



La gramática de COREL: un lenguaje de representación conceptual

The COREL grammar: a conceptual representation language

Carlos Perrián-Pascual

Universidad Católica San Antonio
España

Ricardo Mairal-Usón

Universidad Nacional de Educación a Distancia (UNED)
España

Resumen

Los sistemas informáticos de comprensión del lenguaje natural requieren una base de conocimiento provista de representaciones conceptuales que reflejen la estructura del sistema cognitivo de los seres humanos. Aunque la semántica superficial puede ser suficiente en algunas otras aplicaciones computacionales, la construcción de una base de conocimiento robusta garantiza su reutilización en la mayoría de las tareas de procesamiento del lenguaje natural. En este escenario, FunGramKB se presenta como una base de conocimiento multipropósito cuyo modelo ha sido diseñado de manera específica para tareas de comprensión del lenguaje natural. Precisamente, uno de los elementos que han contribuido en forma notable al éxito de esta base de conocimiento ha sido el poder expresivo de su sistema notacional. El propósito de este artículo es describir la gramática, junto con su fundamentación teórica, del lenguaje de representación conceptual utilizado en FunGramKB.

Palabras clave: FunGramKB; COREL; interlingua; representación del significado.

Afiliación: Carlos Perrián-Pascual. Departamento de Idiomas, Universidad Católica San Antonio de Murcia. España — Ricardo Mairal-Usón. Departamento de Filologías Extranjeras y sus Lingüísticas, Universidad Nacional de Educación a Distancia (UNED), Madrid. España.
Correos electrónicos: jcperinan@pdi.ucam.edu; rmairal@flog.uned.es
Dirección postal: Departamento de Idiomas, Universidad Católica de San Antonio, Campus de los Jerónimos, E - 30107 Guadalupe, Murcia. España.

Fecha de recepción: febrero de 2010
Fecha de aceptación: abril de 2010

Abstract

Natural language understanding systems require a knowledge base provided with conceptual representations reflecting the structure of human beings' cognitive system. Although surface semantics can be sufficient in some other systems, the construction of a robust knowledge base guarantees its reuse in most natural language processing applications. In this scenario, FunGramKB is presented as a multipurpose knowledge base whose model has been particularly designed for natural language understanding tasks. Indeed, one of the features which has remarkably contributed to the success of this knowledge base is the expressive power of its notational system. The aim of this paper is to describe the grammar, together with its theoretical foundation, of the conceptual representation language used in FunGramKB.

Keywords: *FunGramKB; COREL; interlingua; meaning representation.*

Después de todo, ¿qué son las palabras? Las palabras son
símbolos para recuerdos compartidos.

(*Credo de Poeta*, Jorge Luis Borges)

1. Introducción¹

FunGramKB Suite² es un entorno computacional destinado a la construcción semiautomática de una base de conocimiento léxico-conceptual multipropósito para el desarrollo de sistemas automatizados del procesamiento del lenguaje natural (PLN). Por una parte, FunGramKB es multipropósito en el sentido de que es tanto multifuncional como multilingüe. En otras palabras, FunGramKB ha sido diseñado con el fin de ser potencialmente reutilizado en diversas tareas del PLN (recuperación y extracción de información, traducción automática, sistemas basados en el diálogo, etc.) y con diversas lenguas³. Por otra parte, nuestra base de conocimiento comprende tres niveles principales de conocimiento (i.e. léxico, gramatical y conceptual), cada uno de los cuales está constituido por diversos módulos independientes

¹ Este trabajo forma parte del proyecto de investigación financiado por el Ministerio de Ciencia y Tecnología, código FFI2008-05035-C02-01.

² En este artículo utilizamos el término "FunGramKB Suite" para referirnos a nuestra herramienta de ingeniería del conocimiento (www.fungramkb.com) y "FunGramKB" para la base de conocimiento resultante. FunGramKB Suite fue construido en C# utilizando la plataforma ASP.NET 2.0 y la base de conocimiento MySQL.

³ Actualmente, FunGramKB Suite ha sido diseñado para poder trabajar con siete lenguas: alemán, búlgaro, catalán, español, francés, inglés e italiano.

aunque interrelacionados (cf. Perinán-Pascual y Arcas-Túnez, 2010).

Dentro de este contexto, en los últimos años hemos impulsado la tarea de poblar la referida base de conocimiento, lo que necesariamente nos ha conducido al desarrollo de un lenguaje para la representación y gestión del conocimiento conceptual. Este artículo pretende establecer los fundamentos básicos de este lenguaje de representación que constituye una pieza angular para la formación de los nuevos ingenieros del conocimiento. En cuanto a la organización de este trabajo, la sección 2 presenta la tipología de los esquemas conceptuales que distinguimos en FunGramKB, mientras que la sección 3 describe la tipología de las unidades ontológicas, centrándonos principalmente en sus propiedades conceptuales, i.e. marcos temáticos (MT) y postulados de significado (PS). Por otra parte, la sección 4 se ocupa de la descripción de la gramática del lenguaje de representación conceptual. Finalmente, la sección 5 incluye unas breves conclusiones.

2. Los esquemas conceptuales en FunGramKB

En la literatura más relevante de la psicología cognitiva, influida notablemente por el modelo de memoria a largo plazo de Tulving (1985), el conocimiento del sentido común que forma parte del razonamiento humano se divide en:

- la memoria semántica, la cual almacena la información cognitiva sobre las palabras y representa un tesoro mental;
- la memoria procedimental, la cual almacena la información sobre cómo percibimos los eventos que ejecutamos en la vida real, por ej. cómo montar en bicicleta, y
- la memoria episódica, la cual almacena la información sobre eventos biográficos, por ej. el día de tu cumpleaños o de tu boda.

Estos tres tipos de conocimiento tienen una proyección diferente en nuestra base de conocimiento, la cual distingue los siguientes módulos conceptuales:

- la *Ontología*, la cual incluye el conocimiento semántico que viene representado por un conjunto de proto-microestructuras,

- el *Cognición*, el cual recoge el conocimiento procedimental bajo la forma de un conjunto de proto-macroestructuras, y
- el *Onomasticón*, el cual incluye el conocimiento episódico bajo la forma de bio-estructuras.

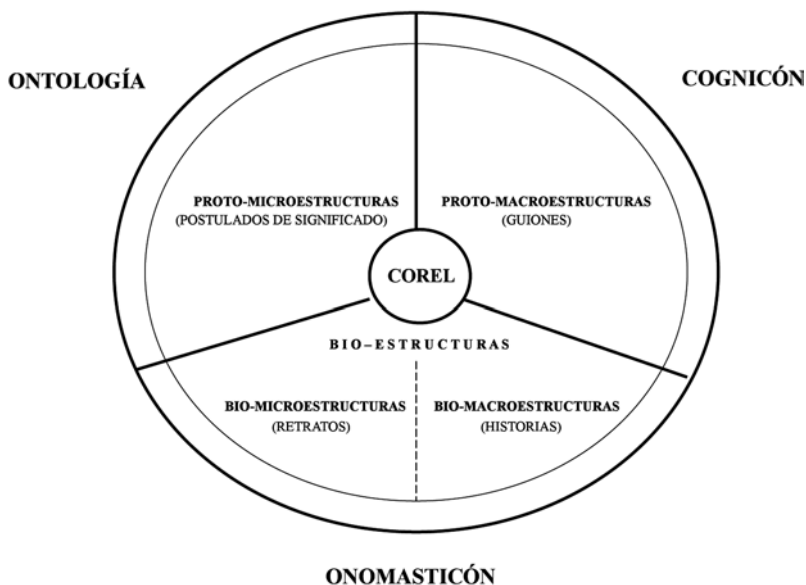
Precisemos un poco más esta clasificación y la terminología que la acompaña. En FunGramKB, los esquemas conceptuales se clasifican atendiendo a dos parámetros: (i) la prototipicidad y (ii) la temporalidad. De un lado, los esquemas conceptuales almacenan conocimiento prototípico, i.e. las proto-estructuras, o bien pueden servir para describir una instancia de una entidad o un evento, i.e. las bio-estructuras. Por ejemplo, la descripción del significado de la unidad léxica *canción* implica describir la proto-estructura del concepto al que va asignada; en cambio, si deseamos proporcionar información sobre la canción *Heartbreak Hotel*, necesitamos hacerlo a través de una bio-estructura. De igual forma, podemos presentar el conocimiento atemporalmente, i.e. las microestructuras, o inserto en un paradigma temporal, i.e. las macroestructuras. Por ejemplo, la descripción de la biografía de Elvis Presley requiere una macroestructura, mientras que una microestructura es suficiente para describir la profesión de cantante. Si combinamos estos dos parámetros, obtenemos la tipología de esquemas conceptuales mostrada en la Tabla 1.

TABLA 1
Tipología de los esquemas conceptuales en FunGramKB

		T E M P O R A L I D A D	
		-	+
P R O T O T I P I C I D A D	+	Proto- microestructura	Proto- macroestructura
	-	Bio- microestructura	Bio- macroestructura

Tulving (1985) mantiene que las memorias semántica, procedimental y episódica trabajan conjuntamente con el fin de facilitar la representación y la recuperación de la información. Por consiguiente, una de las cuestiones claves para el desarrollo de un motor de razonamiento como parte de un sistema para el procesamiento del lenguaje natural consiste en el desarrollo de un único lenguaje de interfaz para la representación de cada uno de estos esquemas conceptuales, lo que permitirá compartir información más fácil y rápidamente entre cada uno de estos tres niveles. A tal efecto, FunGramKB postula COREL (*Conceptual Representation Language*) como lenguaje de representación conceptual común a los tres tipos de conocimiento, tal como ilustramos en la Figura 1.

FIGURA 1
El planeta cognitivo



A diferencia de otras bases de conocimiento en las que los conceptos se definen por la relación que mantienen con otros conceptos, por ej. EuroWordnet (Vossen, 1998), FunGramKB

implementa el lenguaje de representación COREL, el cual nos permite definir todas las unidades conceptuales con la ventaja añadida de que estas definiciones se ven enriquecidas tras la aplicación de los mecanismos de herencia e inferencia sobre las propias proto-microestructuras (cf. Perinián-Pascual y Arcas-Túnez, 2005). Como en todo lenguaje de representación, distinguimos dos componentes:

- 1) Una colección de términos, i.e. las unidades conceptuales que se organizan jerárquicamente para permitir la herencia y así facilitar la brevedad definitoria y aumentar el poder expresivo.
- 2) Un sistema de notación, i.e. la sintaxis a la que se tienen que ajustar las representaciones interlingüísticas para poder considerarlas estructuras bien formadas.

En las siguientes dos secciones, describimos los términos y la notación de COREL respectivamente.

3. Las unidades conceptuales

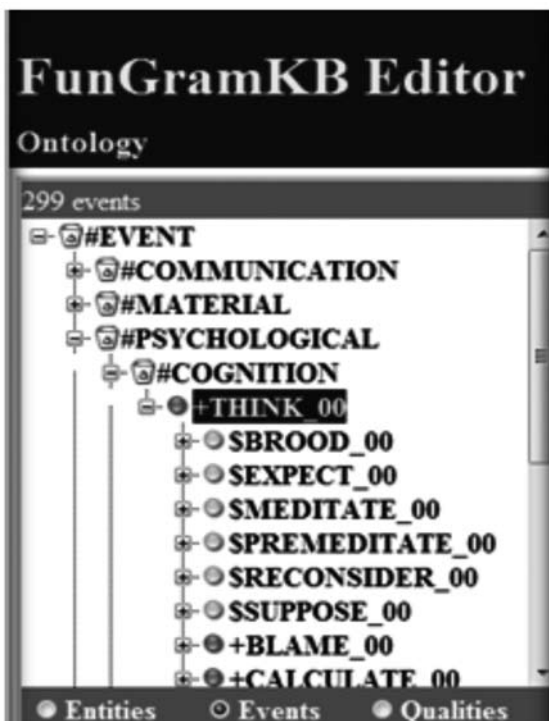
FunGramKB está poblada por tres tipos de unidades conceptuales que aparecen representadas por un sistema de notación diferente:

- (i) los *metaconceptos*, que coinciden con muchas de las unidades ontológicas del nivel superior de otras ontologías, por ej. SUMO, DOLCE, GUM, Mikrokosmos, SIMPLE, etc. Los metaconceptos vienen precedidos por el símbolo # (por ej. #ENTITY, #EVENT y #QUALITY);
- (ii) los *conceptos básicos*, que fueron identificados inicialmente a partir del vocabulario definitorio del *Longman Dictionary of Contemporary English* (Procter, 1978), recurso que, recordemos, tuvo una amplia difusión en el procesamiento del lenguaje natural. Como demostramos en Perinián-Pascual y Mairal-Usón (s. f.), el listado final de conceptos básicos ha sido el resultado de varias fases de proyección: conceptualización, jerarquización, remodelación y refinamiento. Utilizamos el símbolo + para su representación, por ej. +WANT_00, +PERCEIVE_00, +TRUST_00, +FEEL_00, etc;
- (iii) los *conceptos terminales*, que representan los nodos finales de la estructuración jerárquica conceptual, aparecen

representados por el símbolo \$, por ej. \$MEDITATE_00, \$ADAPT_00, \$FLUCTUATE_00, etc.

En las Figuras 2 y 3 incluimos una visión parcial de los conceptos básicos y terminales en las dimensiones conceptuales #EVENT > #PSYCHOLOGICAL > #COGNITION, y #ENTITY > #ABSTRACT respectivamente:

FIGURA 2
Representación parcial de los eventos

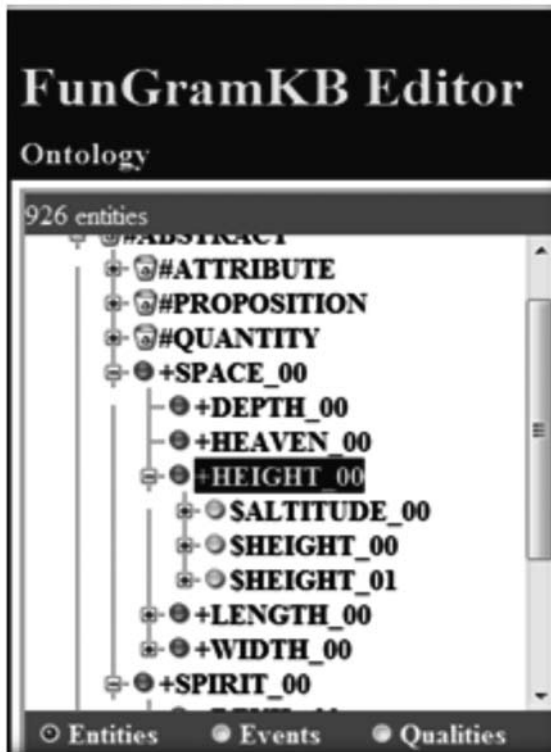


Los conceptos básicos y terminales de la Ontología de FunGramKB poseen tres características que no podemos pasar por alto:

- (i) La mayoría de estos conceptos se conciben como “conceptos de motivación léxica”, en el sentido de que, al menos en una lengua, cada uno de estos conceptos debe estar

- lexicalizado⁴. Por tanto, entendemos la semántica léxica como el estudio de la estructuración conceptual.
- (ii) Estos conceptos no son primitivos semánticos al estilo de los formulados, por ejemplo, en el Metalenguaje Semántico Natural de Goddard y Wierzbicka (2002).
 - (iii) Estos conceptos tienen una serie de propiedades semánticas: los MT y los PS. Ambas propiedades tienen una base conceptual, ya que se construyen con conceptos y no con unidades léxicas, por lo cual tanto los MT como los PS son representaciones del conocimiento semántico independientes de una lengua.

FIGURA 3
Representación parcial de las entidades



⁴ A pesar de que en el caso de los conceptos terminales esta característica es obligatoria, esto no ocurre con algunas unidades conceptuales básicas que sirven como “conceptos paraguas”, cuya función es exclusivamente organizativa.

Siguiendo con la corriente principal en psicología, FunGramKB adopta un enfoque mentalista, por el cual el concepto se concibe como una “representación mental estructurada”⁵. En nuestro caso, COREL es utilizado para describir formalmente estas representaciones mentales. Por tanto, y coincidiendo con “the language of thought hypothesis” (Fodor, 1975), nuestro sistema de representación tiene una sintaxis y una semántica al igual que las lenguas naturales. De todas las teorías más influyentes sobre la estructuración de los conceptos, abogamos por adoptar el enfoque de las *teorías duales* (Osherson y Smith, 1981; Landau, 1982; Armstrong y otros, 1983), teorías que reciben la influencia tanto del modelo clásico de definición como de la teoría de los prototipos (cf. sección 3.2). En los siguientes dos subapartados pasamos a describir las propiedades de las unidades conceptuales básicas y terminales.

3.1. Los marcos temáticos

Cuando el ingeniero del conocimiento comienza a editar las propiedades de la ontología, su primera tarea es especificar el MT, el cual concebimos como un constructo conceptual que especifica los participantes que intervienen típicamente en una situación cognitiva. Por ejemplo, si consideramos la dimensión cognitiva #COMMUNICATION, podríamos identificar básicamente tres papeles temáticos (o conceptuales): un *Tema* (“Theme”), i.e. la entidad que transmite un mensaje, un *Referente* (“Referent”), i.e. el mensaje, y una *Meta* (“Goal”), i.e. la entidad que recibe el mensaje. Nótese que cuando especificamos los participantes de un MT no estamos pensando en el comportamiento léxico de un verbo, por ej. *comunicar*, sino en los participantes que intervienen en la dimensión cognitiva #COMMUNICATION⁶.

Así, los metaconceptos, que estrictamente no son conceptos sino dimensiones cognitivas, vienen representados por una serie de papeles temáticos, a partir de los cuales podremos construir los MT correspondientes a sus conceptos subordinados. Elaboremos un poco más este punto. Si tuviéramos que especificar las propiedades de la unidad conceptual básica

⁵ En cambio, muchos filósofos adoptan el enfoque fregeiano de tratar a los conceptos como entidades abstractas. Véanse Peacocke (1992) y Rey (1994) para la motivación de este enfoque.

⁶ Véase el anexo 1 para una descripción de los papeles temáticos que definen cada dimensión conceptual.

+ANSWER_00, convendríamos que la referida unidad está subordinada al concepto +SAY_00, el cual a su vez depende del metaconcepto #COMMUNICATION. Puesto que la unidad metaconceptual tiene asignado el esquema temático prototípico (1), podemos anticipar que todas las unidades conceptuales subordinadas a este metaconcepto heredarán dicho esquema en forma de MT.

(1) (x1)Theme (x2)Referent (x3)Goal

Por consiguiente, una de las propiedades del evento +ANSWER_00 será precisamente este MT, a lo que deberemos añadir un PS que ofrezca una expresión más detallada del significado de la unidad conceptual propiamente dicha, a saber, “una persona (x1) dice algo (x2) a alguien (x3) en relación a una pregunta (x4) que x3 le ha dicho a x1”.

(2) MT: (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal



PS: +(e1: +SAY_00 (x1)Theme (x2)Referent (x3)Goal (f1: (e2: +SAY_00 (x3)Theme (x4: +QUESTION_00)Referent (x1)Goal)) Scene)

Por tanto, los MT quedan perfectamente integrados en los PS. Los papeles temáticos de cada metaconcepto funcionan como papeles conceptuales que proporcionan una descripción de las instancias prototípicas que participan en una dimensión conceptual. Como podrá apreciar el lector en el anexo 1, en FunGramKB no postulamos papeles temáticos diferentes para cada dimensión conceptual, es decir, los papeles temáticos no están vinculados a un metaconcepto específico, ya que la propia dimensión conceptual enriquece el significado del papel temático. Así, utilizamos un inventario reducido de papeles, si bien estos adquieren una interpretación semántica diferente según el metaconcepto al que pertenece un determinado concepto. En este sentido, postulamos el papel temático central *Tema*, que

aparece en todas las dimensiones metacognitivas y adquiere un valor definicional diferente en cada una de ellas: por ej. en #EMOTION el papel Tema es la entidad que siente una emoción, mientras que en #CREATION el papel Tema es la entidad que crea otra entidad. Puesto que es un requisito fundamental que la entidad Tema aparezca como papel conceptual central en todas las dimensiones conceptuales, resulta bastante lógico que hayamos optado por la denominación de “marco temático” para referirnos a este tipo de mecanismo representacional. Adicionalmente, esta disposición de los papeles temáticos evita una proliferación excesiva de las funciones semánticas al tiempo que preserva la informatividad conceptual. Con el fin de evitar inexactitudes, es fundamental proporcionar a cada papel temático una definición clara y rigurosa según su ubicación en cada dimensión conceptual (cf. anexo 1).

En este punto, cumple preguntarse cómo hemos llegado a este inventario de papeles temáticos. Parece que existe un consenso en torno al número y el tipo de participantes prototípicos en cada dimensión cognitiva. Por este motivo, se desarrolló un protocolo que nos permite formular nuestro propio inventario de papeles temáticos para cada dimensión cognitiva a partir de criterios tales como (a) la obligatoriedad semántica y (b) la especificidad semántica de Koenig y otros (2003), los cuales reproducimos a continuación:

- (a) Criterio de obligatoriedad semántica:
Si r es un papel de participante argumento del predicado P , entonces cualquier situación que P describa incluye el referente de r .
- (b) Criterio de especificidad semántica:
Si r es un papel de participante argumento del predicado P denotado por el verbo V , entonces r es específico a V y a una clase restringida de verbos.

Koenig y otros sugieren que la aplicación de forma complementaria de estos dos criterios validados empíricamente permite determinar el estatus de un participante semántico implicado en la situación descrita por un evento. En otras palabras, un participante es un argumento semántico de un verbo si y solo si satisface tanto el criterio de obligatoriedad semántica como el criterio de especificidad semántica, a lo que denominaron

“Hipótesis de Codificación Léxica”⁷. Con este enfoque llegamos a un inventario de papeles temáticos que no difiere sustancialmente de los *procesos* que Halliday (1985) distingue y de los roles que forman parte de los *tipos semánticos* de Dixon (1991)⁸.

Un paso ulterior consiste en complementar cada MT con una representación del significado de cada unidad conceptual. A tal efecto, proponemos los PS como sistema de representación del significado conceptual, aspecto del que nos ocupamos a continuación. Sin embargo, como antes avanzábamos, cada PS integrará el MT que define al concepto objeto de análisis o, dicho de otra forma, los PS enriquecen los MT con información adicional e idiosincrática a cada unidad conceptual.

3.2. Los postulados de significado

Cuando queremos representar el conocimiento de una unidad ontológica, dotamos a cada concepto de un PS, el cual construimos siguiendo las premisas de una teoría dual, la cual identifica dos tipos de rasgos:

- a) Un conjunto de rasgos que poseen una *función categorial*, los cuales determinan la pertenencia a una clase. Se

⁷ Koenig y otros (2003) realizaron un experimento con unos 5.500 lemas verbales del inglés con el fin de probar empíricamente esta hipótesis. En este experimento, un grupo de sujetos humanos tenía que determinar si cada verbo requería, permitía o excluía un determinado tipo de participante, para lo cual debían responder una serie de preguntas con el fin de observar si la situación descrita por el verbo implicaba la presencia o no de la información de los diversos roles. En definitiva, un participante semántico es obligatorio cuando su presencia es requerida en la situación descrita por el verbo. Estos participantes semánticos obligatorios pueden adquirir el estatus de argumento o satélite dependiendo de la centralidad del papel en el escenario descrito por el evento.

⁸ Véase el anexo 3 para una comparativa de la propuesta de FunGramKB y estos dos enfoques. A este respecto, y dependiendo del grado de obligatoriedad, FunGramKB distingue dos tipos de participantes: los argumentos y los satélites, los cuales anotamos con las variables x y f , respectivamente, siguiendo la convención establecida en la Gramática Funcional de S.C. Dik (1997). Es necesario advertir que esta distinción no es pertinente en los MT, ya que solo recogen aquellos participantes prototípicos (y, por ende, argumentos) que definen un marco conceptual. Cada participante solo puede desempeñar un único papel argumental independientemente de las preferencias de selección a las que esté sujeto. Los anexos 1 y 2 muestran la interpretación semántica de los papeles temáticos de los argumentos y de los satélites respectivamente.

trata de rasgos necesarios, aunque no implica que sean suficientes.

- b) Un conjunto de rasgos que poseen una *función identificativa*, que utiliza el ingeniero del conocimiento para describir el miembro prototípico de la categoría.

Por ejemplo, el concepto GRANDMOTHER se definiría como MOTHER OF A PARENT como su “núcleo conceptual”; en cambio, los rasgos GREY HAIR y WRINKLES tienen una “función de ejemplaridad”. En este modelo híbrido persiste el problema de identificar cuáles son los rasgos pertinentes a cada concepto, pero en realidad éste es un problema inherente a cualquier tipo de teorías de rasgos. En la teoría dual, los rasgos que configuran el núcleo del concepto son más difíciles de identificar que los que describen la función identificativa. En el resto de este apartado presentamos algunas consideraciones derivadas de la aplicación de la teoría dual de la semántica léxica en FunGramKB.

En primer lugar, el PS de un concepto básico o terminal está formado por una o más predicaciones, cada una de las cuales representa un rasgo. La única excepción se encuentra en el caso de dos predicaciones ligadas, las cuales representan un único rasgo (cf. sección 4.1).

En segundo lugar, diferenciamos dos tipos de rasgos: nucleares y ejemplares. Los rasgos nucleares tienen una función categorial, mientras que los ejemplares tienen una función identificativa. Desde un punto de vista formal, los primeros se representan por medio de predicaciones estrictas precedidas del operador +, y los segundos por predicaciones rebatibles precedidas del operador *. Así, los rasgos nucleares son rasgos obligatoriamente necesarios en el mundo real y, por consiguiente, no pueden ser rebatidos bajo ninguna circunstancia: por ej. “un avestruz es un ave” o “una ballena es un mamífero” constituyen dos ejemplos de rasgos nucleares. En cambio, los rasgos ejemplares pueden ser rebatidos, dando lugar incluso a situaciones singulares. Por ejemplo, una predicación como “los perros tienen cuatro patas” es rebatible, ya que es posible ver un perro con tres patas sin que por ello deje de ser un perro⁹. En este sentido, es importante subrayar que no debemos confundir

⁹ Para una descripción detallada de las nociones de predicación estricta y rebatible, remitimos al lector a la sección 4.2.1, donde describimos los operadores de razonamiento.

entre rasgos prototípicos y rasgos estereotípicos, aunque éstos coincidan a veces. Debemos evitar incorporar en el PS rasgos que resulten de convenciones socioculturales, tales como “los zorros son astutos”, ya que serían una fuente potencial de problemas (Papafragou, 1996). Por una parte, a este tipo de proposiciones se les puede asignar el valor de verdad si se conciben como estereotipos, pero en realidad estos rasgos expresan algo falso con respecto al mundo real. Por otra, y como producto de las diversas convenciones socioculturales, podríamos asignar incluso varios rasgos estereotípicos contradictorios a un mismo concepto, lo cual resultaría en un concepto ambiguo. En definitiva, el riesgo que corremos es terminar representando el significado connotativo, el cual además de ser subjetivo, es relativamente inestable y bastante indeterminado (Leech, 1974).

El número de rasgos ejemplares es más numeroso que el de los nucleares. En realidad, los conceptos pueden carecer de alguno de estos dos tipos de rasgos, pero no de los dos. De hecho, los rasgos nucleares no implican que su presencia sea obligatoria en un PS, sino que, en caso de aparecer, el referente de ese concepto tendrá necesariamente ese rasgo. También es importante advertir el hecho de que, en el caso de los rasgos ejemplares, no seguimos ningún procedimiento estadístico para obtener los rasgos, sino generamos las predicaciones manualmente a partir de recursos lexicográficos y la propia introspección del ingeniero del conocimiento. Pero ¿cuánto conocimiento semántico debemos almacenar bajo un concepto? ¿Cuál debe ser el grado de granularidad semántica que debemos imprimir en la descripción del PS de cada concepto? En este sentido, seguimos un posicionamiento muy similar al expresado en Laurence y Margolis (1999: 29), quienes afirman que “on the one hand, a concept should encode a considerable amount of information about its instances and exemplars, but on the other, it shouldn't include so much that the concept becomes unwieldy.”

Con el fin de obtener homogeneidad y consistencia en la relevancia de los rasgos que configuran los PS, hemos confeccionado un inventario de descriptores¹⁰ que servirán al ingeniero del

¹⁰ Nuestro inventario de descriptores está inspirado en la noción de *Estructura Qualia Extendida* utilizada por SIMPLE (Lenci, 2000; Lenci y otros, 2000), donde se han ampliado los valores de las *qualia* del modelo de Pustejovsky (1995) con el fin de expresar distinciones más finas entre los componentes semánticos.

conocimiento como guía semántica para ampliar la información contenida en los recursos lexicográficos durante el proceso de elaboración de los PS. Estos descriptores han sido vinculados manualmente a las dimensiones conceptuales de la ontología conforme a su tipicidad, de tal forma que el editor ontológico de FunGramKB puede mostrarnos automáticamente aquellos descriptores asociados al metaconcepto al que pertenece el concepto cuyo PS estamos editando. Por ejemplo, si estuviéramos elaborando el PS de \$TOASTER_00, nuestra guía estaría configurada por los siguientes diecisiete descriptores, los cuales se vinculan al metaconcepto #SELF_CONNECTED_OBJECT:

- a. This entity has some PARTS.
- b. The COLOUR or other VISUAL ATTRIBUTES of this entity.
- c. The SIZE/LENGTH of this entity.
- d. The TOUCH of this entity.
- e. The TASTE of this entity.
- f. The WEIGHT of this entity.
- g. The SHAPE of this entity.
- h. The TEMPERATURE of this entity.
- i. The SMELL of this entity.
- j. The VALUE of this entity.
- k. This entity is made of some MATERIAL.
- l. This entity is FOUND IN some places.
- m. This entity is used for some PURPOSE/FUNCTION.
- n. Some ACTIVITIES/ACTIONS related to this entity.
- o. This entity is OBTAINED FROM a place.
- p. This entity is PRODUCED BY another entity.
- q. The MANNER in which this entity is OBTAINED/PRODUCED.

El primer paso consistiría en decidir cuál sería el concepto superordinado de \$TOASTER_00, el cual sería MACHINE, i.e. “artefacto que utiliza alguna forma de energía para realizar una tarea”. El segundo paso consistiría en determinar los rasgos más relevantes, para lo cual se tomaría como guía el inventario anterior. Los ingenieros del conocimiento formulan cada uno de los descriptores como preguntas que guían el proceso de introspección: por ej. la pregunta correspondiente al rasgo (a) sería: “¿Cuáles son las partes típicas que tiene la mayoría de x?”. Hacemos hincapié en las nociones de “típico” y “mayoría”, ya que en esta ocasión intentamos averiguar un rasgo ejemplar. En caso de encontrar una respuesta, intentaremos averiguar

si se trata en realidad de un rasgo nuclear, por lo cual nos preguntaremos: “Pero ¿son estas partes típicas necesarias?”. En realidad, en FunGramKB los rasgos nucleares son un subconjunto de los rasgos ejemplares, i.e. $F_{\text{nuclear}} \subset F_{\text{ejemplar}}$, ya que los rasgos necesarios son también los más típicos pero no todos los rasgos típicos son necesarios.

En otras palabras, no es el caso que todos estos descriptores nos permitan identificar un rasgo relevante al concepto que describimos semánticamente, razón por la que afirmamos que estos descriptores solo constituyen una guía y no deben utilizarse como un protocolo. Siguiendo con nuestro ejemplo de \$TOASTER_00, los rasgos (a-j) y (n-q) son descartados, mientras que los rasgos (k), (l) y (m) se representan por último a través de las predicaciones 2, 3 y 4 respectivamente:

- (3) +(e1: +BE_00 (x1: \$TOASTER_00)Theme (x2: +MACHINE_00 Referent)
 +(e2: +BE_01 (x1)Theme (x3: +METAL_00 | +PLASTIC_00 Attribute)
 *(e3: +BE_02 (x1)Theme (x4: +KITCHEN_00)Location)
 *(e4: +TOAST_00 (x5: +HUMAN_00)Theme (x6: +BREAD_00 Referent (f1: x1)Instrument)

Somos conscientes de que es preciso mejorar este procedimiento de identificación de los componentes semánticos de un PS con el fin de facilitar más la tarea de los ingenieros del conocimiento, pero actualmente nadie ha llevado a cabo todavía la automatización de tal procedimiento. Nuestro próximo objetivo consistiría entonces en refinar el inventario de descriptores propuestos para cada concepto. La solución radicaría, por ejemplo, en vincular los descriptores no solo a los metaconceptos sino también a los conceptos básicos; es decir, a los conceptos superordinados inmediatos. De esta forma, rasgos como (e-o) y (p-q) no hubieran aparecido en la lista anterior, ya que no son potencialmente relevantes para MACHINE.

4. El sistema de notación

Según Woods (1983), una gramática formal para la representación del conocimiento debe estar sujeta a tres estándares de adecuación que podemos formular como sigue:

- *Adecuación expresiva*: un lenguaje debe tener la capacidad de representar cualquier tipo de conocimiento que queramos describir.
- *Adecuación notacional*: un lenguaje debe constar de un sistema gramatical que permita expresar eficazmente las propiedades y relaciones entre los elementos.
- *Adecuación natural*: el formalismo para la representación del conocimiento debe ser lo más “natural” posible, por lo que la proyección de una oración en lenguaje natural al formalismo de representación del conocimiento debe ser suficientemente directa, reflejando tanto como sea posible la estructura del texto de entrada. Por consiguiente, entendemos que la representación de una oración a través de una estructura de predicado-argumento es más consistente que la de una regla lógica, a pesar de que ambas representaciones puedan tener un significado idéntico (Ali, 1993).

En el siguiente subapartado presentamos los aspectos más destacados de la gramática de COREL, para lo cual comenzamos con la descripción general de las predicaciones y posteriormente describimos los operadores que podemos incluir en estas predicaciones.

4.1. Las predicaciones

Las proto-estructuras de FunGramKB están formadas por predicaciones genéricas interpretadas bajo un enfoque inductivo. Más concretamente, un esquema conceptual de FunGramKB está formado por una o más predicaciones¹¹, cada una de ellas representada por una variable *e* indizada, seguida por dos puntos (:) y precedida por un operador de razonamiento (cf. sección 4.2.1). A modo de ilustración, presentamos el PS de la entidad +LEAF_00 (4a), del evento +ANSWER_00 (5a) y de la cualidad +SWEET_00 (6a).

- (4a) *(e1: +BE_00 (x1: +LEAF_00)Theme (x2: +PLANT_PART_00)
Referent)
*(e2: +BE_01 (x1)Theme (x3: +FLAT_00 & +THIN_00)
Attribute)

¹¹ Solo unos pocos conceptos básicos de la ontología carecen de PS por tratarse de primitivos semánticos. En estos casos, el valor para la propiedad del PS es “sp” (del inglés, *semantic primitive*), por ej. +SAY_00.

- *(e3: +BE_01 (x1)Theme (x4: +GREEN_00 ^ +BROWN_00) Attribute)
 +(e4: +ABSORB_00 (x5: +PLANT_00)Agent (x6: +AIR_00)Theme (x7)Location (x8)Origin (x5)Goal (f1: x1)Instrument)
- (5a) +(e1: +SAY_00 (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal (f1: (e2: +SAY_00 (x3)Theme (x4: +QUESTION_00)Referent (x1)Goal))Scene)
- (6a) *(e1: +BE_01 (x1)Theme (x2: +SWEET_00)Attribute)
 *(e2: +BE_01 (x1)Theme (x3: +PLEASANT_00)Attribute (f1: (e3: +TASTE_00 (x4)Theme (x1)Referent))Scene | (f2: (e4: +SMELL_00 (x4)Theme (x1)Referent))Scene)

A continuación, presentamos los equivalentes aproximados en lenguaje natural de las predicaciones anteriores:

- (4b) Una hoja es parte de una planta.

Una hoja es típicamente plana y delgada.

Una hoja es típicamente verde o marrón.

Una planta absorbe el aire a través de las hojas.

- (5b) Alguien te dice algo cuando haces una pregunta.

- (6b) Algo que es dulce típicamente tiene un sabor o un olor agradable.

Nótese que las predicaciones están formadas por conceptos básicos o terminales¹². En el caso de las entidades, observamos que cada una de las predicaciones que configuran un PS debe necesariamente referenciar a su *definiendum*, ya sea explícitamente (primera predicación) o por coindización (predicaciones 2-4). Esto es lo que denominamos como “principio de la entidad omnirreferencial”, cuya motivación se encuentra en el hecho de que cada una de estas predicaciones expresa un rasgo de la misma entidad, y por tanto es preciso referenciarla en cada una de ellas. En el caso de las cualidades, ocurre el mismo fenómeno, con la diferencia de que no es el *definiendum* el que

¹² Una de las pocas variaciones que tiene lugar en la gramática de COREL para los diferentes esquemas conceptuales radica precisamente en el tipo de conceptos que se permiten. Así, las proto-estructuras solo permiten conceptos de la Ontología Nuclear, i.e. conceptos básicos para las proto-microestructuras y conceptos básicos y terminales para las proto-macroestructuras, mientras que las bio-estructuras se construyen a partir de los conceptos de toda la Ontología y del Onomasticón.

debe aparecer en todas las predicaciones del PS, sino la entidad a la cual se atribuye típicamente dicha cualidad.

La relación lógica entre dos predicaciones es siempre la conjunción (&), aunque hemos optado por omitir el conector lógico. Por ejemplo, los rasgos genéricos de que una isla es “una porción de tierra” y además “está rodeada de agua” se describen en (7).

- (7) +(e1: +BE_00 (x1: +ISLAND_00)Theme (x2: +LAND_00) Referent)
 +(e2: +SURROUND_00 (x3: +WATER_00)Theme (x1) Location)

Por otra parte, FunGramKB diferencia dos tipos de predicaciones en sus PS: predicaciones libres y predicaciones ligadas. Las predicaciones libres son aquellas que por sí solas expresan una especificación conceptual completa. En cambio, las predicaciones ligadas son aquellas que dependen de otras predicaciones para expresar de forma completa e inequívoca la especificación conceptual. Por ejemplo, podemos definir el evento +LIE_01 y la entidad +SEA_00 como en (8), “decir a alguien algo que no es verdad”, y (9), “un gran receptáculo de agua salada”, respectivamente:

- (8) +((e1: +SAY_00 (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00)Goal)(e2: n +BE_01 (x2)Theme (x4: +TRUE_00) Attribute))
 (9) *((e2: +CONTAIN_00 (x3: m +WATER_00)Theme (x1: +SEA_00) Location)(e3: +COMPRISE_00 (x3)Theme (x4: m +SALT_00) Referent))

Las características más relevantes de las predicaciones ligadas son las siguientes:

- i. Las predicaciones ligadas se presentan dentro de paréntesis redondos, los cuales sirven para marcar el ámbito de actuación del ligamiento cognitivo.
- ii. Solo pueden ligarse predicaciones contiguas. Típicamente, suelen ligarse dos predicaciones, aunque pueden ser más.
- iii. A partir de la segunda de las predicaciones ligadas, debe existir coindización hacia algunos de los participantes de la primera predicación. En realidad, esta es la característica más importante del ligamiento, ya que define la naturaleza

de este fenómeno: en la primera predicación ligada aparece típicamente una entidad que requiere una especificación conceptual adicional que solo puede ser proporcionada por otra predicación.

- iv. Se omite el operador de razonamiento delante de cada una de las predicaciones ligadas. En cambio, el paréntesis de inicio del ligamiento debe ir precedido por dicho operador.

Aunque a primera vista pueda parecer que el ligamiento conceptual implique la existencia de una predicación principal y otra “subordinada conceptualmente”, no es lo que en realidad sucede a nivel cognitivo. En un sintagma complejo, podemos hablar de un sintagma principal, porque por sí solo es gramaticalmente aceptable a la vez que tiene sentido dentro del discurso que lo enmarca, y de un sintagma subordinado. En cambio, ninguna de las predicaciones implicadas en el ligamiento conceptual posee un rango similar al de las predicaciones libres, ya que, de lo contrario, podrían presentarse diversos problemas:

- i. La predicación podría no ser sintácticamente aceptable dentro del esquema conceptual. Por ejemplo, este sería el caso de la segunda predicación en el ligamiento conceptual, ya que esta predicación depende gramaticalmente de la primera debido a la coindización de al menos una de sus variables. Por ejemplo, este es el caso de x^2 en (8), y de x^3 en (9).
- ii. La predicación podría no ser semánticamente aceptable dentro del esquema conceptual. Por ejemplo, si (8) fuera interpretado como dos rasgos diferentes, i.e. “mentir es decir algo” y “algo no es verdad”, pronto advertiríamos que la primera aserción no es cierta, a no ser que se interprete junto con la segunda aserción.

Por estas razones, decimos que las predicaciones implicadas en un ligamiento conceptual son ligadas, y por tanto no expresan varios rasgos diferentes sino tan solo un rasgo complejo.

El ámbito explicativo de las predicaciones incluso se extiende al papel que pueden desempeñar cuando actúan como preferencias de selección de uno de los participantes. Así, podemos expresar las preferencias de selección por medio de una predicación o a través de uno o más conceptos, con la restricción de que los conceptos deben ser necesariamente entidades

o cualidades¹³. Como vemos en (10), el PS de +REQUEST_01 sirve para ilustrar estos casos.

(10) +(e1: +SAY_00 (x1: +HUMAN_00)Theme (x2)Referent (x3: +HUMAN_00 ^ +ANIMAL_00)Goal (f1: (e2: +GIVE_00 (x3) Agent (x2)Theme (x3)Origin (x1)Goal))Purpose (f2: (e3: +NEED_00 (x1)Theme (x2)Referent))Reason)

En (10), las preferencias de selección de los satélites *f1* y *f2* son predicaciones; en cambio, los argumentos *x1* y *x3* se ven enriquecidos con entidades. Podemos incluso conectar múltiples preferencias de selección dentro de un participante mediante un conector lógico (cf. sección 4.2.3), como queda reflejado en *x3*¹⁴.

4.2. Los operadores

En una predicación podemos utilizar una serie de operadores, los cuales, según su alcance, podemos clasificar con la siguiente tipología:

- (i) operadores de razonamiento
- (ii) operadores de evento, los cuales se subcategorizan en operadores ATM (i.e. aspectualidad, temporalidad y modalidad) y la negación
- (iii) operadores de participante, los cuales se subcategorizan en cuantificadores y operadores lógicos

Elaboremos un poco más la descripción de estos tres tipos de operadores.

4.2.1. Los operadores de razonamiento

Los seres humanos deseamos comprender y caracterizar el mundo que nos rodea y, a tal efecto, observamos una serie de regularidades en la naturaleza que utilizamos para predecir las acciones de otras personas y los cambios en nuestro entorno. En este punto, conviene recordar que, al expresar el significado de una unidad conceptual, estamos moldeando el

¹³ Las cualidades solo pueden aparecer como preferencias de selección de los participantes con el papel de *Attribute*, *Comparison*, *Frequency*, *Position* o *Speed*.

¹⁴ Para un tratamiento más exhaustivo de las preferencias de selección en los participantes de las predicaciones, remitimos al lector al trabajo de Jiménez Briones y Pérez Cabello de Alba (2009).

conocimiento del mundo que tenemos sobre la referida unidad conceptual, por lo cual la finalidad de las predicaciones genéricas que forman parte de los PS es precisamente capturar estas regularidades. Sin embargo, una dificultad que debemos resolver es la naturaleza irregular de las supuestas regularidades. En este sentido, FunGramKB permite asignar un operador de razonamiento para cada una de las predicaciones, indicando de esta forma si la predicación en cuestión es estricta [+] o rebatible [*]¹⁵.

El motor de razonamiento de FunGramKB interpreta las predicaciones como reglas que permiten un razonamiento monotónico con las predicaciones estrictas y un razonamiento no monotónico con las predicaciones rebatibles. Por una parte, las predicaciones estrictas se comportan como reglas sin excepciones, por ej. (29) “las ballenas son mamíferos” o (30) “los círculos son redondos”.

(29) +(e1: +BE_00 (x1: \$WHALE_00)Theme (x2: +MAMMAL_00 Referent)

(30) +(e1: +BE_00 (x1: +CIRCLE_00)Theme (x2: +ROUND_00 Referent)

Por otra parte, las predicaciones rebatibles, que forman parte de nuestro sentido común, permiten activar inferencias que son invalidadas cuando se aporta información contradictoria. Por ejemplo, en el PS de +BIRD_00 se indica que “el pájaro típicamente vuela” (31), pero sabemos a través del PS de \$OSTRICH_00 que, a pesar de que “el avestruz es un pájaro”, “ningún avestruz vuela” (32).

(31) *(e3: +FLY_00 (x1)Agent (x1: +BIRD_00)Theme (x4)Origin (x5)Goal)

(32) +(e1: +BE_00 (x1: \$OSTRICH_00)Theme (x2: +BIRD_00 Referent)

...

+(e9: n +FLY_00 (x1)Agent (x1)Theme (x10)Origin (x11) Goal)

Con el fin de dar un tratamiento adecuado a estos casos excepcionales, la capacidad de volar de los pájaros se expresa

¹⁵ Recuerde que, en el caso de las predicaciones ligadas, el operador de razonamiento no se indica.

por medio de una predicación rebatible. Una alternativa al operador de razonamiento sería la incorporación de un índice numérico que expresara el grado de certeza de una predicación. Por ejemplo, la probabilidad de que “los pájaros vuelen” podría ser 0.8. El problema de esta estrategia radica en la dificultad de obtener con precisión ese índice de probabilidad para determinados rasgos prototípicos.

4.2.2. Los operadores de evento

Los operadores de evento adquieren diferentes valores tal como mostramos en la siguiente tabla¹⁶:

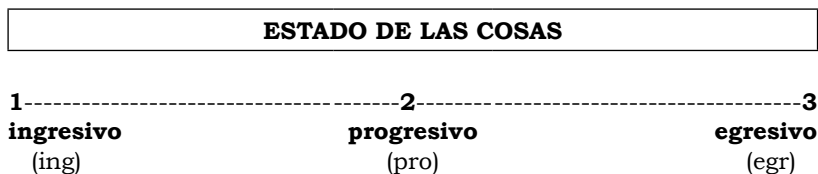
TABLA 2
Operadores de evento

Rasgo		Valor
Aspectualidad		ing pro egr
Temporalidad		rpast past npast pres nfut fut rfut
Modalidad	epistémica	cert prob pos
	no-epistémica	obl adv perm
Polaridad		n

Los operadores de aspectualidad expresan aquellas distinciones que marcan el desarrollo de un evento, i.e. su inicio, su continuidad y su finalización, etapas que expresamos a través de los operadores ingesivo (*ing*), progresivo (*pro*) y egresivo (*egr*). Así, estos operadores expresan un componente temporal que es relevante para el desarrollo interno del estado de cosas, como ilustramos en la Figura 4.

¹⁶ Esta tabla nos ofrece una visión selectiva de los operadores de evento más relevantes, por lo cual no debe ser considerada como un inventario cerrado y exhaustivo.

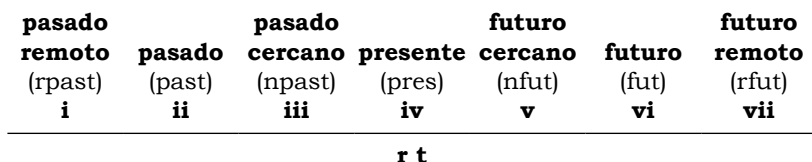
FIGURA 4
Operadores de aspectualidad



- (11) i. Juan empezó a llorar.
 ii. Juan estaba llorando.
 iii. Juan dejó de llorar.

La función de los operadores de temporalidad es ubicar el estado de cosas designado por la predicación en un eje temporal. La Figura 5 nos ofrece las distinciones más relevantes que forman parte de la configuración de este eje temporal donde *rt* representa un punto de referencia temporal.

FIGURA 5
Operadores de temporalidad



- (12) (i) María había cantado.
 (ii) María cantó.
 (iii) María ha cantado.
 (iv) María está cantando.
 (v) María está a punto de cantar.
 (vi, vii) María cantará.

A modo de ilustración, presentamos el PS de +FORGIVE_00 (13), i.e. “dejar de estar enfadado con alguien a quien uno culpó de algo”, que ilustra el uso de los operadores de aspectualidad y temporalidad.

(13) +(e1: egr +FEEL_00 (x1)Agent (x2)Theme (x3: +ANGRY_00) Attribute (f1: (e2: past +BLAME_00 (x1)Theme (x4)Referent (x2)Goal))Scene)

Por lo que se refiere a los operadores que indican modalidad, FunGramKB sigue la distinción tradicional entre los operadores epistémicos y los no-epistémicos, Así, como parte de los primeros, FunGramKB distingue entre certeza (*cert*), probabilidad (*prob*) y posibilidad (*pos*), operadores que expresan los diferentes grados de compromiso que el hablante puede expresar con respecto a la veracidad del contenido proposicional expresado en la predicación. De otro lado, los operadores no-epistémicos tienen un carácter eminentemente deóntico y pueden reflejar los siguientes valores: obligación (*obl*), consejo (*adv*) y permiso (*perm*). Por ejemplo, en el PS de +ADVISE_00 (14), encontramos un operador modal epistémico y otro no epistémico, i.e. “una persona le dice a otra *sugerentemente* que *puede* hacer algo”.

(14) +(e1: adv +SAY_00 (x1: +HUMAN_00)Theme (x4: (e2: pos +DO_00 (x3: +HUMAN_00)Theme (x2)Referent))Referent (x3)Goal)

Como coda final a esta sección, no podemos pasar por alto el operador de polaridad que utiliza FunGramKB para expresar la negación fuerte, la cual expresa explícitamente que, por ejemplo, “A es falso”. Esta es una cuestión fundamental para comprender la naturaleza no monotónica de los operadores de razonamiento (cf. sección 4.2.1), donde la falta de conocimiento no implica su falsedad. Por consiguiente, el operador *n* difiere de lo que se conoce como negación débil, la cual indica que “nada en esta base de conocimiento confirma que A sea cierto”. Por ejemplo, (15) presenta el PS de +FORBID_00, i.e. “le decimos a una persona o animal que está *obligado* a *no* hacer algo”, donde el operador de negación se combina con un operador modal.

(15) +(e1: +SAY_00 (x1: +HUMAN_00)Theme (x4: (e2: obl n +DO_00 (x3: +HUMAN_00 ^ +ANIMAL_00)Theme (x2) Referent))Referent (x3)Goal)

4.2.3. Los cuantificadores y los conectores lógicos

Cada unidad conceptual que actúa como una preferencia de selección en un participante puede estar sometida a operadores que expresen cuantificación. La Tabla 1 nos ofrece los

operadores más relevantes que el ingeniero del conocimiento puede utilizar en FunGramKB.

TABLA 3
Operadores de cuantificación

Rasgo	Valor
Cuantificador absoluto	1 2 3 4 ...
Cuantificador relativo	m s p
Cuantificador indefinido	i

Consideremos los PS (16) y (17) para ilustrar el alcance de estos operadores. Mientras que (16) incluye un cuantificador relativo, por ej. “una cubertería es un grupo de cuchillos, tenedores y cucharas”, (17) ilustra la cuantificación absoluta, por ej. “una cara tiene dos mejillas, un mentón, dos ojos, una nariz y una frente”.

(16) *((e1: +BE_00 (x1: \$CUTLERY_00)Theme (x2: +GROUP_00 Referent)(e2: +COMPRISE_00 (x2)Theme (x3: s +KNIFE_00 & s +FORK_00 & s +SPOON_00)Referent))

(17) *(e2: +BE_02 (x3: 2 +CHEEK_00 & 1 +CHIN_00 & 2 +EYE_00 & 1 +NOSE_00 & 1 +FOREHEAD_00)Theme (x1: +FACE_00 Location)

Si incluimos un cuantificador, utilizaremos solo uno para cada preferencia de selección y, además, deberá ser lo más restringido posible. De esta forma, si nos referimos a una sola entidad, entonces utilizaremos el operador *1*. Sin embargo, si la preferencia de selección se refiere a más de una entidad, debemos recurrir a los valores expuestos en la Tabla 3. Así, si conocemos el número exacto de entidades, entonces utilizaremos un cuantificador absoluto. Si este no es el caso, recurriremos a los cuantificadores relativos que sirven para expresar “muchos” (*m*), “algunos” (*s*) o simplemente unos “pocos” (*p*) referentes. Finalmente, si ninguno de estos criterios es relevante, aplicaremos el operador *i*, que es la opción por defecto y que debe ser interpretado como “*x* se refiere a más de una entidad, pero no sabemos cuántas”. Nótese que estos operadores no son obligatorios, ya que existen casos en los que resulta difícil, e incluso imposible, elegir uno de ellos, por ej. cuando el concepto sujeto

a la cuantificación se refiere a una entidad incontable, o bien a toda una predicación o, simplemente, cuando la cuantificación no es relevante.

Un estudio aparte merece cuando la entidad conceptual sujeta a la cuantificación no es una entidad sino una cualidad. En estos casos, podemos utilizar los cuantificadores relativos *p* y *m* para colocar el concepto como un valor dentro de una dimensión semántica graduable. Así, en la dimensión de la temperatura, podemos describir que una entidad está (18) muy caliente (por ej. *abrasador*), (19) un poco caliente (por ej. *cálido*), (20) muy fría (por ej. *gélido*), o (21) un poco fría (por ej. *fresco*).

(18) +(e2: +BE_01 (x1)Theme (x3: m +HOT_00)Attribute)

(19) +(e2: +BE_01 (x1)Theme (x3: p +HOT_00)Attribute)

(20) +(e2: +BE_01 (x1)Theme (x3: m +COLD_00)Attribute)

(21) +(e2: +BE_01 (x1)Theme (x3: p +COLD_00)Attribute)

Una vez descritas las estructuras sintáctico-semánticas de las predicaciones y los participantes, concluimos este apartado con una breve descripción del uso de los tres conectores lógicos válidos en COREL, i.e. (22) la conjunción &, (23) la disyunción | y (24) la exclusión ^.

(22) a & b & c = a , b y c

(23) a | b | c = a, b, c, ab, ac, bc o abc

(24) a ^ b ^ c = sólo a, sólo b o sólo c

Los conectores lógicos operan sobre estructuras del mismo rango, por ej. entre satélites (25) o entre preferencias de selección (26).

(25) *(e3: +TAKE_00 (x4: +HUMAN_00)Theme (x1: +HANDLE_00) Referent (f1: (e4: +MOVE_00 (x4)Agent (x3)Theme (x5) Location (x6)Origin (x7)Goal))Purpose ^ (f2: (e5: +HOLD_00 (x4)Theme (x3)Referent))Purpose)

(26) *(e4: +BE_01 (x1: +TYRE_00)Theme (x5: +BLACK_00 & +THICK_00 & +HARD_00 & +ROUND_00)Attribute)

Por ejemplo, en la predicación (25) se utiliza el operador de exclusión, a saber, “agarramos un mango para sostener o mover algo”, mientras que la predicación (26) hace uso de la conjunción, i.e. “un neumático es típicamente negro, grueso, duro y redondo”. En el caso de los satélites, los participantes

que conectemos no deben poseer necesariamente la misma función semántica, como ilustramos en el PS del concepto +KISS_00 (27).

(27) *(e1: +TOUCH_00 (x1: +HUMAN_00)Theme (x2)Referent (f1: +LIP_00)Instrument (f2: (e2: +LOVE_00 (x2)Agent (x1)Theme))Reason | (f3: (e3: +RESPECT_00 (x1)Theme (x2)Referent))Reason | (f4: (e4: +GREET_00 (x1)Theme (x3)Referent (x2)Goal))Scene)

Debemos advertir que la conjunción lógica puede tener lugar sin que exista una representación explícita del conector. Por tanto, la conjunción es el conector por defecto, lo que significa que no es necesario incluirlo explícitamente para indicar la concatenación entre dos predicaciones o participantes. Por ejemplo, aunque el PS de +CLIMB_00 (28a) pueda ser teóricamente reformulado como (28b) o (28c), siendo estas tres predicaciones conceptualmente idénticas *a priori*, el validador gramatical¹⁷ de COREL detectará un error sintáctico en (28b), pero no en (28a) y (28c).

(28a) +(e1: +RISE_00 (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal (f1: +FOOT_00 | +HAND_00)Instrument (f2: +CAREFUL_00)Manner (f3: +SLOW_00)Speed)

(28b) +(e1: +RISE_00 (x1)Agent & (x2)Theme & (x3)Location & (x4)Origin & (x5)Goal (f1: +FOOT_00 | +HAND_00)Instrument (f2: +CAREFUL_00)Manner (f3: +SLOW_00)Speed)

(28c) +(e1: +RISE_00 (x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal (f1: +FOOT_00 | +HAND_00)Instrument & (f2: +CAREFUL_00)Manner & (f3: +SLOW_00)Speed)

¹⁷ Con el fin de ayudar en la adquisición y el mantenimiento de la base de conocimiento, revisar su consistencia y soportar interacciones con los ingenieros del conocimiento, se ha desarrollado una serie de herramientas con una interfaz gráfica de fácil manejo que resultan particularmente útiles para la creación, edición, revisión y consulta de FunGramKB. Por ejemplo, en lo concerniente a la revisión automática de los PS, hemos implementado tres tipos de analizadores: léxico, sintáctico y semántico. Por una parte, el analizador léxico comprueba que cada cadena (*token*) que conforma el PS tiene un formato correcto, comprobando en nuestro caso los identificadores conceptuales, los operadores y los signos de puntuación. Por otra parte, el analizador sintáctico comprueba que los *tokens* están en la posición correcta. Finalmente, el analizador semántico detecta los errores que surgen en el PS cuando, por ejemplo, se ha utilizado un concepto que no existe en la ontología, o bien sí existe pero no corresponde al tipo de constructo en el que se ha incorporado.

Puesto que la relación lógica entre dos argumentos siempre es la conjunción, la gramática de COREL no permite el uso explícito de ningún operador en estas situaciones, razón por la cual (28b) está mal formado. En el caso de (28c), los operadores de conjunción son redundantes, ya que, aunque no se explicitaran, se seguiría manteniendo la misma relación lógica. Ahora bien, éstos son válidos porque la gramática de COREL contempla la posibilidad de que en otras predicaciones se pueda establecer alguna otra relación lógica entre dos satélites, como se mostró en (23) y (24).

5. Comentarios finales

Actualmente, la mayoría de bases de datos léxicas multilingües (p. ej. EuroWordNet o SIMPLE, entre otra muchas) adoptan un enfoque relacional para representar los significados léxicos, ya que es más fácil presentar las asociaciones entre palabras en forma de relaciones de significado que describir formalmente el contenido conceptual de las unidades léxicas. No obstante, a pesar de que el desarrollo a gran escala de recursos léxicos dotados de semántica profunda requiere mucho más tiempo y esfuerzo, su poder expresivo no es solo más robusto¹⁸, sino además la gestión del conocimiento resulta más eficiente, como demostraron Perinián-Pascual y Arcas-Túnez (2007).

Dentro del marco de la semántica profunda, este trabajo ofrece una descripción detallada de los aspectos centrales del lenguaje COREL, desarrollado en el ámbito de la base de conocimiento FunGramKB. Como cualquier otro lenguaje, COREL consta de una sintaxis y una semántica, cuyos fundamentos se edifican no solo sobre las unidades conceptuales sino también sobre los constructos de los participantes (i.e. argumentos y satélites) y otros elementos notacionales como los operadores. En definitiva, pretendemos que este trabajo constituya una guía útil

¹⁸ Un caso práctico de integración de nuestra base de conocimiento a un sistema del PLN se describe en Perinián-Pascual y Mairal-Usón (2010), quienes proponen reemplazar la base de datos léxica del traductor automático UniArab (Salem y otros, 2008) por FunGramKB. Precisamente, debido a un tratamiento más profundo en la representación del significado en esta base de conocimiento, UniArab puede adoptar así un auténtico enfoque interlingüístico que favorezca la generación multilingüe.

en la formación del ingeniero del conocimiento que se enfrenta a la labor compleja de representar el conocimiento conceptual y, más concretamente, dentro del marco de FunGramKB.

6. Bibliografía citada

- ALI, Syed, 1993: "A propositional semantic network with structured variables for natural language processing" en *Proceedings of the Sixth Australian Joint Conference on Artificial Intelligence*.
- ARMSTRONG, Sharon Lee, Lila GLEITMAN y Henry GLEITMAN, 1983: "What some concepts might not be", *Cognition* 13, 263-308.
- DIK, Simon, 1997: *The theory of Functional Grammar. Part 1: The structure of the clause*, Berlín / Nueva York: Mouton de Gruyter.
- DIXON, Robert, 1991: *A new approach to English grammar on semantic principles*, Oxford: Clarendon Press.
- GODDARD, Cliff y Anna WIERZBICKA (eds.), 2002: *Meaning and universal grammar*, Ámsterdam: John Benjamins.
- HALLIDAY, Michael, 1985: *An introduction to Functional Grammar*, Londres: Edward Arnold.
- JIMÉNEZ BRIONES, Rocío y María Beatriz PÉREZ CABELLO DE ALBA, 2009: "Selection restrictions within a conceptual knowledge base", comunicación presentada en *The 2009 International Conference in Role and Reference Grammar*.
- KOENIG, Jean-Pierre, Gail MAUNER y Breton BIENVENUE, 2003: "Arguments for adjuncts", *Cognition* 89, 67-103.
- LANDAU, Barbara, 1982: "Will the real grandmother please stand up? The psychological reality of dual meaning representations", *Journal of Psycholinguistic Research* 11(1), 47-62.
- LEECH, Geoffrey, 1974: *Semantics*, Harmondsworth: Penguin.
- LAURENCE, Stephen y Eric MARGOLIS, 1999: "Concepts and cognitive science" en Eric MARGOLIS y Stephen LAURENCE (eds.): *Concepts: Core readings*, Cambridge (Mass.): Bradford Books/MIT Press, 3-81.
- LENCI, Alessandro, 2000: "Building an ontology for the lexicon: Semantic types and word meaning" en *Workshop on Ontology-Based Interpretation of Noun Phrases*.
- LENCI, Alessandro, Nuria BEL, Federica BUSA, Nicoletta CALZOLARI, Elisabetta GOLA, Monica MONACHINI, Antoine OGWONSKI, Ivonne PETERS, Wim PETERS, Nilda RUIMY, Marta VILLEGAS y Antonio ZAMPOLLI, 2000: "SIMPLE: A general framework for the development of multilingual lexicons", *International Journal of Lexicography* 13 (4), 249-263.
- OSHERSON, Daniel y Edward SMITH, 1981: "On the adequacy of Prototype Theory as a theory of concepts", *Cognition* 9, 35-58.
- PAPAFRAGOU, Anna, 1996: "On generics", *UCL Working Papers in Linguistics* 8, 165-98.

- PEACOCKE, Christopher, 1992: *A study of concepts*, Cambridge (Mass.): MIT Press.
- PERINIÁN-PASCUAL, Carlos y Francisco ARCAS-TÚNEZ, 2005: "Microconceptual-Knowledge Spreading in FunGramKB" en *Proceedings of the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, Anaheim-Calgary-Zurich: ACTA Press, 239-244.
- , 2007: "Deep semantics in an NLP knowledge base" en *Proceedings of the 12th Conference of the Spanish Association for Artificial Intelligence*, Salamanca: Universidad de Salamanca, 279-288.
- , 2010: "The architecture of FunGramKB" en *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, Malta.
- PERINIÁN-PASCUAL, Carlos y Ricardo MAIRAL-USÓN, 2010: "Enhancing UniArab with FunGramKB", *Procesamiento del Lenguaje Natural* 44, 19-26.
- , s. f.: "Prototypicality, universality and culturality in an NLP knowledge base".
- PROCTER, Paul (ed.), 1978: *Longman Dictionary of Contemporary English*, Harlow (Essex): Longman.
- PUSTEJOVSKY, James, 1995: *The Generative Lexicon*, Cambridge (Mass.): MIT Press.
- REY, Georges, 1994: "Concepts" en Samuel GUTTENPLAN (ed.): *A companion to the philosophy of mind*, Cambridge (Mass.): Blackwell, 185-193.
- SALEM, Yasser, Arnold HENSMAN y Brian NOLAN, 2008: "Towards Arabic to English machine translation", *ITB Journal* 17, 20-31.
- TULVING, Endel, 1985: "How many memory systems are there?", *American Psychologist* 40, 385-398.
- VOSSSEN, Piek, 1998: "Introduction to EuroWordNet", *Computers and the Humanities* 32 (2-3), 73-89.
- WOODS, William, 1983: "What's important about knowledge representation?", *IEEE Computer* 16 (10), 22-26.

ANEXO 1
**La interpretación semántica de los papeles temáticos
de los argumentos**

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

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

ANEXO 2
**La interpretación semántica de los papeles temáticos
de los satélites**

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.

¹⁹ El satélite con la función Position es el único participante que depende de otro para su interpretación semántica. De acuerdo con su definición, inferimos que (i) la preferencia de selección del participante Position debe ser una cualidad que pertenezca a la dimensión #SPATIAL –preferentemente conceptos como +DOWN, +IN, +ON, +OUT o +UP– y (ii) este participante siempre aparece inmediatamente después de un participante Location, Origin o Goal. Por ejemplo, la predicación (31) describe que “el relleno se coloca dentro de la carne o las verduras”.

(31) +(e2: +BE_00 (x1 \$STUFFING_00)Theme (x4: +MEAT_00 ^ +VEGETABLE_00) Location (f1: +IN_00)Position)

ANEXO 3

La comparación de los papeles temáticos en FunGramKB con los roles semánticos de Halliday (1985) y con los de Dixon (1991)

	Halliday	Dixon	FunGramKB
#COGNITION	Senser + Phenomenon (COGNITION)	Cogitator + Thought (THINKING)	(x1)Agent (x2)Theme (x3)Referent
#COMMUNICATION	Sayer + Saying + Addressee (PROPER VERBALS)	Speaker + Message + Addressee (SPEAKING)	(x1)Theme (x2)Referent (x3)Goal
#CONSTITUTION	Part + Whole (PART-WHOLE)	Resting + Locus (CONTAIN)	(x1)Theme (x2)Referent
#CREATION	Actor + Result (CREATIVE MATERIAL ACTION)	Agent + Target (BUILD)	(x1)Theme (x2)Referent
#EMOTION	Senser + Phenomenon (EMOTION)	Experiencer + Stimulus (LIKING)	(x1)Agent (x2)Theme (x3)Attribute ²
#EXISTENCE	Existent (EXISTENCE)		(x1)Theme
#IDENTIFICATION	Domain + Range (IDENTITY) Attribute + Attribuend (ASCRPTION) Symbol + Symbolized (SYMBOLIZATION)		(x1)Theme (x2)Referent (x3)Attribute
#INTENTION	Senser + Phenomenon (INTENTION)		(x1)Theme (x2)Referent
#LOCATION	Carrier + Location (GENERALIZED POSITIONING)	Resting + Locus (STAY)	(x1)Theme (x2)Location
#MATERIAL	Actor [+ Range] (NON-DIRECTED DOING) Actor (HAPPENING)	Agent + Target (HIT, RUB, STAB, TOUCH, WRAP) Human [+ Substance] (CORPOREAL)	(x1)Theme (x2)Referent
#MOTION	[Agentive +] Actor [+ Range] (MOTION PROCESS)	Moving (DROP, RUN) Moving + Locus (ARRIVE, FOLLOW) Causer + Moving (CARRY, THROW) Causer + Moving + Locus (TAKE)	(x1)Agent (x2)Theme (x3)Location (x4)Origin (x5)Goal
#PERCEPTION	Senser + Phenomenon (PERCEPTION)	Perceiver + Impression (ATTENTION)	(x)Theme (x)Referent
#POSSESSION	Possessor + Possessed (OWNERSHIP)	Owner + Possession (OWNING)	(x1)Theme (x2)Referent
#TRANSFER	Actor + Range + Recipient (NON-DIRECTED DOING)	Donor + Gift + Recipient (GIVING)	(x1)Agent (x2)Theme (x3)Origin (x4)Goal
#TRANSFORMATION	Actor + Result (DISPOSITIVE MATERIAL ACTION)	Agent + Target (BREAK, STRETCH)	(x1)Theme (x2)Referent

²⁰ Nótese que utilizamos corchetes para marcar la opcionalidad de los argumentos.