas funcionales se diferencian del resto en que, además de analizar la estructura gramatical desde una perspectiva formal, prestan atención a la situación comunicativa, ya que, de acuerdo con sus defensores, la situación

comunicativa modela la estructura gramatical. Según Nichols (1984), entre los modelos funcionales que se han desarrollado en las últimas décadas, se encuentran modelos de tipo conservador, moderado y extremo, dependiendo de cómo expresen la interrelación sintáctico-gramatical del lenguaje. Por una parte, los modelos conservadores (p.ej., Kuno, 1975) comparten la percepción de la estructura gramatical con las teorías formales, pero no proponen un análisis lingüístico alternativo, por lo que “they represent an extension of them rather than an alternative to them” (Van Valin 2003, p. 330). Por otra parte, los modelos extremos (p. ej., Hopper, 1987) consideran la estructura discursiva como el elemento principal del análisis y están disconformes con el papel otorgado a la gramática en el análisis lingüístico. Finalmente, los modelos funcionales moderados no coinciden con la noción de estructura gramatical defendida por los modelos conservadores, pero establecen que el análisis de la misma debe hacerse teniendo en cuenta elementos como la pragmática y la semántica. Ejemplos relevantes de este tipo de teorías son, además de la RRG, la Gramática Sistemática Funcional (Halliday, 1967, 1994) y la Gramática Funcional (Dik, 1978, 1989).

La motivación subyacente al nacimiento de la RRG radica esencialmente en la búsqueda de una teoría lingüística capaz de dar explicación a los rasgos lingüísticos de todas las lenguas del mundo a través de una descripción gramatical en la que las propiedades semánticas y pragmáticas de las estructuras morfosintácticas desempeñan un papel importante. En este sentido, la RRG pretende desarrollar un marco lingüístico que permita la descripción estructural de todas las lenguas ya sean lenguas configuracionales, i.e. lenguas cuyos elementos siguen un cierto orden, o no configuracionales, i.e. lenguas en las que el orden de los elementos no está predeterminado (Mairal-Usón & Gil, 2004). Así, el orden por excelencia de lenguas como el inglés o el español es el de SVO – *Peter loves Anna*, mientras que lenguas como el lakota o el japonés mantienen una estructura SOV- *Peter Anna loves*. Otras lenguas, como las amerindias, siguen un patrón en el que el verbo precede al resto de elementos, i.e. VOS- *loves Anna Peter*. Finalmente, lenguas como el ruso tienen una estructura variable que no afecta al rasgo gramatical de la oración.

Además, como afirmaron Mairal-Usón y Gil (2004), la RRG defiende el rol que juega el lenguaje en los procesos de conceptualización y razonamiento, y, por esta

razón, el modelo comparte características comunes con los estudios cognitivos del lenguaje. Así, desde una perspectiva histórica, la RRG 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, siendo el modo en el que tratan la relación sintáctico-semántica la principal diferencia entre ellas.

Para poder llevar a cabo una descripción gramatical a través de la interacción entre la sintaxis, la semántica y la pragmática, la RRG dispone de tres componentes, tal y como se muestra en la Figura 3 (Van Valin, 2005, p. 131):

- a) un lexicón que almacena toda la información semántica de las unidades lingüísticas,
- b) un inventario de plantillas sintácticas que pueden aparecer en la gramática de una lengua concreta, y
- c) un analizador de la estructura oracional del enunciado, la cual proporciona la variación lingüística en función de los elementos discursivo-pragmáticos a partir del enlace sintáctico-semántico.

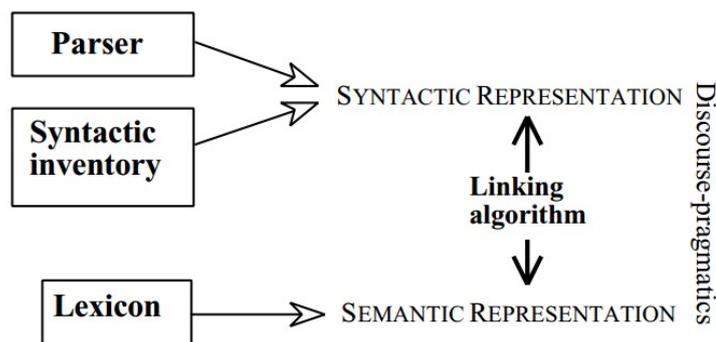


Figura 3. Arquitectura general de la RRG.

En relación a los niveles descriptivos en el análisis de la lengua, la RRG es una teoría monoestratal, donde los elementos sintácticos y semánticos son trazados a partir de la estructura de la oración tal y como se presenta en el discurso. Establece, pues, una

única representación, donde el lexicon y el inventario sintáctico están interrelacionados a través de un algoritmo de enlace que se caracteriza por ser bidireccional, es decir, de la sintaxis a la semántica y de la semántica a la sintaxis.

### **3.1.1 La estructura de la cláusula**

Uno de los conceptos clave en el modelo teórico de la RRG es la estructura de la cláusula. En su búsqueda por la adecuación tipológica, la RRG establece una representación sintáctica en la que diferencia elementos universales de elementos específicos de una lengua. Más concretamente, según la RRG, hay dos focos de interés en los que una teoría sobre la estructura de la cláusula debe centrarse: la consideración de todos los rasgos universales de las cláusulas y la representación de un análisis comparativo de las estructuras de las diferentes lenguas. Desde esta perspectiva, Van Valin (2005) estimó que debemos centrarnos en dos aspectos de la estructura de la cláusula:

- La estructura relacional de la cláusula: las relaciones semánticas, sintácticas y pragmáticas que se establecen entre los elementos de la cláusula, y más concretamente, entre el predicado y sus argumentos.
- La estructura por capas de la cláusula: la organización jerárquica de las oraciones, las cláusulas y los sintagmas.

Para esta última, la teoría propone tres unidades sintácticas:

(a) el núcleo

(b) el centro

(c) la periferia

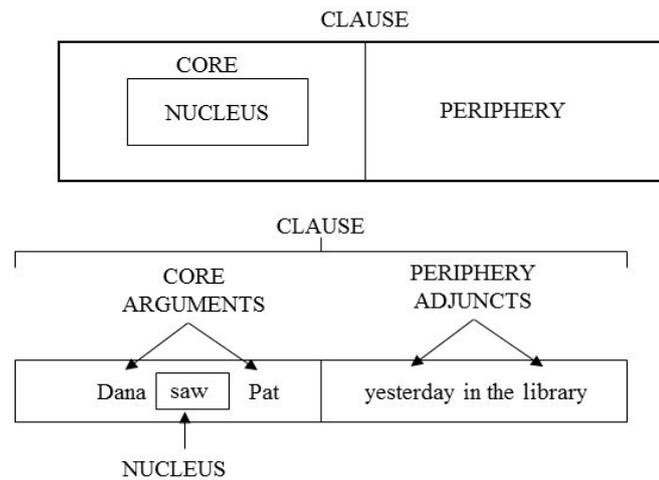


Figura 4. Componentes de la estructura jerárquica de la cláusula.

La estructura jerárquica de la cláusula está construida sobre el contraste, i.e. “between the predicate and non-predicating elements, on the one hand, and, among the non-predicating elements, between arguments and non-arguments, on the other” (Van Valin, 2005, p. 4). Más detalladamente, la dicotomía se plantea en la relación entre los elementos argumentales del predicado y los que no lo son. Según Van Valin (2005), este contraste ocurre en todas las lenguas del mundo. Por tanto, la estructuración de los elementos en la estructura jerárquica de la cláusula propuesta por la RRG, tal y como se puede observar en la Figura 4 (Van Valin, 2005, p. 4), queda de la siguiente manera: por un lado, el núcleo que contiene al verbo; por otro lado, el centro que abarca al núcleo y sus argumentos; y, finalmente, la periferia que contiene los elementos no construccionales de la cláusula.

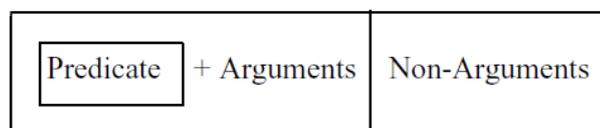


Figura 5. Contrastes universales en la estructura jerárquica de la cláusula.

A modo de ejemplo, la Figura 5 (Van Valin, 2005, p.4) incluye un predicado verbal, el cual es el núcleo de la oración, y tres argumentos nucleares, además de un elemento no construccional en la periferia.

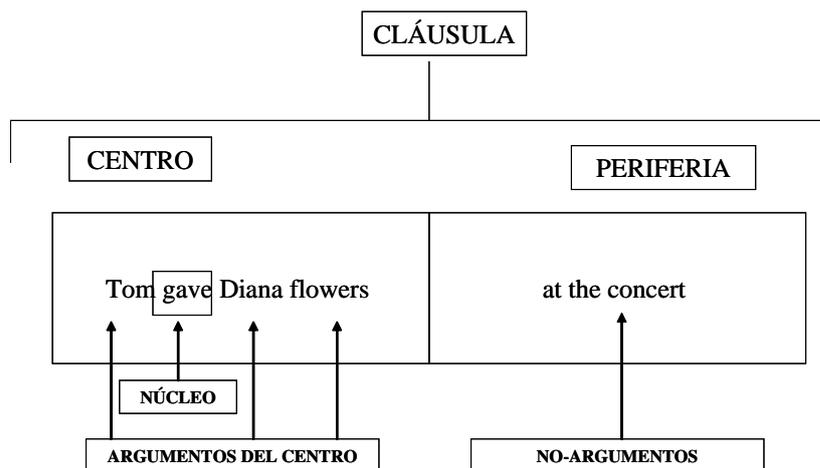


Figura 6. Estructura jerárquica de la cláusula y sus componentes ilustrados

En la Figura 6, el núcleo de la oración es el verbo *give*, el cual representa el predicado verbal, y muestra tres argumentos (i.e. *Tom*, *Diana* y *flowers*). Además, hay un elemento no argumental (i.e. *at the concert*) en la periferia. De este modo, el centro está formado por el predicado y sus argumentos, mientras que la suma del centro y la periferia son los que conforman la estructura total de la cláusula.

El hecho de que los elementos contenidos en la cláusula sean detallados desde un enfoque semántico hace que su presencia en cualquier lengua sea más factible. De este modo, la Figura 7 (Van Valin, 2005, p. 5) resume la proyección ejercida por el algoritmo de enlace que da cuenta de la interfaz semántico-sintáctica que tiene lugar en la estructura jerárquica de la cláusula.

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 7. Interrelación sintáctico-semántica de los elementos que componen la estructura jerárquica de la cláusula.

En los elementos específicos de la lengua, la RRG incorpora, desde la perspectiva pragmática, la posición precentral (PCS, del inglés *PreCore Slot*), en la que

suelen aparecer palabras interrogativas en inglés (p. ej., *what, when, how, etc.*), y la posición dislocada a la izquierda (LDP, del inglés *left-detached position*), en la que se colocan aquellos elementos iniciales de la oración que aparecen separados del resto de elementos por una pausa. Además, para lenguas como el japonés, que posee una estructura de núcleo final, existen las correspondientes posiciones sintácticas postcentral (PoCS, del inglés *PostCore Slot*) y dislocada a la derecha (RDP, del inglés *right-detached position*). A modo de ilustración, la Figura 8 (Van Valin, 1993, p. 7) contiene un adverbio temporal que modifica a toda la oración y está separado de la estructura jerárquica de la cláusula mediante una coma.

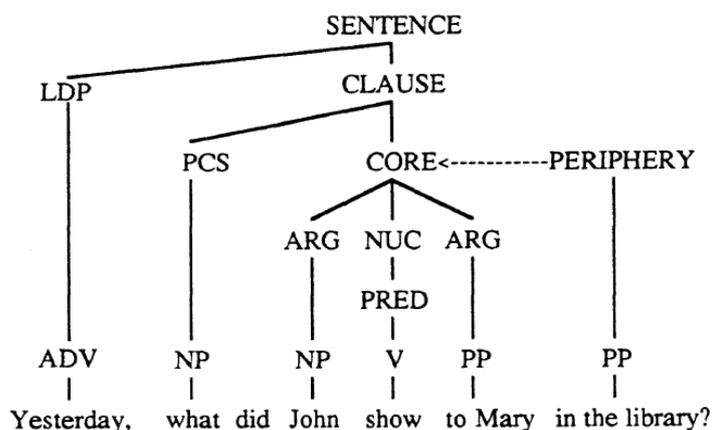


Figura 8. Estructura jerárquica de la cláusula con LDP.

Van Valin (2009) afirmó que el concepto de estructura jerárquica puede también aplicarse a otras categorías como los sintagmas nominales o preposicionales. Por su parte, Rauh (2010) estableció que además de las cláusulas, los sintagmas que contienen adposiciones (i.e. preposiciones o postposiciones) o sustantivos como núcleos también disfrutaban de una estructura jerarquizada. Esta información será comentada en mayor detalle en la sección 3.1.3.

Los operadores son otro componente identificado en la estructura de la cláusula. Se trata de categorías gramaticales (p.ej. aspecto, modo, tiempo, etc.) que modifican diferentes capas de la oración. La Figura 9 muestra los operadores identificados en Van Valin (2005, p. 9):

---

Nuclear operators:
Aspect
Negation
Directionals (only those modifying orientation of action or event without reference to participants)
Core operators:
Directionals (only those expressing the orientation or motion of one participant with reference to another participant or to the speaker)
Event quantification
Modality (root modals, e.g. ability, permission, obligation)
Internal (narrow scope) negation
Clausal operators:
Status (epistemic modals, external negation)
Tense
Evidentials
Illocutionary force

---

Figura 9. Operadores de la estructura jerárquica de la cláusula.

Los operadores están clasificados de acuerdo con la capa de la cláusula a la que modifiquen. Así, el aspecto, la negación y los direccionales son operadores nucleares, mientras que el tiempo, por ejemplo, es un operador clausal. No todas las lenguas expresan los operadores de la misma manera ni siempre mediante signos gramaticales. Los únicos operadores universales a todas las lenguas que se expresan a través de categorías gramaticales son, según Van Valin (2005), el de negación y el de fuerza ilocutiva.

Según Foley y Van Valin (1984), el ámbito de proyección de cada operador está relacionado con su lugar de aparición en la cláusula. De este modo, los operadores que modifican al núcleo aparecen en un lugar más cercano al mismo que los operadores que modifican al centro, a la vez que los operadores que afectan a toda la cláusula, como es el tiempo, tiene una posición más distanciada del núcleo y del centro. Así, en la siguiente oración mostrada en Van Valin (2005, p. 10), los operadores que aparecen son el de fuerza ilocutiva y tiempo (IF/TNS)<sup>15</sup>, modalidad (MOD) y aspecto (ASP):

22) *Will they have to be leaving?*(IF/TNS-MOD-ASP-V)

---

<sup>15</sup>Como indicó Van Valin (2005, p. 10) “the reason that illocutionary force and tense are linked in the English examples is that, in English, illocutionary force is indicated by the position of the tense marker in the main clause”.

La representación formal de los operadores se produce de un modo distinto al resto de componentes de la cláusula, debido a su consideración como modificadores de los elementos de la misma. Así, la representación formal de la proyección de los operadores en la oración (22) es la que se muestra en la Figura 10 (Van Valin, 2005, p. 14).

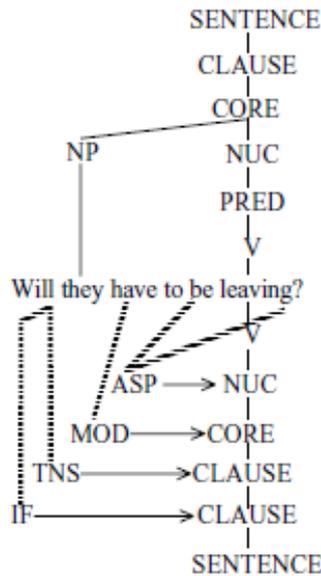


Figura 10. Representación formal de los operadores en una oración en inglés.

En el siguiente apartado, paso a explicar cómo la RRG procede a analizar y representar el significado de las expresiones lingüísticas en su primer nivel, es decir, en el lexicón.

### 3.1.2 La representación semántica en la RRG

El lexicón, donde se almacena la información sobre la representación semántica, se ha convertido en una pieza indiscutible en el análisis lingüístico en los últimos años. La representación semántica en la RRG es producto de la interacción de dos módulos ubicados en el lexicón: las plantillas léxicas y las plantillas construccionales. Por ejemplo, Pérez-Hernández (2009, p. 8) ilustró la estructura de una posible plantilla léxica con el esquema *predicado do'* ( $x$ , [*predicado'*( $x$ )]), la cual está dotada de constantes y variables que permiten la elaboración de la representación semántica de una oración, donde “si el predicado es, por ejemplo, ‘walk’ y la variable  $x$  se sustituye

por la expresión inglesa “my baby” (ej. walk do’ (my baby, [walk’ (my baby)])) esta representación léxica correspondería a la oración “my baby walks””. También a modo de ejemplo, Nolan (2011) describió la plantilla construccional  $[[V] \oplus [_{suffix}]]_{\phi VA}$ , la cual permite formar determinados adjetivos deverbativos en lengua irlandesa a partir de los participios de sus correspondientes verbos, p.ej. *ólta* (bebido) - *ól* (beber). En realidad, la RRG proporciona tres tipos de representación semántica: (a) la representación del significado de las unidades léxicas, como los verbos, (b) la representación de la estructura sintáctica de las oraciones, y (c) la representación de la estructura informativa de una expresión lingüística (Periñán-Pascual & Mairal- Usón, 2009). Siguiendo a Ruiz de Mendoza y Mairal-Usón (2006), parece oportuno detallar cuidadosamente el sistema de descomposición semántica que hace la RRG con el fin de representar la estructura semántica y construccional de los predicados.

En la RRG, cada clase léxica es representada formalmente como una Estructura Lógica (EL), la cual constituye la base de la representación semántica de las unidades léxicas en el Lexicón. Para los verbos, la RRG toma como fuente la clasificación de los verbos de Vendler (1967) basada en el *Aktionsart*. El criterio de la RRG para seguir esta distinción es el hecho de que tiene una base universal para la organización de los predicados verbales. De este modo, los verbos son clasificados en estados, actividades, realizaciones y logros y sus correspondientes versiones causativas. Según Van Valin (2005), en la clasificación de los predicados, los tipos de estado y actividad son los primitivos a partir de los cuales se deriva el resto. Por una parte, tanto los estados como las actividades son temporalmente indefinidos (y por tanto considerados como inherentemente atélicos), aunque los primeros expresan situaciones estáticas (p.ej. amar) mientras que los segundos expresan acciones dinámicas (p.ej. caminar). Por otra parte, tanto los logros como las realizaciones muestran un cambio de estado dentro de una delimitación temporal (y por tanto considerados como inherentemente télicos), aunque los primeros ocurren de forma instantánea (p.ej. explotar) mientras que los segundos tienen una duración en el tiempo (p.ej. aprender). En este contexto, existe otra clase de predicado que se refiere a la combinación de las actividades y los logros, conocida como realización activa, necesaria para la expresión del uso télico de las actividades, tal y como se muestra en (23) (Van Valin 2005, p.33):

23) a. *The soldiers marched in the park.* - actividad

b. *The soldiers marched to the park.* - realización activa

A la clasificación anterior, hay que añadir otra clase de *Aktionsart* propuesta por Smith (1997), que corresponde al esquema temporal puntual sin estado resultante que expresan verbos como *toser*. A esta nueva clase se la conoce como semelfactivo. Asimismo, es relevante comentar que el sistema de descomposición semántica usado en la RRG es una variante del propuesto por Dowty (1979). En resumen, la Tabla 1 (Van Valin, 2005, p. 45) presenta el inventario de las estructuras lógicas utilizadas en la RRG para clasificar cada tipo de predicado verbal, los cuales se definen por medio de los rasgos que se muestran en la Tabla 2 (Van Valin, 2005, p. 33).

Tabla 1. Estructuras lógicas en la RRG para los diferentes tipos de *Aktionsart*.

<i>Aktionsart</i> class	Logical structure
STATE	<b>predicate'</b> (x) or (x, y)
ACTIVITY	<b>do'</b> (x, [ <b>predicate'</b> (x) or (x, y)])
ACHIEVEMENT	INGR <b>predicate'</b> (x) or (x, y), or INGR <b>do'</b> (x, [ <b>predicate'</b> (x) or (x, y)])
SEMELFACTIVE	SEML <b>predicate'</b> (x) or (x, y) SEML <b>do'</b> (x, [ <b>predicate'</b> (x) or (x, y)])
ACCOMPLISHMENT	BECOME <b>predicate'</b> (x) or (x, y), or BECOME <b>do'</b> (x, [ <b>predicate'</b> (x) or (x, y)])
ACTIVE ACCOMPLISHMENT	<b>do'</b> (x, [ <b>predicate</b> <sub>1</sub> ' (x, (y))]) & INGR <b>predicate</b> <sub>2</sub> ' (z, x) or (y)
CAUSATIVE	$\alpha$ CAUSE $\beta$ , where $\alpha$ , $\beta$ are logical structures of any type

Tabla 2. Rasgos para la caracterización de las distintas clases de *Aktionsart*.

a.	State:	[+ static], [– dynamic], [– telic], [– punctual]
b.	Activity:	[– static], [+ dynamic], [– telic], [– punctual]
c.	Achievement:	[– static], [– dynamic], [+ telic], [+ punctual]
d.	Semelfactive:	[– static], [ $\pm$ dynamic], [– telic], [+ punctual]
e.	Accomplishment:	[– static], [– dynamic], [+ telic], [– punctual]
f.	Active accomplishment:	[– static], [+ dynamic], [+ telic], [– punctual]

De esta manera, la RRG permite que cada Estado de Cosas sea representado a través de un metalenguaje semántico universal en formato de EL. Por ejemplo, la EL de la unidad léxica *querer* adopta la forma del *Aktionsart* de estado (24):

24) *querer'* (x, y)



*El niño(x) quiso agua (y)*

Sin embargo, como afirmaban Mairal y Faber (2005), el inventario de EELL presentado en la RRG muestra ciertas carencias que requieren ser mejoradas para poder proporcionar representaciones léxicas tipológicamente adecuadas. Entre sus críticas, destacaron el hecho de que la EL de un predicado en la RRG no contiene un tratamiento interlingüístico de los parámetros semántico-pragmáticos necesarios para la correcta construcción del significado. Del mismo modo, estos autores defendían la creación de un metalenguaje que permitiera representar el predicado de otras lenguas sin hacer uso del inglés y un refinamiento del proceso de descomposición semántica.

Como se mostrará en el siguiente capítulo, los resultados investigadores de esta nueva propuesta coincidían casi en paralelo con la visión de Perrián-Pascual y Arcas-Túnez (2004, 2007) cuando diseñaron e implementaron los principales módulos de FunGramKB, lo cual propició finalmente que el Modelo Léxico Construccional (MLC) (Ruiz de Mendoza & Mairal-Usón, 2008; Mairal-Usón & Ruiz de Mendoza, 2009; Mairal-Usón, Perrián-Pascual & Pérez-Cabello de Alba, 2012; Perrián-Pascual & Mairal-Usón, 2012) se integrase en este último para dar lugar al Gramaticón.

### 3.1.3 Las preposiciones en la RRG

La RRG considera a las preposiciones como elementos activados por principios sistemáticos de la semántica del verbo. El estudio de las preposiciones en este modelo lingüístico ha recibido escasa atención, siendo la obra de Jolly (1991, 1992) una de las pocas aportaciones al respecto. Jolly (1992, p. 275) afirmó que:

*prepositions did not have to be idiosyncratically specified with each verbal lexical entry but rather would be predicted by some general principles related to the semantics of the verb and the assignment of arguments to morphosyntactic positions within a clause.*

En consonancia con este argumento, Jolly diseñó una clasificación de tres tipos de sintagmas preposicionales de acuerdo con la función que realizaban en la oración dentro del marco de la RRG:

- a) Sintagmas preposicionales introducidos por preposiciones no predicativas que establecen argumentos verbales en la EL del predicado verbal a modo de marca de caso:

25) *Brendan showed the newspaper to Bill.*

- b) Sintagmas preposicionales introducidos por preposiciones semipredicativas (O'Dowd, 1998) que establecen argumentos-adjuntos centrales:

26) *Sarah baked a cake for me.*

- c) Sintagmas preposicionales encabezadas por preposiciones predicativas que introducen adjuntos no predicativos y que funcionan como adjuntos en la periferia de la RRG:

27) *Mary prayed after dawn.*

Bajo esta clasificación, la atención de esta tesis se centra en el tipo c, es decir, en los sintagmas preposicionales que funcionan como adjuntos periféricos en la cláusula de la RRG. Como se comentó en la sección 3.1.1, la preposición es considerada un predicado en estos adjuntos. Van Valin (2008, p.170) defendió el carácter endocéntrico de los sintagmas adpositivos, de modo que en los siguientes ejemplos, la preposición es considerada un núcleo predicativo con su correspondiente argumento:

28) a. .Los trabajadores proclamaron una huelga en la fábrica.

b. *They enjoyed their holidays in the summer.*

Mora-Bustos (2012) estableció que la configuración estructural de estos sintagmas preposicionales “implica una disposición específica de todos sus componentes dentro de la estructura de las capas” (p. 205). Así, los sintagmas preposicionales que funcionan como adjuntos en la periferia de la cláusula contienen,

como se explicó en el apartado 3.1.1, una estructura jerarquizada con los mismos componentes que una cláusula. En esta línea, en la oración (29) la estructura del sintagma preposicional, i.e. *in the library*, contendría una estructura jerarquizada con un centro que abarcaría el núcleo, el predicado y el centro, es decir, la preposición y sus argumentos, quedando como muestra la Figura 11 (Van Valin, 2005, p. 23).

29) *Chris saw Pat in the library.*

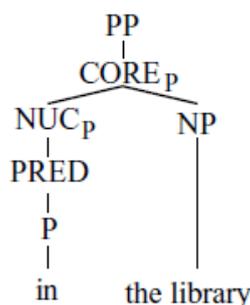


Figura 11. Estructura del sintagma preposicional de una preposición predicativa en la periferia de la RRG.

Puesto que los sintagmas preposicionales introducidos por una preposición predicativa disfrutaban de una estructura jerárquica igual que la de la cláusula, i.e. con su predicado, núcleo y centro, existen operadores que interactúan en el complemento regido por la preposición. Más concretamente, los operadores para un sintagma nominal en la RRG son los que se muestran en la Figura 12 (Van Valin, 2005, p. 24).

---

Nuclear <sub>N</sub> operator:
Nominal aspect (count–mass distinction, classifiers in classifier languages)
Core <sub>N</sub> operators:
Number
Quantification (quantifiers)
Negation
NP operators:
Definiteness
Deixis

---

Figura 12. Operadores en la estructura jerárquica de un sintagma nominal.

Los sintagmas preposicionales son de naturaleza predicativa y, como afirmó Van Valin (2005), funcionan como un modificador, ya sea en el centro de la cláusula

(marcando los argumentos construccionales) o en la periferia. Asimismo, Van Valin estableció la estructura jerarquizada de la cláusula, distinguiendo el centro de la periferia por la presencia de argumentos o no, y consideró que existían dos tipos de realizaciones de los adjuntos periféricos: los sintagmas preposicionales y los adverbios. En esta línea, determinó que los adjuntos periféricos con forma de sintagmas preposicionales podían modificar al centro de la cláusula cuando expresaban rasgos temporales o locativos del estado de cosas expresado por el centro de la cláusula. De hecho, argumentó que los predicados de los sintagmas prepositivos contribuyen a la semántica de la cláusula en la que aparecen, a efectos tanto de su propio significado como del argumento que rigen. En este sentido, Mora-Bustos (2012) consideró que las preposiciones pueden tener tres tipos de representaciones: locativa, temporal y modal.

Con respecto a la relación entre los operadores de la cláusula y los adjuntos periféricos, Navarro y Mora-Bustos (2015, p. 234) afirmaron que “los adjuntos son modificadores sensibles a la estructura de la cláusula, ya que interactúan de manera directa con la proyección de los operadores”. En esta línea, en el capítulo 6, se realizará un análisis del trabajo de Cortés-Rodríguez (2014) con el fin de esclarecer cómo algunos operadores, como el aspecto y el tiempo, afectan a algunos de los elementos que conforman la estructura jerárquica de la cláusula.

Teniendo en cuenta las consideraciones de los autores mencionados en esta sección, la distribución de las dimensiones cognitivas en la creación de la subontología #RELATION se ha realizado en tres taxonomías, i.e. #TIME, #LOCUS y #CIRCUMSTANCE, como se detallará en el capítulo 5 de esta tesis. Otro aspecto relevante que es necesario resaltar es el hecho de que, a diferencia de otros proyectos sobre la descripción semántica de las preposiciones, cuyo principal objetivo ha sido limitarse a asignar el papel temático del sintagma preposicional, en esta investigación se presta también atención a la modelación semántica de la cláusula originada por este tipo de sintagmas dentro del marco de la lingüística computacional. De hecho, este aspecto, el cual ha sido totalmente desatendido por los sistemas del PLN, adquiere un papel primordial en un momento en que “NLP research has been gradually shifting from lexical semantics to compositional semantics” (Cambria & White, 2014, p. 49).



## **CAPÍTULO 4**

### **MARCO DE LA INGENIERÍA DEL CONOCIMIENTO Y EL PROCESAMIENTO DEL LENGUAJE NATURAL**

#### **4.1 FunGramKB**

##### **4.1.1 Introducción**

En este capítulo voy a proceder a la explicación detallada de FunGramKB (Mairal-Usón & Perrián-Pascual, 2009; Perrián-Pascual & Arcas-Túnez, 2004, 2007,2008, 2010a, 2010b, 2011; Perrián-Pascual & Mairal-Usón, 2010b), la base de conocimiento en la que se incluirá la descripción lingüística y conceptual de las preposiciones y relaciones, respectivamente, propuesta en este estudio.

FunGramKB<sup>16</sup> es una base de conocimiento léxico-conceptual que aporta un cambio paradigmático en los sistemas de comprensión del lenguaje natural. Se diseñó sobre fundamentos teóricos de diferentes ciencias y disciplinas tales como la lingüística teórica, la ciencia cognitiva, la ingeniería del conocimiento y la inteligencia artificial, siendo su principal objetivo la aplicación en sistemas del PLN.

Como afirmaron Perrián-Pascual y Arcas-Túnez (2010a), FunGramKB es de carácter multipropósito. Por un lado, es una base de conocimiento multifuncional que se

---

<sup>16</sup>[www.fungramkb.com](http://www.fungramkb.com)

puede aplicar a diferentes tareas del PLN, como la extracción de información, la traducción automática o la recuperación de documentos. Por otro lado, tiene un carácter multilingüe, pues permite desarrollar los lexicones de diversas lenguas; la actual versión de FunGramKB trabaja con siete lenguas (i.e. alemán, búlgaro, catalán, español, francés e italiano), pero, debido al diseño modular del sistema, es posible ampliar el número de lenguas con relativa facilidad.

En este sentido, la modularidad es un concepto de diseño clave en la ingeniería del software (Pressman, 1998), donde una aplicación informática está formada por componentes identificables y tratables por separado aunque integrados de tal forma que todos ellos puedan satisfacer los requisitos del programa. El atributo de la modularidad está íntimamente relacionado con el método "divide y vencerás" en la programación informática, por el cual, a través de un diseño descendente, un problema complejo se descompone en subproblemas más simples, y éstos a su vez en otros problemas más sencillos, hasta que el problema original puede ser resuelto más fácilmente. Además de facilitar la división del trabajo, Cerrada, Collado, Gómez y Estivariz (2000, p. 110) describieron las ventajas de utilizar un diseño modular de la siguiente forma:

**CLARIDAD:** Siempre es fácil de entender y manejar cada una de las partes o módulos de un sistema que tratar de entenderlo como un todo compacto.

**REDUCCIÓN DE COSTOS:** Resulta más barato desarrollar, depurar, documentar, probar y mantener un sistema modular que otro que no lo es. Hay que tener en cuenta que si el número de módulos crece innecesariamente esta afirmación puede no ser correcta. Cuando hay demasiados módulos aumentan también las relaciones entre ellos y las interfaces necesarias.

**REUTILIZACIÓN:** Si los módulos se diseñan teniendo en cuenta otras posibles aplicaciones resultará inmediata su reutilización.

En FunGramKB, esta modularidad se organiza en torno a tres niveles de información destinados a la descripción de las unidades léxicas, las unidades

construccionales y las unidades conceptuales. Por una parte, los niveles léxico y gramatical están contruidos sobre dos modelos lingüísticos: la RRG<sup>17</sup> y el MLC, los cuales comparten características fundamentales para cualquier modelo computacional del lenguaje natural: en primer lugar, su perspectiva funcional del lenguaje posibilita la transmisión del matiz semántico de los fenómenos gramaticales; en segundo lugar, su adecuación tipológica permite la introducción de las distinciones universales como parte del marco lingüístico y, por lo tanto, resulta un ingrediente esencial para los modelos multilingües. Por otra parte, el nivel conceptual, y más concretamente el módulo ontológico, supone innegablemente el eje central de FunGramKB. Es precisamente en este último módulo donde se realiza la descripción conceptual de todas las unidades léxicas incluidas en la base de conocimiento. De hecho, FunGramKB abandona el enfoque lexicista para adoptar uno conceptualista u ontológico (Periñán-Pascual & Mairal-Usón, 2010b), el cual proporciona un nuevo modelo que trata de enlazar la descripción conceptual de la Ontología con los rasgos sintáctico-semánticos almacenados en el Lexicón y el Gramaticón. Para ello, FunGramKB hace uso no sólo de la Estructura Lógica Conceptual sino también de COREL (*Conceptual Representation Language*) (Periñán-Pascual & Arcas-Túnez, 2004; Periñán-Pascual & Mairal-Usón, 2010b), un lenguaje de representación conceptual que permite la descripción de las unidades conceptuales para la inclusión de dicho conocimiento en la Ontología a través de dos propiedades conceptuales: el marco temático y el postulado de significado.

Para llevar a cabo el proceso de comprensión textual y resolver problemas actuales en el PLN, como es la cuestión de la desambiguación, se desarrolló ARTEMIS (*Automatically Representating TExt Meaning via an Interlingua-based System*) (Periñán-Pascual, 2013b; Periñán-Pascual & Arcas-Túnez, 2014), un sistema que permite inferir conocimiento extraído de un repositorio de conocimiento léxico, gramatical y conceptual como FunGramKB.

Este capítulo permite al lector comprender el escenario donde se aplicarán los resultados investigadores de esta tesis. Más concretamente, mi propuesta de descripción sintáctico-semántica podrá ser introducida en FunGramKB en forma de lexicones de

---

<sup>17</sup>Véase el capítulo 3.

preposiciones y una subontología de relaciones que permita a ARTEMIS tratar adecuadamente el problema de la desambiguación de las preposiciones.

#### 4.1.2 Paradigmas influyentes en FunGramKB

##### 4.1.2.1 Teorías cognitivistas

Para el estudio de las teorías que influyeron en el diseño de FunGramKB, debemos remitirnos en primer lugar a la ciencia cognitiva. En este sentido, uno de los impactos más significativos sobre la representación del conocimiento conceptual proviene del modelo de “esquema”, el cual se originó en campos como la psicología cognitiva, la lingüística y la inteligencia artificial, dando lugar a diferentes tipos de constructos cognitivos, p.ej. “esquema” (Rumelhart, 1975), "marco" (Minsky, 1975), "guión" (Schank & Abelson, 1977), "modelo mental" (Johnson-Laird, 1983), o "modelo cognitivo idealizado" (Lakoff, 1987), entre otros. El esquema es fundamentalmente una representación de una determinada entidad, evento o situación, formado generalmente por un conjunto de expectativas sobre cómo será algo. En la psicología cognitiva, por ejemplo, el esquema suele entenderse como una representación mental que contiene un conocimiento basado en experiencias pasadas que nos permite simplificar el proceso de inferencia informativa a partir de nuestra conceptualización de la realidad. Los esquemas representan nuestro saber del mundo en el momento de la comunicación. De esta manera, la comprensión del lenguaje viene determinada por la competencia pragmática de los participantes. De esta forma, al escuchar el enunciado (30), el conocimiento generalizado hace pensar que Juan es un paciente; sin embargo, si el contexto lingüístico proporciona también el enunciado (31), entonces se infiere que Juan es en realidad médico.

30) Juan estuvo esperando en la consulta.

31) Juan atendió a más de treinta pacientes hoy.

Por consiguiente, el modelo de esquema se convierte en una pieza fundamental para la representación del conocimiento conceptual en FunGramKB, ya que es a partir de los esquemas conceptuales cuando se facilita el proceso de inferencia del

conocimiento que tiene lugar durante el proceso de comprensión del lenguaje, como se muestra en la Figura 13.

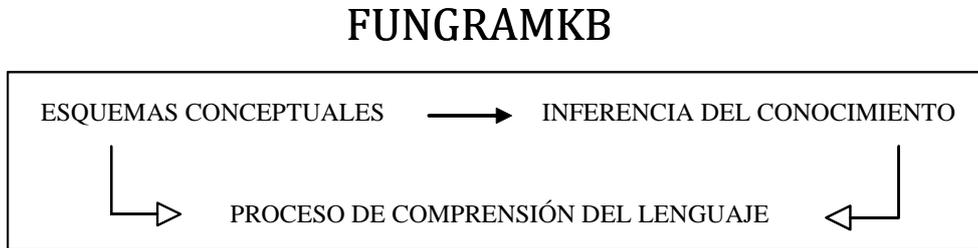


Figura 13. Proceso de comprensión del lenguaje en FunGramKB.

En FunGramKB, los esquemas conceptuales pueden ser clasificados siguiendo dos criterios (Periñán-Pascual & Arcas-Túnez, 2010a):

- a) **Prototipicidad.** Este parámetro origina las “protoestructuras”, es decir, aquellos esquemas conceptuales que contienen un valor prototípico, y las “bioestructuras”, que suponen descripciones de las instancias de eventos y entidades. Por ejemplo, la unidad léxica *novela* implicaría la descripción del concepto prototípico al que va asociado a través de una protoestructura, mientras que la palabra *Cervantes* implicaría la descripción de una entidad con nombre propio y, por lo tanto, requeriría una bioestructura para su descripción.
- b) **Temporalidad.** Este parámetro determina si el concepto necesita ser descrito dentro de un paradigma temporal a través de una “macroestructura” o, si por el contrario, puede ser presentado de forma atemporal a través de una “microestructuras”. Por ejemplo, la descripción del esquema conceptual de la biografía de “John Lennon” necesita de 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, se obtiene la tipología de esquemas conceptuales que muestra la Figura 14 (Periñán-Pascual & Arcas-Túnez, 2010, p. 5).

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

Figura 14. Tipología de los esquemas conceptuales en FunGramKB.

Una de las principales dificultades que se ha encontrado en el tratamiento de la representación conceptual es el grado de influencia que ejerce el contexto en la descripción de las categorías cognitivas (Periñán-Pascual, 2012a). Tradicionalmente se ha considerado que la representación conceptual de una unidad léxica es resultado de la aplicación de ciertos rasgos semánticos a los contextos en los que se suele utilizar dicha unidad, de modo que cada individuo construye un esquema mental de una unidad léxica acorde a las experiencias que haya desarrollado con esa entidad en el mundo real. Así, dependiendo del conocimiento adquirido por un hablante, la representación conceptual puede variar.

De este modo, aunque muchos sistemas adoptan este enfoque en la descripción conceptual, Barsalou (1982) diseñó un modelo para la descripción del significado léxico considerando que las propiedades para la descripción podían depender o no del contexto. Por una parte, las propiedades dependientes del contexto se activan en las situaciones recopiladas para cada unidad léxica. Por otra parte, las propiedades independientes del contexto son aquellas que se activan por las unidades léxicas para cualquier situación. Barsalou (2002, 2003) acuñó el término “conceptualización situada” para referirse a cómo las personas conceptualizan la misma categoría cognitiva de diferente manera en función de no sólo los rasgos intrínsecos de la propia entidad sino también de las propiedades características de la situación en la que el hablante conceptualiza la entidad (i.e. propiedades situacionales) y de otras propiedades

particulares de las situaciones como pueden ser la emoción, la motivación o la intención (i.e. propiedades introspectivas) (Barsalou, 2005). Consecuentemente, su modelo defendía la necesidad de organizar el conocimiento conceptual a partir de situaciones de manera que facilitaran la administración del procesamiento cognitivo. En esta línea, FunGramKB organiza la dimensión de la temporalidad siguiendo la distinción propuesta por Barsalou (1985, 1991) entre categorías taxonómicas (*taxonomic categories*) y categorías derivadas de objetivos. Así, como argumentó Periñán-Pascual (2012a), éstas suponen el fundamento para las macroestructuras usadas en FunGramKB, ya que, según Barsalou (1991), los hablantes comparten estructuras prototípicas que son estables en el tiempo.

Otra teoría de gran influencia en la forma de describir el conocimiento conceptual en FunGramKB es la teoría de Lakoff (1987) sobre los modelos cognitivos idealizados proposicionales. Lakoff afirmó que los modelos cognitivos idealizados guían a los procesos cognitivos en el razonamiento y la categorización. Concretamente, los modelos cognitivos idealizados proposicionales son representaciones mentales resultantes del conocimiento real y se organizan como categorías jerarquizadas entre las que se aplican relaciones semánticas. FunGramKB integra las diferentes estructuras propuestas para los modelos cognitivos idealizados proposicionales con el fin de representar el conocimiento. Por ejemplo, los marcos temáticos de FunGramKB corresponden a los esquemas predicado-argumento en esta teoría de Lakoff.

Finalmente, la arquitectura cognitiva de muchos sistemas inteligentes se ha visto influida por el modelo multimemoria de Tulving (1985), quien afirmó que la memoria a largo plazo se construye a partir de la combinación de tres tipos de componentes que ayudan a los humanos a almacenar y recuperar el conocimiento:

- El componente semántico, el cual almacena la información sobre las palabras.
- El componente procedimental, el cual recoge los datos sobre cómo las acciones se realizan en situaciones cotidianas.
- El componente episódico, el cual almacena información sobre eventos específicos o situaciones biográficas.

Según Perrián-Pascual y Mairal-Usón (2010b), los tres módulos que integran el nivel conceptual se corresponden con los diferentes componentes de la memoria a los que se refería Tulving. De esta forma, como se muestra en la Figura 15 (Perrián-Pascual & Mairal-Usón, 2010b, p. 5), el componente semántico coincide con la Ontología, la cual representa el conocimiento a través de protomicroestructuras, el componente procedimental coincide con el Cognición, el cual representa el conocimiento en forma de protomacroestructuras, y el componente episódico adopta la forma del Onomasticón, donde se almacena el conjunto de bioestructuras. Gracias a esta arquitectura de la base de conocimiento, la interacción entre estos tres módulos cognitivos es precisamente la que permite en FunGramKB la comprensión textual, simulando así el razonamiento humano.

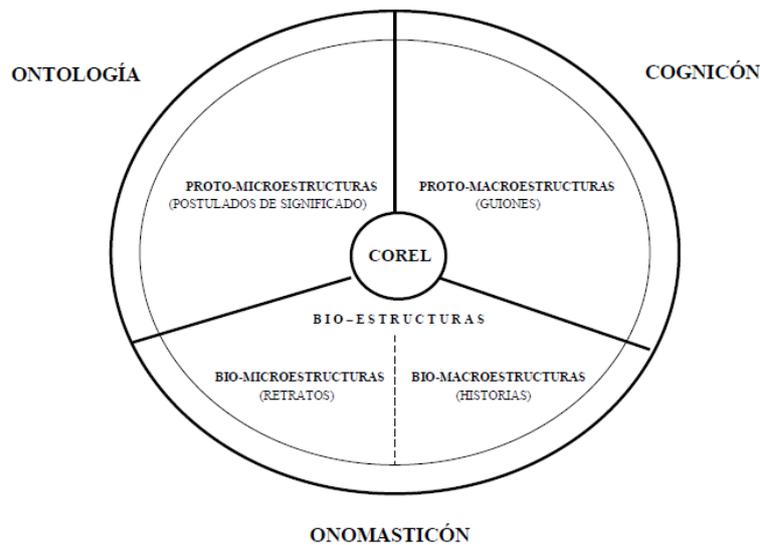


Figura 15. La interrelación de los tres niveles conceptuales de la Ontología según el tipo de conocimiento.

#### 4.1.2.2 Teorías constructivistas

Los modelos lingüísticos sobre los que se sustentan los niveles léxico y gramatical de FunGramKB son el MLC y la RRG, ya comentada en el capítulo anterior. En este apartado, describo brevemente en qué consiste el MLC y su repercusión en FunGramKB.

Como punto de partida para su definición, Pérez-Hernández y Ruiz de Mendoza (2011, p. 103) establecieron que el MLC es “a complex linguistic model that productively combines selected theoretical proposals from a variety of compatible functional, cognitivist and constructional approaches”. En el caso del nivel construccional argumental, las teorías lingüísticas que nutren al modelo para el estudio del significado léxico y construccional son la RRG y los estudios cognitivos como el trabajo de Goldberg (1995, 2005). El MLC tiene como finalidad explicar los procesos lingüísticos que intervienen en la construcción del significado, aportando a la gramática contenido semántico, pragmático y cultural (Mairal-Usón & Periñán-Pascual, 2010b). En esta línea, Jiménez-Briones (2008, p. 1043) afirmó que el MLC “es una propuesta de representación léxica centrada en comprender la relación entre la sintaxis y todos los aspectos del significado, incluyendo la implicatura y el significado ilocutivo”.

Con el fin de proporcionar una representación exhaustiva del significado, el MLC está formado por cuatro niveles (i.e. argumental, implicacional, ilocutivo y discursivo) a través de los cuales se construye el significado mediante el uso de plantillas léxicas y construccionales que interactúan a través de la activación de los procesos de subsunción y pautaación. La Figura 16 (Ruiz de Mendoza & Mairal-Usón, 2008, p. 4) muestra la arquitectura del MLC.

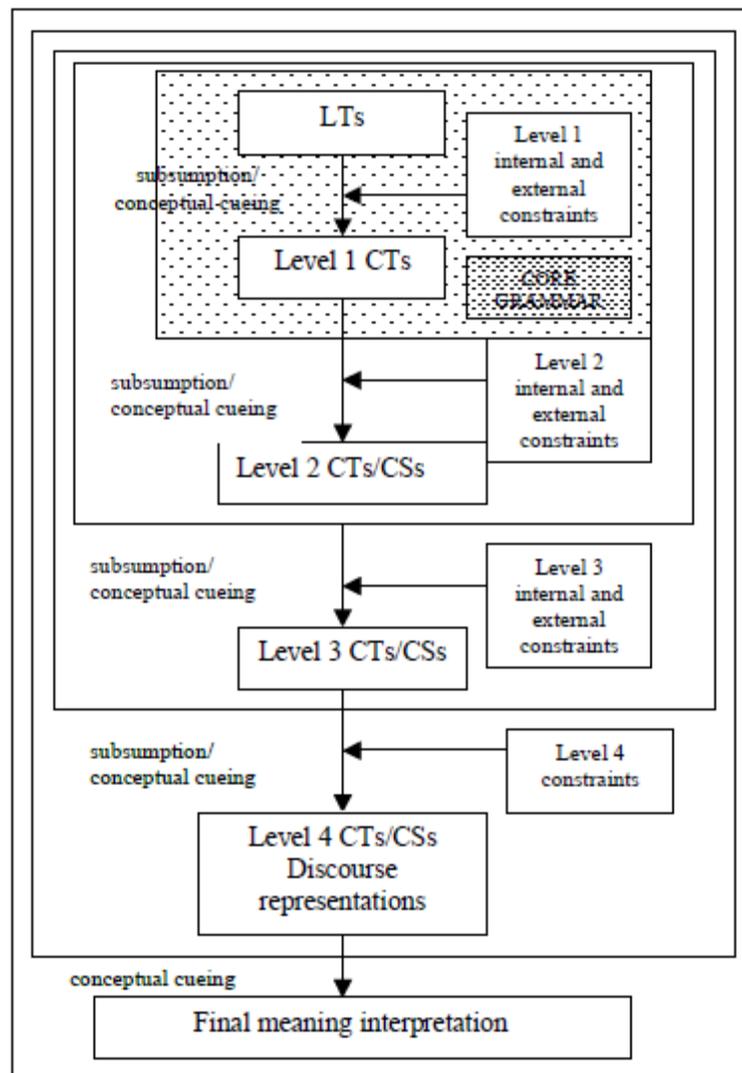


Figura 16. Arquitectura del MLC.

En realidad, la plantilla léxica es una representación semántica mejorada de las EELL mostradas en la RRG (Mairal & Faber, 2002, 2005, 2007). Como estableció Jiménez-Briones (2008, p. 1044), “las plantillas del MLC se delinean usando un metalenguaje preciso y riguroso, basado en el uso de primitivos semánticos, funciones léxicas y distinciones de *Aktionsart*, lo que las confiere un carácter universal y tipológicamente válido”. Así, el uso de plantillas léxicas sirve como un intento para mejorar la adecuación tipológica y proporcionar una mayor precisión en la construcción del significado, ya que no sólo ofrecen contenido léxico-semántico, sino que también proporcionan contenido discursivo-pragmático. Las plantillas constructivas contienen la información semántica de las construcciones que pueden aparecer en la

cláusula como resultado de los argumentos verbales. En consecuencia, la interpretación semántica del significado, según el MLC, es el resultado de un proceso de subsunción léxico-construccional.

Las plantillas léxicas operan en el nivel argumental, mientras que las plantillas construccionales pueden hacerlo en los cuatro niveles del MLC. La interrelación entre las plantillas se activa mediante dos procesos cognitivos. El proceso de subsunción es un mecanismo de construcción de significado que da cuenta de la integración de las estructuras semánticas al integrar bidireccionalmente las plantillas léxicas y las plantillas construccionales. Por su parte, el proceso de pauta proporciona la información de cómo damos sentido a las expresiones lingüísticas, al mostrar indicios de diferentes fuentes que pueden limitar la interpretación. Por ejemplo, en la siguiente interacción, el receptor del mensaje puede inferir cierta información, obviada por el emisor, siempre que tenga información sobre el contexto:

32) *He has just arrived (at the classroom).*

De esta manera, cada nivel del MLC se ocupa de un aspecto determinado para la creación del significado:

- El Nivel 1, o argumental, se preocupa de la representación semántica de los predicados en forma de plantillas léxicas. Como se muestra en Ruiz de Mendoza y Mairal-Usón (2008, p. 8), el formato de una plantilla léxica sería el siguiente:

33) predicate: [SEMANTIC MODULE<lexical functions>]  
[AKTIONSART MODULE<semantic primes>]

- El Nivel 2, o implicacional, trata con las inferencias transmitidas por las plantillas léxicas, como, por ejemplo, en la siguiente interacción:

34) A: *Are you ready for the marathon?*

B: *I broke my leg two weeks ago.*

- El Nivel 3, o ilocutivo, proporciona el significado derivado de situaciones más genéricas donde el proceso de pautación da muestra de implicaciones referidas a convenciones sociales de un nivel superior. Como ilustró Sacramento (2014, p.17) :

35) A: *I'm out of petrol.*

B: *I'll give you a lift.*

El Nivel 4, o discursivo, presta especial atención a los recursos textuales, concretamente, a la coherencia y la cohesión, que ayudan a configurar la estructura discursiva de cualquier texto. Así, como observaron Ruiz de Mendoza y Mairal-Usón (2008, p.15), el ejemplo (36) puede ser parafraseado por medio de la construcción '*Just Because X Doesn't Mean Y*', la cual supone un parámetro convencionalizado de un patrón *EVIDENCE-CONCLUSION/CONCLUSION-EVIDENCE* que suele expresarse con conectores discursivos como *so* y *after all*. De esta manera, el ejemplo (36) resultaría en la oración (37) al aplicar los procesos activados en el Nivel 4.

36) *They live in Berkeley. They must be left wing radicals.*

37) *They live in Berkeley, so they must be left wing radicals.*

Debido a su intención de afinar la descripción de los rasgos de las unidades léxicas a partir de las operaciones que tienen lugar en cada uno de los niveles explicados, el MLC supone un marco idóneo para el desarrollo de aplicaciones del PLN basadas en una semántica construccional más profunda. Por ello, la integración del MLC a FunGramKB dio como resultado la incorporación del nivel gramatical, y más concretamente del Gramaticón.

### 4.1.3 La arquitectura de FunGramKB

La arquitectura modular de FunGramKB no impide que ni los módulos que configuran un mismo nivel ni los módulos que se ubican en diferentes niveles se

comuniquen, siendo la Ontología el pivote de esta base de conocimiento, como se ilustra en la Figura 17 (Periñán- Pascual & Arcas Túnez, 2011, p.3).

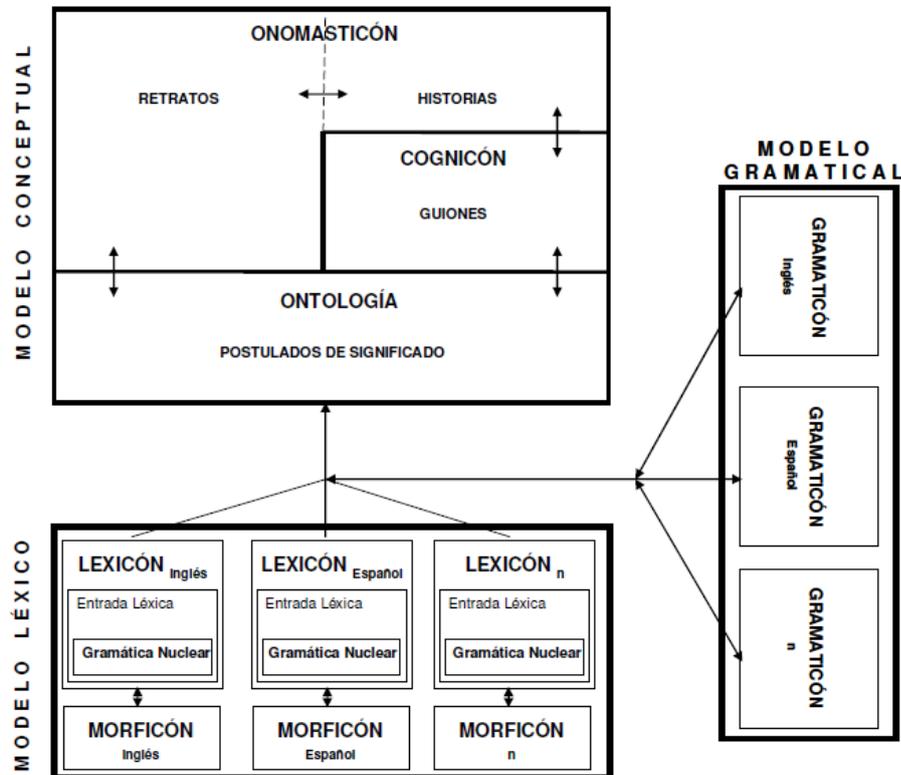


Figura 17. Arquitectura modular de FunGramKB.

En el nivel léxico se almacena la información lingüística (morfosintáctica, principalmente) de las unidades léxicas introducidas en la base de conocimiento. Este módulo está compuesto por dos módulos:

- El Lexicón alberga la información morfosintáctica, pragmática y colocacional, así como las plantillas léxicas que intervienen en el algoritmo de enlace de la RRG.<sup>18</sup>
- El Morficón es el encargado de contener la información propia de cada lengua en relación a la morfología flexiva.

El segundo nivel es el gramatical, el cual está compuesto por un único módulo:

<sup>18</sup>Véase la sección 4.3.1 para una explicación más detallada de este nivel.

- El Gramaticón es el responsable de recopilar la información sobre las construcciones correspondientes a los cuatro niveles identificados en el MLC, el cual proporciona la ayuda necesaria para enriquecer el algoritmo sintáctico-semántico aportado por la RRG.

El tercer nivel es el conceptual, el cual está formado por tres módulos:

- La Ontología consiste en una taxonomía jerarquizada basada en la relación de subsunción (IS-A) en la que se organizan todos los conceptos que sirven para representar el conocimiento semántico. En otras palabras, se trata de una estructura jerárquica donde los conceptos de nivel superior subcategorizan o subsumen a los conceptos de nivel inferior, teniendo lugar un proceso de herencia de las propiedades de los primeros hacia los segundos. En el caso de FunGramKB, el tipo de herencia es múltiple y no monotónica; por una parte, un concepto subordinado puede tener más de un concepto superordinado, siempre y cuando la única relación taxonómica sea la subsunción; por otra parte, los valores de las propiedades heredadas pueden ser invalidados por los valores de los conceptos subordinados.
- El Cognición almacena el contenido procedimental siguiendo el modelo temporal de intervalos de Allen (1983). En otras palabras, este módulo describe los diferentes pasos que conforman los procesos expresados por los eventos dentro de un marco temporal, de tal manera que permita al sistema inferir el orden prototípico de la situación descrita. A modo de ilustración, el ejemplo (38) muestra algunas de las predicaciones que se describen en el esquema cognitivo “comer en un restaurante” (Periñán-Pascual & Arcas-Túnez, 2010a, pp. 2669-2670):

38) \*(e1: +ENTER\_00 (x1: +CUSTOMER\_00)Agent (x1)Theme  
(x2)Location (x3)Origin (x4: +RESTAURANT\_00)Goal (f1:  
(e2: +BE\_01 (x1)Theme (x5: +HUNGRY\_00)Attribute))  
Reason)

*\*(e3: \$ACCOMPANY\_00 (x6: +WAITER\_00)Agent (x6)Theme (x7)Location (x8)Origin (x9: +TABLE\_00)Goal)*

*\*(e4: +SIT\_00 (x1)Theme (x9)Location)*

*\*(e5: +TAKE\_01 (x6)Agent (x10: \$MENU\_00 / \$WINE\_LIST\_00)Theme (x11)Location (x12)Origin (x9)Goal)*

*\*(e6: +REQUEST\_01 (x1)Theme (x13: +FOOD\_00 /+BEVERAGE\_00)Referent (x6)Goal)*

*+(e7: +SAY\_00 (x6)Theme (x14: (e8: +COOK\_00(x15: \$COOK\_D\_00)Theme (x16: +FOOD\_00)Referent))Referent (x15)Goal)*

*\*(e9: +TAKE\_01 (x6)Agent (x17: +BEVERAGE\_00)Theme (x18)Location (x19: \$BAR\_00)Origin (x9)Goal)*

Un cliente (*x1, Agent*) entra (*e1, +ENTER\_00*) a un restaurante (*x4, Location*) porque tiene (*e2, +BE\_01*) hambre (*x5, Attribute*). Seguidamente, el camarero (*x6, Agent*) le acompaña (*e3, \$ACCOMPANY\_00*) a la mesa (*x9, +TABLE\_00*) y el cliente se sienta (*e4, +SIT\_00*). A continuación el camarero le lleva (*e5, +TAKE\_01*) la carta o la carta de vinos (*x10, \$MENU\_00 / \$WINE\_LIST\_00*). El cliente le pide (*e6, +REQUEST\_01*) la comida o la bebida (*x16, +FOOD\_00 / +BEVERAGE\_00*) al camarero, y éste le dice al cocinero (*x15, \$COOK\_D\_00*) que prepare la comida. El camarero lleva la bebida desde la barra (*x19, \$BAR\_00*) a la mesa.

- El Onomasticón contiene datos sobre eventos y entidades nombrados (p.ej. John Lennon), los cuales pueden ser descritos de manera sincrónica o diacrónica.

De los niveles que conforman la arquitectura de FunGramKB, el único que es compartido por todas las lenguas es el nivel conceptual, mientras que el conocimiento almacenado en el nivel léxico y el nivel gramatical es específico de cada lengua. De este modo, el lingüista que trabaje en FunGramKB debe crear un Lexicón, un Morficón y un Gramaticón para cada lengua, mientras que los ingenieros del conocimiento sólo deben desarrollar una Ontología, un Cognicón y un Onomasticón para que sean compartidos por todas las lenguas.

#### 4.1.3.1 El Lexicón

Los modelos OLIF (McCormick, 2002; McCormick, Lieske & Culum, 2004) y EAGLES/ISLE (Calzolari, Lenci & Zampolli, 2001a, 2001b; Monachini, Bertagna, Calzolari, Underwood, & Navarretta, 2003) sirvieron como guía para diseñar y desarrollar el modelo de Lexicón de FunGramKB como un componente suficientemente sólido para proveer a un sistema del PLN. Como se ha observado en la sección anterior, el Lexicón de FunGramKB es el encargado de almacenar la información morfosintáctica, pragmática y colocacional de las unidades léxicas de una determinada lengua. Más concretamente, la información<sup>19</sup> que se contempla en cada entrada léxica es la siguiente:

- a) Información básica: palabra clave, índice e idioma.
- b) Información morfosintáctica: número, género, posición adjetival, información categorial, etc.
- c) Información de la gramática nuclear: *Aktionsart*, plantilla léxica y construcciones sintácticas.<sup>20</sup>
- d) Información miscelánea: estilo, dialecto, ejemplos, traducción, etc.

---

<sup>19</sup>Véase Mairal-Usón y Perrián-Pascual (2009) para una descripción más detallada de cada uno de los elementos incluidos en la entrada léxica.

<sup>20</sup>Es importante recordar que la información sobre los esquemas constructivos es almacenada en el Gramaticón de FunGramKB, a pesar de que la entrada léxica deba contener los marcadores de las construcciones en las que un determinado verbo pueda estar implicado (Perrián-Pascual, 2013b).

Concretamente, la gramática nuclear del Lexicón contiene diferentes parámetros cuyos rasgos permiten al sistema relacionar la información sintáctico-semántica de un verbo con la información conceptual almacenada en la Ontología, dando como resultado una Estructura Lógica Conceptual (ELC) (Periñán-Pascual & Mairal-Usón, 2009). A continuación, se explican brevemente los parámetros más relevantes para este estudio dentro de la gramática nuclear, (Periñán-Pascual & Mairal-Usón, 2012):

- *Aktionsart*: a cada verbo se le asigna uno o más *Aktionsarten* del inventario descrito en la sección 3.1.2 de esta tesis. Por ejemplo, el *Aktionsart* para el verbo *read* es realización activa.
- Plantilla léxica, la cual contiene a su vez los siguientes atributos:
  - o Variables: a cada verbo se le asignan un número de variables (x, y, z) que representan argumentos prototípicos del mismo. Así, en el caso de *read*, el *Aktionsart* de realización activa determina dos variables, por lo que las variables asignadas al verbo *read* son x, y.
  - o Proyección de variables: consiste en relacionar cada una de las variables comentadas anteriormente a un participante en el marco temático del concepto al que está ligado el verbo. De este modo, los participantes que aparecen en la ELC del verbo *read* son (x)Theme (y)Referent.
  - o Rasgos idiosincráticos: tal y como explican Periñán-Pascual y Mairal-Usón (2012, p.341), “estos rasgos, los cuales se expresan por medio del atributo [MR=W] donde W puede ser 0, 1 ó 2, sirven para indicar que una unidad léxica no sigue los principios defectivos para la asignación de los macropapeles”.

Consecuentemente, la ELC de FunGramKB proporciona una descripción más sólida que la ofrecida por la EL en la RRG (Periñán-Pascual & Arcas-Túnez, 2010a), ya que, al estar sustentada por la Ontología, la representación semántica es de carácter conceptual (descrita a través del metalenguaje COREL) y, por ende, independiente de la lengua. Así, como afirma Periñán-Pascual (2012b, pp.40-41):

La utilidad de la estructura lógica conceptual se limita al tratamiento de los fenómenos gramaticales que tienen lugar durante los procesos de análisis o generación lingüísticos. En el caso de que el sistema deba aplicar alguna tarea de razonamiento sobre el texto de entrada, es preciso que la estructura lógica conceptual se traduzca automáticamente a un esquema conceptual en COREL.

La Figura 18 (Mairal-Usón, Periñán-Pascual & Pérez-Cabello de Alba, 2012, p. 94) representa el papel de la ELC como interlingua:

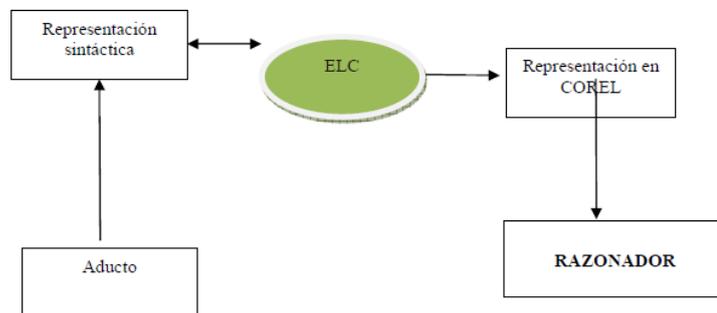


Figura 18. El papel de la ELC como interlingua.

Por consiguiente, se puede concluir que la misión de la ELC es servir como puente entre el nivel abstracto desarrollado en la Ontología y las características lingüísticas descritas en el Lexicón.

De los tres niveles de representación que ofrece la RRG (véase sección 3.1.2), FunGramKB puede crear una ELC para los eventos, como ya se ha explicado, así como un esquema COREL enriquecido con otras propiedades conceptuales inferidas por el motor de razonamiento durante el procesamiento de un texto de entrada (Periñán-Pascual & Mairal-Usón, 2012). En la sección 4.4, se explicará en mayor detalle el proceso empleado por ARTEMIS para la creación de la ELC de una oración.

#### 4.1.3.2 La Ontología

La piedra angular en FunGramKB es la Ontología, pues las unidades léxicas son descritas conceptualmente a través de un lenguaje que posibilita la representación de significado, atendiendo a los dos parámetros sobre los que se sustentan los esquemas

conceptuales en FunGramKB: prototipicidad y temporalidad. Así, COREL surgió para cumplir con las necesidades de un sistema del PLN con respecto a la representación conceptual de esos esquemas conceptuales, aportando mecanismos de herencia e inferencia que mejoran las definiciones de las unidades conceptuales (Periñán-Pascual & Mairal-Usón, 2010b). Además, este metalenguaje fusiona la información de los tres tipos de conocimiento propuestos por Tulving para facilitar el proceso de comprensión y razonamiento. Asimismo, como defienden Periñán-Pascual y Mairal-Usón (2010b, p. 19), “FunGramKB adopta un enfoque mentalista, por el cual el concepto se concibe como una representación mental estructurada”. De este modo, el desarrollo de la Ontología de FunGramKB tiene lugar en tres niveles conceptuales organizados jerárquicamente (Periñán-Pascual & Arcas-Túnez, 2010a):

a) **Los metaconceptos:** configuran el nivel superior de la Ontología y son dimensiones cognitivas a las que no hay asignada ninguna unidad léxica. No poseen una descripción formal por medio de postulados de significado y están precedidos por el símbolo #. Conducen a un modelo metaconceptual que muestra un diseño que contribuye a la integración e intercambio de información con otras ontologías. La configuración de este nivel está distribuida en tres subontologías bajo las categorías de #ENTITY, #EVENT y #QUALITY que se corresponden con los nombres, verbos y adjetivos, respectivamente, en el módulo léxico, dando un total de 42 metaconceptos. La Figura 19 muestra una visión parcial de la subontología de los eventos:

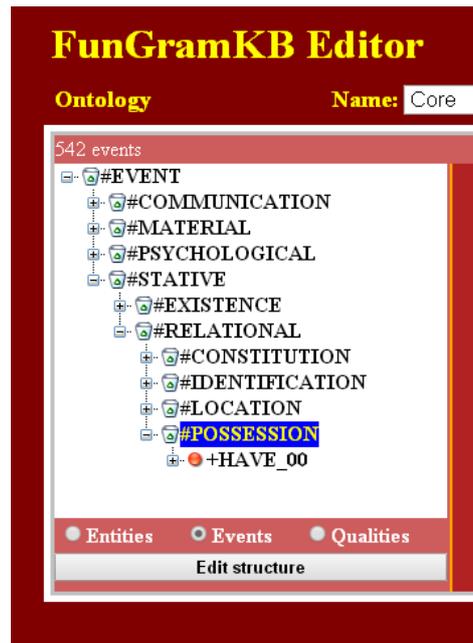


Figura 19. Visión parcial de la ontología de los eventos en FunGramKB.

- b) **Los conceptos básicos:** se encuentran en el segundo nivel de la taxonomía y están precedidos por el símbolo + (p.ej. +MEETING\_00, +SCHOOL\_00, +BORING\_00, etc.). Son usados en FunGramKB como unidades que posibilitan la construcción de los postulados de significado de los conceptos básicos y terminales, así como la descripción de las preferencias de selección en los marcos temáticos (Periñán-Pascual & Arcas-Túnez 2010b). Inicialmente, los conceptos básicos fueron identificados a través de *Longman Dictionary of Contemporary English* (Procter, 1978), creando un inventario total de 5000 conceptos que fue posteriormente revisado y reducido a 1300 conceptos a través de lo que Periñán-Pascual y Mairal-Usón (2011) denominaron como la metodología COHERENT (i.e. *CO*nceptualization, *Hi*erarchization, *RE*modelling and *refinemeNT*). Este nivel actúa de intermediario entre las dimensiones cognitivas representadas por los metaconceptos y los conceptos terminales ubicados en el siguiente nivel. A modo de ilustración, a continuación se muestra la ruta conceptual para el concepto básico +READ\_00:

39) #EVENT>>#PSYCHOLOGICAL>>#PERCEPTION>>

+PERCEIVE\_00>>+SEE\_00>>+EXAMINE\_00>>+READ\_00

- c) **Los conceptos terminales:** son los conceptos más específicos de la ontología y van encabezados por el símbolo \$ (p.ej. \$JUDGE\_00, \$INVESTIGATE\_00, \$FEATURE\_00, etc.). Al contrario que los conceptos básicos, los conceptos terminales no poseen poder definitorio en los postulados de significado. Sin embargo, como afirmaron Perriñán-Pascual y Arcas-Túnez (2010b), la promoción o degradación de conceptos forma parte de la negociación que tiene lugar durante el desarrollo de la base de conocimiento, por lo que algunos conceptos básicos pueden ser degradados a terminales, si su poder definitorio no satisface la descripción, y de manera opuesta, un concepto terminal puede ser clasificado como básico si la introducción de una nueva lengua en el módulo léxico requiere de su semántica para poder definir nuevos conceptos. A continuación, se muestra la ruta conceptual para el concepto terminal \$CONVICTION\_00:

40) #ENTITY>>#ABSTRACT>>#PROPOSITION>>+THOUGHT\_0  
0>>+BELIEF\_00>>\$CONVICTION\_00

Para la descripción de los conceptos en la Ontología, concretamente para los conceptos básicos y terminales, FunGramKB recurre al uso de dos propiedades semánticas: el marco temático y el postulado de significado. Ambos constituyen esquemas cognitivos al estar contruidos a partir de conceptos y dan como resultado representaciones del conocimiento semántico formalizado en lenguaje COREL. De esta manera, a cada evento y cualidad de la Ontología se le otorga un marco temático, es decir, una construcción conceptual que establece el número y tipo de participantes prototípicos implicados en una situación cognitiva descrita por el concepto (Perriñán-Pascual & Arcas-Túnez, 2010b). Por ejemplo, el marco temático para el concepto básico +TRANSFER\_00 es el siguiente:

41) (x1)Agent (x2)Theme (x3)Origin (x4)Goal

un participante ( $x1$ , *Agent*) transfiere una entidad ( $x2$ , *Theme*) desde un origen ( $x3$ , *Origin*) a una meta ( $x4$ , *Goal*).

Como se explicó anteriormente, los conceptos almacenados en FunGramKB están relacionados mediante una relación de subsunción o IS-A. Por ejemplo, en el nivel metaconceptual #TRANSFER, encontramos el concepto básico +TRANSFER\_00 cuyos conceptos básicos subordinados son los conceptos básicos +DISMISS\_00,+GIVE\_00 y +SEND\_00, de modo que estos conceptos básicos comparten el mismo marco temático.

Los marcos temáticos también pueden incluir aquellas preferencias de selección que mejor ayudan a describir la situación cognitiva prototípica, es decir, aquellas que ejercen algún tipo de poder predictivo sobre los participantes. Estas preferencias de selección son descritas mediante conceptos básicos. Como ilustración, se muestra el siguiente ejemplo que encontramos en Mairal-Usón y Perrián-Pascual (2009, p. 223) sobre el marco temático del concepto +GIVE\_00:

42) ( $x1$ : +HUMAN\_00 ^ +ANIMAL\_00)*Agent* ( $x2$ : +CORPUSCULAR\_00)*Theme* ( $x1$ )*Origin* ( $x3$ : +HUMAN\_00 ^+ANIMAL\_00)*Goal*

dos participantes: una persona o animal ( $x1$ ) es el origen para el evento donde algo ( $x2$ ) llega a otra persona o animal ( $x3$ ) como destino.

En este escenario, es importante resaltar que sólo es posible asignar un único marco temático por concepto, mientras que las unidades léxicas vinculadas a un concepto disfrutan de diferentes patrones de subcategorización lingüísticos.

Es importante destacar que las preferencias de selección no funcionan en realidad como “restricciones de selección” en sentido estricto. Tradicionalmente (p.ej. Katz & Fodor, 1963; Chomsky, 1965) las restricciones de selección se han utilizado para emitir juicios absolutos sobre la aceptabilidad de una determinada expresión lingüística. En otras palabras, cuando no se cumplían las restricciones de selección vinculadas a un patrón argumental, la construcción en cuestión se etiquetaba como mal formada. Esta postura ha sido ampliamente criticada, porque la determinación de todas

las posibles restricciones de selección implicaría la tarea imposible de examinar todos los contextos concebibles en que dicha expresión lingüística pudiera utilizarse. Una alternativa más viable, como la propuesta por la Gramática Funcional (Dik, 1978, 1989), es considerar estas supuestas restricciones de selección como preferencias que marcan la interpretación semántica más plausible de los argumentos de un predicado, sin pretender imponer prohibiciones en la inserción de términos en las variables argumentales. De hecho, en el PLN se ha optado por acuñar términos como "preferencias" (Wilks, 1972) y "expectativas" (Schank & Abelson, 1977) para referirse a este fenómeno. Por ejemplo, Wilks utiliza el término "preferencia" porque hace hincapié en que estos marcadores semánticos pueden ser satisfechos o no. Por otra parte, Schank enfatiza el hecho de que estos marcadores se emplean para la predicción descendente durante el procesamiento lingüístico, de ahí el término "expectativa".

Posteriormente, los participantes del marco temático son descritos con respecto a la situación cognitiva que denota el evento o la cualidad, para lo cual se construye un postulado de significado formado por predicaciones conectadas de manera lógica. Así, el postulado de significado del concepto +GIVE\_00 es el siguiente:

43) +((e1: +TRANSFER\_00 (x1)Agent (x2)Theme (x1)Origin  
(x3)Goal (f1: +HAND\_00)Instrument)(e2: +BE\_02  
(x1)Theme (x3)Location (f2: m +NEAR\_00)Position))

Transferir (*e1*) con la mano (*f1*) la posesión de algo concreto a alguien.

Al igual que en la Gramática Funcional de Dik (1989), existen dos tipos de participantes en el postulado de significado, i.e. argumentos y satélites, los cuales se representan a través de las etiquetas indizadas *x* y *f* respectivamente.<sup>21</sup> Como se puede observar en (43), los argumentos del marco temático están coindizados en el postulado de significado, razón por la cual no es preciso incorporar de nuevo cuáles son sus correspondientes preferencias de selección. Además, el postulado de significado incorpora información adicional en forma de satélite sobre el instrumento con el que se

---

<sup>21</sup>Los anexos I y II muestran el inventario de los papeles temáticos de los argumentos y los satélites respectivamente.

realiza la acción, i.e. ( $f1: +HAND\_00$ )*Instrument*), y el lugar en el que se encuentra la meta, i.e. ( $x3$ )*Location* ( $f2: +NEAR\_00$ )*Position*.

Para otorgar consistencia a las descripciones realizadas a través de los postulados de significado, existe un inventario de descriptores asociado a los diferentes metaconceptos que permite a los ingenieros del conocimiento realizar las descripciones de los conceptos de manera más detallada y homogénea. Este inventario se construyó a partir del modelo de Estructura Qualia Extendida diseñado por SIMPLE (Lenci et al., 2000; SIMPLE Specification Group, 2000), donde se extendieron los valores de los papeles Qualia propuestos por Pustejovsky (1991, 1995).

Un aspecto relevante del postulado de significado de una unidad conceptual es que está compuesto por predicaciones, las cuales sirven para representar diferentes tipos de rasgos. De acuerdo con Perrián-Pascual (2013a), y por influencia de las Teorías Dualistas del Significado (Osherson & Smith, 1981; Landau, 1982; Armstrong, Gleitman & Gleitman, 1983; Smith, Osherson, Rips & Keane, 1988), FunGramKB diferencia dos tipos de rasgos que pueden ser representados a través de las predicaciones. Por una parte, los rasgos nucleares tienen una función categorial (p.ej. *los seres vivos respiran*), mientras que los rasgos ejemplares tienen una función identificativa, como se ilustra en la predicación *los pájaros tienen plumas*, ya que puede darse el caso de la existencia de un pájaro que no tenga plumas y no por ello deja de ser un pájaro. En FunGramKB, los rasgos nucleares se representan a través de predicaciones estrictas, las cuales van precedidas del operador +, y los rasgos ejemplares se representan a través de predicaciones rebatibles, las cuales van precedidas del operador \*. En otras palabras, mientras las predicaciones estrictas son similares a las leyes, las cuales no muestran excepciones en el mundo real, las predicaciones rebatibles pueden ser bloqueadas dependiendo de la evidencia que se tenga. De esta forma, FunGramKB permite el razonamiento monotónico con las predicaciones estrictas y el no monotónico con las predicaciones rebatibles. Por otra parte, esto facilita que los postulados de significado puedan recoger, como indicaron Jiménez-Briones, Luzondo y Pérez-Cabello de Alba (2011, p. 24) “tanto nuestro conocimiento semántico como nuestro conocimiento del sentido común que [...] a veces no aparece en ningún

diccionario”. Así, el postulado de significado del concepto +BIRD\_00 quedaría como sigue:

44) +(e1: +BE\_00 (x1: +BIRD\_00)Theme (x2: +VERTEBRATE\_00)Referent)  
 \*(e2: +COMPRISE\_00 (x1)Theme (x3: m +FEATHER\_00)Referent)  
 \*((e3: +COMPRISE\_00 (x1)Theme (x4: 2 +LEG\_00)Referent)  
 \*(e4: +BE\_01 (x4)Theme (x5: +THIN\_00 & +SHORT\_00)Attribute))  
 \*(e5: +COMPRISE\_00 (x1)Theme (x6: 2 +WING\_00)Referent)  
 \*(e6: +COMPRISE\_00 (x1)Theme (x7: 1 +BEAK\_00)Referent)  
 \*(e7: +BE\_01 (x1)Theme (x8: +SMALL\_00)Attribute)  
 \*(e8: +EAT\_00 (x1)Agent (x9: +WORM\_00 ^ +INSECT\_00)Theme (x10)Location(x11)Origin (x12)Goal)  
 Ave (x1, +BIRD\_00) vertebrada (x2, +VERTEBRATE\_00) que está compuesta de plumas (x3, +FEATHER\_00), dos patas (x4, 2 +LEG\_00) finas y cortas (x5, +THIN\_00 & +SHORT\_00), dos alas (x6, 2 +WING\_00), un pico (x7, 1 +BEAK) y que come (e8, +EAT\_00) gusanos o insectos (x9, +WORM\_00 ^ +INSECT\_00).

## 4.2 ARTEMIS

ARTEMIS (Periñán-Pascual, 2013b; Periñán-Pascual & Arcas-Túnez, 2014) es un sistema que infiere automáticamente la representación del significado de una oración en forma de una ELC. Desde sus inicios, este programa, cuya interfaz se muestra en la Figura 20, se vinculó al modelo léxico-semántico de FunGramKB, lo cual permite tratar la semántica textual desde el aparato sintáctico de la RRG y una representación del significado basada en un modelo ontológico (Periñán-Pascual & Arcas-Túnez, 2014).

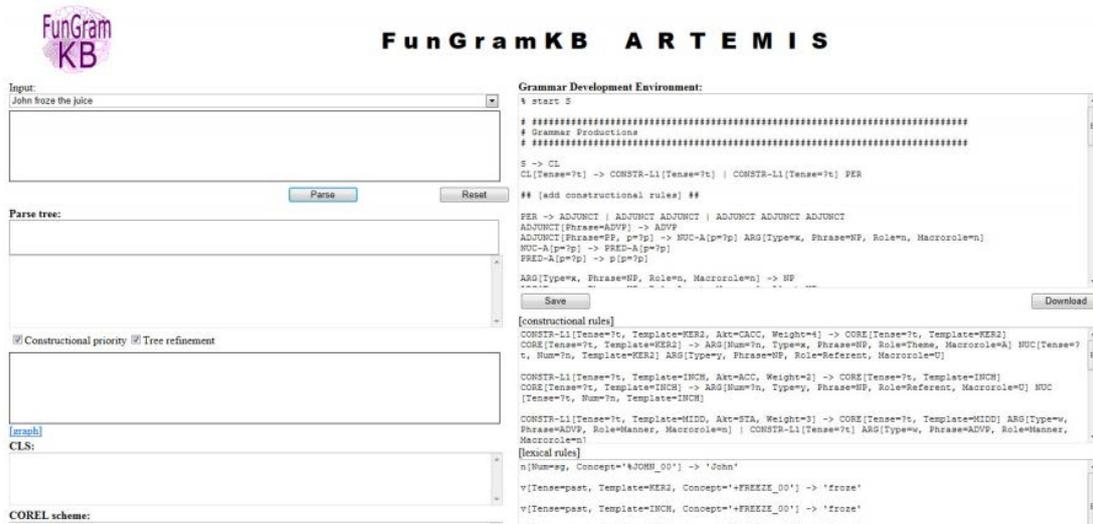


Figura 20. Interfaz de edición de ARTEMIS.

Al principio de este capítulo, se presentó la modularidad como un aspecto clave de la base de conocimiento. Por otra parte, Mahesh (1995) definió tres tipos diferentes de modularidad en el diseño de la tarea de comprensión textual en PLN:

- (a) Modularidad representacional: Cada módulo tiene su propia representación exclusiva de un tipo de conocimiento que no es accesible a ningún otro módulo, con lo cual la interacción no puede ocurrir a través de representaciones compartidas.
- (b) Modularidad procesual: Cada módulo se ocupa de su propio proceso y posee su propia estructura de control, donde la única forma de interacción posible es a través de una representación común.
- (c) Modularidad funcional: Los módulos pueden interactuar siempre que sean funcionalmente independientes. La modularidad funcional permite la interacción a través de representaciones comunes o el intercambio de mensajes entre las estructuras de control de los módulos.

En el caso de ARTEMIS, la modularidad en la tarea de comprensión textual es procesual, donde los componentes vinculados a la Ontología y al Lexicón tienen su propia estructura de control, donde la interacción entre ambos módulos tiene lugar a través de la ELC, tal y como expliqué en el apartado 4.3.1. Debo destacar que no existe

realmente una auténtica independencia funcional entre los módulos que intervienen en el procesamiento, ya que el Lexicón no puede realizar su trabajo de forma autónoma en el caso de que la Ontología fracase, como es el caso cuando una unidad léxica presente en un texto de entrada no tiene su correspondiente unidad conceptual almacenada en la Ontología. No obstante, Crocker (1999) nos recuerda que la modularidad no exige necesariamente que el procesamiento tenga lugar en una determinada dirección, como ocurre en un procesamiento tradicional de dos fases en serie, donde el “procesador sintáctico” construye una representación de estructura de constituyentes y un “procesador semántico” selecciona el papel temático adecuado para cada constituyente en la estructura sintáctica (cf. Frazier & Fodor, 1978; Frazier, 1979). En realidad, ARTEMIS se construyó sobre un modelo interactivo que utiliza la información sintáctica y semántica desde el inicio del análisis a través de un analizador de tablas (*chart parser*) basado en restricciones, el cual es psicológicamente adecuado por ser:

- a. an incremental left-corner parser, where each successive word being encountered is incorporated into a larger structure by combining bottom-up processing with top-down predictions, and
- b. a parallel parser, since multiple parse structures can be generated locally, so there is no need to re-analyse the input if one parse structure proves incorrect (i.e. no backtracking).

(Periñán-Pascual & Arcas-Túnez, 2014, p. 182).

En consecuencia, la colaboración entre FunGramKB y ARTEMIS supone un paso hacia delante en las aplicaciones destinadas a la comprensión del lenguaje. Por otra parte, pocos han sido los intentos de implementar un analizador sintáctico dentro del marco lingüístico de la RRG (cf. Butler 2004; Salem, Hensman & Nolan, 2008, Nola & Salem, 2009; Guest, 2009). No obstante, estos estudios, concretamente Guest (2009), alcanzaron algunos resultados optimistas en el análisis sintáctico de oraciones del inglés y otras lenguas no indoeuropeas como el Dyirbal. Por ejemplo, Salem, Hensman y Nolan (2008) y Nola y Salem (2009) diseñaron un modelo de interlingua para el procesamiento computacional del árabe en tareas de traducción automática, dando como resultado *UniArab (Universal Arabic Machine Translator)*. Sin embargo, este prototipo

presentaba diversos problemas en la proyección léxico-semántica durante el proceso traducción, demostrando así la necesidad de contar con una base de conocimiento más sólida:

On the one hand, the UniArab Lexicon stores a separate lexical entry for every allomorph of a lemma, so inflectional morphology is not handled efficiently. On the other hand, UniArab avoids the problem of word sense disambiguation by adopting a naive one-word one-sense approach to lexical polysemy.

(Periñán-Pascual & Mairal-Usón, 2010a, p. 21)

En este marco de actuación, la ELC creada en FunGramKB a partir del algoritmo léxico-conceptual permite a ARTEMIS construir un esquema COREL, alcanzando un nivel de comprensión textual superior al que se ha logrado en otros estudios. A modo de ilustración, describo brevemente el ejemplo presentado en Periñán-Pascual y Arcas-Túnez (2014, pp. 175-176):

*Peter broke the glass.*

45) **EL:**

$\langle_{IF}^{DEC} \langle_{TNS}^{PAST} \langle_{ASP}^{PERF} \langle [do\text{'}(Peter, \emptyset)] CAUSE [BECOME broken\text{'}(glass)] \rangle \rangle \rangle \rangle$

46) **ELC:**

$\langle_{IF}^{DEC} \langle_{TNS}^{PAST} \langle_{ASP}^{PERF} \langle_{CONSTR-} LI^{KER2} \langle [AKT+BREAK\_00(\%PETER\_00- Theme, \$GLASS\_00 Referent)] \rangle \rangle \rangle \rangle$

47) **Esquema COREL:**

$+ (e1: +DAMAGE\_00 (x1: \%PETER\_00)_{Theme} (x2: \$GLASS\_00)_{Referent} (f1: (e2: +SPLIT\_00 (x1)_{Theme} x2_{Referent}))_{Result})$

Una vez que el usuario ha introducido en ARTEMIS el texto de entrada, en nuestro caso la oración *Peter broke the glass*, el primer paso consiste en reconstruir su EL para luego transformarla inmediatamente en su correspondiente ELC. De este modo, la plantilla léxica vinculada a *romper*, i.e. [do' (x,φ')] CAUSE [BECOME broken' (y)] es representada en la ELC a través del esquema [+BREAK\_00 (x) (y)]. Seguidamente, cada variable de esta infraespecificada ELC se instancia con una unidad conceptual al mismo tiempo que se le atribuye un papel temático obtenido del marco temático del evento al cual el verbo está vinculado en la Ontología. Así, se obtiene una ELC como [AKT<sup>CACC</sup>+BREAK\_00(%PETER\_00-Theme,\$GLASS\_00-Referent)], la cual se enriquece finalmente con los diferentes tipos de operadores de la RRG. Es relevante comentar que la ELC es enriquecida no sólo por la Ontología, sino también por el resto de módulos que configuran el nivel cognitivo (principalmente, el Onomasticón). Por consiguiente, para la ELC en (46), el concepto %PETER\_00 se obtiene del Onomasticón. A partir de esta ELC, ARTEMIS es capaz de realizar el enlace léxico-conceptual y proporcionar el esquema COREL.

Una de las aportaciones del modelo de ARTEMIS a la teoría lingüística de la RRG es la incorporación de las construcciones argumentales a través de un operador construccional, lo cual contribuye no sólo al enriquecimiento semántico del esquema COREL, sino también a la determinación del *Aktionsart* de la oración (Periñán-Pascual & Arcas-Túnez, 2014). Así, en la ELC de las oraciones (46) y (47), ejemplos presentados en Periñán-Pascual y Arcas-Túnez (2014, p. 177), se puede observar cómo el *Aktionsart* del verbo *pound* se modifica al añadir el constructo *into the wall* en el texto de entrada, causando así una modificación en el significado aspectual de la cláusula:

48) *John pounded the nail.*

<IF<sup>DECL</sup><Tense<sup>PAST</sup><CONSTR-  
 LI<sup>KER2</sup><AKT<sup>ACT</sup>[\$POUND\_02(%JOHN\_00-Agent, +NAIL\_01-  
 Goal)]>>>>

*Aktionsart:* Actividad

49) *John pounded the nail into the wall.*

<IF<sup>DECL</sup><Tense<sup>PAST</sup><CONSTR-LI<sup>CMOT</sup><CONSTR-LI<sup>KER2</sup><AKT<sup>CACC</sup>  
 [\$POUND\_02(%JOHN\_00-Agent,+NAIL\_01-Goal,+WALL\_00-  
 Goal)]>>>>

*Aktionsart:* Realización Causativa

Una de las consecuencias de la incorporación de este nuevo operador construccional (L1-CONSTRUCTION) se encuentra en la modificación de la estructura estratificada de la cláusula de la RRG (Figura 21), la cual presenta ahora la forma de la Figura 22:

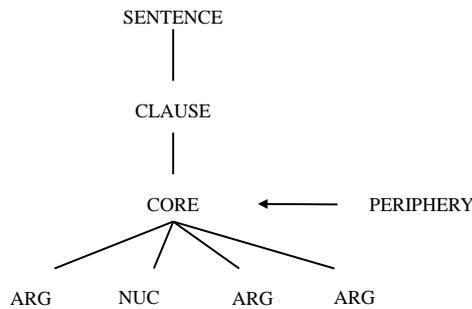


Figura 21. Estructura estratificada de la cláusula en la RRG.

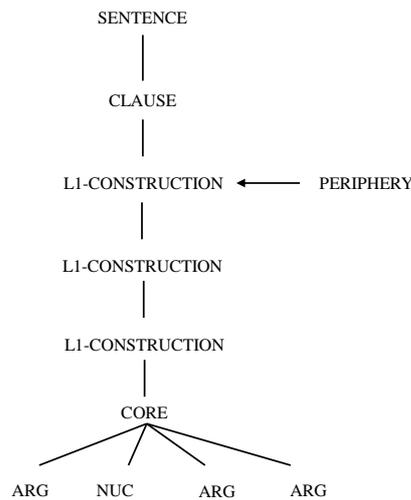


Figura 22. Estructura estratificada de la cláusula en ARTEMIS.

Para el propósito de mi estudio, es importante establecer una clara distinción entre las construcciones argumentales, las cuales vienen determinadas frecuentemente

por el verbo y representadas por el nodo L1-CONSTRUCTION en la estructura estratificada de la cláusula, y los constructos no constructivos que disponen de carga semántica y cierta autonomía predicativa. Mientras la semántica de las primeras es tratada en el nivel gramatical de FunGramKB (i.e. el Gramaticón), la semántica de los segundos es procesada con ayuda de la Ontología. Actualmente, ARTEMIS es capaz de proporcionar un tratamiento adecuado a las construcciones argumentales, como se muestra en la obtenida de Perrián-Pascual (2013, p. 229):

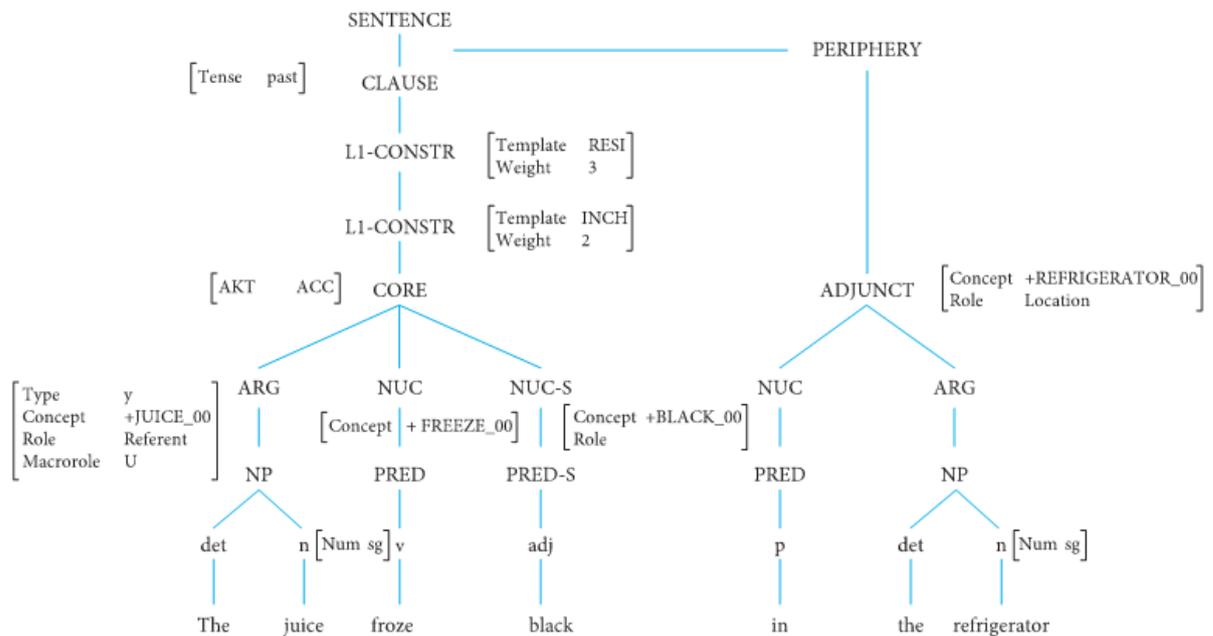


Figura 23. Análisis sintáctico-semántico de *The juice froze black in the refrigerator*.

Sin embargo, esta tesis tiene como objetivo proporcionar un tratamiento más robusto a los sintagmas preposicionales que aparecen en las posiciones periféricas de la cláusula, como es el caso del sintagma *in the refrigerator* en la Figura 23, aportando una descripción léxico-conceptual que se explica en los capítulos 5 y 6 de este estudio.



# CAPÍTULO 5

## MODELIZACIÓN CONCEPTUAL DE LAS RELACIONES

### 5.1 Introducción

A lo largo de este capítulo, se describirán en detalle los preceptos metodológicos que se han llevado a cabo para el diseño y desarrollo para el tratamiento conceptual de las relaciones incorporadas a la base de conocimiento FunGramKB. Al igual que los nombres, verbos y adjetivos son las realizaciones lingüísticas de las entidades, los eventos y las cualidades respectivamente en FunGramKB, las preposiciones se interpretan por medio de las unidades ontológicas que denomino como “relaciones”. Por tanto, mi definición de relación es la siguiente:

- 50) Una relación en FunGramKB es una unidad ontológica que establece una configuración relacional entre dos entidades, o entre una entidad y una cualidad, a través de un evento.

De esta forma, siguiendo con el sistema de organización y descripción utilizado en FunGramKB, las preposiciones almacenan su representación semántica en una subontología de relaciones encabezada por el metaconcepto raíz #RELATION.

En este sentido, el objetivo de este estudio no sólo radica en presentar una propuesta sobre el marco teórico de las preposiciones y las relaciones en una base de conocimiento para el PLN, sino también ilustra dicha propuesta con el análisis de las ocho preposiciones más frecuentes en lengua inglesa. Por este motivo, se consultó la lista de frecuencias ofrecida por Leech et al. (2001)<sup>22</sup> a partir del *British National Corpus* (BNC).<sup>23</sup> La Tabla 3 muestra las diez preposiciones más frecuentes del inglés.

Tabla 3. Las diez preposiciones más frecuentes en inglés hablado y escrito según el BNC.

Preposición	Frecuencia
of	29391
in	18214
to	9343
for	8412
with	6575
on	6475
by	5096
at	4790
from	4134
as	1774

Debido a que el objetivo de esta tesis se centra en la descripción léxico-conceptual de los sintagmas preposicionales que funcionan como adjuntos predicativos en la periferia de la cláusula según la RRG, finalmente se eligieron las ocho preposiciones con mayor número de sentidos en el marco de actuación mencionado. Consecuentemente, las preposiciones seleccionadas fueron *at*, *between*, *for*, *from*, *in*, *on*, *to* y *with*.

## 5.2 Metodología del estudio

La metodología usada en FunGramKB para el desarrollo y diseño de la Ontología se compone de cuatro fases: conceptualización, jerarquización, remodelación

---

<sup>22</sup> Esta lista puede descargarse en [http://ucrel.lanacs.ac.uk/bncfreq/lists/5\\_8\\_all\\_rank\\_preposition.txt](http://ucrel.lanacs.ac.uk/bncfreq/lists/5_8_all_rank_preposition.txt)

<sup>23</sup> Este corpus puede consultarse en <http://corpus.byu.edu/bnc/>

y refinamiento (Periñán-Pascual & Mairal-Usón, 2011). Asimismo, Periñán-Pascual y Arcas-Túnez (2010b) describieron siete compromisos metodológicos que deben aplicarse en la construcción de los niveles conceptuales que conforman la Ontología en FunGramKB. En un principio, mi intención fue aplicar fielmente los mismos fundamentos metodológicos durante el proceso de construcción de la ontología #RELATION, con el fin de mostrar una consistencia en el diseño de este módulo conceptual. Sin embargo, debido a la idiosincrasia de la semántica de las preposiciones, fue necesario reformular dichos fundamentos, tal y como describiré a lo largo de este capítulo.

El primer paso de este estudio consistió en establecer las herramientas lexicográficas que ayudarían a llevar a cabo el análisis de las preposiciones. En este caso, además del BNC, se consultaron los siguientes diccionarios en su versión online<sup>24</sup>:

- *Cambridge Dictionary Online (CAD)*
- *Collins British English Dictionary Online (COD)*
- *Longman Dictionary of Contemporary English Online (LDOCE)*
- *Merriam-Webster Dictionary Online (MWD)*

Durante el análisis semántico de las preposiciones, se observó que los diccionarios ofrecían una amplia clasificación de sentidos con respecto a las preposiciones establecidas, tal y como se muestra en la Tabla 4.

Tabla 4. Ejemplo del número de los sentidos para las ocho preposiciones seleccionadas proporcionados por los cuatro diccionarios consultados.

	<b>at</b>	<b>between</b>	<b>for</b>	<b>from</b>	<b>in</b>	<b>on</b>	<b>to</b>	<b>with</b>
<b>CAD</b>	8	5	15	4	16	25	9	9
<b>COD</b>	8	5	25	7	13	19	11	9
<b>LDOCE</b>	18	10	25	20	28	30	22	23
<b>MWD</b>	6	6	10	3	17	9	8	11

<sup>24</sup> *Cambridge Dictionary Online*: [www.dictionary.cambridge.org](http://www.dictionary.cambridge.org); *Collins British English Dictionary Online*: [www.collinsdictionary.com](http://www.collinsdictionary.com); *Longman Dictionary of Contemporary English Online*: [www.ldoconline.com](http://www.ldoconline.com); *Merriam-Webster Dictionary Online*: [www.merriam-webster.com](http://www.merriam-webster.com).

El amplio número de significados proporcionados por los diccionarios está relacionado con una de las características fundamentales de las preposiciones. Como ya se ha comentado, su capacidad de establecer relaciones con otras categorías gramaticales favorece el carácter polisémico que presentan las preposiciones. Siguiendo el ejemplo de otros autores (O'Hara & Wiebe, 2003; Alam, 2004; Tratz & Hovy, 2009), se consideró necesario realizar un análisis más exhaustivo sobre el contexto lingüístico que acompaña a las preposiciones, ya que la información proporcionada por los recursos lexicográficos arriba mencionados era insuficiente. El propósito era poder encontrar marcadores lingüísticos que ayudaran a obtener datos sobre el comportamiento sintáctico, semántico y pragmático de esta clase léxica. Esta información ayudaría a etiquetar semánticamente las preposiciones (y por ende, los sintagmas preposicionales) presentes en una oración, y, por tanto, contribuiría al proceso de desambiguación. De este modo, tras la consulta de los diccionarios, se hizo uso del BNC para estudiar, además de la propia preposición, el verbo regente de la cláusula y el complemento preposicional.

Otros recursos de gran utilidad en este análisis han sido PDEP (Litkowski, 2014) y la taxonomía *Supersenses* creada por Schneider (2014) y Schneider et al. (2015). Los trabajos de Litkowski y Schneider resultan de gran utilidad desde un punto de vista lingüístico. Sin embargo, para el propósito perseguido en esta tesis, ambos recursos suponen un inventario excesivamente desgranado de sentidos, en el que se incluyen sentidos metafóricos, p.ej. el sentido 1(1)-1 de la preposición *in* que muestra (51), y sentidos que sirven para introducir argumentos construccionales, como el sentido 2(1a) en (52):

51) 1(1)-1: *Trouserless men looked absurd in socks.*

52) 2(1a): *I drove to Reno with my son and all the things I could fit in my car.*

Así, en PDEP la preposición *in* contiene quince sentidos que se ven incrementados en la taxonomía *Supersenses* a diecinueve.<sup>25</sup>

El segundo paso de este estudio consistió en filtrar mi lista de sentidos preposicionales con los sentidos pertinentes al objetivo del estudio, excluyendo así los sentidos derivados de construcciones gramaticales o extensiones semánticas metafóricas, para luego aglutinar los sentidos que comparten un alto grado de contenido semántico. A modo de ilustración, la entrada léxica de la preposición *on* en CAD contiene un sentido que se refiere a *Travel* y otro diferente a *Tool*, como muestro en (53) y (54) respectivamente.

53) *We went to France on the ferry. – Travel*

54) *I do all my household accounts on computer. - Tool*

Estos sentidos se fusionaron bajo un mismo concepto, considerando que el uso de un transporte en (53) puede conceptualizarse como instrumento del mismo modo que ocurre con el uso del ordenador en la oración (54). La Tabla 5 muestra el número de sentidos de las ocho preposiciones después de este filtrado.

Tabla 5. Número final de sentidos de las ocho preposiciones.

	at	between	for	from	in	on	to	with
FunGramKB	3	4	6	6	7	4	3	3

Como se observó en la Tabla 4, el número de sentidos listado para cada entrada lexicográfica es bastante alto, lo cual no facilita las tareas computacionales: cuantos más sentidos tenga una entrada léxica, más complejo será el procedimiento de desambiguación, ya que se introducirán importantes efectos combinatorios (Ide & Véronis, 1998). Por tanto, como explica Buitelaar (2001), el nivel de granularidad de la ontología que subyace a un lexicón está estrechamente conectado al nivel de ambigüedad de cada uno de los términos definidos. En este sentido, la semántica

<sup>25</sup> A modo de ilustración, los anexos III y IV ilustran todos los sentidos de *in* en PDEP y *Supersenses* respectivamente.

ontológica tiene descripciones de significado de granularidad variable en sus diversas implementaciones (Nirenburg & Raskin, 2004), con lo cual, como apunta Hovy (2002), los ingenieros de las ontologías deben ser capaces de indicar las directrices que se siguen con respecto a la granularidad de la semántica de las palabras. En el caso de FunGramKB, Perrián-Pascual y Mairal-Usón (2011, p.24) reconocieron que la Ontología es “coarse-grained in comparison with standard lexicography. However, it is fine-grained in comparison with the axioms in other formal ontologies”.

En el tercer paso, se implementaron las cuatro fases establecidas en Perrián-Pascual y Mairal-Usón (2011) para la construcción de la subontología de las relaciones. Así, se procedió a elaborar los niveles metaconceptual y básico que conforman la ontología (i.e. fases de conceptualización y jerarquización) y finalmente se atendió a la remodelación y refinamiento de los conceptos básicos creados. Siguiendo los principios de FunGramKB, se comenzó estableciendo un esquema metaconceptual fundamentado en estudios sobre taxonomías ontológicas (Reichenbauch, 1947; Jackendoff, 1983; Brée, 1992; Pratt & Brée, 1995; Zhang, 1996; Brée & Pratt-Hartmann, 2002; Schneider, 2014; Schneider et al., 2015) con el propósito de crear un inventario reducido que facilitara la tarea de la desambiguación para un sistema de PLN. En consecuencia, se crearon tres dimensiones conceptuales bajo el nodo #RELATION: #TIME, #LOCUS y #CIRCUMSTANCE. A pesar de que en esta tesis no se va a realizar un análisis conceptual y léxico de todos los sentidos de las preposiciones del inglés, los niveles conceptuales de las tres dimensiones se crearon desde una perspectiva más global que permitiera seguir trabajando en la Ontología posteriormente con otras preposiciones del inglés y con preposiciones de otras lenguas, debido al carácter multilingüe de FunGramKB.

El cuarto paso fue la creación de las entradas conceptuales mediante la asignación de papeles temáticos a los argumentos del marco temático de los conceptos de las ocho preposiciones elegidas, así como el estudio de la preferencia de selección en dichos argumentos, decidiendo qué unidades conceptuales ya creadas en FunGramKB podían desempeñar tal función. Los pasos tercero y cuarto se describen más detalladamente en los siguientes apartados.

Finalmente, se construyeron las entradas léxicas a partir de las unidades conceptuales identificadas en los pasos anteriores, almacenando la información sintáctica relevante con la cual las preposiciones contribuyen a la semántica de las oraciones en las que aparecen como adjunto periférico en la RRG. Debido a su naturaleza, este último paso será descrito en el capítulo 6.

### **5.3 Principios metodológicos para el diseño de la subontología de las relaciones**

En este apartado, se detallan los criterios metodológicos que guían la creación de la subontología de relaciones en FunGramKB:

**CRITERIO 1:** la subontología de relaciones se compone de dos niveles conceptuales formados por metaconceptos y conceptos básicos.

**CRITERIO 2:** la jerarquización de la subontología, la cual está basada en la subsunción, sólo tendrá lugar en el nivel metaconceptual.

**CRITERIO 3:** la entrada conceptual de la relación debe contener un marco temático que incluya la preferencia de selección correspondiente al argumento interno de la preposición.

En cuanto a las propiedades conceptuales que ayudan a definir los conceptos básicos, se tomó la decisión de que las relaciones carecieran de postulado de significado, a diferencia de lo que ocurre con los eventos, las entidades y las cualidades. La decisión fue tomada tras observar el papel que desempeñan las preposiciones en la ELC derivada de una oración y su correspondiente esquema COREL. Al contrario que los nombres, verbos y adjetivos, las preposiciones no contribuyen a la extensión del esquema COREL tras la proyección léxico-conceptual. Como explicó Perrián-Pascual (2013b), en ARTEMIS, la ELC funciona como una representación entre el texto de entrada y el esquema COREL, es decir, como representación formal del significado textual entre la realización lingüística y el dominio conceptual. Así, siguiendo con el ejemplo de la Figura 23, (55) muestra la ELC creada para la oración *The juice froze black in the refrigerator* (Perrián-Pascual, 2013b, p. 23).

55) <<sub>IF</sub> DECL <<sub>Tense</sub> past <<sub>CONSTR-LI</sub> RESI <<sub>CONSTR-LI</sub> INCH  
 <<sub>AKT</sub> ACC [+FREEZE\_00 (+JUICE\_00-Referent,  
 +BLACK\_00-Result)] (+REFRIGERATOR\_00-Location)  
 >>>>

Para el procesamiento del significado textual aportado por la ELC, ARTEMIS necesita un esquema COREL que sirva como puente para la expansión conceptual del significado, ayudando así a construir posibles inferencias durante el proceso de razonamiento. Por este motivo, la ELC mostrada en (55) se traduce al esquema COREL mostrado en (56).

56) +(e1: past +FREEZE\_00 (x1)Theme (x2: +JUICE\_00)Referent  
 (f1:(e2:+BECOME\_00 (x2)Theme (x3:  
 +BLACK\_00)Attribute)) Result(f2: +REFRIGERATOR\_00)  
 Location)

De este modo, el esquema COREL en (56) puede enriquecerse mediante un mecanismo de inferencia a través del postulado de significado extendido del evento +FREEZE\_00, como se muestra en (57):

57) +(e1: past +COOL\_00 (x1)Theme (x2: +JUICE\_00)Referent (f1:  
 +MUCH\_00)Quantity (f2: (e2: +BECOME\_00 (x2)Theme (x3:  
 +SOLID\_00 & +BLACK)Attribute))Result (f3:  
 +REFRIGERATOR\_00)Location)

Se observa que, durante la proyección del sintagma preposicional *in the refrigerator* a la ELC y, posteriormente, al esquema COREL, la aportación semántica de la preposición *in* se limita a la asignación del papel temático al constructo correspondiente a dicho sintagma preposicional. De hecho, la proyección de la preposición no se materializa de la misma forma como lo hacen los nombres, los verbos y los adjetivos, es decir, no se proyectan a unidades conceptuales, sino a papeles temáticos, p.ej. *Location* en el ejemplo (56). Es por esta razón que las preposiciones carecen de postulado de significado en la subontología de relaciones que propongo, ya que conceptualmente sólo se proyectan a papeles temáticos, los cuales no se expanden

en el proceso de *Microconceptual-Knowledge Spreading*, o *MicroKnowing*, el cual funciona de la siguiente forma:

This multi-level process is performed by means of two types of reasoning mechanisms: inheritance and inference. Our inheritance mechanism strictly involves the transfer of one or several predications from a superordinate concept to a subordinate one in the ontology. On the other hand, our inference mechanism is based on the structures shared between predications linked to conceptual units which do not take part in the same subsumption relation within the ontology.

(Periñán-Pascual & Arcas-Túnez, 2005, p. 241).

Por tanto, la aplicación del CRITERIO 1 viene motivada por la decisión de no construir un postulado de significado para cada relación, dando lugar a que en la nueva subontología de relaciones no exista una jerarquía conceptual similar a la del resto de subontologías. Más concretamente, al no existir postulados de significado para las relaciones, no tiene sentido que exista una diferenciación entre conceptos básicos y conceptos terminales, donde la función de estos últimos consiste en participar en la construcción de los postulados de significado de cualquier unidad conceptual. Por tanto, es lógico pensar que la subontología de las relaciones deba estar formada únicamente por metaconceptos y conceptos básicos.

Con respecto al CRITERIO 2, se estudió la relación taxonómica existente entre los dos niveles conceptuales que componen la subontología de relaciones, es decir, el nivel metaconceptual y el nivel conceptual básico, y entre las unidades que configuran cada nivel conceptual. Como se ha explicado en el capítulo 4, la única relación taxonómica permitida en FunGramKB es la subsunción, o relación IS-A. Debido a la ausencia de postulados de significado, los cuales integran los dos componentes aristotélicos de la definición (i.e. *genus* y *diferentia*), planteo que, por el momento, no hay razón alguna para estructurar jerárquicamente los conceptos básicos. En cambio, el nivel metaconceptual, que está compuesto por amplias dimensiones cognitivas y no por unidades conceptuales específicas, es el único que podría disfrutar de cierta jerarquía

con el fin de facilitar la integración y el intercambio de información con otras ontologías del PLN.

Para aplicar el CRITERIO 3, es decir, la construcción de las preferencias de selección en el marco temático de los conceptos básicos, se estudiaron los complementos de la preposición en los ejemplos extraídos de los corpus<sup>26</sup>. En relación al complemento de la preposición, estoy de acuerdo con Hindle y Rooth (1993) cuando afirmaron que ignorar la información proporcionada por el argumento preposicional puede provocar que el sistema del PLN no logre los resultados esperados. Por ejemplo, en las oraciones (58a) y (58b) (Resnik & Hearst, 1993, p. 59), se puede comprobar que el sentido de la preposición es diferente a través de su complemento, por lo que la información acerca de este complemento ayudaría a un analizador a desambiguar semánticamente el sintagma preposicional. Más concretamente, (58a) representa un sentido temporal, mientras que (58b) representa un valor locativo.

58) a. *Britain reopened its embassy in December.*

b. *Britain reopened its embassy in Teheran.*

En ambos ejemplos, el sentido de la preposición sólo puede ser determinado por la preferencia de selección de la relación. Esta información debe ser almacenada en el marco temático ubicado en las entradas conceptuales de la subontología #RELATION de FunGramKB. Por ejemplo, el sustantivo *December* está vinculado a la unidad conceptual \$DECEMBER\_00 en la subontología de las entidades y *Teherán* tiene la unidad conceptual %TEHERAN\_00 almacenada en el Onomástico, al tratarse de una entidad nombrada. Como puede observarse en las rutas conceptuales obtenidas de la base de conocimiento, el concepto \$DECEMBER\_00 está categorizado debajo del concepto básico +TIME\_00, mientras que %TEHERAN\_00 está conceptualizado como +PLACE\_00, como muestran los ejemplos (59) y (60) respectivamente.

---

<sup>26</sup> Cabe recordar que PDEP contiene ejemplos de tres corpus *SemEval Corpus*, *Oxford English Corpus* y *CPA Corpus of Prepositions* desarrollados a partir de *FrameNet*, *Oxford Dictionary* y *BNC* respectivamente.

59) **\$DECEMBER\_00**

+*TIME\_00* >> +*PERIOD\_00* >> +*MONTH\_00* >> *\$DECEMBER\_00*

60) **%TEHERAN\_00**

+*PLACE\_00* >> +*ARTIFICIAL\_AREA\_00* >> +*TOWN\_00* >>  
+*CITY\_00* >> *\$CAPITAL\_00* >> *%TEHERAN\_00*

De este modo, concluyo que para el CRITERIO 3, las preferencias de selección de los argumentos en los marcos predicativos de las relaciones se representarán por medio de conceptos de la Ontología de FunGramKB.

Tampoco se debe olvidar que las preferencias de selección de los argumentos de las preposiciones pueden utilizarse a veces para resolver ambigüedades léxicas de otras categorías gramaticales, p.ej. adjetivo, verbo y nombre. De hecho, Higinbotham (1990) presentó un algoritmo en pseudocódigo que incluye los métodos más característicos para la resolución de las ambigüedades léxicas en el marco de la traducción automática, donde las preferencias de selección se tienen en cuenta en la tercera fase:

SI la palabra ambigua ocurre en una locución ENTONCES

traduzca la frase entera con su significado idiomático

SI tras aplicar restricciones sintácticas, sólo queda un sentido ENTONCES

utilice la traducción apropiada para ese sentido

SI tras aplicar las preferencias de argumentos de verbos y preposiciones, sólo queda un sentido ENTONCES

utilice la traducción apropiada para ese sentido

SI la palabra ya ha aparecido en el texto (memoria discursiva) ENTONCES

utilice la misma traducción que usó en la ocasión anterior

SI tras eliminar todas aquellas palabras que no pertenecen al mismo dominio técnico, sólo queda un sentido ENTONCES

utilice la traducción apropiada para ese sentido

SI NO

aplique métodos de frecuencia y activación extensiva

FIN SI

En definitiva, este algoritmo sirve para ilustrar que las preferencias de selección introducidas por las preposiciones siempre han desempeñado un papel fundamental en el proceso de desambiguación léxica.

## **5.4 Modelización conceptual de la subontología de las relaciones**

El primer paso de la modelización conceptual de la subontología de las relaciones fue la creación del metaconcepto raíz #RELATION, bajo el cual se construyeron los metaconceptos #TIME, #LOCUS y #CIRCUMSTANCE. Posteriormente, se organizaron jerárquicamente estos tres metaconceptos, atendiendo a los resultados investigadores provenientes de trabajos en lingüística cognitiva y lingüística computacional sobre taxonomías preposicionales. Una vez establecidas las dimensiones cognitivas de las tres taxonomías, se procedió a incorporar los conceptos básicos y a asignar las unidades léxicas en inglés representativas de cada unidad conceptual. Por tanto, el resto de este apartado muestra en detalle los resultados investigadores de aplicar los CRITERIOS 1 y 2, mientras el apartado 5.5 de este capítulo se ocupa del CRITERIO 3.

### **5.4.1 La dimensión cognitiva temporal: #TIME**

Tal y como argumentaron O'Hara y Wiebe (2009, p. 151), "the senses of prepositions tend to be closely related to one another". Por este motivo, para la creación de la dimensión metaconceptual #TIME, se consideraron las sutilezas semánticas expresadas en los sentidos de las preposiciones consultadas tanto en los recursos lexicográficos como en los inventarios del PDEP y *Supersenses*. Consecuentemente, la

estructura metaconceptual de la subontología #TIME contiene una jerarquía profunda que permite reflejar esas sutilezas semánticas.

Tras revisar la literatura concerniente a las taxonomías de preposiciones temporales creadas para su aplicación computacional en la comprensión del lenguaje natural y que pueden proporcionar un adecuado marco para las distinciones aspectuales defendidas en la RRG, mi modelación ontológica se fundamenta principalmente en los estudios realizados por Brée y sus colegas (Brée, 1992; Brée & Pratt-Hartmann, 2002; Pratt & Brée, 1995), ya que su investigación se centró en un análisis que “is likely to be reflected in the use of temporals in all languages, or at least all European languages” (Brée, 1992, p. 347). A partir de estos trabajos, distribuí los diferentes metaconceptos dentro de la dimensión cognitiva #TIME teniendo en cuenta las propuestas de Reichenbach (1947) con respecto a las nociones temporales que facilitan una mejor comprensión de la semántica de los sintagmas preposicionales de tiempo, es decir:

- a) el momento en el que sucede el evento expresado por el predicado (TOE, del inglés *time of occurrence*),
- b) el momento a partir del cual TOE se lleva a cabo (TOR, del inglés *time of reference*), y
- c) la duración de tiempo introducida por el complemento de la preposición (*t*).

A partir de estos tres parámetros temporales, comencé a modelar la dimensión cognitiva #TIME. A continuación, se realizó una descripción detallada de los metaconceptos incluidos en esta dimensión junto con la asignación de los conceptos básicos, las unidades léxicas vinculadas a dichos conceptos básicos y de los ejemplos que sirven para ilustrar cada una de las configuraciones.

A este respecto, la dimensión metaconceptual #TIME se organiza en dos metaconceptos más específicos, i.e. #DURATION e #INTERVAL, que a su vez disfrutan de ramificaciones metaconceptuales más profundas que permiten la descripción conceptual de ciertos matices semánticos. En primer lugar, describo la organización interna de la dimensión cognitiva representada por #DURATION:

▪ #DURATION hace referencia a  $t$  cuando es considerado un momento que no contiene una representación explícita en la percepción mental del hablante. Este metaconcepto es a su vez distribuido en otros dos metaconceptos:

- #DURATION\_F1: donde TOE es igual a  $t$  (i.e.  $TOE = t$ ).
- #DURATION\_F2: donde TOE ocurre con respecto a TOR.

Dentro de #DURATION\_F1, las dimensiones metaconceptuales que se pueden encontrar son:

➤ #DURATION\_F1\_1: representa aquellas relaciones que expresan la duración de la acción con sentido atético. El concepto básico asignado a este metaconcepto es +FOR\_00, el cual se instancia lingüísticamente a través de la preposición *for*:

+FOR\_00

61) *This mixture is stable at room temperature for at least 8 weeks.*

En (61), la acción denotada por el evento del predicado supone que el momento en el que sucede la acción está expresado por la duración que indica el complemento de la preposición, es decir,  $TOE=t$ , y se combina con una predicación que resulta atética.

➤ #DURATION\_F1\_2: representa aquellas relaciones que expresan la duración de la acción con sentido télico, es decir, que aportan una delimitación temporal a la acción denotada por el evento de la predicación. Así, el concepto básico vinculado a este metaconcepto es +IN\_00, el cual se lexicaliza mediante la preposición *in*.

+IN\_00

62) *He walked two miles in eight days.*

Por consiguiente, el metaconcepto #DURATION\_F1 queda estructurado como muestra la Figura 24:

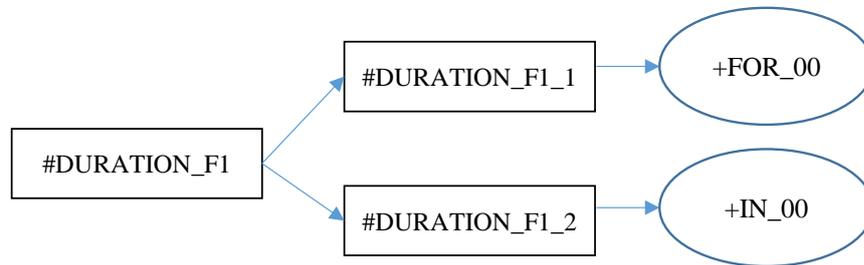


Figura 24. Taxonomía jerarquizada del metaconcepto #DURATION\_F1.

Dentro de #DURATION\_F2, se distribuyeron, asimismo, otras dos dimensiones metaconceptuales:

- #DURATION\_F2\_1: donde se ubican los conceptos que indican el espacio temporal en el que TOE ocurre en un momento específico tras TOR, es decir,  $TOE = TOR + t$ . El concepto básico introducido en esta dimensión cognitiva es +IN\_01, el cual se representa léxicamente en la preposición *in*.

+IN\_01

63) *I'll have some breakfast ready in a few minutes.*

- #DURATION\_F2\_2: contiene las relaciones cuyos sentidos implican que TOE sucede en un rango de tiempo comprendido entre TOR y  $t$  como fin. En otras palabras,  $TOE < TOR + t$ . El concepto básico vinculado a este metaconcepto es +WITHIN\_00, el cual puede lexicalizarse mediante las preposiciones *in* o *within*.

+WITHIN\_00

64) a. *They promised to come back in 60 days if nothing had been done to redress their grievances.*

b. *About 180,000 doses are expected within the next few days.*

De este modo, el nivel metaconceptual #DURATION\_F2 queda jerarquizado como muestra la Figura 25.

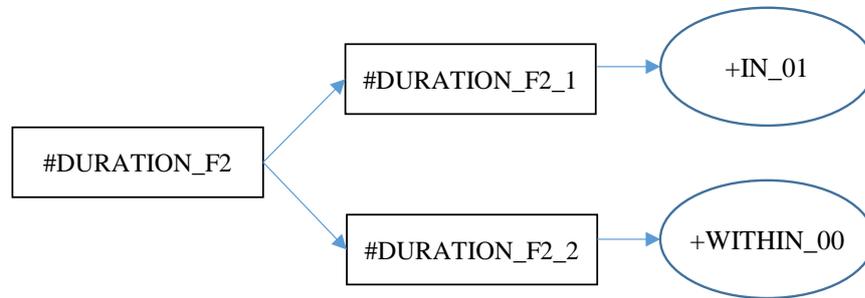


Figura 25. Taxonomía jerarquizada del metaconcepto #DURATION\_F2.

El segundo metaconcepto (junto a #DURATION) que configura la dimensión metaconceptual inmediatamente inferior a #TIME es #INTERVAL:

▪ #INTERVAL hace referencia a cuando  $t$  es un momento que sí tiene representación explícita en la mente del hablante. Del mismo modo que sucede con #DURATION, #INTERVAL se estructura en niveles metaconceptuales de mayor profundidad que permiten incluir conceptos básicos que aportan matices semánticos a la dimensión cognitiva a la que pertenecen. Así, #INTERVAL se divide en:

- #INTERVAL\_F1: donde TOE =  $t$ , es decir, donde la acción denotada por el evento ocurre en el tiempo establecido por  $t$ .
- #INTERVAL\_F2: donde TOE sucede en un rango de tiempo en el que TOR marca un inicio y  $t$  un fin.

Dentro de #INTERVAL\_F1, la jerarquía metaconceptual ha quedado distribuida en:

- #INTERVAL\_F1\_1: abarca aquellas relaciones que introducen a  $t$  como el punto específico temporal en el que sucede la acción. El concepto básico vinculado a este metaconcepto es +AT\_00, asociado a la unidad léxica *at*:

+AT\_00

65) *Clubs closed at 1am and it is understood local police liaised with managers before Saturday.*

➤ #INTERVAL\_F1\_2: las relaciones incluidas bajo este metaconcepto expresan a *t* como el periodo de tiempo en el que sucede la acción. Así, el concepto básico vinculado a este metaconcepto es +IN\_02, el cual se representa léxicamente en las preposiciones *in* y *during*.

+IN\_02

66) *My neighbours were moving furniture in the evening.*

67) *So to catch these spring beasts you obviously need a water that is open during the spring period.*

Asimismo, se ha creado el concepto básico +ON\_00, que es ubicado directamente bajo el metaconcepto #INTERVAL\_F1, ya que expresa la conceptualización de un momento temporal que reúne las características de +AT\_00 y +IN\_02, es decir, expresa un momento específico en el tiempo:

+ON\_00

68) *He expected Mr Major to turn up at another party on Thursday night.*

De este modo la distribución para la dimensión cognitiva #INTERVAL\_F1 queda como se observa en la Figura 26.

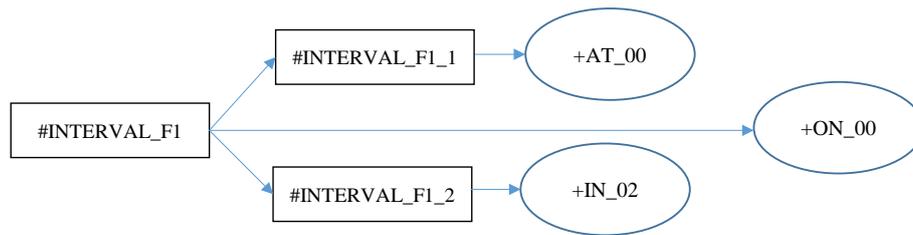


Figura 26. Taxonomía jerarquizada del metaconcepto #INTERVAL\_F1.

Por otra parte, el metaconcepto #INTERVAL\_F2 se organiza a su vez en tres metaconceptos:

- INTERVAL\_F2\_1: incluye aquellas relaciones que consideran a  $t$  como comienzo de TOR por defecto, además de que TOE tiene que durar todo  $t$ . El concepto básico vinculado a este metaconcepto es +UNTIL\_00, el cual se representa léxicamente en las preposiciones *until* y *till*.

+UNTIL\_00

69) *They serve food until 7 pm.*

70) *Why do we have to wait till the next central committee meeting?*

- INTERVAL\_F2\_2: incluye las relaciones que expresan a  $t$  como el momento inicial de TOR y donde TOE puede ocupar todo el tiempo indicado por  $t$  u ocurrir en algún momento dentro de  $t$ . Los conceptos básicos incluidos en esta dimensión cognitiva son +BETWEEN\_00 y +BY\_00, asignados a las unidades léxicas *between* y *by*, respectivamente:

+BETWEEN\_00

71) *This study took place between 1981 and 1985 in a large northern, metropolitan social services department<sup>27</sup>*

<sup>27</sup> Durante el proceso de modelización conceptual, el sentido descrito por +BETWEEN\_00 presentó cierta dificultad al necesitar dos entidades que delimiten el intervalo de tiempo en el que sucede TOE. Finalmente, se decidió que la primera entidad introducida por  $t$  establecería TOR y que TOE podría ocurrir a partir de TOR y el fin delimitado por la segunda entidad.

+BY\_00

72) *The work must be finished by May 1.*

- INTERVAL\_F2\_3: incluye las relaciones que expresan a *t* como el punto temporal en el que comienza el evento denotado por la predicación. Los conceptos básicos incluidos bajo este metaconcepto son +FROM\_00 y +SINCE\_00, lexicalizados en las preposiciones *from* y *since* respectivamente:

+FROM\_00

73) *Does it really matter about people being there from 8 in the morning?*

+SINCE\_00

74) *Arnold has been a member of parliament since 1987.*

Por consiguiente, la dimensión conceptual #INTERVAL\_F2 queda distribuida como muestra la Figura 27.

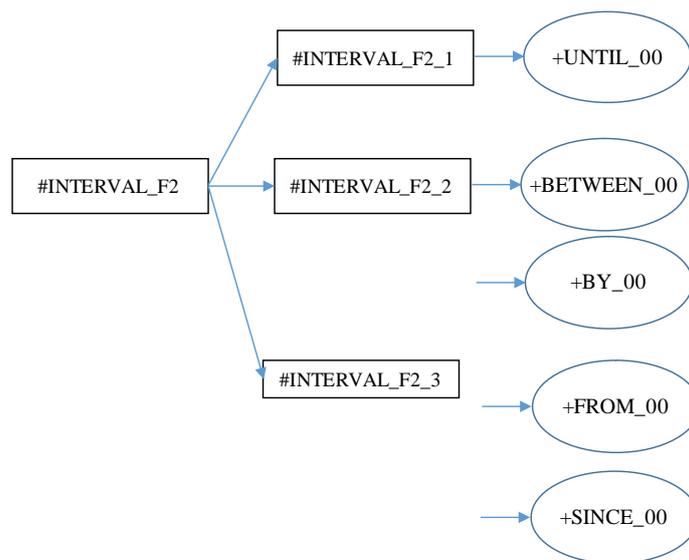


Figura 27. Taxonomía jerarquizada del metaconcepto #INTERVAL\_F2.

A modo de resumen, la Figura 28 muestra la organización taxonómica completa de la dimensión cognitiva #TIME dentro de la subontología de relaciones creada en FunGramKB. Asimismo, en el Anexo V se puede consultar un resumen de los metaconceptos, conceptos básicos y unidades léxicas creadas para la taxonomía #TIME.

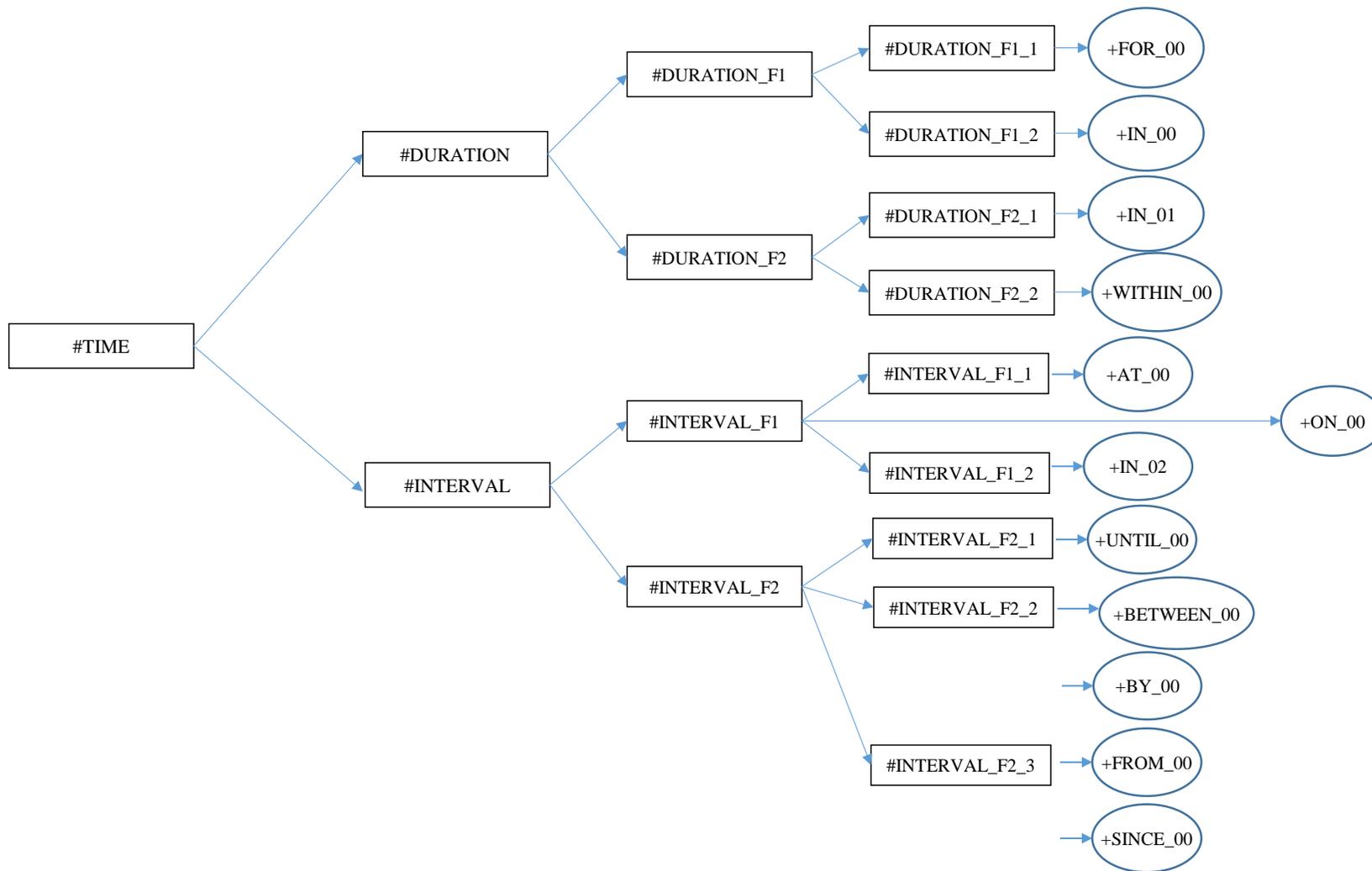


Figura 28. Taxonomía de relaciones #TIME

### 5.4.2 La dimensión cognitiva espacial: #LOCUS

En la literatura sobre las preposiciones espaciales, diversos estudios (cf. Clark, 1973; Bennett, 1975; Jackendoff, 1983, 1990; entre otros muchos) han realizado simplemente una distinción entre preposiciones locativas y preposiciones direccionales. En el caso de FunGramKB, la subontología de las relaciones de la dimensión espacial ha sido construida fundamentalmente teniendo en cuenta dos aproximaciones sobre la semántica de estos sintagmas preposicionales. Más concretamente, se integró el enfoque de Jackendoff (1983), quien propuso el diseño de una taxonomía apoyada en los ejes de lugar (*place*) y trayectoria (*path*) con el enfoque de Zhang (1996), quien hizo una descripción de la composición semántica de los sintagmas preposicionales a través de tres elementos semánticos: la relación locativa, la entidad referencial y el valor locativo. A continuación, procedo a comentar brevemente ambos estudios con el fin de justificar la estructura de la subontología de las relaciones espaciales que aquí se presenta.

Por una parte, Jackendoff (1983) defendía la distinción tradicional entre lugar y trayectoria. Según este autor, esta distinción conceptual tiene su justificación en el modo en que, por ejemplo, lenguas como el alemán, a través de los casos dativo y acusativo, o el húngaro, mediante postposiciones que añaden un sufijo, lexicalizan dichos conceptos. Por esta razón, trató de explicar cómo los sintagmas preposicionales podían corresponder a conceptos locativos diferentes al considerar la composición semántico-funcional de la estructura conceptual de un sintagma preposicional atendiendo al complemento de la preposición. En relación a aquellos conceptos categorizados como lugar, Jackendoff manifestó su percepción de que la naturaleza del complemento de la preposición u objeto de referencia (como él lo denominó) está regido por ciertas limitaciones conceptuales en la correspondiente preposición. De este modo, afirmó que la preposición *in* rige un complemento que represente conceptualmente un área o volumen delimitado, como en *the mouse lives in the box*, mientras que la preposición *on* necesita que su complemento permita conceptualizar una superficie exterior, como en el ejemplo *the book is on the table*. Asimismo, indicó que los sintagmas preposicionales formados por este tipo de preposiciones pueden aparecer dentro de la estructura de un verbo (75a) o pueden expresar la localización del evento de una cláusula (75b):

75) a. *The mouse lives in the box.*

b. *In London, we lived in an apartment.*

Sin embargo, para este autor, los sintagmas preposicionales formados por conceptos categorizados en el tipo trayectoria intervienen en la estructura de los eventos con un mayor número de interpretaciones, a la vez que suelen tener una estructura interna más variada. Así, indicó que la típica estructura interna de un sintagma preposicional de trayectoria puede ser simple, al estar formada por una preposición cuya función sea trayectoria y un objeto de referencia como en (76), o compleja, si está formada por una preposición de trayectoria y un sintagma preposicional de lugar como en (77):

76) *They were walking towards the mountain.*

77) *The mouse ran from under the table.*

De este modo, para el sintagma *from under the table*, la estructura conceptual propuesta por Jackendoff (1983, p. 163) es la siguiente:

78) [Path FROM ([Place UNDER ([Thing TABLE]))]]

De acuerdo con estos argumentos, el autor propuso una distribución taxonómica de relaciones que pueden expresar funciones de lugar o de trayectoria. Además, Jackendoff organizó la categoría de trayectoria en tres tipos de conceptos: trayectorias delimitadas (*bounded paths*), direcciones (*directions*) y rutas (*routes*). La categoría de las trayectorias delimitadas contiene aquellos conceptos que expresan el origen (*source*) o destino (*goal*) de una trayectoria. La lexicalización de estos sentidos se lleva a cabo a través de las preposiciones *from* y *to* respectivamente. La categoría de las direcciones abarca el concepto de aquellas preposiciones cuyo complemento determina la dirección hacia o desde la cual se articula la acción del verbo. Las típicas preposiciones dentro de esta categoría son *towards* y *away from*. Finalmente, el tercer grupo incluido dentro de trayectoria, i.e. rutas, incluye conceptos cuyas preposiciones determinan un modo para la trayectoria de acuerdo con el complemento de la preposición. Dentro de esta categoría se encuentran las preposiciones *by*, *along* o *through*.

Por otra parte, el segundo estudio que ha resultado relevante para la construcción de la subontología espacial de las relaciones en FunGramKB es el análisis de Zhang (1996) sobre la composición semántica de los sintagmas preposicionales, quien concluye que un sintagma preposicional de lugar proporciona información relativa a tres elementos. En primer lugar, establece la relación locativa (LR, del inglés *Locative Relation*), que define la categoría semántica resultante de la relación entre una entidad X y un lugar:

79) X PP → *There is an apple on the shelf*

X: *apple*

PP: *on the shelf*

Para esta categoría, distingue cuatro elementos primitivos: posición (*position*), origen (*source*), destino (*goal*) y ruta (*route*). Zhang definió estos elementos de la siguiente manera (1996, p. 145):

80) In X [LR Z [PLACE [PV][RE Y]]]

If Z is POSITION, it shows that the PLACE is the position of X;

if Z is GOAL, it shows that the PLACE is the terminus of X;

if Z is SOURCE it shows that the PLACE is the starting point of X;

if Z is ROUTE it shows that the PLACE is the space intervening from SOURCE to GOAL taken by X.

El segundo elemento considerado por Zhang es la entidad referencial (RE, del inglés *Referent Entity*), la cual está representada por el complemento de la preposición:

81) RE: *the shelf*

Al igual que sucede con la categoría semántica de LR, el concepto de RE corresponde con la noción expresada por otros autores (Jackendoff, 1983; Talmy, 1985; Langacker, 1986; Lakoff, 1987) con respecto a la entidad dentro del marco de la preposición.<sup>28</sup>

---

<sup>28</sup> Véase apartado 2.2.

Por último, esta autora afirmó que la relación entre la preposición y RE debe categorizarse conforme a un valor locativo (PV, del inglés *Place Value*). De este modo, basándose en el análisis semántico realizado por Bennett (1975), propuso la clasificación de elementos primitivos que se presenta en la Tabla 6 (Zhang, 1996) para la representación de los diferentes tipos de PV:

Tabla 6. Lista de elementos primitivos de PV para algunas preposiciones inglesas según Zhang.

HIGHER	WARDS	LOWER	SURFACE	TRANSVERSE
INTERIOR	SURROUND	OFF	PROXIMITY	SUPERIOR
EXTERIOR	ANTERIOR	LENGTH	SPACE BELOW	SPACE BEHIND

Zhang consideró, de forma similar a Jackendoff (1983), que pueden existir estructuras complejas de sintagmas preposicionales donde aparezcan preposiciones que impliquen una estructura conceptual de lugar y trayectoria a la vez. No obstante, afirmó que las lenguas contienen ciertas limitaciones en este sentido y que un sintagma preposicional sólo puede expresar un único tipo de LR y PV. La autora consideró como marcadores locativos puros a las preposiciones *at*, *from*, *to* y *via*, a las que además no asignó ningún PV. Por esta razón, descartó la posibilidad de que dos preposiciones que expresen relaciones locativas distintas puedan aparecer conjuntamente en la misma oración. Por consiguiente, cuando una preposición como *behind*, cuya LR es POSITION y su PV es SPACE BEHIND, aparece en un mismo sintagma con una preposición locativa pura, p.ej. *from*, cuya LR es SOURCE, la estructura conceptual que prevalece en el sintagma es la correspondiente a *from*, por considerarse un marcador locativo puro:

82) *Trevor appeared from behind the desk.*

83) [LR SOURCE[PLACE[PV SPACE BEHIND]][RE desk]]

(Zhang, 1996, p.146)

De igual manera, un sintagma preposicional sólo puede contener un único LV, impidiendo así que dos preposiciones con LV diferente puedan aparecer juntas en un mismo sintagma preposicional (Zhang, 1996, p.144):

84) *\*In on the table*

Por este motivo, la organización conceptual de la subontología espacial está determinada por una distribución metaconceptual que permita identificar la LR de la relación con respecto a la limitación de la función locativa y el PV, siguiendo las indicaciones de Zhang. Además de Jackendoff (1983) y Zhang (1996), otros autores como Lakoff (1987), Coventry y Garrod (2004), Evans y Green (2006) o Lindstromberg (2010) afirmaron que, en ocasiones, una misma preposición puede ser usada para expresar lugar o trayectoria. Por ejemplo:

85) a. The picture is over the sofa. [ABOVE]

b. The ball landed over the wall. [ON THE OTHER SIDE]

c. The car drove over the bridge. [ACROSS]

(Evans & Green, 2006, p.329)

En definitiva, a partir de la información obtenida en el análisis de los estudios de Jackendoff (1983) y Zhang (1996), comencé a modelar la dimensión cognitiva #LOCUS, la cual se estructuró en otras dos dimensiones cognitivas: #PLACE y #PATH. En el resto de este apartado, realizo una descripción detallada de los metaconceptos incluidos en esta taxonomía y la asignación de los diferentes conceptos básicos, que servirán para crear las entradas conceptuales de las relaciones espaciales.

Por una parte, la dimensión cognitiva #PLACE ha sido estructurada a partir de una serie de metaconceptos que en su mayoría toman su nombre de los valores locativos expuestos por Zhang (1996). Como establecieron Jiménez-Briones, Luzondo y Pérez-Cabello de Alba (2011, p.20), cada metaconcepto en FunGramKB “contiene un número determinado de participantes prototípicos a los que se les asigna un papel temático concreto que se define de forma diferente dependiendo del metaconcepto al que pertenezcan”. Por esta razón, la jerarquía metaconceptual de #PLACE contiene un número de doce metaconceptos que permiten vincular conceptos básicos relacionados con la descripción de los valores locativos de Zhang. Para ello, ha sido preciso realizar modificaciones que permitieran una distribución más homogénea de los metaconceptos con el fin de evitar redundancia en las descripciones. Así, se han creado los metaconceptos #OFF\_LIMITS e #INTERLOCAL, se han aglutinado HIGHER y

SUPERIOR en el metaconcepto #SUPERIORITY, se ha modificado LOWER por #INFERIORITY y se ha eliminado el valor locativo OFF. Igualmente, los valores locativos TRANSVERSE Y WARDS se han eliminado con el fin de crear conceptos básicos que aparecerán bajo los metaconceptos #ROUTE y #DIRECTION respectivamente. La Figura 29 muestra la taxonomía metaconceptual resultante de la dimensión cognitiva #PLACE.



Figura 29. Taxonomía metaconceptual de la dimensión cognitiva #PLACE

#EXTERIOR expresa la localización de una entidad que se encuentra situada en una posición distanciada y exterior con respecto a la entidad denotada por el argumento interno de la preposición, al que denomino entidad referenciada (ER). El concepto básico vinculado a este metaconcepto es +OUT\_00, el cual es lexicalizado en las preposiciones *out (of)* y *outside*:

+OUT\_00

86) *But it's because of the Church, of course, that people like Joyce and, for a great deal of his time, Yeats, had to live out of Ireland.*

87) *This venue provides substantial meals outside our price range.*

#INFERIORITY hace referencia a una posición inferior, en un eje vertical, en la que se encuentra una entidad con respecto a la ER. El concepto básico creado para esta dimensión cognitiva es +BELOW\_00. Las unidades léxicas asociadas a +BELOW\_00 son *below*, *beneath*, *under* y *underneath*, las cuales pueden considerarse como cuasisinónimos desde un punto de vista lingüístico, ya que, a pesar de la sutileza semántica que en ocasiones pueden transmitir, son intercambiables en la mayoría de casos:

+BELOW\_00

88) a. *Sea level was 120 metres below the present level during the Ice Age.*

b. *The bandit stopped directly beneath the tree, and looked around.*

c. *Additional lighting was also installed under the platform.*

d. *Each book in the pile depends upon the one underneath it to hold it up.*

#INTERIOR expresa la localización de una entidad que se encuentra dentro de los límites del concepto expresado por la ER. Los conceptos básicos creados para este metaconcepto son +AT\_01, representado por la preposición *at*, y +IN\_03, lexicalizado en las preposiciones *in* e *inside*:

+AT\_01

89) *The Ulysses probe will also feature at Expo.*

+IN\_03

90) a. *But there was a sound of boots in the passage outside and my door was unlocked.*

b. *Lights are on inside the house but no one can hear her silent cries for help*

#INTERLOCAL expresa la posición central de una entidad que se sitúa en el espacio intermedio con respecto a un conjunto expresado por la ER. Los conceptos básicos vinculados a este metaconcepto son +AMONG\_00 y +BETWEEN\_01, cuyas unidades léxicas son *among* para el primero y *between* para el segundo:

+AMONG\_00

91) *Mozart immediately found himself among friends, especially Cannabich (...)*

+BETWEEN\_01

92) *Does the report have a translucent blank page between the front cover and the first page, or a separate `dust jacket`?*

#LENGTH expresa la posición de una entidad que se encuentra distribuida con respecto a la ER. Los conceptos básicos creados son +ALONG\_00, el cual está asociado a la preposición *along*, y +FOR\_01, representado léxicamente por la preposición *for*:

+ALONG\_00

93) *People travelled to and from the town and the villages along the river.*

+FOR\_01

94) *I only used it for a few hundred yards, perilously close to a stretch of canal.*

#OFF\_LIMITS expresa la posición de una entidad que se localiza a una distancia más allá del límite establecido por la ER. Esta dimensión incluye el concepto básico +BEYOND\_00, el cual está asociado a la unidad léxica *beyond*:

+BEYOND\_00

- 95) *The greatest advantage of infrared astronomy, however, comes when looking at objects well beyond the Solar System.*

#PROXIMITY expresa el lugar de una entidad que se localiza a una distancia próxima de otra entidad formulada por la ER. Existe un único concepto básico en este metaconcepto, i.e. +NEAR\_00, el cual puede representarse léxicamente en las preposiciones *near*, *next to*, *by* o *beside* en inglés, que, al igual que sucede en el concepto básico +BELOW\_00, estas preposiciones pueden ser consideradas cuasisinónimos:

+NEAR\_00

- 96) a. *Six of my men were killed in one spot near me.*  
b. *At one point outside the pavilion I was sat next to five other players in our side who were all aged under 18.*  
c. *Just sit by Heather.*  
d. *He rode beside you in the street car and complained about the traffic and the weather with you.*

#SPACE\_BEHIND expresa la posición de una entidad que se encuentra exactamente detrás de otra. El concepto básico para este metaconcepto es +BEHIND\_00, lexicalizado en la preposición *behind*:

+BEHIND\_00

- 97) *Behind the living chamber the rest of the shell is divided into smaller chambers.*

El metaconcepto #SPACE\_FRONT expresa la posición de una entidad que se encuentra exactamente delante de otra. El concepto básico creado es +IN\_FRONT\_OF\_00, asociado a las unidades léxicas *in front of* y *ahead of*:

+FRONT\_00

98) a. *This would allow residents to use the areas in front of their houses for parking.*

b. *But they believed they were protected by a front line of fighters ahead of them.*

El metaconcepto #SUPERIORITY hace referencia a una posición superior, en un eje vertical, en la que se encuentra una entidad con respecto de la ER. El concepto básico creado para esta dimensión cognitiva es +ABOVE\_00, el cual se lexicaliza en las preposiciones *above* y *over*. Al igual que sucede en otras ocasiones, las unidades léxicas pueden ser consideradas cuasisinónimos:

+ABOVE\_00

99) a. *He was only 100 or so feet above the sea and we closed on him rapidly.*

b. *The arms positioned alongside the body turn in at the elbows with the hands over the groin.*

#SURFACE expresa la posición de una entidad que se encuentra en contacto con la superficie de la entidad conceptualizada en la ER. En esta dimensión cognitiva, el concepto básico asignado es +ON\_01, el cual está asociado a las preposiciones *on*, *upon*, *atop* y *on top of*:

+ON\_01

100) a. *The bed height should allow the patient to touch the floor with his feet when he sits on the side on the bed.*

- b. *Two attachment points were defined in the reference space, upon which the two main arms were positioned.*
- c. *It is sited unusually, on flat land rather than atop one of the encircling hills.*
- d. *There were two GIs in charge -- the driver and a sentry guard -- and Eric had to sit on top of a pile of waterproof mailbags.*

Finalmente, el último metaconcepto de esta categoría es #SURROUND, el cual expresa la posición de una entidad que rodea a otra entidad. El concepto básico de este nivel es +AROUND\_00. La lexicalización de este concepto está representada por las preposiciones *around* y *round* en inglés:

+AROUND\_00

101) *Then we squeezed round the table.*

Por otra parte, el metaconcepto #PATH indica la noción de trayectoria. Teniendo en cuenta los estudios comentados anteriormente, se ha seguido la distinción hecha por Jackendoff (1983), de modo que esta dimensión conceptual está a su vez distribuida en cuatro metaconceptos (i.e. #DIRECTION, #GOAL, #ROUTE y #SOURCE), tal y como se muestra en la Figura 30.

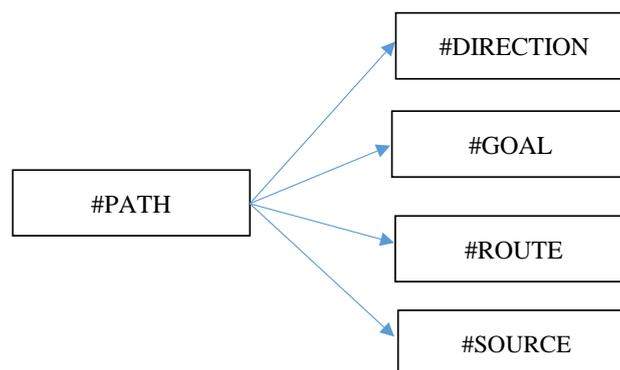


Figura 30. Distribución metaconceptual de la dimensión cognitiva #PATH.

#DIRECTION hace referencia a aquellas relaciones que expresen dirección con respecto de la ER a la que acompañan. Así, vinculados a este metaconcepto, se crearon

los conceptos básicos +AWAY\_FROM\_00 y +TOWARDS\_00, lexicalizados en las preposiciones *away from* y *towards* y *for* respectivamente:

+AWAY\_FROM\_00

102) *I need to get away from this house.*

+TOWARDS\_00

103) a. *At gunpoint, the woman walked slowly towards the main door.*

b. *As we left the stadium for the airport, bottles and bricks smacked against our coach*

#GOAL expresa un punto en el espacio en el que un movimiento o acción comienza. Este metaconcepto tiene vinculado al concepto básico +TO\_00, el cual es representado léxicamente por la preposición *to*:

+TO\_00

104) *John pushed the cart a mile to the edge of the park in an hour.*

El metaconcepto #ROUTE indica el modo en el que transcurre la trayectoria. Aquí se encuentran los conceptos +ACROSS\_00, +ALONG\_00, +AROUND\_00, +BY\_01, +OVER\_00 y +THROUGH\_00. Las unidades léxicas asignadas a cada uno de los conceptos básicos son las siguientes:

+ACROSS\_00: across

105) *As the bombs in the hangars started to go off they beat a hasty retreat across the field.*

+ALONG\_01: along

106) *It snakes in and out of ports, along our busiest highways and through our most crowded cities.*

+AROUND\_01: around y round

107) *Mrs M. darted around the waters with great vivacity.*

+BY\_01: by

108) *I usually go by NY when I fly to Europe*

+OVER\_00: over

109) *The UN Security Council was planning to vote last night to adopt a resolution enforcing its ban on flights over Bosnia.*

+THROUGH\_00: through y via

110) *It can only be dried out by diffusing it slowly through the tube walls.*

111) *We went via Kings Cross and I got a taxi across to Baker Street for Hertfordshire.*

Finalmente, el último metaconcepto de esta categoría es #SOURCE, el cual expresa un punto en el espacio para el comienzo de una trayectoria. El concepto básico incluido bajo este metaconcepto es +FROM\_01, que se manifiesta en la unidad léxica *from*:

+FROM\_01

112) *From Valencia, El Cid began to show his true strength, gradually extending his sway until he was virtual ruler of the great part of eastern Spain.*

En las Figuras 31, 32 y 33, se puede observar la distribución conceptual de la taxonomía #LOCUS. Asimismo, los anexos VI y VII muestran las taxonomías jerarquizadas de #PLACE y #PATH, al igual que las unidades léxicas en inglés asignadas a los conceptos básicos de cada metaconcepto.

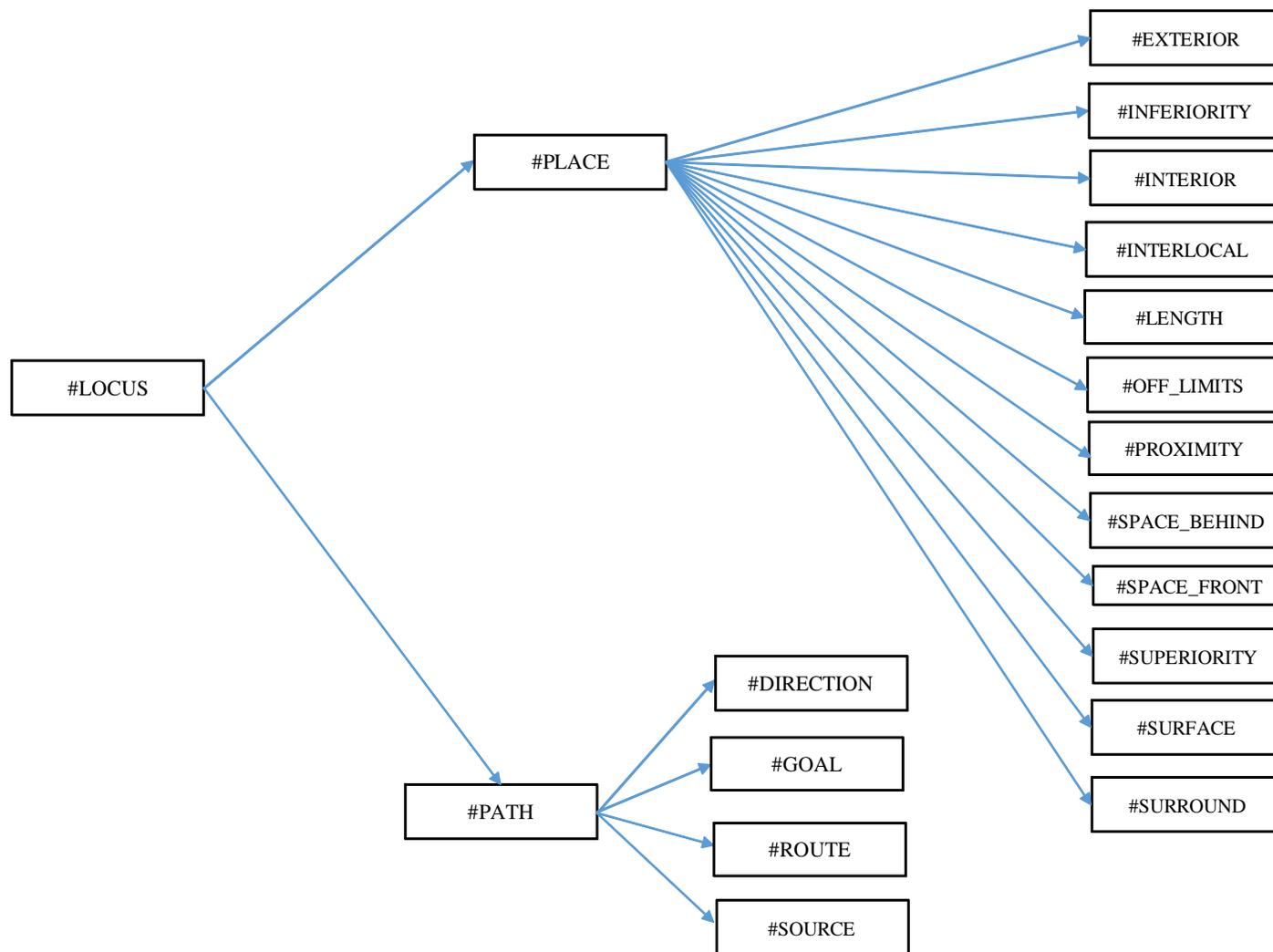


Figura 31. Distribución metaconceptual de la taxonomía #LOCUS

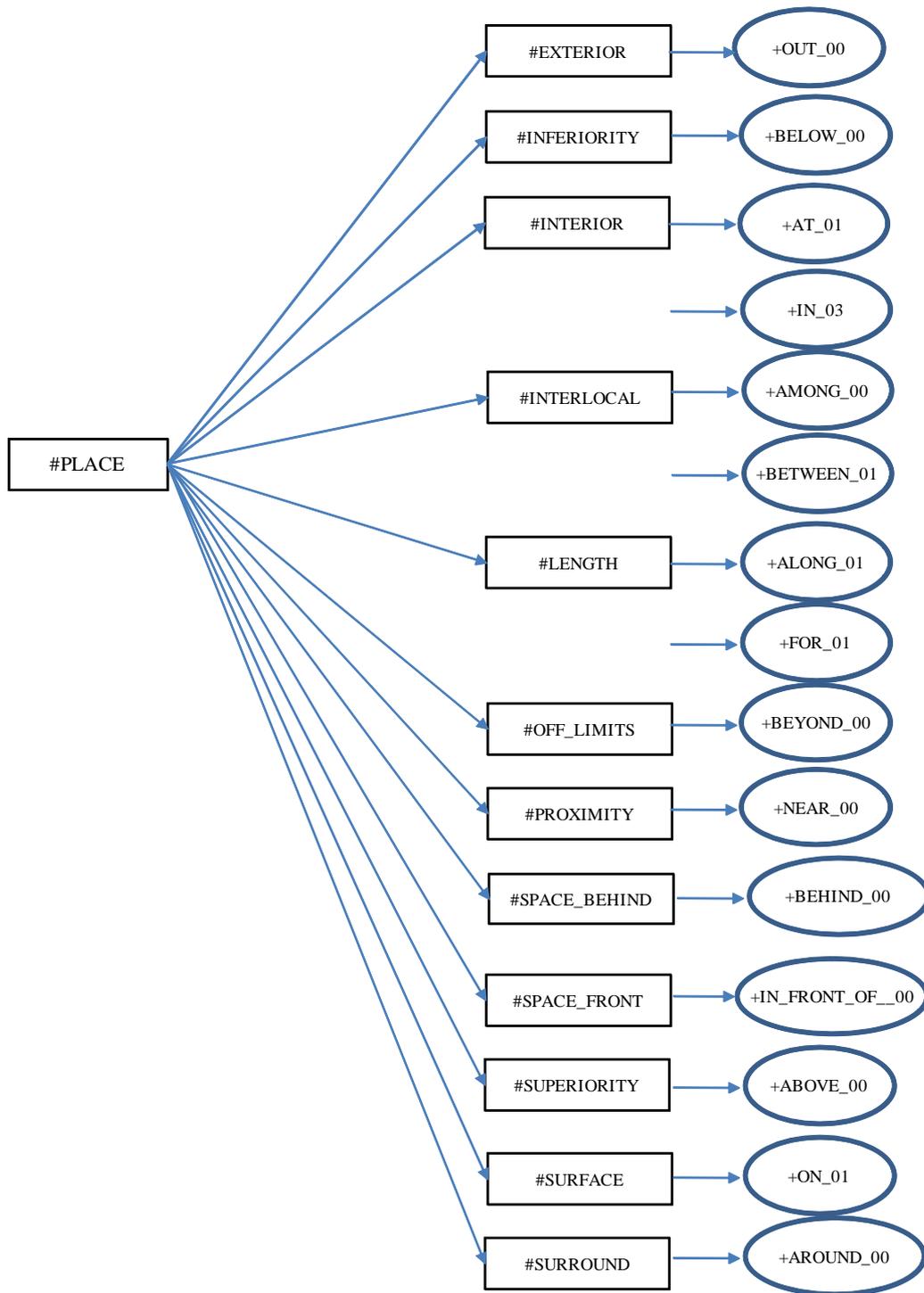


Figura 32. Distribución de los niveles conceptuales en la dimensión cognitiva #PLACE.

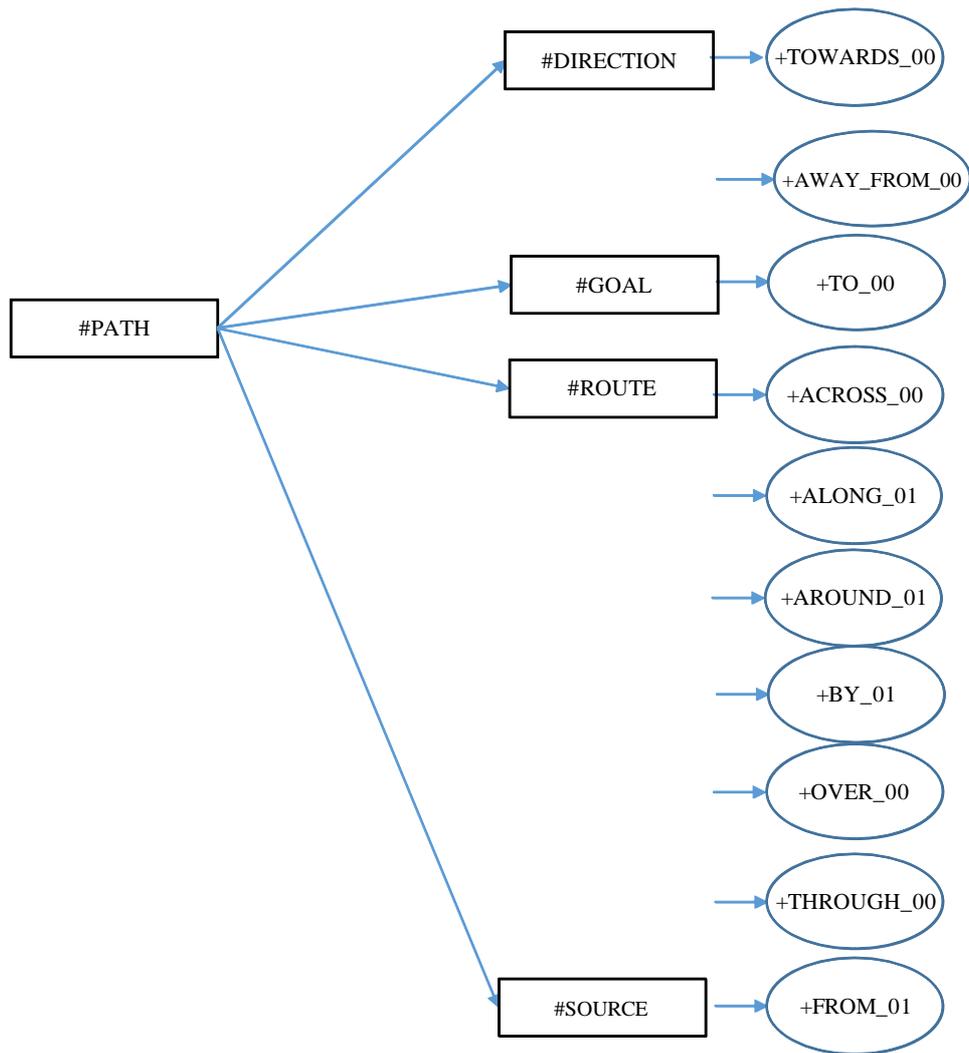


Figura 33. Distribución de los niveles conceptuales en la dimensión cognitiva #PATH

### 5.4.3 La dimensión cognitiva circunstancial: #CIRCUMSTANCE

Para la creación de la taxonomía #CIRCUMSTANCE, se consultaron estudios recientes en relación a inventarios semánticos que pretenden dar respuesta a la asignación de papeles temáticos en las relaciones expresadas por las preposiciones. Por ejemplo, Schneider (2014) y Schneider et al. (2015) propusieron un enfoque para el tratamiento de la semántica de las preposiciones basado en la descripción lingüística, creando un inventario que permitiera una mejor integración con otros recursos léxicos. Para ello, su análisis se centró en la etiquetación de todos los sintagmas preposicionales presentes en diferentes corpus<sup>29</sup>, formados por una o más preposiciones, permitiendo así la construcción de la taxonomía *Supersenses*. Para las etiquetas que caracterizan a las preposiciones, tomó como punto de partida el inventario de relaciones propuesto por Srikumar y Roth (2013a, 2013b). Estos autores propusieron un inventario de 32 clases de relaciones siguiendo los sentidos anotados en TPP para las 34 preposiciones usadas en *SemEval-2007*. Su objetivo fue la especificación de las relaciones semánticas expresadas por las preposiciones transitivas y el desarrollo de un sistema computacional que permitiera la predicción automática de dichas relaciones y la identificación de los argumentos de las mismas y el tipo al que pertenecían. Para la creación del inventario de etiquetas, su metodología consistió en agrupar los sentidos de las preposiciones y, a partir de los ejemplos extraídos de TPP, modelar el sintagma preposicional (i.e. la relación de la preposición y sus argumentos). De este modo, su inventario está formado por las relaciones presentadas en la Figura 34 (Srikumar & Roth, 2013b, p.234).

---

<sup>29</sup>Véase la nota 13.

Relation Name	Example
Activity	good at boxing
Agent	opened by Annie
Attribute	walls of stone
Beneficiary	fight for Napoleon
Cause	died of cancer
Co-Participants	pick one among these
Destination	leaving for London
Direction	drove towards the border
EndState	driven to tears
Experiencer	warm towards her
Instrument	cut with a knife
Journey	travel by road
Location	living in London
Manner	scream like an animal
MediumOfCommunication	new show on TV
Numeric	increase by 10%
ObjectOfVerb	murder of the boys
Opponent/Contrast	fight with him
Other	<i>all others</i>
Participant/Accompanier	steak with wine
PartWhole	member of gang
PhysicalSupport	lean against the wall
Possessor	son of a friend
ProfessionalAspect	works in publishing
Purpose	tools for making it
Recipient	unkind to her
Separation	ousted from power
Source	purchased from the shop
Species	city of Prague
StartState	recover from illness
Temporal	arrived on Monday
Topic	books on Shakespeare

Figura 34. Lista de relaciones de Srikumar y Roth.

A partir de este estudio, Schneider (2014) y Schneider et al. (2015) presentaron la construcción de la taxonomía *SuperSenses*. En ella, la lista de relaciones de Srikumar y Roth (2013a, 2013b) es revisada y extendida tras su comparación con otras taxonomías, como *VerbNet* (Kipper et al., 2008) y *AMR* (Banarescu et al., 2013). El objetivo era mejorar el nivel descriptivo, incluyendo categorías que no habían sido introducidas por Srikumar y Roth y aplicando una taxonomía jerarquizada como la de *VerbNet* para los nombres y los verbos. El resultado es una extensa taxonomía que contiene la descripción de 865 sentidos clasificados en 63 categorías. La necesidad de que exista un gran número de etiquetas radica en el hecho de que las preposiciones

pueden ser lexicalizadas o gramaticalizadas y, por ende, pueden expresar más dimensiones cognitivas que los nombres o verbos (Schneider, 2014). Asimismo, argumentan que este tipo de taxonomía guía a los lingüistas en el complejo proceso de etiquetación, cuando encuentran dificultades en el momento de asignar las relaciones semánticas a las preposiciones. Como se explicó en el apartado 5.2, existe un debate en relación a la granularidad con la que debe realizarse la descripción semántica de las categorías conceptuales utilizadas por los sistemas del PLN. En lo que se refiere a la descripción semántica de las dimensiones cognitivas de las relaciones, esta tesis aboga por un enfoque menos granular que en las descripciones lingüísticas pero más preciso que el de la mayoría de sistemas de PLN, permitiendo así obtener beneficios de las ventajas ofrecidas por la perspectiva de la semántica profunda comentadas en el capítulo 4. Además, se debe recordar que la ontología #RELATION trata únicamente aquellas relaciones expresadas por preposiciones transitivas y lexicalizadas, dejando el resto de casos a otros módulos de FunGramKB.

Aunque la intención inicial fue crear un inventario de metaconceptos para la taxonomía #CIRCUMSTANCE con el fin de satisfacer la asignación de los papeles temáticos establecidos para los satélites en FunGramKB, exceptuando lógicamente los papeles vinculados a la temporalidad y a la espacialidad, pronto observé que, tras el estudio de la información almacenada en PDEP y en *SuperSenses*, era preciso incorporar diversos metaconceptos que permitieran la descripción de nuevas dimensiones cognitivas. Asimismo, a pesar de que FunGramKB ya contaba con un inventario de papeles temáticos para los satélites que intervenían en los postulados de significado, propongo añadir nuevos papeles como consecuencia de los nuevos metaconceptos para #CIRCUMSTANCE. Por ejemplo, uno de los papeles temáticos que propongo incorporar es *Topic*, que expresa la entidad que describe el tema o asunto con el que trata el evento.

Como resultado, los metaconceptos construidos para la taxonomía #CIRCUMSTANCE son los que muestra la Figura 35.

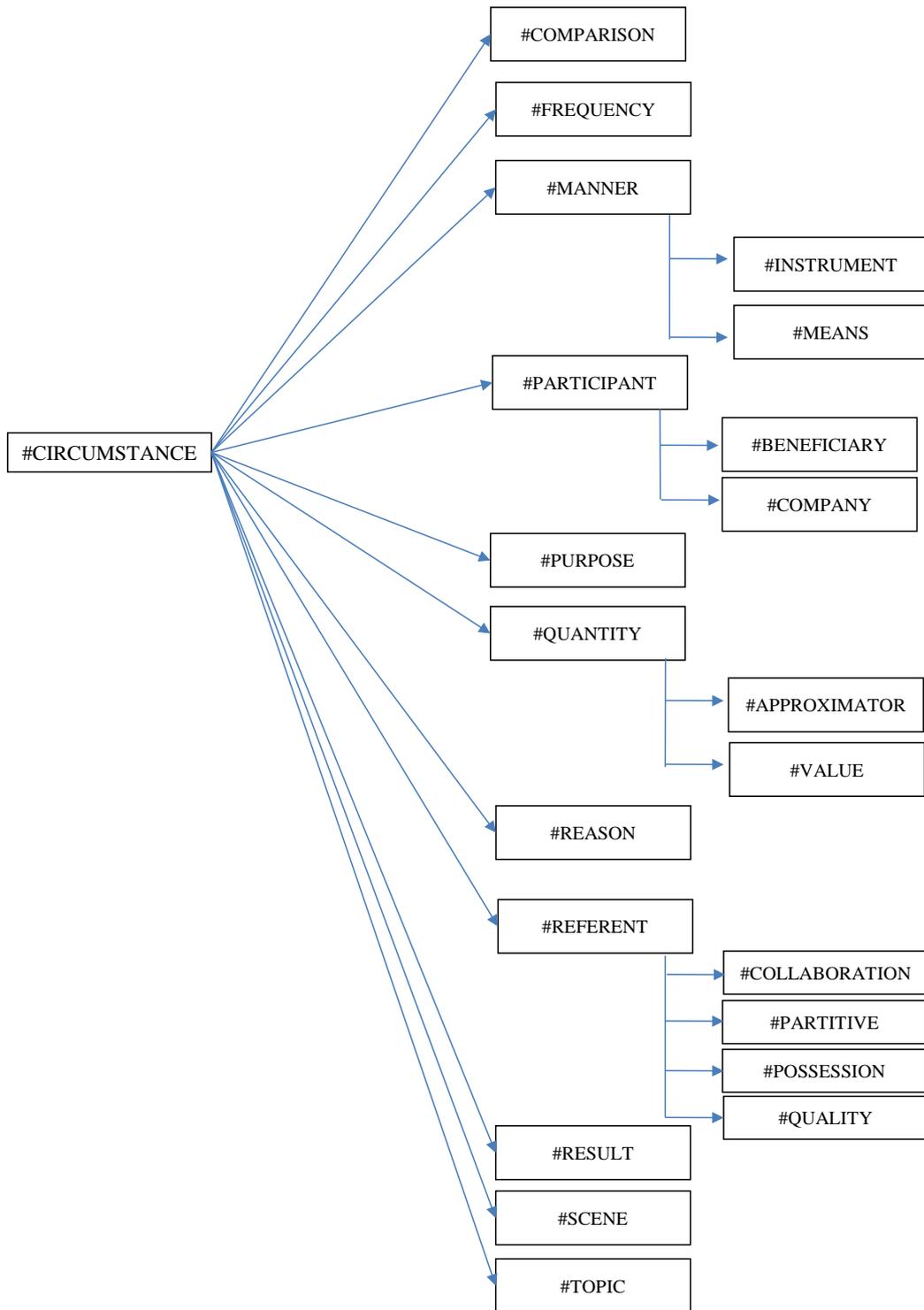


Figura 35. Distribución de los niveles metaconceptuales de la taxonomía #CIRCUMSTANCE.

El metaconcepto #COMPARISON describe una de las entidades que participa en la comparación entre dos entidades. Los conceptos básicos incluidos en este nivel

son +BESIDE\_00, +FROM\_02 y +TO\_01, lexicalizados en las preposiciones *beside*, *from* y *to* respectivamente:

+BESIDE\_00

113) *They might just feel stupid beside those who are much better than themselves.*

+FROM\_02

114) *Writers often differ quite widely from each other over ethical issues and questions.*

+TO\_01

115) *The club's nothing to what it once was.*

El metaconcepto #FREQUENCY representa el grado de frecuencia con la que sucede la acción expresada por el evento. Esta dimensión ha sido incluida dentro de esta taxonomía a pesar de que la relación cognitiva que describe podría incorporarse a la taxonomía #TIME. Sin embargo, semánticamente no es posible describirlos a través de los parámetros usados en #TIME (i.e. TOE, TOR y *t*). El concepto básico creado en este nivel es +FOR\_02, que se ejemplifica en la unidad léxica *for*:

+FOR\_02

116) *So for the third year running here's the low-down on Melas happening in your area.*

La siguiente dimensión cognitiva diseñada es #MANNER, la cual describe el modo en el que se produce el evento. Los conceptos básicos que se han incorporado a esta dimensión son +IN\_04, +WITH\_00 y +UNDER\_00, los cuales se instancian léxicamente como *in*, *with* y *under* respectivamente:

+IN\_04

117) *If at any point they lose the scent they fly in zig-zags.*

+WITH\_00

118) *The people shouted with pleasure.*

+UNDER\_00

119) *He said that management had done its best to retain a viable share of the market under difficult trading conditions.*

Tras el estudio de las preposiciones, se llegó a la conclusión de que los sentidos que expresan instrumento o medio están íntimamente relacionados con aquellos sentidos que expresan modo. Por esta razón, se decidió estructurar el metaconcepto #MANNER con los metaconceptos #INSTRUMENT y #MEANS, pues consideré que la expresión de modo puede ser matizada semánticamente por el instrumento o por el medio que se utilice. Ha de clarificarse que, según la descripción de los papeles temáticos de los satélites en FunGramKB, el papel temático MEANS hace referencia a una entidad que junto con el instrumento se usa para realizar la acción. Por ejemplo, en la oración *he built the bust in wood with an adze*, la preposición *in* introduciría el medio y la preposición *with* introduciría el instrumento. Por consiguiente, la estructura de este nivel queda como muestra la Figura 36.

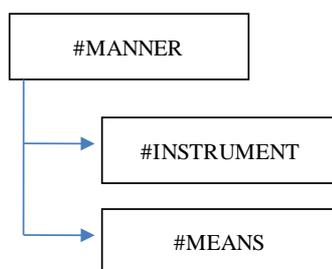


Figura 36. Estructura del nivel metaconceptual #MANNER

El metaconcepto #INSTRUMENT expresa la relación que representa una entidad que se usa para realizar la acción del evento. Así, los conceptos básicos incluidos en esta dimensión cognitiva son +BY\_02, +ON\_02 +THROUGH\_01, +WITH\_01 y +WITHOUT\_00. Sus unidades léxicas están expuestas al lado de cada concepto en los siguientes ejemplos:

+BY\_02: *by* y *by means of*

120) a. *John of Brittany and much of the garrison fled by ship (...)*

b. *The places of ascent or descent are reached by means of stairways.*

+ON\_02: *on*

121) *Marc used to say ' If it's not on the Internet, it doesn't exist.'*

+THROUGH\_01: *through, via*

122) a. *Rehabilitating older public-sector estates through a programme of co-operation with local authorities.*

b. *If the file is too large to send via email, you're left trying to find other ways to send it.*

+WITH\_01: *with*

123) *He had a seat belt on and we cut it with a knife and tried to pull him out but we couldn't.*

+WITHOUT\_00: *without*

124) *They have to light a fire without matches.*

El metaconcepto #MEANS expresa la relación representada por una entidad que, conjuntamente con el instrumento, se usa para realizar la acción del evento. Los conceptos básicos incorporados en este nivel son +FROM\_03 y +IN\_05, a los que se asignaron las unidades léxicas *from* e *in* respectivamente:

+FROM\_03

125) *A paint made from a natural resin.*

+IN\_05

126) *At that time I painted mostly in watercolor.*

La dimensión cognitiva #PARTICIPANT está estructurada a su vez en los metaconceptos #BENEFICIARY y #COMPANY. #PARTICIPANT describe a una entidad, diferente a la denotada por los argumentos, que participa de manera activa o pasiva en la acción del evento. Así, #BENEFICIARY especifica que el participante se beneficia de la acción del evento. El concepto básico para este metaconcepto es +FOR\_03, lexicalizado en la preposición *for*:

+FOR\_03

127) *This present is for you.*

Por otra parte, #COMPANY expresa que la entidad participa de manera coordinada en la realización de la acción del evento. El concepto básico asignado es +WITH\_02, cuya lexicalización se realiza a través de la preposición *with*:

+WITH\_02

128) *Boris still lives with his parents.*

#PURPOSE es el metaconcepto que expresa el propósito del evento e incluye al concepto +FOR\_04, el cual se lexicaliza mediante la preposición *for*.

+FOR\_04

129) *Networks for the exchange of information*

La dimensión cognitiva #QUANTITY expresa una entidad que describe una cifra, valor o cantidad. Al considerar las distintas maneras de expresar cantidad que existen en diversas lenguas europeas, se decidió realizar una división metaconceptual entre #APPROXIMATOR y #VALUE. Debajo de #APPROXIMATOR, se incluyeron los conceptos básicos +ABOUT\_00 y +BETWEEN\_03, que permiten expresar una entidad que describe una cifra o tiempo aproximado relacionado con la acción del evento. De este modo, para el concepto básico +ABOUT\_00 se asignaron las

preposiciones *about*, *around* y *by*, mientras que para el concepto +BETWEEN\_03 la unidad léxica asignada es *between*:

+ABOUT\_00

130) a. *It will take about an hour.*

b. *It is highly unlikely that you will be doing this around January but diaries are available from many stores at most times of the year.*

c. *Total recorded figures on these offences have increased by 378 per cent.*

+BETWEEN\_03

131) *The girl was able to tell detectives the car driver was white and aged between 30 and 40.*

Por otra parte, #VALUE expresa el valor descrito por la ER. Bajo este metaconcepto se han incluido los conceptos básicos que presento a continuación, indicando también sus correspondientes unidades léxicas:

+AT\_02: at

132) *Prices start at £18,500.*

+BY\_03: by

133) *A map measuring 400 by 600 mm.*

+FOR\_05: for

134) *Copies are available for £1.20.*

+FROM\_04: from

135) *If we have a scale from one to ten.*

+IN\_06: in

136) *A local income tax running at six pence in the pound.*

+TO\_02: to

137) *My car only does ten miles to the gallon.*

#REASON expresa una entidad que describe la causa del evento. El concepto básico creado para este metaconcepto es +FROM\_05, al que se le asignaron las unidades léxicas *from* y *because of*:

+FROM\_05

138) a. *A child suffering from asthma.*

b. *He had to stop work as a retail assistant at Iceland in Rochdale because of ill health.*

La dimensión cognitiva #REFERENT ha sido creada para organizar aquellos sentidos que expresan la configuración de una relación establecida por la ER con respecto a otras entidades de la situación descrita. Por este motivo, se decidió crear una estructura jerarquizada bajo este metaconcepto que permitiera especificar semánticamente el tipo de relación expresada. Los metaconceptos en cuestión son #COLLABORATION, #PARTITIVE, #POSSESSION y #QUALITY.

El metaconcepto #COLLABORATION expresa una relación configurada por dos o más entidades. Los conceptos básicos incluidos en este nivel son +AMONG\_01 y +BETWEEN\_02, cuyas unidades léxicas son *among* y *between* respectivamente.

+AMONG\_01

139) *She noted that lawyers have always been among the most avid participants in the survey.*

+BETWEEN\_02

- 140) *Negotiations will then start between the parties to see if they can work.*

El metaconcepto #PARTITIVE expresa que la entidad involucrada en la relación se refiere sólo a una parte o fracción de un todo. El concepto básico para este nivel es +OF\_00, el cual está lexicalizado en la preposición *of*:

+OF\_00

- 141) *Well-dressed and outgoing, Naina made the most of her trip, starting with breakfast at the Hyde Park Hotel.*

El metaconcepto #POSSESSION expresa que la ER es un poseedor. El concepto básico creado es +OF\_01, el cual se instancia léxicamente en la preposición *of*:

+OF\_01

- 142) *He was about five years older than me but I knew him as he was the son of one of my mum's friends.*

Finalmente, el último metaconcepto de tipo #REFERENT es #QUALITY, el cual indica que la ER expresa una característica del referente denotado por el nombre al que modifica la preposición. El concepto básico para esta dimensión cognitiva es +OF\_02, el cual se puede lexicalizar en las preposiciones *of* y *with*:

+OF\_02

- 143) a. *Diana's smile of delight is truly radiant.*  
b. *After all, people with money and status employ other people to clean up after them.*

El metaconcepto #RESULT expresa una entidad que describe la consecuencia de la acción del evento. El concepto básico asignado es +INTO\_00, el cual se representa en la preposición *into*:

+INTO\_00

144) *The project is already over a year into its three-year life.*

El metaconcepto #SCENE expresa una entidad que describe la situación o el contexto en el que ocurre el evento. El concepto básico creado para este metaconcepto es +OVER\_01, al cual se le ha asignado la unidad léxica *over*:

+OVER\_01

145) *He told me the truth over a glass of wine.*

Finalmente, la dimensión cognitiva #TOPIC expresa la entidad que describe el tema o asunto sobre el que trata otra entidad. El concepto básico incluido bajo este metaconcepto es +ON\_03, el cual puede representarse con las preposiciones *about*, *on* y *over*:

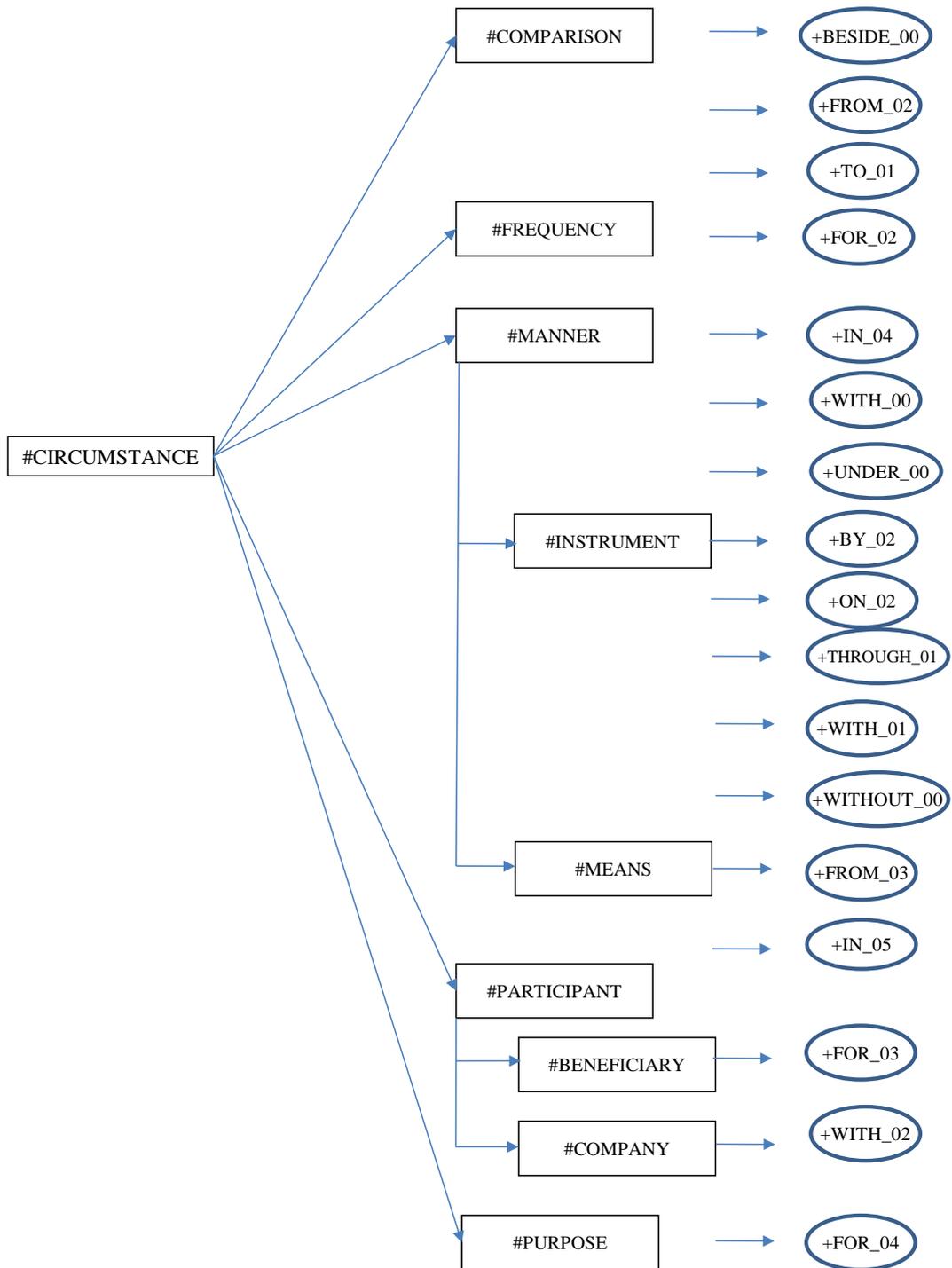
+ON\_03

146) a. *This is a subject of modern concern about which classical Buddhist sources have little to say.*

b. *I want to write a book on how to eat properly.*

c. *Right now, the world is faced with a debate over how to deal with the water crisis.*

A modo de conclusión, la Figura 37 muestra la organización taxonómica de #CIRCUMSTANCE. Asimismo, en el anexo VIII, se recogen las unidades léxicas para cada concepto básico vinculado a este metaconcepto.



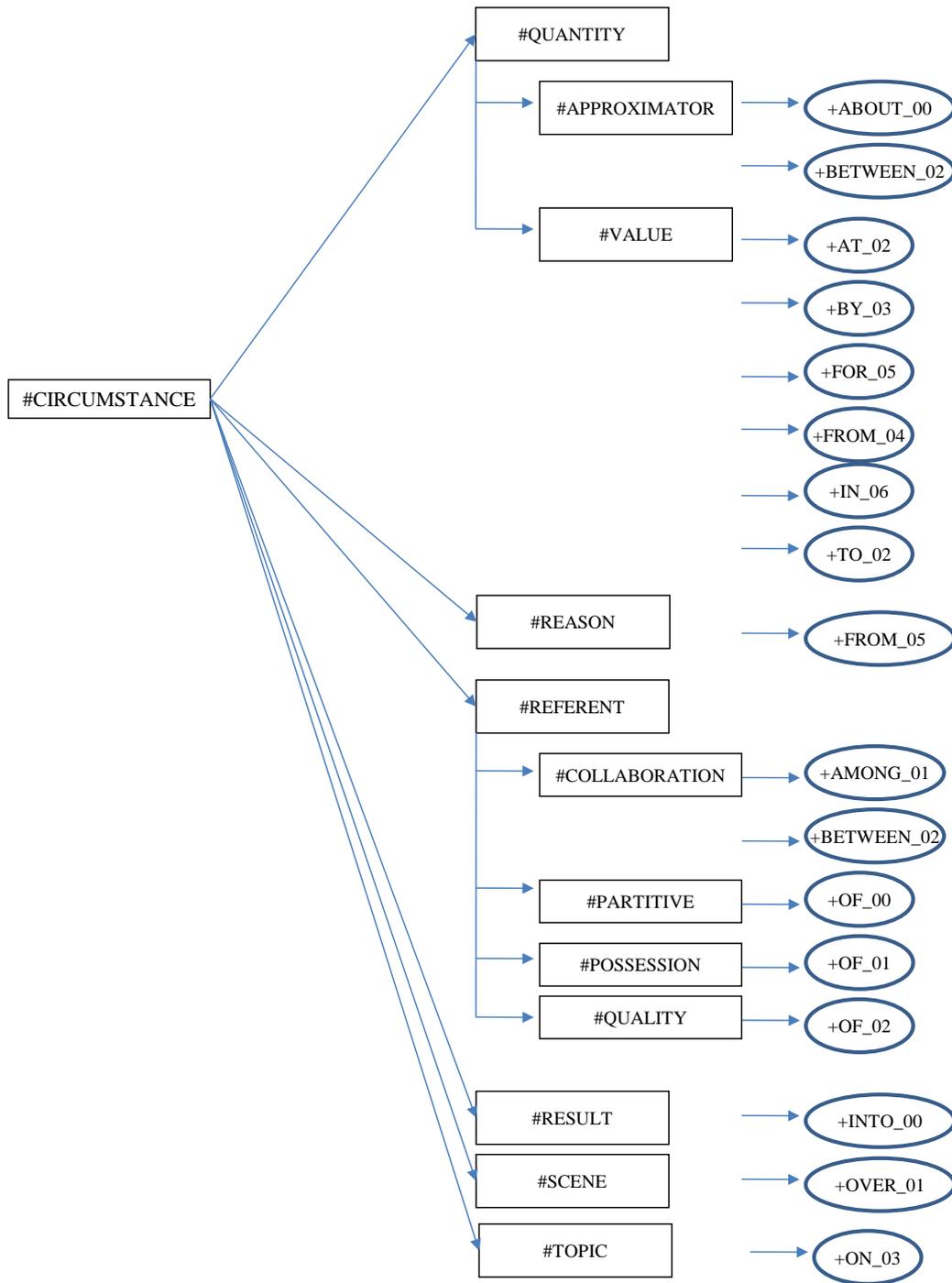


Figura 37. Distribución de los niveles conceptuales para la taxonomía #CIRCUMSTANCE.

Tras la organización ontológica de las tres taxonomías creadas para la subontología #RELATION, procedí al desarrollo de la entrada conceptual para cada uno de los conceptos básicos (CRITERIO 3). Antes de pasar al siguiente apartado, es importante recordar que la estructuración ontológica de #RELATION, al igual que

cualquier otra subontología en FunGramKB, es independiente de la lengua. Esto significa que esta misma taxonomía (y en particular, el nivel metaconceptual) podría ser reutilizada para asignar a los conceptos básicos unidades léxicas de otras lenguas, o bien crear nuevos conceptos básicos si los sentidos de las nuevas preposiciones lo requiriesen.

## 5.5 Entrada conceptual de las relaciones

A pesar de que el modelo de entrada conceptual que propongo, el cual fundamenté en el apartado 5.3, sólo requiere información acerca de la preferencia de selección del argumento al que se asignará un papel para el marco temático de la relación, es relevante considerar la ruta conceptual desde el metaconcepto raíz hasta el concepto básico para poder proporcionar claridad en la construcción del significado, ya que la distribución metaconceptual que se ha planteado tiene como objetivo la transmisión de matices semánticos a las preposiciones vinculadas a un mismo concepto básico, las cuales comparten cierta proximidad semántica.

La entrada conceptual de las relaciones se ha elaborado tras haber analizado y construido las diferentes taxonomías descritas en el apartado anterior. Se debe recordar que la organización conceptual de las taxonomías creadas bajo la subontología #RELATION se ha realizado teniendo en cuenta el CRITERIO 1, que establece dos niveles conceptuales formados por metaconceptos y conceptos básicos, y el CRITERIO2, el cual permite organizar internamente el nivel metaconceptual a través de la relación taxonómica de la subsunción. Finalmente, de acuerdo con el CRITERIO 3, en esta fase del estudio procedí al desarrollo de la entrada conceptual de cada concepto básico creado para el grupo de preposiciones estudiadas (i.e. *at*, *between*, *for*, *from*, *in*, *on*, *to* y *with*). Este apartado describe el número y tipo de participantes asignados a los marcos temáticos de estos conceptos, junto a las preferencias de selección de sus participantes. Hay que recordar que, siguiendo las directrices de FunGramKB, los metaconceptos no disponen de marcos temáticos.

Con el objetivo de exponer claramente los resultados de este trabajo, las entradas conceptuales se presentarán ordenadas por taxonomías en las Tablas 7, 8 y 9.

Tabla 7. Entradas conceptuales para las relaciones creadas en la taxonomía #TIME

<b>Entradas conceptuales para las relaciones creadas en la taxonomía #TIME</b>	
+AT_00	(f1: +HOUR_00 ^ +MEAL_00 ^ +NIGHT_00 ^ +OCURRENCE_00)Time
+BETWEEN_00	(f1: +TIME_00)Time (f2: +TIME_00)Time
+FOR_00	(f1: +CENTURY_00 ^ +DAY_00 ^ +HOUR_00 ^ +MINUTE_00 ^ +MOMENT_00 ^ +MONTH_00 ^ +SECOND_00 ^ +WEEK_00 ^ +YEAR_00)Time
+FROM_00	(f1:+CENTURY_00 ^ +DAY_00 ^ +MOMENT_00 ^ +MONTH_00 ^ +TIME_01 ^ +YEAR_00)Time
+IN_00	(f1 : +DAY_00 ^ +HOUR_00 ^ +MINUTE_00 ^ +MONTH_00 ^ +SECOND_00 ^ +WEEK_00 ^ +YEAR_00)Time
+IN_01	(f1:+DAY_00 ^ +HOUR_00 ^ +MINUTE_00 ^ +MONTH_00 ^ +SECOND_00 ^ +WEEK_00 ^ +YEAR_00)Time
+IN_02	(f1: +AFTERNOON_00 ^ +CENTURY_00 ^ +EVENING_00 ^ +MONTH_00 ^ +MORNING_00 ^ +SEASON_00 ^ +YEAR_00)Time
+ON_00	(f1: +DATE_00 ^ +DAY_00)Time

Tabla 8. Entradas conceptuales para las relaciones creadas en la taxonomía #LOCUS

<b>Entradas conceptuales para las relaciones creadas en la taxonomía #LOCUS</b>	
+AT_01	(f1: +PLACE_00 ^ +BUILDING_00 )Location
+BETWEEN_01	(f1: +PLACE_00)Location (f2: +PLACE_00)Location
+FOR_01	(f1:+LENGTH ^ +UNIT_OF_MEASURE_00)Location

+FROM_01	(f1: +PLACE_00)Origin
+IN_03	(f1: +BUILDING_00 ^ +CLOTHING_00 ^ +CONTAINER_00 ^ +COUNTRY_00 ^ +FAMILY_00)Location
+ON_01	(f1:+BARRIER_00 ^ +FURNITURE_00 ^ +SURFACE_00)Location
+TO_00	(f1: +PLACE_00)Goal

Tabla 9. Entradas conceptuales para las relaciones creadas en la taxonomía #CIRCUMSTANCE

<b>Entradas conceptuales para las relaciones creadas en la taxonomía #CIRCUMSTANCE</b>	
+AT_02	(f1: +UNIT_OF_VALUE_00)Quantity
+BETWEEN_02	(f1: +CORPUSCULAR_00)Referent (f2: +CORPUSCULAR_00)Referent
+BETWEEN_03	(f1:+NUMBER_00 ^ +UNIT_OF_MEASURE_00)Quantity (f2:+NUMBER_00 ^ +UNIT_OF_MEASURE_00)Quantity
+FOR_02	(f1:+DAY_00 ^ +TIME_01 ^ +YEAR_00 + ^ +WEEK_00)Frequency
+FOR_03	(f1: +CORPUSCULAR_00)Beneficiary
+FOR_04	(f1:+OCCURRENCE_00)Purpose
+FOR_05	(f1: +NUMBER_00 ^ +PRICE_00 ^ +UNIT_OF_VALUE_00)Quantity
+FROM_02	(f1)Comparison
+FROM_03	(f1:+MATERIAL_00)Means
+FROM_04	(f1: +NUMBER_00 ^ +UNIT_OF_MEASURE_00)Quantity

+FROM_05	(f1:+ILLNESS_00 ^ +OCCURRENCE_00)Reason
+IN_04	(f1: +COLOUR ^ +SHAPE_00 ^ +WAY_00)Manner
+IN_05	(f1: +LANGUAGE_00)Means
+IN_06	(f1:+UNIT_OF_VALUE_00)Quantity
+ON_02	(f1:+CALCULATOR_00 ^ +CAMERA_00 ^ +CLOCK_00 ^ +COMPUTER_00 ^ +MUSICAL_INSTRUMENT_00 ^ +RADIO_00 ^ +TELEPHONE_00 ^ +TELEVISION_00 ^ +TRANSPORT_00)Instrument
+ON_03	(f1: +TOPIC_00)Topic
+TO_01	(f1)Comparison
+TO_02	(f1:+UNIT_OF_CAPACITY_00 ^ +UNIT_OF_VALUE_00)Quantity
+WITH_00	(f1:+FEELING_00)Manner
+WITH_01	(f1:+ARTEFACT_00 ^ +FEELING_00)Instrument
+WITH_02	(f1: +CORPUSCULAR_00)Company

A continuación, presento una serie de observaciones a raíz de la identificación de las preferencias de selección arriba expuestas:

- El inventario de preferencias de selección asignado a un participante no es necesariamente exhaustivo. Como indica su nombre, se trata únicamente de especificar aquellas entidades con las cuales el referente conceptual del complemento preposicional está más estrechamente vinculado.
- Los marcos temáticos de +FROM\_02 y +TO\_01 no incluyen preferencia de selección alguna. Esto debe entenderse como que, en un principio, cualquier tipo de entidad podría servir como referente del complemento preposicional.

- Las preferencias de selección de +AT\_01 son +PLACE\_00 y +BUILDING\_00, lo cual podría resultar un poco extraño a primera vista, ya que comúnmente se puede pensar en los edificios como lugares. La cuestión es que, desde el punto de vista de la modelización conceptual de las entidades en FunGramKB, +PLACE\_00 pertenece al metaconcepto #REGION y +BUILDING\_00 al metaconcepto #SELF\_CONNECTED\_OBJECT.
- En el caso del marco temático de +FOR\_00, aunque todas las preferencias de selección pertenecen a +TIME\_00, fue preciso especificar qué conceptos básicos subordinados a éste mejor reflejan dichas preferencias de selección. De lo contrario, si hubiera optado por +TIME\_00 como única preferencia de selección, hubiera estado afirmando que conceptos como +DAY\_00 o +SEASON\_00, los cuales también pertenecen a +TIME\_00, podrían representar conceptualmente al referente del complemento preposicional de *for*, lo cual no sería cierto.

Sin duda alguna, el papel que desempeñan las preferencias de selección durante el procesamiento lingüístico es muy relevante. Por una parte, durante la generación lingüística, ayudan a la máquina a decidir por una determinada preposición; por ejemplo, cuando el referente del complemento preposicional denote una fecha (i.e. +DATE\_00), la máquina optará por la preposición *on*, frente a otras como *at* o *in*. Por otra parte, durante el análisis lingüístico, las preferencias de selección ayudan a solventar el problema de la ambigüedad léxica; por ejemplo, entre los diferentes sentidos de la preposición *in*, la máquina etiquetará el sintagma *in English* con el papel temático *Means*, ya que el referente conceptual (i.e. +LANGUAGE\_00) es una preferencia de selección de +IN\_05. No obstante, a diferencia de la mayoría de sistemas del PLN, todo el potencial de la desambiguación no debe recaer únicamente en las preferencias de selección. De lo contrario, y a modo de ejemplo, la desambiguación de la preposición *in* en el sintagma *in two hours* no sería posible, ya que las preferencias de selección de los conceptos +IN\_00 y +IN\_01 son idénticas. Como se presenta en el siguiente capítulo, la conjugación del conocimiento proporcionado por las entradas conceptuales y las entradas léxicas es decisiva para un procesamiento lingüístico más robusto.

## CAPÍTULO 6

# MODELIZACIÓN LINGÜÍSTICA DE LAS PREPOSICIONES

### 6.1 Motivación teórica

Como ya se ha comentado, este estudio propone un enfoque léxico–conceptual para el tratamiento de las preposiciones en sistemas del PLN. Con respecto a la información léxica que debe ser incorporada en el Lexicón, se decidió crear un modelo de plantilla léxica para las preposiciones que permitiera al sistema almacenar y modelar el conocimiento incluido en FunGramKB para hacer frente al problema de la desambiguación léxica.

Para poder decidir sobre la estructuración de esta plantilla léxica, hubo que revisar las implicaciones sintácticas de las preposiciones objeto de este estudio. Se debe recordar que esta tesis está orientada a las preposiciones no seleccionadas y transitivas (Baldwin et al., 2009) que funcionan como adjuntos no predicativos de la cláusula en la RRG. Así, en la oración (147), la preposición *in* puede ser considerada no seleccionada y transitiva pues no es regida por ningún elemento de la oración y contiene un complemento, introduciendo un sintagma prepositivo, i.e. *in the workshop*, que funciona como adjunto periférico al representar un elemento no construccional que proporciona un valor locativo al sentido de la cláusula.

147) *Jane and Tom talked about the situation in the workshop.*

Van Valin (2005, p. 19) afirmó que “PP adjuncts modify the core when they express locational or temporal features of the state of affairs coded by the core. Consequently, the periphery containing them is a core modifier”. Teniendo en cuenta este enfoque, se estableció la hipótesis de que el sintagma preposicional que funcionara como adjunto periférico podría ayudar a modelar el significado aspectual de la cláusula, yendo más allá del puro significado aportado por la preposición, el cual es descrito en el caso de esta tesis a través de una organización ontológica de las relaciones. Esta perspectiva aporta un nuevo enfoque al tratamiento computacional de las preposiciones, pues permite una consideración más sofisticada de esta categoría gramatical, yendo de la semántica léxica a la semántica oracional. Desde un punto de vista estrictamente lingüístico, Cortés-Rodríguez (2014, p. 33) afirmó que en la RRG “the aspectual features of an event cannot be treated as a predicate-level fact alone, but as a higher-level phenomenon”. En otras palabras, los rasgos aspectuales deben distribuirse a lo largo de las cuatro capas de la cláusula de modo que la información aspectual de una de las capas pueda extenderse al resto mediante una interfaz léxico-conceptual. De este modo, este autor defendió que la interpretación de los rasgos aspectuales en el marco de la RRG tendría lugar en el predicado (i.e. el *Aktionsart*), en el núcleo (i.e. el operador aspectual), en el centro (i.e. la aspectualidad representada por el cuantificador del argumento verbal) y en la cláusula (i.e. el operador de tiempo), como se infiere de la Tabla 10 (Cortés-Rodríguez, 2014, p.32).

Tabla 10. Rasgos aspectuales en las capas de la cláusula según la RRG.

RRG Layer:	Predicate	Nucleus	Core	Clause
Domain	Aktionsart	Aspect	Aspectuality	Tense/Status &IF
Entity	Event-type	Extended event-type)	Event(event-type reification)	SoA/Speech Act
Features	Stativity	Progressive	E.Quantification	
	Dynamicity	Perfective	Incrementality	
	L-Telicity	Perfect...	S-Telicity	
	Punctuality (Causativity)			

Por tanto, el concepto de aspectualidad defendido por Cortés-Rodríguez (2014) hace referencia al componente télico del evento elegido por un hablante. Esto es, este autor resalta la diferencia entre el componente télico inherente a los diferentes tipos de eventos, representado en el *Aktionsart*, y la telicidad expresada en un evento concreto como resultado de la interacción de diversos rasgos extendidos en las distintas capas de la cláusula. A modo de ilustración, presentó el siguiente ejemplo (p. 34):

148) a. Chona come galletas

b. Chona se come unas galletas

En (148a) el verbo *comer* describe un estado de cosas (*State of Affairs*) sin delimitación temporal, es decir, atélico, cuyo evento tiene un *Aktionsart* de actividad. En cambio, en la oración (148b) el *Aktionsart* es de realización activa, debido a la introducción de un complemento definido y cuantificado, y, consecuentemente, la aspectualidad de la cláusula se transforma en télica. En este punto, su propuesta trata de dar respuesta a la dicotomía establecida entre telicidad léxica y sintáctica. Así, defendió que la telicidad léxica (i.e. *L[exical]-(a)telicity*) es aquella que se refiere al rasgo aspectual del *Aktionsart* y que sirve para describir el tipo de evento de un predicado y que la telicidad sintáctica (i.e. *S[yntactic]-(a)telicity*) es la aspectualidad de la cláusula, como muestran los siguientes ejemplos:

149) *Larry painted.*

*Aktionsart*: actividad

(L-atelic)

150) *Larry painted three portraits in two hours.*

*Aktionsart*: realización activa

Incremento del complemento: *three portraits*

SP adjunto: *in two hours*

(S-telic)

En el ejemplo (150) se puede observar cómo la interrelación de la información semántica proporcionada en el centro (i.e. *three portraits*) modela la aspectualidad sintáctica de la cláusula en (149), resultando en un *Aktionsart* de realización activa que posibilita la aparición del sintagma preposicional en la periferia (i.e. *in two hours*). Por consiguiente, también estoy de acuerdo con Suppes (1979) cuando afirmó que, para la correcta interpretación de un sintagma preposicional, se debe atender a la interrelación sintáctico-semántica de los elementos oracionales.

A colación de estos fundamentos, Zwarts (2005) defendió que los sintagmas preposicionales no son meros modificadores, sino que pueden contribuir a la estructura de los eventos a nivel aspectual. Así, en el siguiente ejemplo extraído de Biały (2013, p. 108), se puede observar que la presencia de un tipo u otro de sintagma preposicional puede determinar el aspecto de la cláusula:

151) a. *John ploughed the field in an hour. [culmination]*

b. *John ploughed the field for an hour. [no culmination]*

En ambas oraciones, los rasgos aspectuales de todas las capas de la cláusula, es decir, el operador aspectual del núcleo, el tipo de evento con su argumento en el centro y el operador de tiempo en la cláusula, son el mismo (i.e. *John ploughed the field*). Sin embargo, en la capa del predicado, el *Aktionsart* es diferente para cada ejemplo al incorporar un sintagma preposicional u otro (i.e. *in an hour* o *for an hour*); el *Aktionsart* del ejemplo (151a) es una realización activa, mientras que en la oración (151b) es una actividad. El rasgo diferenciador entre los *Aktionsarten* de actividad y realización activa es la telicidad, según el test lingüístico mostrado en la sección 3.1.2. Esta situación es

explicada por Cortés-Rodríguez (2014) como casos excepcionales en los que la telicidad léxica de un predicado no está en consonancia con la telicidad sintáctica de la cláusula por la introducción de un complemento definido (i.e. *the field*). Para esta situación, la explicación viene dada por los estudios de Krifka (1998) y la revisión hecha por Rothstein (2004), cuando se refieren a que la relación entre el evento y el argumento es correlativa en el sentido de que sólo es posible saber la culminación de la acción cuando se conoce la cantidad del argumento. De este modo, la clave para conocer la aspectualidad sintáctica en (151), según estos autores, reside en la presencia del artículo definido *the*.

Otro ejemplo más claro de esta problemática es el ofrecido por Beavers (2013, p. 683):

- 152) a. Mary drank a glass of wine in an hour/\*for an hour.  
b. Mary drank wine \*in an hour/for an hour.

En (152a) la cuantificación del nombre incontable *wine* a través del cuantificador *a glass of* permite determinar que la aspectualidad de la cláusula es télica, siendo el *Aktionsart* de tipo realización activa. En realidad, a este tipo de eventos se les conoce como “realizaciones activas de verbos de consumo y creación”, los cuales resultan en una aspectualidad télica por la cuantificación del argumento. Lo mismo sucede con eventos que expresan movimiento cuando el argumento es una trayectoria con un final explícito (i.e. *run to the house*), y con eventos de tipo logro graduado (i.e. *degree achievements*), los cuales expresan un cambio de estado, como se ejemplifica en (153) obtenido de Beavers (2013, p. 683):

- 153) *The vase cooled in 5 minutes.*

Por tanto, y de acuerdo con Cortés-Rodríguez (2014, p. 43), la telicidad sintáctica de la cláusula es resultado de la interacción de los componentes semánticos, i.e. “S-Telicity is semantically compositional: S-Telicity is a feature of Events constructed by a calculus from the *Aktionsart* properties of a predicate, and/or semantic co-composition from the meanings of core constituents”. Consecuentemente, se consideró que la aparición de ciertos sintagmas preposicionales dependía

fundamentalmente de la combinación entre el predicado y su argumento en el centro, la cual sirve para modelar la aspectualidad sintáctica, determinando así que sea necesario introducir información sobre la telicidad en forma de marcadores o rasgos lingüísticos en la entrada léxica de las preposiciones. De hecho, como defendió Filip (2011, p. 1209-1210), “telicity is a feature of some lexical verbs which explains a range of phenomena as is the resistance of some verbs to form part of some unbounded eventualities together with the durative for NP adverbial”.

Cabe mencionar que, con respecto a las preposiciones de tipo espacial, diversos investigadores (Zwarts, 2005; Tungseth, 2006; Svenonius, 2008; Biały, 2013) defendieron la postura de que la combinación de un mismo evento con distintos sintagmas preposicionales de trayectoria puede suponer una variedad de interpretaciones semánticas, lo que demuestra que es el sintagma introducido por la preposición el que determina la telicidad aspectual de la cláusula. En esta línea, Tungseth (2006, p. 30) afirmó:

it is the properties of the prepositional phrase which determine whether the event is telic or not (...). If PathP contains a PP like *til*, ‘to’, which denotes a bounded path, the interpretation of the event as a whole is bounded. If the PP contains a preposition like *mot*, ‘towards’, which can never denote a bounded path, the interpretation of the event is atelic.

Por ejemplo, en las siguientes oraciones se puede apreciar cómo la aspectualidad resulta télica en (154a), mientras que en (154b) resulta atélica:

154) a. *He swam to the other side of the river.*

b. *He swam towards the other side of the river.*

La explicación a este hecho está en consonancia con lo defendido por Krifka (1998) y Rothstein (2004) sobre la relación entre el evento y su argumento, si se tiene en cuenta que la preposición *to* implica el final de la trayectoria.

En relación a esta concepción de las preposiciones espaciales como elementos que pueden proporcionar telicidad al estado de las cosas expresado por el evento, diversos estudios (Jackendoff, 1983, 1991; Nam, 2000; Zwarts, 2005; Piñón, 1993; Svenonius, 2008) reclamaron la importancia del rasgo de la delimitación espacial (*spatial boundedness*) en las preposiciones espaciales. Como se comentó en la sección 5.4.2, Jackendoff (1983) hizo una distinción sobre las preposiciones de trayectoria entre aquellas cuyo objeto de referencia determina el lugar de comienzo o fin de la trayectoria, i.e. las trayectorias delimitadas, las que expresan la dirección hacia la cual se dirige la acción, i.e. las direcciones, y las que consideró como rutas, pues expresan el modo en el que se sucede la trayectoria. Posteriormente, Jackendoff (1991) estableció la distinción entre delimitado (*bounded*) y no delimitado (*unbounded*), incluyendo en esta última clase a las preposiciones que expresan dirección o ruta. Fundamentalmente, esa distinción hace referencia a la misma distinción evocada en los tests lingüísticos para los *Aktionsarten* sobre la telicidad. De este modo, la delimitación espacial hace alusión al rasgo télico proyectado en el espacio, mientras que la no delimitación espacial se relaciona con el rasgo atélico. Piñón (1993) revisó la distinción hecha por Jackendoff (1991) e incluyó dentro de la categoría de trayectorias delimitadas una nueva clase llamada rutas delimitadas, como se puede observar en la Figura 38 (Piñón, 1993, p. 299).

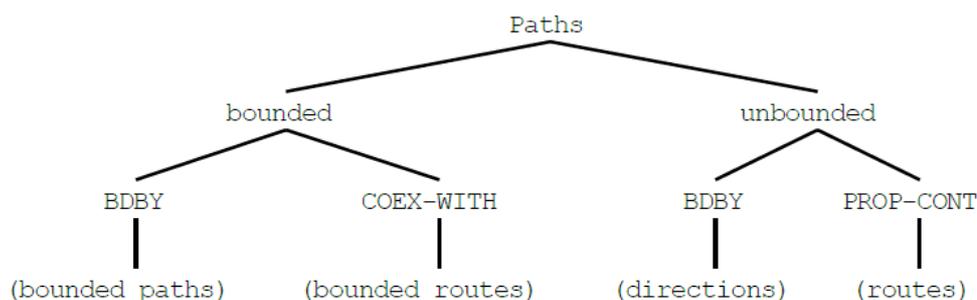


Figura 38. Revisión hecha por Piñón de la división entre trayectorias delimitadas y no delimitadas de Jackendoff.

Esta distinción fue hecha con la finalidad de explicar cómo es posible que, en oraciones como las del ejemplo (155) en Piñón (1993, p. 299), se permita la presencia de sintagmas preposicionales de tiempo télicos y atélicos:

- 155) *The procession walked by the church* {for 45 minutes, in 45 minutes}.

Según el autor, no existe claridad en este hecho si se considera al sintagma introducido por *by* como *route* y, por lo tanto, no delimitado espacialmente, de acuerdo con la distinción hecha por Jackendoff. Así, Piñón afirmó que el hecho de que un sintagma preposicional de trayectoria pudiera considerarse como delimitado o no dependía de que hubiera un punto inicial o final de la trayectoria, ya fuera de manera explícita o implícita, lo cual coincide con las ideas de Krifka (1998). Nam (2000) trató de explicar este tipo de ambigüedad aspectual a través de las siguientes oraciones, en las que consideró el papel del sintagma preposicional según la semántica del verbo (Nam, 2000, p. 701);

- 156) V.*walk*: e1[process] + PP.*in the room*

a. VP.*walk in the room*: e1[process]

b. VP.*walk in the room*: e1[process] + e2\*[state]

El verbo *walk* denota un *Aktionsart* actividad que, al combinarlo con un sintagma preposicional cuyo núcleo nominal denota lugar, p.ej. *in the room*, expresa que la acción de caminar se realiza dentro de la habitación, permaneciendo así el rasgo atélico. Sin embargo, si se considera al sintagma preposicional de trayectoria, el cual conllevaría un cambio del estado de las cosas expresado por la oración, el *Aktionsart* pasaría a ser de tipo realización activa, donde la cláusula tendría una aspectualidad sintáctica télica.

En este sentido, es interesante comentar la diferencia entre los sintagmas preposicionales que intervienen en las construcciones argumentales de los verbos y los adjuntos periféricos que se tratan en esta tesis. Desde esta perspectiva, los sintagmas preposicionales de trayectoria que contribuyen a la telicidad de la cláusula son típicamente de tipo construccional. De este modo, tal como estableció Cortés-Rodríguez (2014), las construcciones argumentales pueden hacer que un predicado de actividad se perciba como una realización activa al considerar la contribución de elementos como la cuantificación del argumento o el tipo de preposición. Por esta razón, para un

procesamiento más adecuado del lenguaje, se consideró la necesidad de incorporar información acerca de los elementos aspectuales que configuran la telicidad sintáctica de la cláusula, en la línea que defiende Cortés-Rodríguez (2014), de modo que los cambios de aspectualidad de la cláusula puedan ser considerados como un fenómeno sintáctico en ARTEMIS.

Por otro lado, y como se ha comentado en la sección 3.1.3, Van Valin (2005, p. 21) afirmó que “(p)redicative adpositions function like predicates in that they contribute substantive 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”. Por este motivo, otro aspecto importante a tener en cuenta es la información acerca del argumento de la preposición. Debido a su naturaleza como núcleo predicativo del sintagma, la preposición restringe el tipo de argumentos que pueden ocurrir en el sintagma del mismo modo que hacen los verbos. Así, como analizó Rauh (1993, p. 107), los siguientes ejemplos dan muestra de cómo los complementos son restringidos léxicamente a causa de cuestiones determinadas por las propiedades semánticas de un núcleo predicativo verbal:

- 157) *Bill read the book.*  
      \**the boy.*

Del mismo modo sucede con las preposiciones. La activación de un sentido u otro de la preposición supone restricciones semánticas al argumento preposicional en la periferia de la cláusula. Por ejemplo, en las siguientes oraciones, la preposición *at* puede aparecer con *the airport*, cuando su sentido es locativo, o *1 o'clock*, cuando su valor es temporal:

- 158) a. *She woke up at the airport.*  
      b. *She woke up at 1 o'clock.*

Sin embargo, dentro del dominio temporal o locativo que puede determinar la preposición, existen restricciones léxicas al propio complemento estipuladas por el sentido de la preposición. De este modo, la preposición *at* en su valor temporal requiere complementos que definan a *t* como el punto específico temporal en el que sucede la

acción, lo cual corresponde a la descripción hecha del concepto básico +AT\_00 incluido bajo el metaconcepto #INTERVAL\_F1\_1 en la taxonomía temporal propuesta en el apartado 5.4.1. Así, en el siguiente ejemplo, la oración (159b) no es correcta:

- 159) a. *She woke up at 1 o'clock.*  
b. *She woke up at the summer.\**

No obstante, para la información acerca del complemento de la preposición, se adoptó el enfoque ontológico de FunGramKB (cf. Sección 5.5), el cual proporciona ventajas con respecto a la redundancia informativa que provocaría un enfoque lexicista en un entorno multilingüe. Por tanto, los rasgos semánticos prototípicos de los argumentos preposicionales no serán tratados en la entrada léxica de la preposición, sino en la entrada conceptual de la relación, y más concretamente en su marco predicativo. De esta manera, a través de una interfaz léxico-sintáctica-conceptual, el sistema podrá vincular la información de la entrada léxica (p. ej. *Aktionsart*) con el tipo de aspectualidad sintáctica presente en la cláusula a través del conocimiento semántico que la Ontología asigna a cada unidad conceptual; de esta forma, al proporcionar un texto de entrada a la aplicación, el sistema será capaz de comprender el sentido de la preposición y la coocurrencia con determinados rasgos aspectuales.

Como conclusión de este apartado, establezco que la entrada léxica de las preposiciones quedará alojada en el Lexicón y requerirá de un único criterio para su diseño:

**CRITERIO 4:** la entrada léxica debe contemplar el modo en el que la preposición interactúa con los rasgos aspectuales en las cuatro capas de la oración (i.e., predicado, núcleo, centro y cláusula).

## **6.2 Análisis de los rasgos lingüísticos para los sentidos de las preposiciones**

Una de las últimas fases de mi investigación fue estudiar los marcadores lingüísticos que ayudarán, junto con el conocimiento almacenado en la Ontología, a activar los diversos sentidos de las preposiciones. En este punto es importante recordar

que la finalidad es ahora la creación de plantillas léxicas para aquellos sentidos de las preposiciones que funcionen como adjuntos predicativos en la periferia de la RRG, es decir, aquellos que contribuyan semánticamente a algún aspecto de la cláusula, dejando al Gramaticón, y más concretamente al L1-Constructicón, el tratamiento de los sintagmas preposicionales que funcionan como argumentos construccionales de los predicados. Asimismo, y en línea con lo que afirmó Van Valin (2005), los sintagmas preposicionales predicativos de tiempo y lugar ubicados en la periferia pueden modificar al centro de la cláusula, por lo cual, para la identificación de los marcadores lingüísticos me limité a estudiar los sentidos desarrollados en esas dos dimensiones cognitivas. De este modo, considerando que en el caso de los adjuntos periféricos no hay contribución aspectual, pero sí una necesidad de establecer qué tipo de aspectualidad sintáctica permite la aparición de un tipo u otro de sintagma preposicional periférico, analicé la presencia de estos sintagmas en un conjunto de oraciones obtenidas de diferentes fuentes utilizadas a lo largo de este estudio.

Tras realizar el análisis, llegué a la conclusión de que los sintagmas preposicionales introducidos por una preposición de lugar que se conceptualice en algún concepto básico ubicado en el metaconcepto #PLACE dentro de #LOCUS son compatibles con una aspectualidad sintáctica télica o atélica y no muestran ninguna restricción lingüística con otros rasgos de la cláusula, por lo que decidí no crear ninguna entrada léxica para este tipo de preposiciones, pues su contribución semántica puede incorporarse a la entrada conceptual a través de los papeles temáticos de las relaciones. Con respecto a las preposiciones encontradas bajo el metaconcepto #PATH, sólo *to* parece interactuar con otros rasgos de la cláusula en los casos que actúa como construcción. Por lo tanto, por los motivos comentados en la sección anterior, tampoco creé entradas léxicas para ese conjunto de preposiciones.

Por el contrario, para las preposiciones con sentido temporal, sí existen restricciones que, en parte, coinciden con los rasgos característicos para los tipos de *Aktionsart*. De este modo, por ejemplo, para las cláusulas cuya aspectualidad sintáctica es estado o actividad, el tipo de sintagma preposicional debe contener un sentido de la preposición que exprese durabilidad, coincidiendo con el carácter atélico de las mismas, mientras que para las predicaciones de realización y realización activa, el sentido de la

preposición debe expresar un punto final temporal para la acción. Además, se encontraron casos en los que un mismo sentido de la preposición podía aparecer con predicaciones télicas y atélicas, donde el marcador lingüístico era proporcionado por el aspecto gramatical o el tiempo verbal.

En consecuencia, el número de entradas léxicas necesarias para el tratamiento léxico-conceptual de los sintagmas preposicionales periféricos en la RRG se redujo a ocho, las cuales corresponden a los conceptos temporales creados para las preposiciones *at*, *between*, *for*, *from*, *in* y *on*, ya que el resto de sentidos no presentan una interacción semántico-sintáctica que pueda interferir en el correcto procesamiento de la cláusula.

Con el fin de mostrar una mayor claridad descriptiva, a continuación se muestra una serie de ejemplos a través de los cuales se explica el proceso de creación de las entradas léxicas. Considérese la preposición *in* en el siguiente ejemplo:

160) *John ran to the park in an hour.*

En (160), el adjunto temporal (i.e. *in an hour*) es concebido como #DURATION y el concepto básico asignado para su representación sería +IN\_00, ya que TOE es igual a *t* y la predicación expresa un evento télico. Como se ha explicado al principio de este capítulo, la aspectualidad de ciertos eventos es resultado de la interacción de los rasgos aspectuales distribuidos en todas las capas de la cláusula, de manera que los valores de los rasgos presentes en las capas superiores pueden invalidar a aquellos que están presentes en las capas inferiores. De hecho, desde el enfoque jerárquico de la estructura estratificada de la cláusula, puedo afirmar que tiene lugar una herencia no monótonica de los rasgos aspectuales. Así, en el ejemplo (160), el *Aktionsart* del predicado corresponde a un tipo de evento de actividad que expresa atelicidad léxica (*L-atelic*), pero el argumento construccional proporciona una delimitación al evento que se traduce en telicidad sintáctica (*S-telic*) en la predicación, modelando al *Aktionsart* en realización activa. El ejemplo demuestra cómo la aspectualidad del centro de la cláusula invalida la atelicidad mostrada por el *Aktionsart* en la entrada léxica del verbo. Consecuentemente, la entrada léxica para el sentido expresado por la preposición *in* (i.e. +IN\_00) debe

incluir las siguientes restricciones en forma de marcadores lingüísticos, los cuales se presentan a través de una matriz atributo-valor de manera parentética:<sup>30</sup>

161) [[HEAD: in]  
 [RELATION: +IN\_00]  
 [CONSTRAINTS:  
     [CLAUSE:  
         [CORE:  
             [NUC:  
                 [PRED: [AKT: aca]]]]]]]]

Igualmente, en las oraciones en (162), el sentido de la preposición es percibido como #DURATION. Sin embargo, en este caso el concepto asociado a la preposición es +IN\_01, debido a la identificación del referente del objeto preposicional como +MINUTE\_00 a través de la interfaz léxico-conceptual.

- 162) a. *The train arrives in ten minutes.*  
 b. *We shall run in five minutes.*

Además, la explicación de que TOE está localizado en un momento específico después de TOR (i.e.  $TOE = TOR + t$ ) está en consonancia con lo que establece Filip (1999, p. 23), i.e. “with achievements and activities the time indicated by the time-span adverbial is calculated from a contextually given reference point”, de modo que +IN\_01 es una relación donde TOE ocurre con respecto de TOR. La idiosincrasia lingüística de esta relación reside en el hecho de su incompatibilidad con el aspecto perfectivo o con los operadores de tiempo pasado, puesto que la preposición se centra en el comienzo del evento. Por consiguiente, la entrada léxica de la preposición *in* asociada al concepto +IN\_01 debe incluir las siguientes restricciones en forma de marcadores lingüísticos:

---

<sup>30</sup> En las matrices representadas en este capítulo, los atributos se presentan en mayúscula, al igual que el identificador del concepto y el operador de negación, y los valores en minúscula. Además, se puede observar que, dentro del atributo CONSTRAINTS, los marcadores lingüísticos se distribuyen entre las diferentes capas de la estructura estratificada de la cláusula (i.e. CLAUSE, CORE, NUC y PRED).

- 163) [[HEAD: in]  
[RELATION: +IN\_01]  
[CONSTRAINTS:  
[CLAUSE: [TNS: pres || fut]  
[CORE:  
[NUC: [ASP: NOT perf]  
[PRED: [AKT: act || ach]]]]]]]]

Se puede llegar a pensar que lo lógico sería introducir estas restricciones en la entrada conceptual de la relación y no en la entrada léxica de la preposición. No obstante, como se ha comentado en el capítulo 4, la información sobre el comportamiento lingüístico de las unidades léxicas y el conocimiento semántico de las unidades conceptuales está claramente separado, aunque interconectado desde diferentes módulos de la arquitectura de FunGramKB, i.e. el Lexicón y la Ontología respectivamente. Esta modularidad del conocimiento es particularmente adecuada cuando algunos marcadores léxico-sintácticos de las preposiciones son parámetros específicos de la lengua, por lo cual todo ellos deben ser descritos en las entradas léxicas. En esta línea, Slabakova (2000, p. 748) presentó un ejemplo que claramente refleja el fundamento de esta decisión:

- 164) *Tja z-gotvijadeneza 3 c`asa.*

Ella cocinó comida en 3 horas.

*She cooked some food in 3 hours.*

En la oración en búlgaro, a pesar de que el complemento es un nombre incontable, el morfema del verbo indica que la acción es télica, de modo que “mass nouns or bare plurals in form take on an interpretation of specified quantity or number” (Slabakova, 2000, p. 749). Asimismo, la oración (164) en español ilustra otro caso donde las idiosincrasias lingüísticas justifican que estos marcadores deban estar reflejados en la entrada léxica de la preposición, en lugar de en la entrada conceptual de la relación; aquí, (165), por ejemplo, la presencia del pronombre reflexivo clítico *se* aparece con un elemento cuantificable para enfatizar la telicidad de la cláusula:

165) Juan se comió una hamburguesa en 10 minutos.

A modo de conclusión de este apartado, se presentan las entradas léxicas creadas para los sentidos comentados anteriormente:

166) [[HEAD: at]  
[RELATION: +AT\_00]  
[CONSTRAINTS:  
[CLAUSE:  
[CORE:  
[NUC: [ASP: NOT progr]  
[PRED: [AKT: act || ach]]]]]]]]

La preposición *at* con el sentido +AT\_00 puede aparecer en una cláusula con un verbo de tipo actividad o logro que no esté en progresivo.

167) [[HEAD: between]  
[RELATION: +BETWEEN\_00]  
[CONSTRAINTS:  
[CLAUSE: [TNS: NOT presperf]  
[CORE:  
[NUC:  
[PRED: [AKT: NOT state]]]]]]]]

La preposición *between* con el sentido +BETWEEN\_00 no puede aparecer en una cláusula con un verbo cuyo tiempo verbal sea pretérito perfecto o cuyo *Aktionsart* sea de tipo estado.

168) [[HEAD: for]  
[RELATION: +FOR\_00]  
[CONSTRAINTS:  
[CLAUSE:  
[CORE:  
[NUC:  
[PRED: [AKT: sta|| act]]]]]]]]

La preposición *for* con el sentido +FOR\_00 puede aparecer en una cláusula con un verbo de tipo estado o actividad.

169) [[HEAD: from]  
[RELATION: +FROM\_00]  
[CONSTRAINTS:  
[**CLAUSE: [TNS: NOT presperf]**  
[CORE:  
[NUC:  
[PRED: [**AKT: NOT sta || NOT  
ach**]]]]]]]]

La preposición *from* con el sentido +FROM\_00 no puede aparecer en una cláusula con un verbo cuyo tiempo verbal sea pretérito perfecto o cuyo *Aktionsart* sea estado o logro.

170) [[HEAD: in]  
[RELATION: +IN\_00]  
[CONSTRAINTS:  
[CLAUSE:  
[CORE:  
[NUC:  
[PRED: [**AKT: acc || aca**]]]]]]]]  
  
[CLAUSE: [TNS:]]  
[CORE:  
[NUC: [**ASP: perf**]  
[PRED: [**AKT: ach**]]]]]]]]

La preposición *in* con el sentido +IN\_00 puede aparecer en una cláusula con un verbo de tipo realización o realización activa. También puede aparecer en una cláusula con un verbo de tipo logro cuando el aspecto gramatical sea perfectivo.

171) [[HEAD: in]  
[RELATION: +IN\_01]  
[CONSTRAINTS:  
[CLAUSE: [TNS: pres || fut]  
[CORE:  
[NUC: [ASP: NOT perf]  
[PRED: [AKT: NOT sta]]]]]]]]

La preposición *in* con el sentido +IN\_01 sólo puede aparecer en una cláusula con verbos en presente o futuro con aspecto imperfectivo y cuyo *Aktionsart* no sea de tipo estado.

172) [[HEAD: in]  
[RELATION: +IN\_02]  
[CONSTRAINTS:  
[CLAUSE: [TNS: NOT presperf]  
[CORE:  
[NUC:  
[PRED: [AKT: NOT sta]]]]]]]]

La preposición *in* con el sentido +IN\_02 no puede aparecer en una cláusula con un verbo cuyo tiempo verbal sea pretérito perfecto o cuyo *Aktionsart* sea de tipo estado.

173) [[HEAD: on]  
[RELATION: +ON\_00]  
[CONSTRAINTS:  
[CLAUSE: [TNS: NOT presperf]  
[CORE:  
[NUC:  
[PRED: [AKT: NOT sta]]]]]]]]

La preposición *on* con el sentido +ON\_00 no puede aparecer en una cláusula con un verbo cuyo tiempo verbal sea pretérito perfecto o cuyo *Aktionsart* sea de tipo estado.



## CAPÍTULO 7

### CONCLUSIONES Y FUTURAS INVESTIGACIONES

Esta tesis presenta un modelo que permite una descripción conceptual y un análisis lingüístico de ocho de las preposiciones más frecuentes en lengua inglesa (i.e. *at, between, for, from, in, on, to* y *with*) que pueden actuar como preposiciones predicativas en adjuntos periféricos según la RRG. El objetivo es dar una solución efectiva al problema de la desambiguación semántica de este tipo de preposiciones dentro del campo del PLN. Los resultados investigadores se ubican tanto en el nivel de la representación conceptual del conocimiento como en el nivel lingüístico. Por una parte, se construyó la subontología #RELATION, la cual se organizó internamente en las taxonomías #TIME, #LOCUS y #CIRCUMSTANCE, requiriendo finalmente la elaboración de 71 conceptos y de 36 entradas conceptuales, las cuales corresponden a los sentidos de las ocho preposiciones. De esta manera, se aportaron las propiedades conceptuales tanto de las propias preposiciones, a través de la ubicación de los conceptos básicos en la taxonomía conceptual, como de los referentes de los objetos preposicionales, a través de las preferencias de selección en los marcos predicativos. Por otra parte, el análisis lingüístico de las ocho preposiciones dio como resultado la construcción de ocho entradas léxicas, las cuales se alojan en el Lexicón del inglés.

En definitiva, los resultados investigadores sobre el modelo propuesto en esta tesis sirven para mejorar el algoritmo léxico-conceptual de los procesadores lingüísticos, y más concretamente de ARTEMIS. De esta manera, he diseñado un modelo que proporciona un tratamiento más adecuado en la desambiguación de los sintagmas preposicionales que funcionan como adjuntos periféricos en el marco de la RRG, al mismo tiempo que enriquece la semántica oracional al modelar los diferentes rasgos aspectuales que se encuentran en las cuatro capas de la estructura estratificada de la cláusula. Debo destacar que una de las contribuciones más importantes de este trabajo radica en la formalización de la interfaz sintáctico-semántica dentro de la base de conocimiento en torno a los comportamientos de esta clase gramatical. De hecho, prácticamente ningún estudio previo ha sabido conjugar los resultados investigadores en lingüística teórica, lingüística cognitiva, lingüística computacional e ingeniería del conocimiento para proporcionar un tratamiento más robusto al procesamiento de las preposiciones.

A lo largo de esta tesis se han mostrado diferentes perspectivas de análisis de las preposiciones, una clase léxica que suele ser subestimada en el estudio de las categorías gramaticales. A pesar de no componer un conjunto de palabras tan amplio como el caso de los verbos, nombres, adjetivos o adverbios, las preposiciones poseen la particularidad de mostrar un comportamiento que está enormemente influido por el entorno lingüístico que les rodea, lo cual les infiere un alto grado de polisemia y de diversidad en su funcionamiento sintáctico, exigiendo prestar una mayor atención si cabe cuando se trata de áreas en las que se precisa cierta consistencia en la representación formal, como es el PLN.

El marco de actuación para la realización de esta tesis fue la base de conocimiento FunGramKB, la cual resultó de gran ayuda debido a que incorpora de manera integrada un modelo de ingeniería del conocimiento con un modelo de lingüística teórica, i.e. la RRG, para su aplicación al PLN. De hecho, la búsqueda de la tipicidad lingüística universal por parte del aparato lingüístico de la RRG otorga un carácter multilingüe a esta base de conocimiento léxico-conceptual, facilitando así que el modelo de análisis propuesto en esta tesis pueda ser aplicado en futuros estudios de otras lenguas. Por consiguiente, tanto la modelización conceptual de las relaciones

como el análisis lingüístico de las preposiciones en lengua inglesa adoptan un enfoque igualmente válido para el estudio de otras preposiciones en la misma lengua o en otras lenguas. Igualmente, aunque esta tesis ha utilizado el modelo de FunGramKB como marco donde ubicar mis investigaciones, debo destacar que los datos que se pueden obtener tras la aplicación de mi modelo pueden ser reutilizables en otras bases de conocimiento del PLN.

La cuestión de la desambiguación de las preposiciones ha sido objeto de estudio en diversas publicaciones. Desde el punto de vista gramatical, se aprecia un tratamiento bastante localista de las preposiciones, donde los resultados investigadores son difíciles de extrapolar a otras lenguas. En cambio, esta tendencia no coincide con los numerosos estudios encontrados en la perspectiva cognitiva, donde los análisis de corte cognitivo suelen proporcionar información semántica sobre esta clase léxica, aunque obviando su caracterización sintáctica. Finalmente, en el campo del PLN, existen diversos estudios y recursos orientados al tratamiento conceptual de los nombres y los verbos, pero son pocos los acercamientos al estudio de las preposiciones, bien sea por su complejidad o por la falta de interés que se les ha concedido tradicionalmente. Los proyectos que se han preocupado del tratamiento computacional de esta clase gramatical lo han hecho desde una perspectiva semántica, lo cual no ha sido suficiente para un tratamiento computacional más robusto de las mismas. En este sentido, y como se ha argumentado a lo largo de esta tesis, existía la necesidad de considerar la sintaxis de las preposiciones y sus elementos contextuales desde un enfoque conceptualista pero con una clara conexión con las entradas léxicas, con el fin de concederles un tratamiento más riguroso que pudiera ser útil para una aplicación del PLN.

Una de las dificultades que encontré al inicio de este estudio fue el alto número de sentidos ofrecidos por los recursos lexicográficos. Con el fin de proporcionar un tratamiento computacional adecuado que permitiera maximizar la información almacenada en la base de conocimiento y minimizar la complejidad del proceso de desambiguación léxica, adopté un enfoque donde el número de conceptos no fuese desmedido. De esta forma, evitaba que el procesamiento de las preposiciones no resultara en una explosión combinatoria poco viable, al mismo tiempo que no entorpeciera el proceso de inferencia semántica que se aplica en FunGramKB. Por esta

razón, frente al alto número de sentidos que suelen identificarse para las ocho preposiciones objeto de este estudio, con una media de 16 sentidos entre los diccionarios más populares de lengua inglesa, este trabajo proporciona 36 conceptos básicos distribuidos a lo largo de la subontología de las relaciones. De hecho, fueron dos las razones por las cuales se redujo enormemente el inventario de conceptos: (i) la fusión de sentidos similares y (ii) la eliminación de los sentidos asociados a construcciones argumentales, los cuales son tratados desde el Gramaticón. Con respecto al primer punto, debo señalar que usos como el de la preposición *in* en *He is in a red jacket* no han sido descritos en este trabajo porque se consideraron sentidos resultantes de extensiones metafóricas a partir de un significado prototípico. Por tanto, no distinguí entre, por ejemplo, el sentido de *in* en la oración anterior y el que posee en la oración *He is in the kitchen*. No obstante, la entrada conceptual puede servir como instrumento para la incorporación de aquellas preferencias de selección que pueden reflejar dichas extensiones semánticas. De esta forma, el objetivo de mi estudio se reducía igualmente a aquellas preposiciones que no expresaran sentidos metafóricos o construccionales.

La presente tesis ofrece, asimismo, algunas futuras líneas de investigación. Como apuntó Saint-Dizier (2006), existen numerosas regularidades interlingüísticas en el uso que las lenguas hacen de las preposiciones, y especialmente en el caso de las lenguas que pertenecen a una misma familia. En este sentido, este trabajo aporta un modelo para el tratamiento de las preposiciones a nivel computacional, de modo que, aunque en esta tesis me he ocupado del análisis lingüístico de algunas preposiciones en lengua inglesa, es posible continuar en la misma línea de este estudio pero considerando esta vez las idiosincrasias lingüísticas de otras lenguas que también son trabajadas en FunGramKB, como pueden ser el alemán, el búlgaro, el catalán, el español, el francés o el italiano. Por otra parte, de acuerdo con Coventry y Garrod (2004), la comprensión de los usos metafóricos de las preposiciones puede contribuir a un mejor manejo de su uso en los sentidos literales, por lo cual sería interesante analizar la aportación que esos usos metafóricos podrían realizar sobre el procesamiento de las preposiciones. Por último, habiendo incorporado durante el proceso de construcción de la subontología de las relaciones un número significativo de conceptos básicos vinculados a los diferentes usos de las preposiciones, se podría analizar la interacción entre la información almacenada en las entradas conceptuales y la información almacenada en las entradas

construccionales del Constructicón de nivel 1 de FunGramKB, relativa a las construcciones argumentales. De esta forma, y en la línea de los trabajos de Perinián-Pascual (2013b) y Perinián-Pascual y Arcas-Túnez (2014), se contribuiría a mejorar la interfaz sintáctico-semántica propuesta por la RRG integrando los diferentes conocimientos que FunGramKB posee en la Ontología, el Lexicón y el Gramaticón.



## BIBLIOGRAFÍA

- Agirre, E., Lersundi, M., & Martínez, D. (2002). A multilingual approach to disambiguate prepositions and case suffixes. En *Proceedings of the SIGLEX/SENSEVAL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, pp. 1-8, Philadelphia, USA.
- Alam, Y.S. (2004). Decision trees for sense disambiguation of prepositions: case of over. En *Proceedings of the HLT-NAACL, Computational Lexical Semantic Workshop Conference*, pp. 52-59, Boston, USA.
- Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832-843.
- Armstrong, S., Gleitman L., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, 13, 263-308.
- Asbury, A., Gehrke, B., Van Riemsdijk, H., & Zwarts, J. (2008). Introduction: syntax and semantics of spatial P. En A. Asbury, B. Gehrke & R. Nouwen (Eds.), *Syntax and semantics of spatial P* (Vol. 120, pp. 1-32). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Ávila, J. H. M. (2007). Estructura sintáctico-semántica de la relación preposicional: una aproximación al significado básico de las preposiciones de la lengua española. *Cuadernos de Lingüística Hispánica*, 10, 47-69.
- Baker, C.F., Fillmore, C.J., & Lowe, J. B. (1998). The Berkeley FrameNet project. En *Proceedings of the 17th International Conference on Computational Linguistics*, (Vol. 1) pp. 86–90, Montreal, Canada.
- Baldwin, T. (2006a). Distributional similarity and preposition semantics. En P. Saint-Dizier (ed.), *Computational linguistics dimensions of syntax and semantics of prepositions* (pp. 197-210). Netherlands: Springer.

- Baldwin, T. (2006b). Automatic identification of English verb particle constructions using linguistic features. En *Third ACL-SIGSEM Workshop on Prepositions*, pp. 65–72, Trento, Italy.
- Baldwin, T., Kordoni, V., & Villavicencio, A. (2009). Prepositions in applications: a survey and introduction to the special issue. *Computational Linguistics*, 35 (2), 119–149.
- Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). Abstract meaning representation for sembanking. En *Proceedings of the 7<sup>th</sup> Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 178-186, Sofia, Bulgaria.
- Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*, 10 (1), 82-93.
- Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 629-654.
- Barsalou, L.W. (1991). Deriving categories to achieve goals. En G. Bower (ed.) *The psychology of learning and motivation: advances in research and theory* (Vol. 27, pp. 1–64). San Diego: Academic Press.
- Barsalou, L.W. (2002). Being there conceptually: simulating categories in preparation for situated action. En N.L. Stein, P.J. Bauer & M. Rabinowitz (Eds.), *Representation, memory and development: essays in honor of Jean Mandler* (pp. 1-15). Mahwah: Lawrence Erlbaum.
- Barsalou, L.W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18, 513-562.
- Barsalou, L.W. (2005). Situated conceptualization. En H. Cohen, H., & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619-650). Netherlands: Elsevier.

- 
- Beavers, J. (2013). Aspectual classes and scales of change. *Linguistics*, 51 (4), 681-706.
- Bennet, D. (1975). *Spatial and temporal uses of English prepositions. An essay in stratificational semantics*. London: Longman.
- Biały, A. (2013). On the construal of prepositional phrases. En Z. Wasik, P. Czajka, & M. Szawerna (Eds.), *Alternate construals in language and linguistics* (pp.105-120). Wrocław: Philological School of Higher Education in Wrocław Publishing.
- Bonial, C., Corvey, W., Palmer, M., Petukhova, V.V., & Bunt, H. (2011). A hierarchical unification of LIRICS and VerbNet semantic roles. En *Fifth IEEE International Conference on Semantic Computing*, pp. 483-489, Palo Alto, USA.
- Boonthum, C., Toida, S., & Levinstein, I. (2005). Sense disambiguation for preposition 'with'. En *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pp.153-162, Colchester, England.
- Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75 (1), 1–28.
- Brée, D. S. (1992). Words for time. En F. Macar, V. Pouthas & W. Friedman (Eds.), *Time, action and cognition* (pp.337-348). Dordrecht, NL: Kluwer.
- Brée, D.S., & Pratt- Hartmann, I.E. (2002). Temporal semantics of prepositions in context. En S. Feigenbaum, & D. Kurzon (Eds.), *Prepositions in their syntactic, semantic and pragmatic context* (Vol. 50, pp. 75-115). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Brill, E., & Resnik, P. (1994). A rule-based approach to prepositional phrase attachment disambiguation. En *Proceedings of the 15th Conference on Computational Linguistics*, (Vol. 2) pp. 1198–1204, Morristown, USA.
- Brugman, C. (1981). *The story of over*. (Tesis Doctoral, Universidad de Berkeley).

- Brugman, C., & Lakoff, G. (1988). *The story of over: polysemy, semantics and the structure of the lexicon*. New York: Garland.
- Buitelaar, P. (2001). Semantic lexicons: between ontology and terminology. En *Proceedings of OntoLex 2000: Ontologies and Lexical Knowledge Bases*, pp. 16-24, Sofia, Bulgaria.
- Butler, C. (2004). Notes towards an incremental implementation of the Role and Reference Grammar semantics-to-syntax mapping rules for English. En M. Hannay, & G. J. Steen, (Eds.), *Structural-functional studies in English grammar* (pp. 275-307). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Calzolari, N., Lenci, A., & Zampolli, A. (2001a). The EAGLES/ISLE computational lexicon working group for multilingual computational lexicons. En *Proceedings of the First International Workshop on Multimedia Annotation*, pp. 34-39, Tokyo, Japan.
- Calzolari, N., Lenci, A., & Zampolli, A. (2001b). International standards for multilingual resource sharing: the ISLE computational lexicon working group. En *Proceedings of the ACL 2001 Workshop on Human Language Technology and Knowledge Management*, pp. 71-78, Morristown, USA.
- Cambria, E., & White, B. (2014). Jumping NLP curves: a review of natural language processing research. *IEEE Computational Intelligence Magazine*, 9 (2), 48–57.
- Cambridge Dictionary Online* [Versión electrónica]. Recuperado el 20 de noviembre de 2015, desde <http://dictionary.cambridge.org/>
- Cerrada, J. A., Collado, M., Gómez, S. R., & Estivariz, J. F. (2000). *Introducción a la ingeniería de software*. Madrid: Ramón Areces.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Massachusetts: MIT Press.

- Cifuentes, J. L. (1998). Verbos con incorporación conceptual direccional. En J. L. Cifuentes (ed). *Estudios de lingüística cognitiva*. 2 vol. (pp. 479-505). Alicante: Universidad de Alicante.
- Clark, H. H. (1973). Space, time, semantics, and the child. En T. E. Moore (ed.), *Cognitive development and the acquisition of language* (pp. 27-64). New York: Academic.
- Collins British English Dictionary Online* [Versión electrónica]. Recuperado el 20 de noviembre de 2015, desde <http://www.collinsdictionary.com/>
- Cortés-Rodríguez, F. J. (2014). Aspectual features in Role and Reference Grammar: a layered proposal. *Revista Española de Lingüística Aplicada*, 27 (1), 23-53.
- Coventry, K.R., & Garrod, S. (2004). *Saying, seeing and acting: the psychological semantics of spatial prepositions*. Hove: Psychology Press.
- Crocker, M. W. (1999). Mechanisms for sentence processing. En S. Garrod & M. J. Pickering (eds.) *Language processing* (pp. 191-232) .Londres: Psychology Press.
- Díaz-Galán, A., & Fumero-Pérez, M.C. (2016). Developing parsing rules within ARTEMIS: the case of DO auxiliary insertion. En C. Perrián-Pascual & E. M. Mestre-Mestre (eds.) *Understanding meaning and knowledge representation: from theoretical and cognitive linguistics to natural language processing* (pp. 283-304). Newcastle: Cambridge Scholars Publishing.
- Diedrichsen, E. (2016). Does NLP need theoretical linguistics?. En C. Perrián-Pascual & E. M. Mestre-Mestre (eds.) *Understanding meaning and knowledge representation: from theoretical and cognitive linguistics to natural language processing* (pp. 249-260). Newcastle: Cambridge Scholars Publishing.
- Dik, S. C. (1978). *Functional grammar*. Amsterdam: North-Holland.
- Dik, S. C. (1989). *The theory of functional grammar. Part 1. The structure of the clause*. Dordrecht: Foris.

- Dixon, R. M. W. (1991). *A new approach to English grammar on semantic principles*. Oxford: Clarendon Press.
- Dorr, B.J. (1993). *Machine translation: a view from the lexicon*. Cambridge, Massachusetts: MIT Press.
- Dorr, B. J. (1994). Machine translation divergences: A formal description and proposed solution. *Computational Linguistics*, 20 (4), 597-633.
- Dorr, B.J. (1997). Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. *Machine Translation*, 12 (4), 271–322.
- Dorr, B.J. (2001). *LCS Verb Database*. (Online Software Database of Lexical Conceptual Structures and Documentation). Universidad de Maryland. Recuperado de:  
[http://www.umiacs.umd.edu/~bonnie/Demos/LCS\\_Database\\_Documentation.html](http://www.umiacs.umd.edu/~bonnie/Demos/LCS_Database_Documentation.html)
- Downing, A., & Locke, P. (2006). Semantic features of the prepositional phrase. En A. Downing & P. Locke (Eds.), *English grammar: a university course* (2° ed., pp. 546-559). London/New York: Routledge.
- Dowty, D. (1979). *Word meaning and montague grammar*. Dordrecht: Foris.
- EAGLES (1999). *EAGLES LE3-4244 Preliminary recommendations on lexical semantic encoding final report*. Recuperado de:  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.5838&rep=rep1&type=pdf>
- Evans, V., & Green, M. (Eds.). (2006). *Cognitive linguistics. An introduction*. Edinburgh: Edinburgh University Press.
- Filip, H. (1999). *Aspect, eventuality types and nominal reference*. New York: Garland Publishing.

- 
- Filip, H. (2011). Aspectual class and Aktionsart. En C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: an international handbook of natural language meaning* (pp. 1186-1217). Berlin/Nueva York: Mouton de Gruyter.
- Fillmore, C.J. (1968). The case for case. En E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). New York: Holt, Rinehart, & Winston.
- Fillmore, C. (1982). Frame semantics. En Linguistic Society of Korea (ed.), *Linguistics in the morning calm* (pp. 111–138). Seoul: Hanshin.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6 (2), 222-253.
- Fillmore, C.J., Wooters, C., & Baker, C.F. (2001). Building a large lexical databank which provides deep semantics. En *Proceedings of the Pacific Asian Conference on Language, Information and Computation*, pp. 3-25, Hong Kong, China.
- Foley, W.A., & Van Valin, R.D. Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.
- Frazier, L. (1979). *On comprehending sentences: syntactic parsing strategies*. (Tesis Doctoral, Universidad de Connecticut).
- Frazier, L., & Fodor, J.D. (1978). The sausage machine: a new twostage parsing model. *Cognition*, 6, 1-34.
- Gibbon, D. (2000). Computational lexicography. En F. van Eynde & D. Gibbon (eds.) *Lexicon development for speech and language processing* (pp. 1-42). Dordrecht: Kluwer Academic Publishers.
- Girju, R. (2009). The syntax and semantics of prepositions in the task of automatic interpretation of nominal phrases and compounds: a cross-linguistic study. *Computational Linguistics*, 35 (2), 185-228.
- Goldberg, A.E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: Chicago University Press.

- Goldberg, A.E. (2005). Argument realization: the role of constructions, lexical semantics and discourse factors. En J.O. Oostman & M. Fried (Eds.), *Construction grammar(s): cognitive and cross-language dimensions* (pp. 17-40). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Guest, E. (2009). Parsing Using the Role and Reference Grammar Paradigm. The *13th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI*, Orlando, USA. Recuperado de: <http://eprints.leedsbeckett.ac.uk/778/6/Parsing%20Using%20the%20Role%20and%20Reference%20Grammar%20Paradigm.pdf>
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, 3, 37–81.
- Halliday, M. A. K. (1994). *An introduction to functional grammar* (2<sup>o</sup> Ed). London: Edward Arnold.
- Hanks, P. (2004). Corpus pattern analysis. En *Proceedings of the 11<sup>th</sup> EURALEX International Conference*, pp. 87-98, Lorient, France.
- Harabagiu, S. (1996). An application of WordNet to prepositional attachment. En *Proceedings of the 34th Annual Meeting on Association for Computational Linguistics*, pp. 360-363, Santa Cruz, USA.
- Hartrumpf, S., Helbig, H., & Osswald, R. (2006). Semantic interpretation of prepositions for NLP applications. En *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*, pp. 29-36, Trento, Italy.
- Herskovits, A. (1986). *Language and spatial cognition. An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.
- Herskovits, A. (1997). Language, spatial cognition, and vision. En O. Stock (ed.), *Temporal and spatial reasoning* (pp.155-202). Dordrecht: Kluwer Academic Press.

- 
- Higinbotham, D.W. (1990). *Semantic cooccurrence networks and the automatic resolution of lexical ambiguity in machine translation*. (Tesis Doctoral, Universidad de Texas).
- Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, 19 (1), 103-120.
- Hopper, P. (1987). Emergent grammar. En *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, pp.139–57, Berkeley, USA.
- Hovy, E. (2002). Comparing sets of semantic relations in ontologies. En R.Green, C. A. Bean & S. H. Myaeng (eds.) *Semantics of relationships: an interdisciplinary perspective* (pp. 91-110). Dordrecht: Kluwer Academic Publishers.
- Hovy, D., Tratz, S., & Hovy, E. (2010). What's in a preposition? Dimensions of sense disambiguation for an interesting word class. En *Proceedings of the 23<sup>rd</sup> International Conference on Computational Linguistics: Posters*, pp. 454-462, Beijing, China.
- Hovy, D., Vaswani, A., Tratz, S., Chiang, D., & Hovy, E. (2011). Models and training for unsupervised preposition sense disambiguation. En *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*, (Vol. 2) pp. 323-328, Oregon, USA.
- Huddleston, R., & Pullum, G.K. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Ide, N., & Véronis, J. (1998). Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*, 24 (1), 1-40.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, Massachusetts: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge, Massachusetts: MIT Press.
- Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41, 9–45.

- Jensen, K., & Binot, J. L. (1987). Disambiguating prepositional phrase attachments by using on-line dictionary definitions. *Computational Linguistics*, 13 (4), 251-260.
- Jiménez-Briones, R. (2008). Los metalenguajes y las representaciones léxico-semánticas: una propuesta desde el Modelo Léxico-Construccional. En *El valor de la diversidad (meta) lingüística: Actas del VIII Congreso de Lingüística General*, pp. 1041-1060, Madrid, España.
- Jiménez-Briones, R., Luzondo, A., & Pérez-Cabello de Alba, M.B. (2011). FunGramKB y la organización ontológica. *Anglogermánica Online*, 8, 16-36.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge: Harvard University Press.
- Jolly, J. (1991). *Prepositional analysis within the framework of Role and Reference Grammar*. Berlin: Peter Lang.
- Jolly, J. (1992). Preposition assignment in English. En J.R. Van Valin (ed.), *Advances in Role and Reference Grammar* (pp. 275-310). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Katz, J. J., & Fodor, J. A. (1963). The structure of a semantic theory. *Language*, 39 (2), 170-210.
- Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, 42, 21-40.
- Krifka, M. (1998). The origins of telicity. En S. Rothstein (ed.), *Events and grammar* (pp. 197-235). Dordrecht: Kluwer Academic Press.
- Kuno, S. (1975). Three perspectives in the functional approach to language. En R. E. Grossman, L.J. San, & T.J. Vance (Eds.), *Papers from the parasession on functionalism* (pp. 276-336). Chicago: Chicago University Press.
- Lakoff, G. (1987). *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: Chicago University Press.

- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Landau, B. (1982). Will the real grandmother please stand up? The psychological reality of dual meaning representations. *Journal of Psycholinguistic Research*, 11 (1), 47-62.
- Landau, B., & Jackendoff, R. (1993). Whence and whither in spatial language and spatial cognition?. *Behavioral and Brain Sciences*, 16 (2), 255-265.
- Langacker, R. W. (1986). An introduction to cognitive grammar. *Cognitive Science*, 10 (1), 1-40.
- Leech, G., Rayson, P., & Wilson, A. (2001). *Word frequencies in written and spoken English: based on the British National Corpus*. London: Longman.
- Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowski, A., Peters, I., Peters, W., Ruimy, N., Villegas, M., & Zampolli, A. (2000). SIMPLE: a general framework for the development of multilingual lexicons. *International Journal of Lexicography*, 13 (4), 249-263.
- Lersundi, M., & Agirre, E. (2003). Semantic interpretations of postpositions and prepositions: a multilingual inventory for Basque, English and Spanish. En *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and Their Use in Computational Linguistics Formalisms and Applications*, pp. 56–65, Toulouse, France.
- Levin, B. (1993). *English verb classes and alternations*. Chicago: Chicago University Press.
- Lin, J. (2004). *Event structure and the encoding of arguments: the syntax of the mandarin and English verb phrase* (Tesis Doctoral, Instituto Tecnológico de Massachusetts).
- Lindstromberg, S. (2010). *English prepositions explained*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Litkowski, K. C., & Hargraves, O. (2005). The Preposition Project. En *Proceedings of the 2nd ACL-SIGSEM Workshop on Prepositions*, pp. 171-179, Colchester, England.
- Litkowski, K. C., & Hargraves, O. (2007). The SemEval-2007 Task 06: word sense disambiguation of prepositions. En *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 24-29, Prague, Czech Republic.
- Litkowski, K.C. (2013a). *The Preposition Project corpora*. Technical report 13-01, Damascus, MD: CL Research. Recuperado de: <http://www.clres.com/online-papers/TPPCorpora.pdf>
- Litkowski, K.C. (2013b). *Preposition disambiguation: still a problem*. Technical report 13-02, Damascus, MD: CL Research. Recuperado de: <http://www.clres.com/online-papers/PrepWSD2013.pdf>
- Litkowski, K.C. (2014). Pattern dictionary of English prepositions. En *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, pp. 1274-1283, Baltimore, USA.
- Longman Dictionary of Contemporary English* [Versión electrónica]. Recuperado el 20 de noviembre de 2015, desde <http://www.ldoceonline.com/>
- Lytinen, S.L. (1987). Integrating syntax and semantics. En *Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, pp. 167-178, Hamilton, New York.
- Mahesh, K. (1995). Syntax-semantics interaction in sentence understanding (Tesis Doctoral, Instituto Tecnológico de Georgia).
- Mairal-Usón, R., & Faber, P. (2002). Functional grammar and lexical templates. En R. Mairal-Usón & M.J., Pérez-Quintero (Eds.), *New perspectives on argument structure in functional grammar* (pp. 41-98). Berlin: Mouton de Gruyter.

- Mairal-Usón, R., & Faber, P. (2005). Decomposing semantic decomposition. En *Proceedings of the 2005 International Conference on Role and Reference Grammar*, pp. 279-308. Taiwan, China.
- Mairal-Usón, R., & Faber, P. (2007). Lexical templates within a functional cognitive theory of meaning. *Annual Review of Cognitive Linguistics*, 5, 137–172.
- Mairal-Usón, R., & Gil, J. (Eds.). (2004). *En torno a los universales lingüísticos*. Madrid: Ediciones Akal.
- Mairal-Usón, R., & Perrián-Pascual, J.C. (2009). The anatomy of the lexicon within the framework of an NLP knowledge base. *Revista Española de Lingüística Aplicada*, 22, 217-244.
- Mairal-Usón, R., & Perrián-Pascual, J.C. (2010). Teoría lingüística y representación del conocimiento: una discusión preliminar. En D. García Padrón & M. C. Fumero Pérez (Eds.), *Tendencias en lingüística general y aplicada* (pp. 155-168). Berlín: Peter Lang.
- Mairal-Usón, R., & Ruiz de Mendoza, F. (2009). Levels of description and explanation in meaning construction. En C. Butler & J. Martín Arista (Eds.), *Deconstructing Constructions* (pp.153-198). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Mairal-Usón, R., Perrián-Pascual, C., & Pérez-Cabello de Alba, M. B. (2012). La representación léxica. Hacia un enfoque ontológico. En R. Mairal-Usón, L. Guerrero & C. González (Eds.), *El funcionalismo en la teoría lingüística. La Gramática del Papel y la Referencia. Introducción, avances y aplicaciones* (pp. 85-102). Madrid: Ediciones Akal.
- Mandler, J. (2004). *The foundations of mind: origins of conceptual thought*. Oxford: Oxford University Press.
- Marcus, M., Marcinkiewicz, M.A., & Santorini, B. (1993). The Penn Treebank: Annotating predicate argument structure. *Computational Linguistics*, 19(2), 313-330.

- McCormick, S. (2002). *The Structure and Content of the Body of an OLIF v.2.0/2.1*. The OLIF2 Consortium. Recuperado de: <http://www.olif.net/documents/NewOLIFstruct&content.pdf>
- McCormick, S., Lieske, C., & Culumm, A. (2004). *OLIF v.2: A Flexible Language Data Standard*. The OLIF2 Consortium. Recuperado de: [http://www.olif.net/documents/OLIF\\_Term\\_Journal.pdf](http://www.olif.net/documents/OLIF_Term_Journal.pdf)
- Merriam-Webster Dictionary* [Versión electrónica]. Recuperado el 20 de noviembre de 2015, desde <http://www.merriam-webster.com/>
- Miller, G., & Teibel, D. (1991). A proposal for lexical disambiguation. En *Proceedings of the DARPA Speech and Natural Language Workshop*, pp. 395-399, Washington D.C., USA.
- Minsky, M. (1975). *A framework for representing knowledge*. MIT-AI Laboratory Memo 306. Recuperado de: <https://dspace.mit.edu/handle/1721.1/6089>
- Mokhov, S.A., & Rudzic, F. (2003). *Towards a heuristic categorization of prepositional phrases in English with WordNet*. Technical report. Cornell University. Recuperado de: <http://arxiv.org/pdf/1002.1095v1.pdf>
- Monachini, M., Bertagna, F., Calzolari, N., Underwood, N., & Navarretta, C. (2003). *Towards a Standard for the Creation of Lexica*. En Istituto di linguistica computazionale "Antonio Zampolli". European Language Resources Association (ELRA). Recuperado de: [http://elra.info/media/filer\\_public/2013/12/04/standard\\_lexica.pdf](http://elra.info/media/filer_public/2013/12/04/standard_lexica.pdf)
- Montiel-Ponsoda, E. (2011). *Multilingualism in Ontologies: Multilingual Lexico-Syntactic Patterns for Ontology Modeling and Linguistic Information Repository for Ontology Localization*. (Tesis Doctoral, Universidad Politécnica de Madrid).
- Montiel-Ponsoda, E., Aguado de Cea, G., Gómez-Pérez, A., & Peters, W. (2008). Modelling multilinguality in ontologies. En *Proceedings of the 21st International Conference on Computational Linguistics*, pp. 67-70, Manchester, United Kingdom.

- Montiel-Ponsoda, E., Aguado-de-Cea, G., Gómez-Pérez, A. & Peters, W. (2011). Enriching ontologies with multilingual information. *Natural Language Engineering*, 17 (3), 283-309.
- Mora-Bustos, A. (2012). Los adjuntos periféricos. En R. Mairal-Usón, L. Guerrero & C. González (Eds.), *El funcionalismo en la teoría lingüística: La Gramática del Papel y la Referencia. Introducción, avances y aplicaciones* (pp. 203-223). Madrid: Ediciones Akal.
- Nam, S. (2000). A typology of locatives and event composition in English. *Language Research- Seoul*, 36 (4), 689 – 714.
- Navarro, G. Z. L., & Mora-Bustos, A. (2015). Estructura gramatical de los adjuntos. *Estudios de Lingüística Aplicada*, 61, 225-252.
- Nichols, J. (1984). Functional theories of grammar. *Annual Review of Anthropology*, 13, 97-117.
- Nirenburg, S., & Raskin, V. (2004). *Ontological semantics*. Cambridge: Massachusetts, MIT press.
- Nolan, B. (2011). The layered structure of the Irish word: an RRG account of inflectional and derivation processes. En W. Nakamura (ed.), *New Perspectives in Role and Reference Grammar* (pp. 64-103). Cambridge: Cambridge Scholars Publishing.
- Nolan, B. (2016). What can theoretical linguistics do for natural language processing research En C. Perrián-Pascual & E. M. Mestre-Mestre (eds.) *Understanding Meaning and Knowledge Representation: From Theoretical and Cognitive Linguistics to Natural Language Processing* (pp. 235-248). Newcastle: Cambridge Scholars Publishing.
- Nolan, B., & Salem, Y. (2009). UniArab: an RRG Arabic-to-English machine translation software. En *Proceedings of the Role and Reference Grammar International Conference*, pp. 243-270, Berkeley, USA.

- O'Dowd, E. M. (1998). *Prepositions and particles in English: a discourse-functional account*. Oxford: Oxford University Press.
- O'Hara, T., & Wiebe, J. (2003). Preposition semantic classification via Penn Treebank and FrameNet. En *Proceedings of Conference on Natural Language Learning - 03*, pp. 79–86, Edmonton, Canada.
- O'Hara, T., & Wiebe, J. (2009). Exploiting semantic role resources for preposition disambiguation, *Computational Linguistics*, 35 (2), 151-184.
- Osherson, D., & Smith, E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35-58.
- Pérez-Hernández, L. (2009). Análisis léxico-construccional de verbos de habla. *Círculo de Lingüística Aplicada a la Comunicación*, 40, 62-93.
- Pérez-Hernández, L., & Ruiz de Mendoza, F. J. (2011). A lexical-constructional model account of illocution. *Vigo International Journal of Applied Linguistics*, 8, 99-137.
- Periñán-Pascual, C. (2012a). The situated common-sense knowledge in FunGramKB. *Review of Cognitive Linguistics*, 10 (1), 184-214.
- Periñán-Pascual, C. (2012b). En defensa del procesamiento del lenguaje natural fundamentado en la lingüística teórica. *Onomázein*, 26, 13-48.
- Periñán-Pascual, C. (2013a). A knowledge-engineering approach to the cognitive categorization of lexical meaning. *VIAL: Vigo International Journal of Applied Linguistics*, 10, 85-104.
- Periñán-Pascual, C. (2013b). Towards a model of constructional meaning for natural language understanding. En B. Nolan & E. Diedrichsen (Eds.), *Linking constructions into functional linguistics: the role of constructions in a functional grammar* (pp. 205-230). Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Periñán-Pascual, C., & Arcas-Túnez, F. (2004). Meaning postulates in a lexico-conceptual knowledge base. En *Proceedings 15th International Workshop on Databases and Expert Systems Applications*, pp. 38-42, Los Alamitos, USA.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2005). Microconceptual-Knowledge spreading in FunGramKB. En *9th IASTED International Conference on Artificial Intelligence and Soft Computing*, ACTA Press, pp. 239-244, Anaheim-Calgary-Zurich.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2007). Cognitive modules of an NLP knowledge base for language understanding. *Procesamiento del Lenguaje Natural*, 39, 197-204
- Periñán-Pascual, C., & Arcas-Túnez, F. (2008). A cognitive approach to qualities for NLP. *Procesamiento del Lenguaje Natural*, 41, 137-144.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2010a). The Architecture of FunGramKB. En *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pp. 2667-2674, Valetta, Republic of Malta.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2010b). Ontological commitments in FunGramKB. *Procesamiento del Lenguaje Natural*, 44, 27-34.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2011). Introduction to FunGramKB. *Anglogermánica Online*, 8, 1-15.
- Periñán-Pascual, C., & Arcas-Túnez, F. (2014). The implementation of the FunGramKB CLS constructor in ARTEMIS. En B. Nolan & C. Periñán-Pascual (Eds.), *Language processing and grammars: the role of functionally oriented computational models* (pp. 165-196). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Periñán-Pascual, C., & Mairal-Usón, R. (2009). Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural*, 43, 265-273.

- Periñán-Pascual, C., & Mairal-Usón, R. (2010a). Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural*, 44, 19-26.
- Periñán-Pascual, C., & Mairal-Usón, R. (2010b). La gramática de COREL: un lenguaje de representación conceptual. *Onomázein*, 21, 11-45.
- Periñán-Pascual, C., & Mairal-Usón, R. (2011). The COHERENT methodology in FunGramKB. *Onomázein*, 24, 13-33.
- Periñán-Pascual, C., & Mairal-Usón, R. (2012). La dimensión computacional de la Gramática del Papel y la Referencia: la estructura lógica conceptual y su aplicación en el procesamiento del lenguaje natural. En R. Mairal-Usón, L. Guerrero & C. González (Eds.), *El funcionalismo en la teoría lingüística: la Gramática del Papel y la Referencia. Introducción, avances y aplicaciones* (pp. 333-348). Madrid: Ediciones Akal.
- Piñón, C. J. (1993). Paths and their names. En *Proceedings of the 29<sup>th</sup> Regional Meeting of the Chicago Linguistic Society*, 29 (2), pp. 287-303, Chicago, USA.
- Popescu, O., Tonelli, S., & Pianta, E. (2007). IRST-BP: preposition disambiguation based on chain clarifying relationships contexts. En *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 191-194, Prague, Czech Republic.
- Pratt, I.E., & Brée, D. S. (1995). *Representing time in natural language*. Cambridge, Massachusetts: MIT Press.
- Pratt, I., & Francez, N. (2001). Temporal prepositions and temporal generalized quantifiers. *Linguistics and Philosophy*, 24 (2), 187-222.
- Pressman, R. S. (1998). *Ingeniería del software. Un enfoque práctico*. Madrid: McGraw-Hill.
- Procter, P. (1978). *Longman dictionary of contemporary English*. Harlow: Longman.

- 
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17 (4), 409-441.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, Massachusetts: MIT Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Rauh, G. (1993). On the grammar of lexical and non-lexical prepositions in English. En C. Zelinsky-Wibbelt (ed.), *The semantics of prepositions: from mental processing to natural language processing*, (pp. 99- 150). Berlin: Mouton de Gruyter.
- Rauh, G. (2010). *Syntactic categories: their identification and description in linguistic theories*. Oxford: Oxford University Press.
- Reichenbach, H. (1947). *Elements of symbolic logic*. New York: MacMillan & Co.
- Resnik, P., & Hearst, M. (1993). Structural ambiguity and conceptual relations. En *Proceedings of the Workshop on Very Large Corpora: Academic and Industrial Perspectives*, pp. 58-64, Columbus, Ohio.
- Rothstein, S. (2004). *Structuring events: A study in the semantics of lexical aspect*. Oxford: Blackwell.
- Ruiz de Mendoza, F. J., & Mairal-Usón, R. (2006). Levels of semantic representation: where lexicon and grammar meet. *Interlingüística*, 17, 26-47.
- Ruiz de Mendoza, F.J., & Mairal-Usón, R. (2008). Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica*, 42 (2), 355-400.
- Rumelhart, D.E. (1975). Notes on a schema for stories. En D.G. Bobrow & A.M. Collins (eds.), *Representation and understanding: studies on cognitive science* (pp. 211-236). New York: Academic Press.

- Sacramento, E. (2014). *Computational modelling of sensory perception qualities in an NLP system: the metaconcept #QUALITY in FunGramKB* (Tesis Doctoral, Universidad de La Laguna).
- Saint-Dizier, P. (2005). PrepNet: a framework for describing prepositions. Preliminary investigation results. En *Proceedings of the Sixth International Workshop on Computational Semantics*, pp. 145-157, Tilburg, Netherlands.
- Saint-Dizier, P. (2008). Syntactic and semantic frames in PrepNet. En *Third International Joint Conference on Natural Language Processing*, pp. 763-768, Hyderabad, India.
- Salem, Y., Hensman, A., & Nolan, B. (2008). Implementing Arabic-to-English machine translation using the Role and Reference Grammar linguistic model. En *Proceedings of the Eighth Annual International Conference on Information Technology and Telecommunication*, pp.103-110, Galway, Ireland.
- Schank, R.C., & Abelson, R.P. (1977). *Scripts, plans, goals and understanding: an inquiry into human knowledge structures*. Hillsdale: Lawrence Erlbaum.
- Schneider, N. (2014). *Lexical semantic analysis in natural language text* (Tesis Doctoral, Universidad de Melbourne).
- Schneider, N., Srikumar, V., Hwang, J.D., & Palmer, M. (2015). A hierarchy with, of and for preposition Supersenses. En *Proceedings of the 9th Linguistic Annotation Workshop*, pp. 112-123, Denver, USA.
- Slabakova, R. (2000). L1 transfer revisited: the L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics*, 38, 739-770.
- Smith, C. (1997). *The parameter of aspect*. Dordrecht: Reidel.
- Smith, E., Osherson, D., Rips, L., & Keane, M. (1988). Combining prototypes: a selective modification model. *Cognitive Science*, 12, 485-527.
- Spivey, M. (2007). *The continuity of mind*. New York: Oxford University Press.

- 
- Srikumar, V., & Roth, D. (2011). A joint model for extended semantic role labeling. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 129-139, Edinburgh, Scotland.
- Srikumar, V., & Roth, D. (2013a). *An inventory of preposition relations*. Technical report. Cornell University. Recuperado de: <http://arxiv.org/abs/1305.5785v1>
- Srikumar, V., & Roth, D. (2013b). Modelling semantic relations expressed by prepositions. *Transactions of the Association for Computational Linguistics*, 1, 231–242.
- Suppes, P. (1979). Logical inference in English: a preliminary analysis. *Studia Logica*, 38 (4), 375-391.
- Suppes, P. (2005). The syntax and semantics of English prepositional phrases. En D. Oderberg (ed.), *The old new logic: essays on the philosophy of Fred Sommers* (pp. 101-109). Cambridge, Massachusetts: MIT Press.
- Svenonius, P. (2008). Projections of P. En A. Asbury, J. Dotlačil, B. Gehrke & R. Nouwen (eds.), *Syntax and semantics of spatial P* (pp. 63-84). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. En T. Shopen (ed.), *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon* (pp. 66-168). Cambridge: Cambridge University Press.
- Talmy, L. (1991). Path to realization: a typology of event conflation. En *Proceedings of the Berkeley Linguistic Society*, v. 17, pp. 480-519, Berkeley, California.
- Talmy, L. (2000). *Toward a cognitive semantics. Typology and process in concept structuring*. Vol. 2. Cambridge, Massachusetts: MIT Press.
- The new Oxford dictionary of English* (1997). J. Pearsall, & P. Hanks (eds.). Oxford: Clarendon Press.

- The Oxford dictionary of English* (2003). A. Stevenson, & C. Soanes (eds.). Oxford: Clarendon Press
- Tratz, S., & Hovy, D. (2009). Disambiguation of preposition sense using linguistically motivated features. En *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium*, pp. 96–100, Boulder, USA.
- Tratz, S., & Hovy, D. (2011). A fast, accurate, non-projective, semantically-enriched parser. En *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1257–1268, Edinburgh, Scotland.
- Tulving, E. (1985). How many memory systems are there?. *American Psychologist*, 40, 385-398.
- Tungseth, M. E. (2006). Verbal prepositions in Norwegian. Paths, places and possession (Tesis Doctoral, Universidad de Tromsø).
- Tyler, E., & Evans, V. (2003). *The semantics of English prepositions: spatial scenes, embodied meaning and cognition*. Cambridge: Cambridge University Press.
- Van Valin, R. D. Jr. (1993). *Advances in Role and Reference Grammar*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Van Valin, R. D. Jr. (2003). Functional linguistics. En M. Aronoff & J. Rees-Miller (eds.), *Handbook of Linguistics* (pp. 319-336). Oxford: Blackwell Publishers
- Van Valin, R. D. Jr. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
- Van Valin, R. D. Jr. (2008). *Investigations of the syntax-semantics-pragmatics surface*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Van Valin Jr, R. D. (2009). *An overview of Role and Reference Grammar*. Recuperado de: [http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/RRG\\_overview.pdf](http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/RRG_overview.pdf)

- Van Valin, R. D. Jr., & LaPolla, R. J. (1997). *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.
- Velardi, P., Fasolo, M., & Pazienza, M. T. (1991). How to encode semantic knowledge: a method for meaning representation and computer-aided acquisition. *Computational Linguistics*, 17 (2), 153-170.
- Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca: Cornell University Press.
- Von Stechow, A. (2002). Temporal prepositional phrases with quantifiers: some additions to Pratt and Francez (2001). *Linguistics and Philosophy*, 25, 755-800.
- Wilks, Y. (1972). *Grammar, meaning and the machine analysis of language*. London/Boston: Routledge.
- Wintner, S. (2009). What science underlies natural language engineering?. *Computational Linguistics*, 35 (4), 641-644.
- Ye, P., & Baldwin, T. (2007). MELB-YB: preposition sense disambiguation using rich semantic features. En *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 241-244, Prague, Czech Republic.
- Yuret, D. (2007). KU: word sense disambiguation by substitution. En *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 207-213, Prague, Czech Republic.
- Zelinski-Wibbelt, C. (ed.) (1993). *The semantics of prepositions: from mental processing to natural language processing*. Berlin: Mouton de Gruyter.
- Zhang, N. (1996). Semantic structures of English locative prepositional phrases. *Toronto Working Papers in Linguistics*, 15, 141-152.
- Zwarts, J. (2005). Prepositional aspect and the algebra of paths. *Linguistics and Philosophy*, 28 (6), 739-779.



## **ANEXOS**



## Anexo I. La interpretación semántica de los papeles temáticos de los argumentos en FunGramKB.<sup>31</sup>

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.

<sup>31</sup> La tabla de este anexo fue originalmente publicada por Periñán-Pascual & Mairal-Usón (2010b, pp. 42-43).



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. La interpretación semántica de los papeles temáticos de los satélites en FunGramKB.<sup>32</sup>

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.

<sup>32</sup> La tabla de este anexo fue originalmente publicada por Periñán-Pascual & Mairal-Usón (2010b, p. 44).



### Anexo III. Sentidos ofrecidos por PDEP para la preposición *in*.

#	FN	OEC	TPP	Pattern & primary implicature
1(1)	769	16	114	[[Governor]] in [[Complement]] [[Governor]] expressing the situation of something that is or appears to be enclosed or surrounded by something else [[Complement]]
1(1)-1	33	null	2	[[Governor]] in [[Complement]] [[Governor]] used to indicate articles being worn [[Complement]]
2(1a)	252	18	4	[[Governor]] in [[Complement]] [[Governor]] expressing motion with the result that something ends up within or surrounded by something else [[Complement]]
3(2)	99	12	88	[[Governor]] in [[Complement]] [[Governor]] expressing a period of time during which an event happens or a situation remains the case [[Complement]]
4(3)	2	7	null	[[Governor]] in [[Complement]] [[Governor]] expressing the length of time before a future event is expected to happen [[Complement]]
5(4)	142	10	36	[[Governor]] in [[Complement]] [[Governor]] (often followed by a noun without a determiner) expressing a state or condition [[Complement]]
6(4a)	214	2	41	[[Governor]] in [[Complement]] [[Governor]] indicating the quality or aspect with respect to which a judgement is made [[Complement]]
6(4a)-1	125	null	29	[[Governor]] in [[Complement]] [[Governor]] indicating the manner in which or the degree to which something happens [[Complement]]
7(5)	334	9	111	[[Governor]] in [[Complement]] [[Governor]] expressing inclusion or involvement [[Complement]]
8(6)	28	6	9	[[Governor]] in [[Complement]] [[Governor]] indicating a profession or a specialization within a profession [[Complement]]
9(7)	61	7	12	[[Governor]] in [[Complement]] [[Governor]] indicating the language or medium used [[Complement]]
9(7)-1	14	null	1	[[Governor]] in [[Complement]] [[Governor]] used to indicate a variable quality in clothing or personal items [[Complement]]
10(7a)	null	3	null	[[Governor]] in [[Complement]] [[Governor]] indicating the key in which a piece of music is written [[Complement]]
11(8)	16	6	21	[[Governor]] in [[Complement]] [[Governor]] as an integral part of (an activity) [[Complement]]
12(9)	null	4	2	[[Governor]] in [[Complement]] [[Governor]] expressing a value as a proportion of (a whole) [[Complement]]



## Anexo IV. Sentidos ofrecidos por *Supersenses* para la preposición *in*.

Definition	Example	Supersense
In 1(1)	expressing the situation of something that is or appears to be enclosed or surrounded by something else:	SST-Location,Theme
In 1(1)-1	used to introduce someone's clothing	SST-Attribute
In 10(7a)	indicating the key in which a piece of music is written:	SST-Attribute
In 11(8)	as an integral part of (an activity):	SST-Circumstance
In 12(9)	expressing a value as a proportion of (a whole):	SST-Superset,Value
In 2(1a)	expressing motion with the result that something ends up within or surrounded by something else:	SST-Location
In 3(2)	expressing a period of time during which an event happens or a situation remains the case:	SST-DeicticTime,Time
In 4(3)	expressing the length of time before a future event is expected to happen:	SST-DeicticTime
In 5(4)	(often followed by a noun without a determiner) expressing a state or condition:	SST-Age,State
In 6(4a)	indicating the quality or aspect with respect to which a judgement is made:	SST-Activity,Attribute,Patience,Stimulus,Topic
In 6(4a)-1	indicating manner or degree	SST-Contour,Manner
In 7(5)	expressing inclusion or involvement:	SST-Activity,Locus,Superset
In 8(6)	indicating someone's occupation or profession:	SST-ProfessionalAspect
In 9(7)	indicating the language or medium used:	SST-Attribute
In 9(7)-1	used to indicate set descriptions for clothing and other personal items	SST-Attribute
In X1	expressing a duration of an event, or an upper limit on the duration (not an interval from the present time)	SST-Duration
In X2	directional particle	SST-Direction
In X3	inside an unspecified place; on the premises	SST-Location
In X4	Wiktionary: "Indicates, connotatively, a place-like form of someone's (or something's) personality, as his, her or its psychic and physical characteristics."	SST-Instance



**Anexo V. Dimensión cognitiva #TIME ilustrada con unidades léxicas**

METACONCEPTO	METACONCEPTO	METACONCEPTO	METACONCEPTO	CONCEPTOS BÁSICOS	UNIDADES LÉXICAS
#TIME	#DURATION	#DURATION_F1	#DURATION_F1_1	+FOR_00	For
			#DURATION_F1_2	+IN_00	In
		#DURATION_F2	#DURATION_F2_1	+IN_01	In
			#DURATION_F2_2	+WITHIN_00	Within
	#INTERVAL	#INTERVAL_F1	-	+ON_00	On
			#INTERVAL_F1_1	+AT_00	At
			#INTERVAL_F1_2	+IN_02	In, during
		#INTERVAL_F2	#INTERVAL_F2_1	+UNTIL_00	Until, till, to
			#INTERVAL_F2_2	+BETWEEN_00	Between
				+BY_00	By
			#INTERVAL_F2_3	+FROM_00	From
				+SINCE_00	Since



### Anexo VI. Dimensión cognitiva #PLACE ilustrada con unidades léxicas

METACONCEPTO	METACONCEPTO	METACONCEPTO	CONCEPTOS BÁSICOS	UNIDADES LÉXICAS
#LOCUS	#PLACE	#EXTERIOR	+OUT_00	Out (of), outside
		#INFERIORITY	+BELOW_00	Below, beneath, under underneath
		#INTERIOR	+AT_01	At
			+IN_03	In, inside
		#INTERLOCAL	+AMONG_00	Among
			+BETWEEN_01	Between
		#LENGTH	+ALONG_00	Along
			+FOR_01	For
		#OFF_LIMITS	+BEYOND_00	Beyond
		#PROXIMITY	+NEAR_00	By, near, beside, next to
		#SPACE_BEHIND	+BEHIND_00	Behind
		#SPACE_FRONT	+IN_FRONT_OF_00	In front of, ahead of
		#SUPERIORITY	+ABOVE_00	Above, over
		#SURFACE	+ON_01	On, upon, atop, on top of
		#SURROUND	+AROUND_00	(a)Round



**Anexo VII. Dimensión cognitiva #PATH ilustrada con unidades léxicas**

METACONCEPTO	METACONCEPTO	METACONCEPTO	CONCEPTOS BÁSICOS	UNIDADES LÉXICAS
#LOCUS	#PATH	#DIRECTION	+AWAY_FROM_00	Away from
			+TOWARDS_00	Towards, for
		#GOAL	+TO_00	To
		#ROUTE	+ACROSS_00	Across
			+ALONG_01	Along
			+AROUND_01	(A)round
			+BY_01	By
			+OVER_00	Over
			+THROUGH_00	Via, through
		#SOURCE	+FROM_01	From



**Anexo VIII. Dimensión cognitiva #CIRCUMSTANCE ilustrada con unidades léxicas**

METACONCEPTO	METACONCEPTO	METACONCEPTO	CONCEPTOS BÁSICOS	UNIDADES LÉXICAS	
#CIRCUMSTANCE	#COMPARISON		+BESIDE_00	Beside	
			+FROM_02	From	
			+TO_01	To	
		#FREQUENCY	+FOR_02	For	
	#MANNER		+IN_04	In	
				+WITH_00	With
				+UNDER_00	Under
			#INSTRUMENT	+BY_02	By, by means of
				+ON_02	On
				+THROUGH_01	Through, via
				+WITH_01	With
				+WITHOUT_00	Without
			#MEANS	+FROM_03	From
		+IN_05	In		
	#PARTICIPANT	#BENEFICIARY	+FOR_03	For	
		#COMPANY	+WITH_02	With	
	#PURPOSE		+FOR_04	For	



METACONCEPTO	METACONCEPTO	METACONCEPTO	CONCEPTOS BÁSICOS	UNIDADES LÉXICAS
	#QUANTITY	#APPROXIMATOR	+ABOUT_00	About, around, by
			+BETWEEN_03	Between
		#VALUE	+AT_02	At
			+BY_03	By
			+FOR_05	For
			+FROM_04	From
			+IN_06	In
			+TO_02	To
	#REASON		+FROM_05	From, because of
	#REFERENT	#COLLABORATION	+AMONG_01	Among
			+BETWEEN_02	Between
		#PARTITIVE	+OF_00	Of
		#POSSESSION	+OF_01	Of
		#QUALITY	+OF_02	Of, with
	#RESULT		+INTO_00	Into
	#SCENE		+OVER_01	Over
	#TOPIC		+ON_03	On, about, over



## Anexo IX. Relación de preposiciones con sus conceptos básicos

Preposición	Conceptos	Taxonomía	Metaconcepto
<b>At</b>	+AT_00	#TIME	#INTERVAL_F1_1
	+AT_01	#LOCUS	#INTERIOR
	+AT_02	#CIRCUMSTANCE	#VALUE
<b>Between</b>	+BETWEEN_00	#TIME	#INTERVAL_F2_2
	+BETWEEN_01	#LOCUS	#INTERLOCAL
	+BETWEEN_02	#CIRCUMSTANCE	#APPROXIMATOR
	+BETWEEN_03	#CIRCUMSTANCE	#COLLABORATION
<b>For</b>	+FOR_00	#TIME	#DURATION_F1_1
	+FOR_01	#LOCUS	#LENGTH
	+FOR_02	#CIRCUMSTANCE	#FREQUENCY
	+FOR_03	#CIRCUMSTANCE	#BENEFICIARY
	+FOR_04	#CIRCUMSTANCE	#PURPOSE
	+FOR_05	#CIRCUMSTANCE	#VALUE
<b>From</b>	+FROM_00	#TIME	#INTERVAL_F2_3
	+FROM_01	#LOCUS	#SOURCE
	+FROM_02	#CIRCUMSTANCE	#COMPARISON
	+FROM_03	#CIRCUMSTANCE	#MEANS
	+FROM_04	#CIRCUMSTANCE	#VALUE
	+FROM_05	#CIRCUMSTANCE	#REASON
<b>In</b>	+IN_00	#TIME	#DURATION_F1_2
	+IN_01	#TIME	#DURATION_F2_1
	+IN_02	#TIME	#INTERVAL_F1_2
	+IN_03	#LOCUS	#INTERIOR
	+IN_04	#CIRCUMSTANCE	#MANNER
	+IN_05	#CIRCUMSTANCE	#MEANS
	+IN_06	#CIRCUMSTANCE	#VALUE
<b>On</b>	+ON_00	#TIME	#INTERVAL_F1
	+ON_01	#LOCUS	#SURFACE
	+ON_02	#CIRCUMSTANCE	#INSTRUMENT
	+ON_03	#CIRCUMSTANCE	#TOPIC
<b>To</b>	+TO_00	#LOCUS	#GOAL
	+TO_01	#CIRCUMSTANCE	#COMPARISON
	+TO_02	#CIRCUMSTANCE	#VALUE
<b>With</b>	+WITH_00	#CIRCUMSTANCE	#MANNER
	+WITH_01	#CIRCUMSTANCE	#INSTRUMENT
	+WITH_02	#CIRCUMSTANCE	#COMPANY