

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

FACULTAD DE FILOLOGÍA

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

Trabajo de Fin de Máster

Resolución de la anáfora correferencial con FunGramKB

Autora: María José Ruiz Frutos
Tutora: Dra. Rocío Jiménez Briones
Procesamiento del Lenguaje Natural
Madrid, Septiembre 2012

Resolución de la anáfora correferencial con FunGramKB

María José Ruiz Frutos

A Pedro, el referente en mi vida

Agradecimientos

Este trabajo no habría sido posible sin el constante apoyo, la amable disponibilidad y la valiosa orientación de Rocío Jiménez Briones. Le estoy muy agradecida por todas sus recomendaciones, sugerencias, ánimos y especialmente por haber contribuido a aumentar mi motivación según iba elaborando el estudio. A Carlos Perrián Pascual le agradezco su introducción al mundo del Procesamiento del Lenguaje Natural y el haberme facilitado el acceso a la base de conocimiento FunGramKB. Igualmente quiero expresar mi agradecimiento a los profesores que me acercaron a este tema: a Hans-Jörg Schmid y Manfred von Roncador por sus seminarios de Lingüística Cognitiva y Referencia; a Teresa Solías Arís por despertar mi interés por la Lingüística. En último lugar, gracias a los familiares y amigos que me han acompañado a lo largo de este camino.

El lenguaje es un ordenamiento eficaz de esa enigmática abundancia del mundo.

(“Palabrería para versos”, *El tamaño de mi esperanza*, Jorge Luis Borges)

Índice

ÍNDICE DE ABREVIATURAS, ACRÓNIMOS Y SIGLAS	iii
INTRODUCCIÓN	1
1. MARCO TEÓRICO	4
1.1. ANÁFORA Y CORREFERENCIA	4
1.1.1. CONCEPTOS Y TERMINOLOGÍA	4
1.1.2. TIPOS DE ANÁFORA	7
1.1.3. RESOLUCIÓN AUTOMÁTICA DE ANÁFORAS	12
1.2. ESTADO DE LA CUESTIÓN EN PLN	16
1.2.1. SISTEMAS DE CONOCIMIENTO LIMITADO Y ENRIQUECIDO	16
1.2.2. SISTEMAS DE CONOCIMIENTO ALTERNATIVO	18
1.2.3. APORTACIONES RECIENTES Y RETOS	18
1.3. MODELOS TEÓRICOS	20
1.3.1. LA GRAMÁTICA DEL PAPEL Y LA REFERENCIA	20
1.3.2. EL MODELO LÉXICO CONSTRUCCIONAL	30
1.3.3. LA BASE DE CONOCIMIENTO LÉXICO-CONCEPTUAL FUNGRAMKB	37
2. OBJETIVOS	54
2.1. OBJETIVO PRINCIPAL	54
2.2. OBJETIVOS SECUNDARIOS	54
3. METODOLOGÍA	57
3.1. HIPÓTESIS	57
3.2. HERRAMIENTAS	57
3.3. PROCEDIMIENTO	58
4. DATOS	60
4.1. DESCRIPCIÓN DE DATOS OBTENIDOS	60
4.2. ANÁLISIS DE RESULTADOS	62
4.2.1. IDENTIFICACIÓN DE ELEMENTOS CORREFERENCIALES	63
4.2.2. LOCALIZACIÓN DE LOS CANDIDATOS A ANTECEDENTE	64
4.2.3. SELECCIÓN DEL ANTECEDENTE	68
4.2.4. PROPUESTA DE ALGORITMO PARA RESOLVER LA CORREFERENCIA	69
5. CONCLUSIONES	71
5.1. FUTURAS INVESTIGACIONES	72
ANEXOS	75
BIBLIOGRAFÍA	117

Índice de abreviaturas, acrónimos y siglas¹

- ACS** – *Accessible* (referente del SN identificable y accesible)
- ACV** – *Activated* (referente del SN identificable y activo)
- AFD** – *Actual Focus Domain* (Dominio Real de Foco)
- AGX** – *Agreement index* (índice de concordancia)
- AGX_G/AGX_P** – *Agreement index-genus/-person* (índice de concordancia de género/persona)
- BioKnowing** – *Bioconceptual-Knowledge Spreading* (expansión bio-conceptual)
- BNA** – *Brand New Anchored* (referente del SN no identificable y nuevo, anclado a otro)
- BNU** – *Brand New Unanchored* (referente del SN no identificable y nuevo, no anclado)
- CLS** – *Conceptual Logical Structure* (Estructura Lógico Conceptual)
- COREL** – *Conceptual Representation Language* (Lenguaje de Representación Conceptual)
- FunGramKB** – *Functional Grammar Knowledge Base* (base de conocimiento basada en la Gramática Funcional)
- INA** – *Inactivated* (referente del SN identificable e inactivo)
- LCM** – *Lexical Constructional Model* (Modelo Léxico-Construccional)
- LCS** – *Layered Clause Structure* (*Estructura Estratificada de la Cláusula*)
- LDP** – *Left Detached Position* (Posición Dislocada a la Izquierda)
- LNPS** – *Layered NP Structure* (Estructura Estratificada del SN)
- LS** – *Logical Structure* (Estructura Lógica)
- MacroKnowing** – *Macroconceptual-Knowledge Spreading* (expansión macro-conceptual)
- MicroKnowing** – *Microconceptual-Knowledge Spreading* (expansión micro-conceptual)
- MP** – *Meaning Postulate* (Postulado de Significado)
- NFPF** – *NP-final position* (posición de final del SN)

¹ La terminología está aclarada en español y en inglés a lo largo del trabajo y recopilada en el anexo 3 con el glosario. Debido a que la mayor parte de la bibliografía sobre este tema está redactada en inglés, en este trabajo se emplean las abreviaturas en ese idioma (excepto *PLN*, *SN* y *SP*); indicándose la traducción al español en este índice de abreviaturas, acrónimos y siglas.

NPIP – *NP-initial position* (posición de comienzo del SN)
PFD – *Potential Focus Domain* (Dominio Potencial de Foco)
PLN – Procesamiento del Lenguaje Natural (*Natural Language Processing*)
PoCS – *PostCore Slot* (Espacio Post-Central)
PrCS – *PreCore Slot* (Espacio Pre-Central)
PSA – *Privileged Syntactic Argument* (Argumento Sintáctico Privilegiado)
RDP – *Right Detached Position* (Posición Dislocada a la Derecha)
RP-d – *Uninterrupted reference* (referencia ininterrumpida)
RP-f – *Referential frequency* (frecuencia referencial)
RRG – *Role and Reference Grammar* (Gramática del Papel y la Referencia)
SN – Sintagma Nominal (*Noun Phrase*)
SP – Sintagma Preposicional (*Prepositional Phrase*)
TF – *Thematic Frame* (Marco Temático)

INTRODUCCIÓN

Actualmente los sistemas de Procesamiento del Lenguaje Natural se enfrentan a la dificultad de interpretar correctamente las anáforas y conectar las correferencias de los textos con los que trabajan. En diversos proyectos a nivel internacional se investiga sobre las posibilidades de incluir conocimiento del mundo, del sentido común y/o una semántica profunda para poder simular el razonamiento humano en el proceso de resolución de la anáfora y de la correferencia donde se enmarca este trabajo. Por tanto, nos situamos dentro del ámbito del Procesamiento del Lenguaje Natural desde el cual, como parte de la Inteligencia Artificial, se quiere plantear mecanismos que computacionalmente logren la comprensión del lenguaje humano y proponer así productos de Ingeniería Lingüística. Concretamente, en esta indagación se plantea la base de conocimiento FunGramKB para resolver la anáfora y la correferencia, centrándose en los elementos correferenciales en forma de Sintagma Nominal Léxico definido, indefinido o con un nombre propio. Por tanto, quedan fuera del alcance de este estudio las anáforas pronominales, adverbiales, verbales y las elisiones anafóricas, a excepción de su revisión teórica y ciertas referencias durante el análisis. Se comenta también el tratamiento de anáforas indirectas, cuya resolución supone todo un desafío debido a las asociaciones semánticas que deben tenerse en cuenta. Los idiomas en los que investigamos la resolución correferencial están limitados a aquellos de los que más información dispone FunGramKB hasta el momento: el inglés y el español.

El objetivo en este análisis es examinar a un nivel teórico qué elementos, tanto del conocimiento almacenado en FunGramKB, como del razonamiento llevado a cabo en esta base de conocimiento, pueden contribuir a resolver con éxito anáforas correferenciales. Para ello contamos con un corpus de dos artículos periodísticos en español y en inglés, incluyendo los comentarios escritos por los lectores.

El sustento teórico que guía esta indagación se basa en una investigación documental de los estudios más recientes hasta el año 2012 sobre la resolución automática de la anáfora, la Gramática del Papel y la Referencia, el Modelo Léxico Construccional y FunGramKB.

La tarea de resolver anáforas y correferencias juega un papel importante en diferentes aplicaciones del Procesamiento del Lenguaje Natural y es por ello por lo que ha suscitado nuestro interés para realizar esta investigación. En traducción automática, por ejemplo, resulta fundamental a la hora de traducir adecuadamente los géneros de los pronombres o cuando la lengua de destino presenta un sistema de género diferente a la original. Respecto a la extracción de información, al tener que identificar determinados datos en un corpus dado, se vuelve primordial la resolución de la correferencia para poder integrar la información referida a la misma entidad que aparece en diferentes partes del texto o conjunto de textos. Tanto en la síntesis automática de un texto como de varios textos son relevantes la resolución de la correferencia textual y trans-documental, ya que si no son resueltos, los elementos anafóricos pueden no ser tenidos tanto en cuenta a la hora de calcular su importancia informativa. La resolución de la correferencia resulta también valiosa para conseguir resultados más precisos en la extracción automática de términos y la clasificación de textos; así como en los programas de pregunta-respuesta para ser capaz de interpretar las diferentes formas de referirse a la misma entidad. La minería de opiniones puede aprovechar muy bien la resolución correcta de correferencias para extraer la información sobre las distintas opiniones. Otro ámbito de empleo que puede beneficiar, entre otros, a posibles lectores con dislexia es una aplicación creada para sustituir los pronombres de un texto por sus antecedentes, de modo que personas con dificultades a la hora de resolver anáforas puedan interpretarlas correctamente. Dentro del terreno médico, la acertada resolución de la anáfora se convierte en una herramienta especialmente útil para el desarrollo y la investigación, ya que la minería de textos biomédicos y la extracción de información pueden así obtener datos más precisos de estas publicaciones científicas. Vemos por tanto cómo, en diferentes ámbitos, numerosas aplicaciones del Procesamiento del Lenguaje Natural pueden aprovechar los avances aportados en el campo de la resolución de la anáfora y la correferencia.

Esta investigación consta de cinco capítulos. En el primero se presenta la terminología básica y los conceptos fundamentales sobre la anáfora y la correferencia, se explican los diferentes tipos de anáforas con ejemplos en inglés y español seleccionados de los corpus COCA, BNC y CREA. Tras presentar ciertas diferencias respecto a la anáfora entre los idiomas español e inglés y mencionar los distintos niveles de conocimiento implicados en la resolución de anáforas y

correferencias, se trata el estado de la cuestión en Procesamiento del Lenguaje Natural, nombrando destacados sistemas de conocimiento limitado, enriquecido y alternativo, para finalmente hacer referencia a los avances y retos actualmente en este ámbito. Seguidamente se presenta el marco teórico fundamentado en diversos aspectos de la Gramática del Papel y la Referencia, el Modelo Léxico Construccional y la base de conocimiento FunGramKB.

En el segundo capítulo se expone el objetivo principal y los secundarios según la información léxica, ontológica, procedural, cultural y relacional almacenada en FunGramKB.

En el tercer capítulo se enuncia la hipótesis planteada, se comentan las herramientas usadas y se explica la metodología de investigación seguida.

En el cuarto capítulo se describen los datos en los que se asienta la investigación llevada a cabo y se analizan los resultados obtenidos en base a las diferentes fases que se pueden distinguir en un sistema de resolución automática de correferencias, proponiendo al final un algoritmo para ello.

En el quinto y último capítulo se recopilan las conclusiones a las que ha conducido este estudio y se proponen temas para futuras investigaciones.

Se incluyen al final tres anexos con una propuesta de plantillas basada en la Gramática del Papel y la Referencia para el Sintagma Nominal anafórico y no anafórico, el corpus de artículos periodísticos sobre el que se ha trabajado y un glosario con la definición de los términos de uso frecuente en Procesamiento del Lenguaje Natural y los modelos teóricos. En último lugar, las referencias a las diferentes publicaciones están recogidas en la bibliografía.

1. MARCO TEÓRICO

1.1. Anáfora y correferencia

En este subcapítulo se introduce la terminología básica, se expone la clasificación de los tipos de anáfora, en la cual se enmarca la anáfora indirecta, y se tratan los aspectos a tener en cuenta en el proceso de resolución automática de anáforas y correferencias.

1.1.1. Conceptos y terminología

El fenómeno de la **anáfora** (Halliday y Hasan, 1976; Hirst, 1981; Kuno, 1987) consiste en señalar a un elemento mencionado anteriormente denominado *antecedente*. Por tanto, el **antecedente** es la entidad a la cual se refiere el elemento anafórico. Cuando este se encuentra delante del antecedente se trata de una **catáfora**; de hecho, los términos *anáfora* y *catáfora* provienen del compuesto griego *ἀνά* 'hacia arriba' o *κατά* 'hacia abajo' y *φέρω* 'llevar'. Al hacer referencia a un elemento dentro del discurso, ambos fenómenos son clasificados como una **endófora**², mientras que la **exófora** se produce al señalar a un elemento extralingüístico, que requiere conocimiento del contexto físico en la **deixis** y conocimiento cultural en la **homófora**.

La **correferencia**, en cambio, consiste en que dos elementos aludan al mismo referente en el mundo real. Si los elementos correferenciales aparecen dentro de un mismo texto, entre ellos se establecen además relaciones anafóricas y catafóricas. De modo que, en el caso de un solo texto, la diferencia entre anáfora y correferencia

² Los dos tipos de endófora, anáfora y catáfora, tienen otras denominaciones como *retrospective and anticipatory anaphora* en Huddleston y Pullum (2002).

radica en que, en esta última, antecedente y elemento anafórico comparten un mismo referente extralingüístico, como queda reflejado en la figura 1:

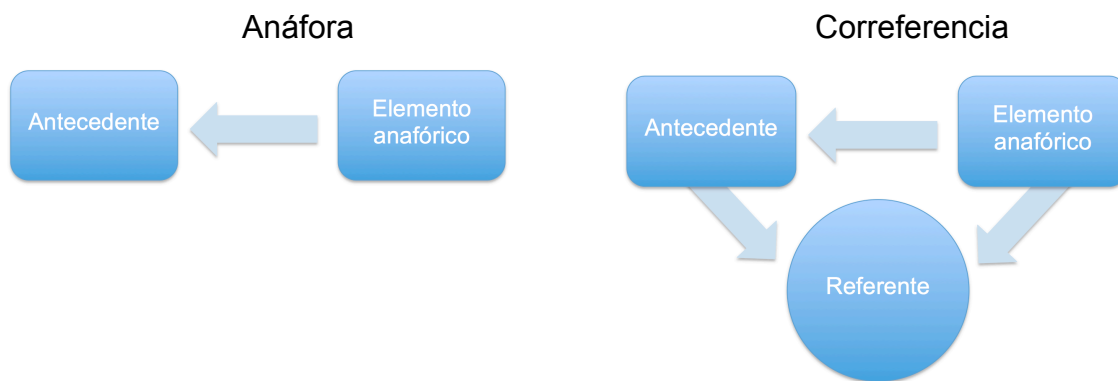


Figura 1. Diferencia entre anáfora y correferencia dentro de un mismo texto

Con el ejemplo aportado en la siguiente figura 2 se puede apreciar cómo lo anafórico señala a otro elemento sin compartir referente: *Todo escritor ... el autor ...*; mientras que la correferencia señala a otro elemento con el que sí comparte el mismo referente: *Carlos Fuentes ... el autor...* Por otro lado, al compartir tanto el antecedente como el elemento anafórico dentro de un texto el mismo referente en el mundo real, se producen **cadena correferenciales** (*correferential chains*) compuestas por todas las entidades con idéntico referente. Por último, la **correferencia trans-documental** (*cross-document coreference*) consiste en dos alusiones al mismo referente a través de elementos correferenciales sin ser anafóricos debido a que aparecen en textos diferentes.

Por otra parte, la **resolución de la anáfora** consiste en seleccionar el antecedente del elemento anafórico, mientras que en la **resolución de la correferencia** se deben identificar las (cadena de) correferencias (Mitkov, 2002) resolviendo diferentes descripciones o denominaciones para la misma entidad (Hendrickx, Hoste y Daelemans, 2007).

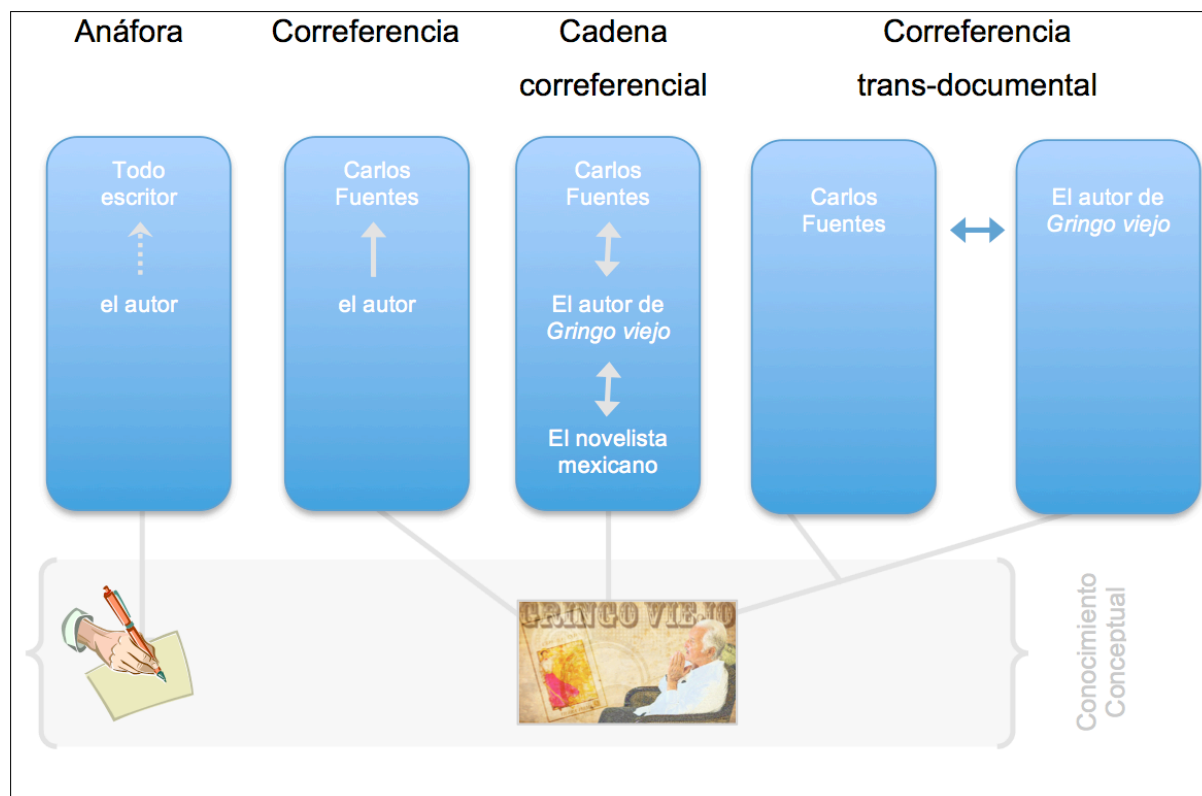


Figura 2. Conocimiento conceptual en la resolución de la anáfora y de la correferencia³

En esta figura podemos observar cómo la resolución de la anáfora consiste en determinar el antecedente correspondiente. Para ello resulta imprescindible disponer del conocimiento conceptual necesario que nos permita seleccionar el antecedente adecuado. Por ejemplo, si el elemento anafórico *el autor* tuviera los siguientes dos candidatos a antecedente: *todo escritor* y *todo libro*, conociendo el significado de *el autor* sabríamos que hace referencia a un ser humano, gracias a lo cual descartaríamos el candidato no humano *todo libro*. Por otro lado, debido a que la interpretación de un elemento depende del otro, de esta manera se favorece la cohesión textual.

También en la figura 2 vemos que, para resolver con éxito las correferencias intra- o intertextuales, es indispensable contar con cierto contenido cultural de modo que podamos identificar el referente en el mundo real: Carlos Fuentes, escritor con nacionalidad mexicana, con varias novelas publicadas (entre otras *Gringo viejo*).

³ Las imágenes de la figura 2 fueron extraídas de:

<<http://office.microsoft.com/en-us/images/results.aspx?qu=pen&ex=1#ai:MC900251063>> y

<http://www.wawis.com.mx/wp-content/uploads/2010/07/wawis_gringo_viejo_carlos_fuentes.jpg>.

1.1.2. Tipos de anáfora

En este apartado se presentan las variedades de la anáfora siguiendo la clasificación de Mitkov (2002). La anáfora se puede clasificar según su forma, tipos y localización de antecedentes, localización del elemento anafórico y antecedente, tipo de inferencia y, por último, dependiendo de si lo que se comparte es el referente o el sentido.

<p>Forma del elemento anafórico</p> <ul style="list-style-type: none"> • Nominal: <ul style="list-style-type: none"> • <i>Pronominal</i> • <i>SN léxico</i> • <i>Pro-forma no léxica</i> • Verbal • Adverbial • Cero 	<p>Forma del antecedente</p> <ul style="list-style-type: none"> • Nombre • Verbo • Sintagma Verbal • Frase • Conjunto de frases • Coordinados 	<p>Localización del antecedente</p> <ul style="list-style-type: none"> • Anáfora • Catáfora • Deixis
<p>Localización del elem. anaf. y el antecedente</p> <ul style="list-style-type: none"> • Intra-oracional • Inter-oracional 	<p>Tipo de inferencia</p> <ul style="list-style-type: none"> • Directa • Indirecta 	<p>Mismo o diferente referente</p> <ul style="list-style-type: none"> • Mismo referente • Mismo sentido

Tabla 1. Clasificación de los diferentes tipos de anáfora
Adaptado de Mitkov (2002)

En los subapartados 1.1.2.1 – 1.1.2.4 se ilustra esta clasificación con ejemplos de lengua oral de distintos países extraídos de varios corpus. Para el inglés británico se ha usado el *British National Corpus* (BNC)⁴ y para el inglés americano el *Corpus of Contemporary American English* (COCA)⁵. Respecto al español, dentro del *Corpus de Referencia del Español Actual* (CREA)⁶ se ha tratado de recopilar ejemplos de diferentes países latinoamericanos (Argentina, Chile, Colombia, Costa Rica, Cuba, México y Venezuela) y de España. De esta manera se pretende no ceñirse únicamente al español peninsular, sino incluir una rica muestra de otras variedades idiomáticas. El hecho de que todos los ejemplos provengan de

⁴ Davies, Mark. *BYU-BNC*. (Basado en el *British National Corpus* de Oxford University Press). [En línea] <<http://corpus.byu.edu/bnc/>> [19.07.12].

⁵ Davies, Mark. *The Corpus of Contemporary American English: 450 million words, 1990-present*. [En línea] <<http://corpus.byu.edu/coca/>> [19.07.12].

⁶ REAL ACADEMIA ESPAÑOLA. Banco de datos (CREA). *Corpus de referencia del español actual*. [En línea] <<http://www.rae.es>> [19.07.12].

textos orales busca reflejar el uso cotidiano y continuo que todo hablante hace de las anáforas y las correferencias.

1.1.2.1. Según la forma

En cuanto a la forma, antes de hablar de la relativa a los antecedentes, se presentará la clasificación respecto a la **forma de los elementos anafóricos**, según la cual se puede distinguir entre la anáfora nominal, verbal, adverbial y cero. De todas estas anáforas la más frecuente es la anáfora **nominal**, ya sea a través de un pronombre, un Sintagma Nominal (SN) léxico o una pro-forma no léxica. Dentro de la anáfora nominal, la anáfora **pronominal** muestra un uso más extendido, empleando pronombres personales, posesivos, demostrativos o relativos⁷:

- (1) [...] and on your screen, you will get a whole list of places where this is used with the beginning bit of the sentence [...]. (Gran Bretaña).
- (2) [...] pero en cambio había otras [partes] donde eso ni se nombraba, sino sus brindis, eso sí, las hallacas siempre, y la gente bailando. Yo te digo que [...]. (Venezuela).

Hay que tener en cuenta que deben exceptuarse los **usos pleonásticos** de algunos pronombres no anafóricos:

- (3) You know my hair frizzes when it rains. (EEUU).
- (4) [...] creo que este año cuando ha llovido al principio de año [...]. (España).

La anáfora de **SN léxico** se lleva a cabo con descripciones definidas y nombres propios. Es necesario resaltar el hecho de que aportan un significado extra y que además tienen un significado aparte del de su antecedente.

- (5) [...] while *Karl May* somehow captured the spirit of the Wild West, he never actually went there. The German writer wrote most of his best work from a German jail [...]. (EEUU).
- (6) *Fuentes* es un mexicano habituado a mirar lejos. [...] Es el escritor que mejor retrata y recrea nuestra vida urbana en este último medio siglo del milenio. (México).

Otra variedad es la anáfora de **pro-forma no léxica** (*noun anaphora*), con la que se refiere a un núcleo o grupo nominal:

- (7) [...] a single *cake* for everybody. Not like a big one to share. (Gran Bretaña).

⁷ Debemos recordar que, si se señala a un elemento no presente en el texto, sino extralingüístico, se trata de una exófora.

- (8) [...] construyó varios *edificios* más, uno grande que hay también en la Plaza de Vigo [...]. (España).

La anáfora **verbal** se ejemplifica con los siguientes casos:

- (9) [...] but *there would be no fundamental change in the regime*, exactly as it happened in Yemen. (EEUU).
- (10) [...] cuando *arrancaron de nuestras casas a nuestros padres*, como ocurrió en mi caso personal. (Argentina).

La anáfora **adverbial** puede observarse en estos ejemplos:

- (11) With us from *New Orleans*, Congressman Adam Putnam. He is there because his Florida Gators played LSU last night in New Orleans [...]. (EEUU).
- (12) [...] él solicitó exilio en *Francia*, allá está, no puede regresar [...]. (México).

Antes de continuar, es fundamental recalcar un hecho que no debemos dejar de tener presente: “[...] la anáfora se basa en la repetición total o parcial de la forma y/o del contenido, la elipsis se basa precisamente en la eliminación de elementos”. (Moreno Cabrera, 1998, p. 361). Los siguientes ejemplos ilustran esta distinción y muestran varios tipos de anáfora **cero** o **elipsis**.

- **Elipsis pronominal**⁸

- (13) I excused myself and _____ went to bed. (Gran Bretaña).
- (14) Ya _____ estaba casada pero _____ estaba joven. (Venezuela).
- (15) Yo estaba casada. Ya _____ estaba casada pero _____ estaba joven. Sí, yo tendría unos treinta y cinco años [...]. (Venezuela).

- **Elipsis nominal**

- (16) Still most *Cubans* say they're catholic but very few ___ are practicing. (EEUU).
- (17) [...] para que aparezcan allí las ganancias, donde no hay *impuestos* o hay muy pocos _____, mientras que acá, normalmente, arrojan pérdidas. (Argentina).

- **Elipsis verbal**

- (18) You *were looking forward* to a cataclysm, and I _____ to the natural process of things. (EEUU).
- (19) [...] pero uno no se *da cuenta*, pero yo sí _____. (Venezuela).

⁸ Branco (2007) propone la consideración de la elipsis del elemento con el macropapel de Actor (función de sujeto) como un reflexivo más que como un pronombre, debido a que comparte más similitudes con aquel que con este último.

- **Elipsis del Sintagma Verbal**

(20) - I've got such a headache. *My head's killing me.*
- Oh me too, what is it? (Gran Bretaña).

(21) And he's always saying, *he's going to see her* and he never does.⁹
(Gran Bretaña).

(22) Bueno, yo no soy *celoso*. Pero él sí ____ . (España).

Respecto a la **forma del antecedente**, este puede ser tanto un **nombre** (*Karl May, Fuentes, cake, edificio, Francia, New Orleans, Cubans, impuestos*), un **verbo** (*se da cuenta*), un **Sintagma Verbal** (*arrancaron de nuestras casas a nuestros padres, soy celoso*) una **frase** (*there would be no fundamental change in the regime*), un **conjunto de frases** (véase el ejemplo 25) o un grupo de antecedentes **coordinados** (véase los ejemplos 23 y 24):

(23) We can reveal exclusively that when *Tom Cruise and Nicole Kidman* were in London a few weeks ago they went shopping for baby clothes.
(Gran Bretaña).

(24) Junto a él y detrás de el [sic] vienen *Tim, otro labrador, Marrón, Luna, una samoyedo, Lara y Gretel* y ya todos ellos son perros que durante años, entre las siete y nueve de la mañana, han paseado por este parque [...]. (Perú).

(25) [...] entonces yo veía el camión de La Lagunita que era un camioncero para cargar mil quinientos kilos, más o menos ¿no?, un camión que se llenaba con cuarenta sacos de piña y cuarenta sacos de naranja yo lo veía tan grande era de doble rueda, por fin un camión doble rueda. Yo lo veía tan grande, porque era un gigante, que casi tocaba la los tejados de las calles del pueblo. Ahora, después de eso he visto camiones que ni con una vara logro alcanzar el extremo superior de las ruedas. (Venezuela).

1.1.2.2. Según la localización

En relación con la **localización del antecedente**, por un lado, se distingue entre **anáfora** y **catáfora**. Aquella es el “act of pointing back to a previous mentioned item” (Mitkov, 2002, p. 23), mientras que esta última tiene lugar cuando el antecedente aparece después. En los siguientes ejemplos 26 y 27 se puede apreciar el empleo catafórico y posteriormente anafórico del adverbio de lugar *there* y *allá*:

(26) And then when the earthquake hit there *in Mexico City* in' 85, I was sent there and ran the relief operations. (EEUU).

⁹ Varios autores como Greenbaum y Quirk (1990) o Huddleston y Pullum (2002) denominan *pro-form* a este segundo Sintagma Verbal abreviado.

- (27) Pero los que están allá, *en la otra sala*, son de material italiano que es bastante diferente del nuestro. Y allá se ve, pues, la diferencia enorme y mucho más lucido el material italiano [...]. (Colombia).

Por otro lado, en el caso de que no haya antecedente en el texto, sino que en realidad se esté haciendo alusión a un elemento extra-textual, se trata de **deixis**:

- (28) I mean, look at that shoe. It's so glorious. (EEUU).
(29) ¿Dónde estacionó el carro del que bajaron las personas? Ahí justamente donde esta [sic] ese ese auto marrón. (Perú).

Respecto a la **localización del elemento anafórico y el antecedente**, se puede distinguir entre **intra-oracional** (ambos en la misma oración) e **inter-oracional** (cada uno en una oración distinta). Ejemplos del primer caso son el posesivo *my* y el reflexivo *se* junto a la morfología verbal de tercera persona plural. Ejemplos del segundo caso son el pronombre *it* y el SN *estas reformas*:

- (30) I want my voice back. I have to get it back. (EEUU).
(31) No abundaré en los contenidos de *la reforma*, a los cuales se han referido en detalle el señor Ministro y el señor Senador informante. Pero quiero hacer presente que estas reformas son necesarias y justas. (Chile).

1.1.2.3. Según la inferencia

Dependiendo del tipo de inferencia se distingue entre anáfora directa e indirecta. El anterior ejemplo 31 muestra una **anáfora directa** que comparte el mismo núcleo del SN: *la reforma* ← *estas reformas*. En el presente trabajo será especialmente relevante la **anáfora indirecta**¹⁰, la cual tiene lugar cuando “a reference becomes part of the hearer’s or reader’s knowledge indirectly rather than by direct mention” (Mitkov, 2002, p. 15); debido a que “bridging reference resolution often requires some form of world knowledge – or at least information that is not explicitly represented in the textual context of the reference” (Hendrickx, De Clercq y Hoste, 2011, p. 2).

- (32) I can't read. But I can order a pizza and *watch TV* / Oh, no / Where is that remote control. (EEUU).
(33) *India's Taj Mahal*, built in the 17th century by an emperor to honor his late wife, a reminder of the past, but also a reminder of modern-day problems. It sits in Agra, one of the poorest and most polluted places in

¹⁰ Esta *referencia puente* también es denominada *anáfora asociativa* por la Real Academia Española en su *Nueva gramática de la lengua española* (2009), donde se destaca el carácter léxico o pragmático de la relación asociativa entre los elementos vinculados.

India. Just a mile from the monument, waste fills the city's main drain, fumes flow out of diesel-fueled cars and smog hangs in the air, discoloring the white-marbled wonder of the world. (EEUU).

(34) Well, you told me *this bag* was indestructible. Look at *it!* The handle's broke. *It's* all shredded, and the leather's all torn. (EEUU).

(35) También le gustó. Mick Jagger. ¿Le gustaban a ella los *Rolling Stones* o no? (España).

1.1.2.4. Según el referente

Por último, dependiendo de si se comparte el mismo referente o no, puede distinguirse entre **anáfora con el mismo sentido** o **anáfora con el mismo referente**. En la primera el elemento anafórico tiene un referente similar al de su antecedente, mientras que en la segunda ambos comparten el mismo referente:

(36) And *the president* has to get his hair cut. *Everybody* has to get their hair cut. (EEUU).

(37) [Contexto: diálogo entre Kauffman y Assuras sobre las actividades que hace Kauffman siempre que está en Nueva York]

- KAUFFMAN: Right. Get my hair cut.
- ASSURAS: Get your hair cut here. (EEUU).

(38) [...] ellos me dicen: “¿pero usted no se aburre de estar todo el día que no le duele *la cabeza*”, ¿no?, a mí nunca me duele la cabeza. (Venezuela).

En el primer ejemplo observamos que se refiere al cabello de diferentes personas: el del presidente por un lado y el de todo el mundo por otro. En el diálogo que aparece en segundo lugar, ambos están hablando del cabello de la misma persona: el del hablante llamado Kauffman. Lo mismo ocurre con el ejemplo en español: en ambos casos se refieren a la cabeza de la interlocutora. El caso contrario lo encontramos en el ejemplo 20, donde un hablante dice que le duele la cabeza y otro responde que a él también; ya que a un ser humano solamente le puede producir dolor su propia cabeza y no la de otra persona.

1.1.3. Resolución automática de anáforas

A lo largo de este trabajo hemos visto ejemplos de anáforas y correferencias, pudiendo apreciar algunas **diferencias en inglés y español**. Una de ellas es la distinción de género en español que no hay en inglés, hecho que tiene consecuencias en la resolución automática de anáforas; pues se tiene más éxito con

lenguas que sí diferencian género, como el español o el búlgaro, que con lenguas que no lo hacen, como el inglés (Mitkov, 2002). Este hecho ya lo observaron Peral, Palomar y Ferrández (1999), señalando que, gracias a la concordancia morfológica, en español se consigue reducir la lista de candidatos a antecedente mucho más que en inglés. Estos autores advirtieron que para los pronombres en español existía una cantidad mayor de posibles antecedentes.

En último lugar, cabe mencionar otros aspectos diferentes en español y en inglés que deben tenerse en cuenta a la hora de resolver anáforas y correferencias:

There is a need to identify elliptical pronouns in Spanish because, unlike overt pronouns, they get their number from the verb, have no gender, and always appear in subject position. = [...] Since Spanish has grammatical gender, two non-pronominal nouns with different gender might still corefer, e.g. *el incremento* 'the increase' (masc.) and *la subida* 'the rise' (fem.). Gender agreement is an appropriate constraint only for pronouns. = [...] = Attributive NPs are non-referential, hence non-markables.

(Recasens y Hovy, 2009, pp. 34-35)

Dentro del **proceso de resolución automática de anáforas**, en primer lugar deben identificarse los elementos anafóricos. Después hay que localizar los candidatos a antecedentes y por último seleccionar el antecedente adecuado (estos pasos son tenidos en cuenta en el algoritmo que proponemos en el apartado 4.2.4). Mitkov (2002) menciona seis niveles de conocimiento necesario para este proceso automático de resolución de anáforas dentro del Procesamiento del Lenguaje Natural (PLN), ejemplificados en las siguientes muestras orales:

- (39) Sí, compré unas *figuritas* de papel así con huequitos. Bueno, ahora te las muestro. (Venezuela).
- (40) Ahora, nuestro suministrador fue *el ejército de Batista*, organizado, equipado, entrenado y, además, asesorado durante todo ese tiempo por oficiales norteamericanos. No era un ejército menospreciable, ni mucho menos, y se creían los dueños del mundo. (Cuba).
- (41) Mexican people, I think, are the ones that like *American products* over any other country in the world. If it's American, it's going to be good, and they prefer it over anything else. (EEUU).
- (42) [...] era *el coche de Óscar Gálvez*, yo lo compré en el cincuenta y uno. (España).
- (43) I like *fish* but I wouldn't, I wouldn't choose it. (Gran Bretaña).
- (44) Have to get Uncle Bert round to have a look. Just get a little present for Kieran and we'll pop it off to Larry's. We've gone past it now. Present for *Kieran*, why? It was his birthday last week and I bought him got him those slippers and they were too big so I'll get something else instead. (Gran Bretaña).

En el primer ejemplo vemos que gracias al conocimiento **morfológico** y **léxico** podemos descartar *papel y huequitos* como candidatos a antecedente del elemento anafórico *las* debido a que no concuerdan en género y número. Sin embargo, en el segundo ejemplo advertimos que la elipsis del pronombre *ellos* en la anáfora cero *se creían* indica tercera persona del plural, pero no concuerda con el único SN que sí muestra esas características: *oficiales norteamericanos*, sino que concuerda con un SN que aparece en tercera persona del singular *el ejército de Batista*. Esta relación entre una entidad anafórica plural y un antecedente singular se debe a que el elemento anafórico concuerda con la entidad del discurso asociada con su antecedente, en este caso *los miembros del ejército de Batista*.

En el tercer ejemplo observamos cómo gracias al conocimiento **sintáctico** ha sido posible distinguir los SSNN *Mexican people, the ones that [...], American products, other country, the world*. De esta manera se han descartado las otras palabras del texto como posibles candidatos a antecedente del pronombre *it*.

En el siguiente ejemplo vemos que los candidatos a antecedente del pronombre *lo* podrían ser no solo *el coche de Óscar Gálvez* sino también *Óscar Gálvez*. No obstante, gracias al conocimiento **semántico** descartamos la opción *Óscar Gálvez*, ya que la información semántica del verbo *comprar* no está relacionada con una persona como Padecedor (*undergoer*); al contrario que ocurre con un objeto como *el coche*.

En el penúltimo ejemplo, parece evidente que el antecedente del pronombre *it* es el único SN que aparece: *fish*. Sin embargo, debemos tener siempre presente que las frases no suelen aparecer aisladas, sino dentro de un texto en el cual hay un tema central sobre el que se está hablando:

(45) Mm. anyway, and I sort of said to the waitress, why what was it? She said *octopus*. Urgh! You wouldn't have had it if you'd have known? No. I don't think I would. No. It's not something No[sic]. I would have picked. No. I wouldn't. I certainly Well wouldn't. it [sic] was it were I mean I like fish but I wouldn't, I wouldn't choose it. (Gran Bretaña).

Como puede observarse, basándonos en el conocimiento **discursivo** podemos detectar el antecedente adecuado en torno al cual gira el diálogo, que en este caso es *octopus*.

En el último ejemplo, el hecho de contar con el conocimiento **del mundo** sobre el comportamiento esperado de un ser humano cuando es el cumpleaños de otra persona, contribuye a seleccionar el antecedente correcto: no *Uncle Bert* ni

Larry's sino Kieran, ya que anteriormente se menciona que es su cumpleaños e incluso que se planea comprarle un regalo.

Respecto a los diferentes niveles tenidos en cuenta, Mitkov (2002) reconoce que se obtendrán resultados más precisos si el algoritmo de resolución de anáforas se activa, no después del análisis sintáctico, sino tras haber llevado a cabo un análisis semántico; manifestando incluso que “if discourse, pragmatic and real-world analysis were available [...] the substitution would be done after the last stage of analysis” (Mitkov, 2002, p. 31). También señala que la mayoría de sistemas de resolución de anáforas no incorporan conocimiento extralingüístico debido a la cantidad de tiempo y trabajo que ello requeriría. Sin embargo, él mismo reconoce que, aunque los conocimientos morfológico, léxico, sintáctico y semántico son eficaces para localizar los posibles candidatos a antecedente, ya que pueden rechazar aquellos que sean inapropiados, realmente los conocimientos discursivo y del mundo llegan a ser decisivos a la hora de seleccionar el antecedente adecuado.

Por tanto vemos que, a pesar de ser necesarios, algunos conocimientos no son aprovechados en los sistemas de resolución automática de anáfora:

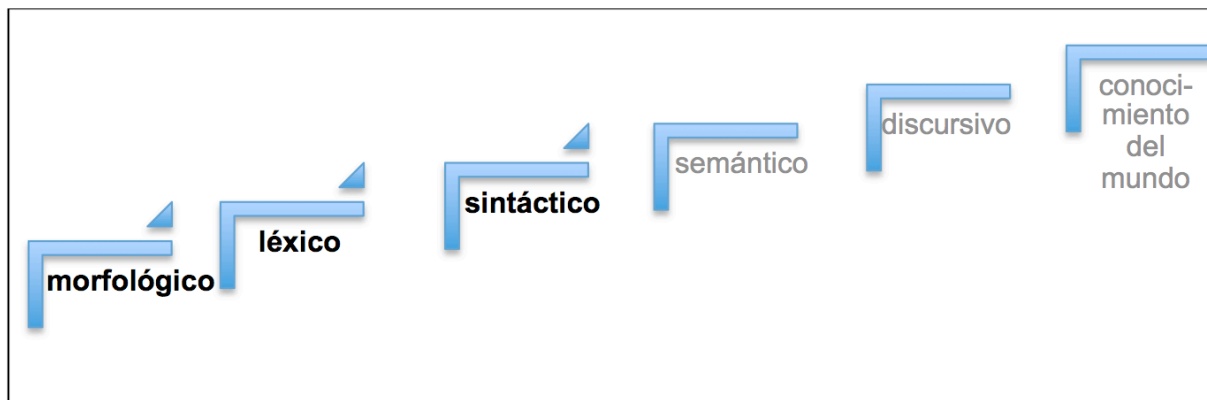


Figura 3. Conocimiento necesario para la resolución de anáforas

En el siguiente subcapítulo 1.2 veremos cómo se han aplicado estos conocimientos en la resolución de la anáfora; pues dependiendo de qué niveles sean tenidos en cuenta podemos distinguir, como hace Saiz Noeda (2002), entre *métodos de conocimiento limitado* que no integran la semántica, *métodos de resolución enriquecidos* que sí la introducen y *métodos alternativos* que tienen en cuenta otra información como la co-ocurrencia estadística.

1.2. Estado de la cuestión en PLN

Dentro de la Lingüística Computacional, la anáfora nominal es sobre la que más investigaciones y proyectos se han llevado a cabo. La mayoría se centra en la resolución de pronombres, SSNN definidos (también la anáfora indirecta) o la elisión de pronombres. Es más, la resolución de la anáfora no-nominal no ha sido prácticamente tratada.

Mitkov (2002) ofrece un detallado informe sobre las diferentes propuestas hechas en el S. XX para la resolución de la anáfora y la correferencia del cual en los apartados 1.2.1 - 1.2.2 se expondrán los datos principales¹¹. En el apartado 1.2.3 se mencionarán las aportaciones realizadas hasta la actualidad¹².

1.2.1. Sistemas de conocimiento limitado y enriquecido

Desde los **años 60** del S. XX se han desarrollado programas para resolver anáforas. Lo que caracteriza a las propuestas de esa década y la posterior es que no se basaban en variados niveles de conocimiento lingüístico, sino que se limitaban a incluir información léxica y sintáctica (cfr. Bobrow, 1964 para localización de antecedentes; para resolver pronombres Winograd, 1972 y Hobbs, 1976) e incluso semántica (véase Woods, Kaplan y Nash-Webber, 1972 para resolver anáforas completas y parciales; Hobbs, 1978 para resolución de pronombres); requiriendo por tanto mucho *input* humano.

A pesar de que a comienzos de **los 80** disminuyó el interés de los investigadores, seguían haciéndose propuestas que incluían el conocimiento sintáctico y semántico (*vid.* Rich y LuperFoy, 1988 para la resolución de pronombres). Por otro lado, se comenzó a incluir el conocimiento discursivo a nivel pragmático: para resolver pronombres Günther y Lehmann (1983) tuvieron en cuenta la posición cercana o alejada del elemento anafórico, así como si el candidato a antecedente tiene macropapel de Actor o no. Brennan, Friedman y Pollard (1987) en su algoritmo prestaron atención al tópico (*topic*) y al foco (*focus*). Además, se

¹¹ En esta obra, el autor explica que él se centra más en la investigación de la última década del S. XX porque las aportaciones anteriores son tratadas en profundidad en Hirst (1981) y Carter (1987).

¹² A nivel internacional destaca el congreso bianual *Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)* que reúne a grandes especialistas en este ámbito.

empezó a añadir el conocimiento del mundo: Carter (1987) en el programa *Shallow Processing Anaphor Resolver* lo aplicó para la resolución de anáforas nominales no resueltas. Carbonell y Brown (1988) también lo usaron para resolver la anáfora interoracional.

En cambio, en **los 90** se recupera el interés en la comunidad investigadora. Se puede observar una fascinación por cuestiones discursivas y por el uso de corpus, aunque lamentablemente los corpus anotados correferencial y anafóricamente no eran muy accesibles ni extensos¹³: para resolver pronombres de tercera persona varios investigadores se basaron en colocados (Dagan e Itai, 1991; Lappin y Leass, 1995). Vemos por tanto que la resolución de la anáfora se centró en los pronombres, para cuya resolución Baldwin (1997) aportó el programa CogNIAC. Más adelante se amplía el horizonte de investigación a la interpretación de la anáfora indirecta (Gelbukh y Sidorov, 1999; Murata y Nagao, 2000) y de SSNN definidos con función anafórica (Vieira y Poesio, 2000).

En esta última década del S. XX aumenta el deseo de basarse en sistemas de conocimiento pobre (información limitada a pocos niveles de conocimiento lingüístico), de manera que se pudieran satisfacer las demandas de soluciones efectivas y no caras en diversos ámbitos de PLN. Por otro lado, esto fue favorecido por el acceso a herramientas de pre-procesamiento de texto en PLN como etiquetadores gramaticales (*part-of-speech taggers*) que asignan la categoría gramatical correcta; analizadores sintácticos ligeros (*shallow parsers*) que encuentran nombres; corpus anotados o no y recursos léxicos como la base de datos léxica *WordNet* que ofrece *synsets* (grupos de palabras con relaciones de sinonimia: hiperonimia, hiponimia, meronimia, holonimia). A pesar de ello, no se logró que los sistemas de resolución de anáfora funcionaran de forma totalmente automática. Por otro lado, hubo proyectos para la resolución de anáforas basados en otras lenguas diferentes al inglés como p. ej. el alemán, japonés, portugués o turco.

También se mostró interés por parte de los investigadores en la resolución plurilingüe de la anáfora y la correferencia, como ocurrió con el algoritmo de conocimiento pobre para la resolución de anáforas de Mitkov (1996) que luego fue

¹³ Algunos de los corpus son *Lancaster Anaphoric Treebank*, *MUC coreference task* (MUC-6 y MUC-7) o el corpus anotado de la University of Wolverhampton (Mitkov et ál., 2000). Más información sobre los corpus en el S. XX puede encontrarse en Mitkov (2002). Dentro de los corpus actuales para el español destaca AnCora-ES (<<http://clic.ub.edu/corpus/es/ancora>>) con 500.000 palabras de textos periodísticos que incluye un nivel de anotación para las relaciones de correferencia.

adaptado para otras lenguas como el polaco, el árabe y el francés. Este algoritmo estaba basado en una lista de indicadores de antecedente que señalaban las preferencias para seleccionar el antecedente dentro de los SSNN de las dos frases anteriores, comprobando entre otros la compatibilidad de género y número; de hecho siguió desarrollándose dando lugar a *Mitkov's Anaphora Resolution System* (Mitkov et ál., 2002) en el que a su vez se basó un sistema de resolución automática de pronombres para el búlgaro.

1.2.2. Sistemas de conocimiento alternativo

Respecto a **finales del S. XX**, cabe mencionar que se hicieron propuestas de aprendizaje mecánico basadas en corpus anotados y no anotados capaces de aprender de ejemplos, p. ej. con la modalidad árboles de decisión. Así, Aone y Benett (1996) aportan un sistema de resolución de anáfora para el japonés basándose en aspectos léxicos, sintácticos, semánticos y posicionales. También hubo propuestas de sistemas para clasificar SSNN (McCarthy y Lehnert, 1995; Soon, Ng y Lim, 1999).

En estos años aparecieron igualmente propuestas probabilísticas enmarcadas en el ámbito estadístico: Ge, Hale y Charniak (1998) resaltan que, cuanto más lejos estén el elemento anafórico y el antecedente, menos opciones hay de que sea el antecedente correspondiente. Basándose en colocados, también estudian las probabilidades que hay de que un candidato ocurra en la misma función sintáctica que el elemento anafórico. Por último, reconocen que aquellos SSNN más frecuentes que aparecen en el texto, tienen más posibilidades de ser el antecedente en las anáforas. Respecto a la resolución de la correferencia, Cardie y Wagstaff (1999) llegan también a la conclusión de que cuanto más cerca estén antecedente y elemento anafórico más probabilidades hay de que sean correferenciales.

1.2.3. Aportaciones recientes y retos

Dentro de la resolución de la correferencia, en **el S. XXI** resalta el hecho de basarse en diferentes tipos de conocimiento: morfológico (género y número), léxico (nombres propios, alias), sintáctico (distancia) y semántico (información en Wikipedia, WordNet o roles semánticos, distancia en WordNet o concordancia de clases para SN en WordNet, como p. ej. *persona, organización, objeto, lugar* o

fecha). No obstante, las investigaciones al respecto han estado muy centradas en el idioma inglés, con excepciones para el español (Recasens y Hovy, 2009), el holandés (Hendrickx, De Clercq y Hoste, 2011) o el checo (Novák y Žabokrtský, 2011).

La anáfora indirecta ha sido muy estudiada respecto al inglés (véase Poesio, Vieira y Teufel, 1997; Poesio y Vieira, 1998; Poesio, Uryupina, Vieira, Alexandrov-Kabadjov y Goulart, 2004); sin embargo no ha habido tantas investigaciones con otras lenguas: apenas en alemán (Versley, 2007; Klenner, Tuggener, Fahrni y Sennrich, 2010; Kimm y Horn, 2011), en francés (Lassalle y Denis, 2011) o en japonés (Sanano, Kawahara y Kurohashi, 2007). En esta última publicación, sus autores propusieron usar la resolución de anáfora indirecta (junto al conocimiento de sinónimos basado en corpus y diccionarios) para mejorar la resolución de la correferencia, señalando la importancia de hipónimos e hiperónimos a la hora de resolver relaciones de correferencia.

En torno a la resolución de la anáfora, han surgido agrupaciones de investigadores como el *Grupo de Procesamiento del Lenguaje y Sistemas de Información* de la Universidad de Alicante, del cual destacan las tesis doctorales y publicaciones de autores respecto a la resolución de la anáfora en diferentes ámbitos de PLN: Martínez Barco en diálogos, Muñoz Guillena en sistemas de extracción de información, Peral Cortés en traducción automática y Ferrández Rodríguez y Vicedo González en sistemas de búsquedas de respuestas. Por otro lado, el ya mencionado sistema de conocimiento limitado *Mitkov's Anaphora Resolution System* (2002) volvió a ser mejorado a comienzos de este siglo, dando como resultado una versión más reciente llamada MARS06. Mitkov se ha convertido en uno de los referentes dentro de la resolución de la anáfora, publicando varios estudios sobre ello (2002, 2003, 2010).

En relación a los retos para el futuro, se ha hecho hincapié en la necesidad de la información semántica para la correcta resolución de la anáfora (Palomar et ál., 2001), aunque también se ha reconocido que una base de datos sobre conocimiento del mundo sería un esfuerzo caro y complicado (Caselli, 2009). Por otro lado, el etiquetado de anotación correferencial no está muy desarrollado (Bisbal Asensi et

ál., 2003; Recasens, Martí y Taulé, 2007)¹⁴. Asimismo se ha reconocido la importancia de poder también identificar las anáforas y antecedentes improbables, ya que un texto alberga bastantes elementos que son no anafóricos y/o no candidatos a antecedente (Uryupina, 2009). Por último, se ha señalado la falta de una semántica profunda para la correcta identificación del antecedente en la resolución de anáforas y se ha propuesto el empleo del reconocimiento de nombres de entidades junto a una ontología que permita detectar elementos con rasgos de *animacy* (Devi, 2008). Además de recordar la carencia de una herramienta enciclopédica que proporcione información del mundo (Martínez Barco, 1999), en los últimos años diversos autores han ratificado esta necesidad de incluir conocimiento ontológico para la resolución de (cadenas de) anáforas y correferencias (cfr. Lech y Smedt, 2007; Jegatha-Deborah, Baskaran y Kannan, 2011)¹⁵.

En el siguiente subcapítulo 1.3 se presentan los modelos teóricos en los que se asienta esta investigación y se comenta la relevancia de varios aspectos tratados en la Gramática del Papel y la Referencia, el Modelo Léxico Construccional y FunGramKB respecto a la anáfora.

1.3. Modelos teóricos

1.3.1. La Gramática del Papel y la Referencia

La **Gramática del Papel y la Referencia** (Van Valin y LaPolla, 1997; Van Valin, 2005) o *Role and Reference Grammar* (RRG) es una teoría funcionalista que integra la semántica, la sintaxis y la pragmática en el proceso productivo y receptivo del lenguaje humano a través de un algoritmo de enlace¹⁶:

¹⁴ Sin embargo, recientemente se ha desarrollado en el ámbito de Búsqueda de Respuestas el modelo de anotación anafórico multilingüe AQA (véase Boldrini, Puchol-Blasco, Navarro, Martínez-Barco y Vargas-Sierra, 2009).

¹⁵ No faltan autores que, a la inversa, señalen la importancia de la resolución de la anáfora en la creación de base de datos de conocimiento u ontologías para la Web Semántica (cfr. Delmonte, 2003).

¹⁶ En <<http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html>> puede encontrarse información actualizada sobre la RRG. Para una perspectiva histórica de otras teorías influyentes en la RRG véase Van Valin (2001, pp. 218-221). Respecto al nombre de la RRG, Van Valin dice: “‘Role’ refers to the semantic role properties of arguments, and ‘Reference’ refers to the discourse-pragmatic functions” (2009, p. 66).

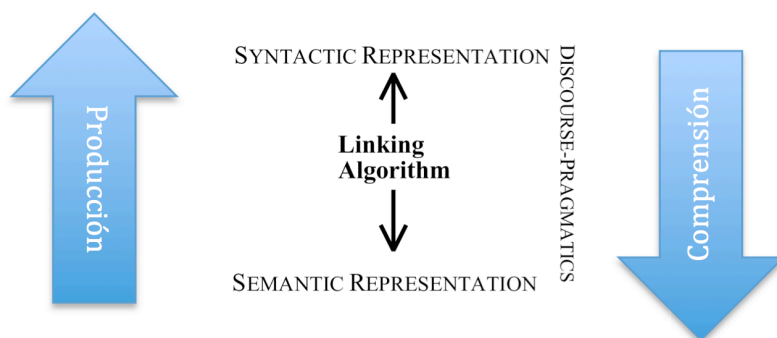


Figura 4. Organización de la RRG
Tomado y adaptado de Van Valin (2005, p. 2)

Según esta estructura, el proceso de comprensión lingüística desciende de la sintaxis acompañado de la pragmática hasta llegar a la semántica; mientras que el proceso de producción lingüística asciende en sentido inverso. La RRG especifica los pasos a dar en ambos casos con el algoritmo de enlace de la semántica a la sintaxis (cfr. Van Valin, 2005, p. 279 para el inglés y Belloro, 2007, pp. 187-188 para el español) y con el algoritmo de enlace de la sintaxis a la semántica (cfr. Van Valin, 2005, pp. 280-281 para el inglés y Belloro, 2007, pp. 188-189 para el español). En la figura 4, *discurso-pragmática* aparece en vertical por su omnipresencia, pudiendo influir en aspectos semánticos, sintácticos y pragmáticos (cfr. Van Valin, 2005, p. 182).

En relación con la anáfora y la correferencia, lo más estudiado en la RRG ha sido la anáfora cero, los reflexivos y los pronombres tanto intra- como inter-oralmente (Van Valin, 1990; Van Valin y LaPolla, 1997; Van Valin, 2005). Respecto al desarrollo de la RRG para el idioma español, son fundamentales los estudios sobre los pronombres de González Vergara (2006b, 2009) y de Belloro (2007, 2009). En los subapartados 1.3.1.1 – 1.3.1.3 se comentarán los datos de la RRG relevantes para la anáfora y la correferencia.

1.3.1.1. Representación semántica

En primer lugar, la **representación semántica** de las oraciones se basa en la representación semántica de los verbos¹⁷, los cuales tienen una representación léxica diferente dependiendo de su tipo de *Aktionsart*. Cada uno de los diferentes tipos es representado como una **estructura lógica** o *Logical Structure* (LS) compuesta por las variables para los argumentos especificadas entre paréntesis tras

¹⁷ Como señalan Mairal Usón y Cortés Rodríguez (2006), en esta gramática los **predicados** no son solo verbos, sino también adjetivos o preposiciones.

los operadores semánticos y los predicados. Dependiendo del lugar que ocupen los argumentos, tendrán diferentes relaciones semánticas con su predicado. Respecto a la anáfora, destaca la especificación de los argumentos en la LS y la coindización de los pronombres correferentes:

(46) *Molly saw herself.* **see'** (Molly_i, herself_i)

(47) *Molly_i's brother saw her_i.* **see'** ([**have.as.kin'** (Molly_i, brother)], her_i)¹⁸

Hasta donde sabemos, aún queda por resolver en la RRG cómo especificar en la LS el elemento correferente dentro de una anáfora indirecta, p. ej.:

(48) *Fuentes murió en el año 2012. Esta voz fue clave en la literatura.*

Ambos elementos de la anáfora estarían coindizados con (Fuentes_i), pero no está claro si para el argumento subrayado debería aparecer el elemento anafórico DEI(voice_i) o el antecedente aludido DEI(producción literaria de Fuentes_i)¹⁹.

Segundo, la RRG contempla una relación semántica que agrupa los diferentes papeles temáticos formando un *continuum* (véase Van Valin y LaPolla, 1997, p. 127) en cuyos extremos se distinguen dos **macropapeles semánticos** fundamentales: *Actor (actor)* y *Padecedor (undergoer)* (véase Van Valin y LaPolla, 1997, p. 146). Según el número de macropapeles de un predicado, se distingue entre verbos (a) *atransitivos* (sin macropapeles), (b) *intransitivos* (con un macropapel) y (c) *transitivos* (con dos macropapeles). Esto es relevante para nuestro tema de investigación, pues la información sobre la atrantividad de ciertos verbos puede resultar de gran utilidad a la hora de reconocer los usos pleonásticos, p. ej. en los verbos meteorológicos, de manera que se evite la búsqueda de referentes.

1.3.1.2. Representación sintáctica

Primeramente se tratará la representación de la cláusula y después del SN. Para mostrar la organización jerárquica de las oraciones, cláusulas y sintagmas, la RRG se basa en una **estructura estratificada de la cláusula** o *Layered Clause Structure* (LCS) con unos elementos universales determinados semánticamente: el **núcleo** (predicado), el **centro** (argumentos y núcleo) y la **periferia** (no-argumentos del predicado que son opcionales). Además, algunas lenguas presentan un **espacio pre-central** o *PreCore Slot* (PrCS) y otras un **espacio post-central** o *PostCore Slot*

¹⁸ Ejemplos extraídos de Van Valin y LaPolla (1997, p. 393).

¹⁹ Se incluye la representación del operador de deixis en la RRG DEI porque en el SN del ejemplo 48 se usa el demostrativo *esta*.

(PoCS). Por último, fuera ya de la cláusula pueden encontrarse tanto una **posición dislocada a la izquierda** o *Left Detached Position* (LDP), como una **posición dislocada a la derecha** o *Right Detached Position* (RDP):

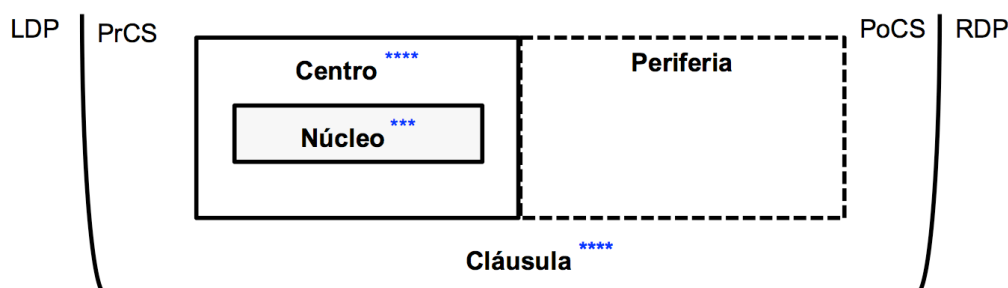


Figura 5. Constituyentes de la LCS
Adaptado de Van Valin y LaPolla (1997, p. 38)²⁰

Esta estructura ayuda a explicar la anáfora y catáfora de reduplicación correferencial con pronombres en inglés y español:

(49) I know them, THOSE BOYS.²¹

(50) LA LECHE yo la tomo fría.

Van Valin aclara que si el LDP o RDP es un argumento semántico del verbo, en el centro reaparecerá en forma de pronombre (2005, p. 6)²² porque si los argumentos son lexicalizados en la oración fuera de [la cláusula], en el centro faltaría un argumento obligatorio: *LA LECHE* [yo __ tomo fría], por ello aparece *la*, correferencial con el LDP *la leche*. Así, no solo pueden explicarse reduplicaciones como *A PEDRO* [lo vi ayer], sino también *A PEDRO* [le dan miedo las polillas]²³.

Por último, del minucioso análisis de las oraciones compuestas realizado en Van Valin y LaPolla (1997), resalta que en las anáforas cero se añade un elemento *pro* en la proyección de constituyentes y también en la LS, señalando la correferencia:

(51) Robin_i drove out of Phoenix this morning and pro_i will arrive in Atlanta tomorrow.
[this morning´ [do´ (Robin_i, [drive´ (Robin_i, Ø))] & BECOME NOT be-

²⁰ Los asteriscos de la figura 5 representan operadores en varios niveles, unidos a los constituyentes por el núcleo (véase Van Valin, 2005, pp. 9, 12).

²¹ Ejemplo de Van Valin (2005, p. 6).

²² Para los elementos LDP que funcionan como adjuntos (no argumentos) no hay un elemento reanudado en la cláusula: *AT THE PARTY*, [who did John see]? (cfr. Van Valin, 1990, p. 196). Cuando el elemento aparece en el PrCS, en la cláusula hay un hueco: *Sam I´ve known ___ for years* (cfr. Van Valin y LaPolla, 1997, p. 228).

²³ Para el análisis de reduplicación de objetos, véase Belloro (2007, pp. 192, 199).

in´ (Phoenix, Robin_i)] and´ [tomorrow´ (BECOME be-in´ (Atlanta, pro_i))]²⁴

Estas oposiciones universales de la LCS de toda lengua pueden transportarse al SN²⁵, que tanta importancia tiene para la anáfora y la correferencia, dando lugar a la **estructura estratificada del SN** o *Layered Structure of the Noun Phrase* (LSNP): formada universalmente por un **núcleo_N**, un **centro_N** y una **periferia_N**; además en algunas lenguas también hay una **posición de comienzo del SN** o *NP-initial position* (NPIP) y/o una **posición de final del SN** o *NP-final position* (NPFP):

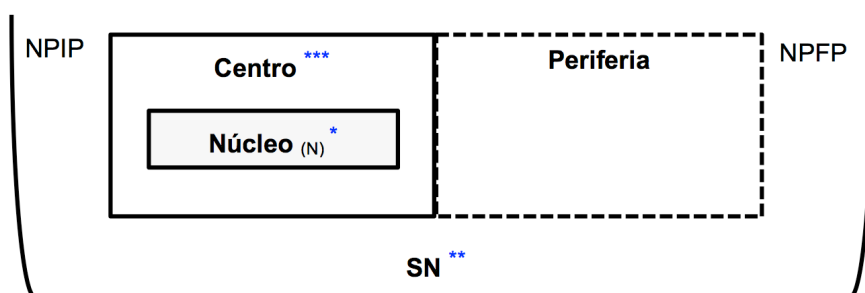


Figura 6. Constituyentes de la LNPS
Adaptado de Van Valin (2005, p. 25)

En esta figura los asteriscos simbolizan los operadores que modifican los constituyentes del SN a diferentes niveles: *operadores nucleares, centrales y del SN* (véase Van Valin, 2005, p. 24). En este trabajo resultan fundamentales los que modifican a todo el SN: *deixis* (DEI) y *definitud*²⁶.

La RRG propone de forma preliminar una plantilla léxica para el SN en inglés, con operadores unidos por el núcleo_N²⁷:

²⁴ Ejemplo de Van Valin y LaPolla (1997, p. 520).

²⁵ Van Valin usa el término NP (*Noun Phrase*), pero recomienda hablar de RP (*Referential Phrase*) porque en algunas lenguas el núcleo de una expresión referencial pertenece a diferentes categorías gramaticales (2005, p. 28).

²⁶ *Definitud* se entiende como una categoría gramatical, lexicalizada en algunas lenguas a través de artículos, para marcar la existencia y generalmente también lo identificable del SN (cfr. Pavey, 2006).

²⁷ Van Valin (2005, pp. 25, 27) ilustra el SN con adjetivo, genitivo o demostrativo. Para SSNN complejos véase Van Valin y LaPolla (1997, pp. 492-505).

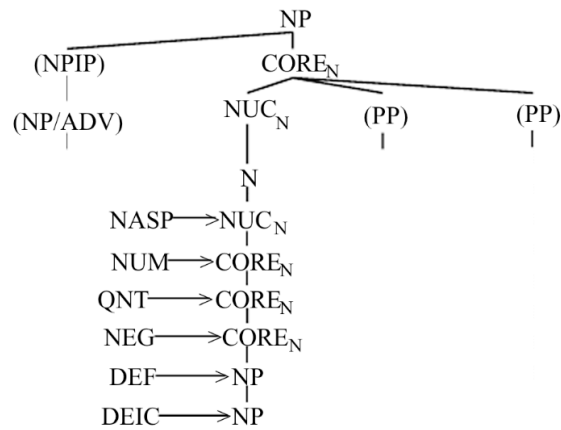


Figura 7. LNPS (preliminar)
Tomado de Van Valin (2005, p. 25)

Teniendo en cuenta que la propuesta para la LSNP de Van Valin (2005) es preliminar y que él mismo advierte sobre la necesidad de desarrollar en el futuro la estructura estratificada para el Sintagma Referencial (cfr. Van Valin, 2005, p. 28), en este trabajo se propone una LNPS en el idioma español adaptada de Van Valin de tal manera que pueda explicar el fenómeno de la anáfora en los términos de la RRG.

En primer lugar, se propone un **índice de concordancia de género** (AGX_G) para morfemas de género y determinantes (artículos, demostrativos). Este AGX_G está inspirado en el **índice de concordancia** o *agreement index* (AGX) **de persona** en el núcleo verbal (denominado en este trabajo AGX_P) propuesto por Belloro (2004) para los morfemas personales verbales y los pronombres átonos en español²⁸:

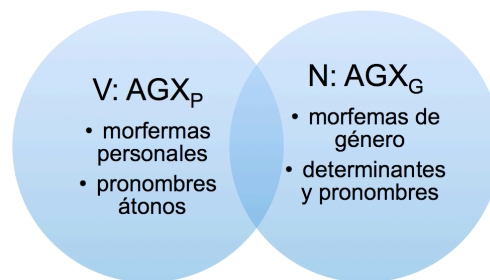


Figura 8. Índice de concordancia verbal e índice de concordancia nominal

En segundo lugar, en el anexo 1 se proponen de forma preliminar **plantillas del SN con AGX_G** para el español, en las cuales se incluye la representación de la anáfora cero²⁹:

²⁸ En base a este AGX_P , González Vergara (2006a, pp. 123-125) propone plantillas sintácticas para la oración simple en español.

²⁹ Hasta donde sabemos, no existen plantillas basadas en la RRG para el SN en español.

- (a) Núcleo_N formado por un nombre común, propio o pronombre y su AGX_G.
- (b) Centro_N cuyo núcleo_N es un nombre común, propio o pronombre y su AGX_G.
- (c) SN con un centro_N cuyo núcleo_N es un nombre común, propio o pronombre y su AGX_G.
- (d) SN con un centro_N cuyo núcleo_N es un nombre común y su AGX_G expandido al artículo definido/indefinido o al determinante demostrativo.
- (e) SN de anáfora cero con un centro_N cuyo núcleo_N es un pronombre y su AGX_G (proveniente del nombre elidido).
- (f) SN con un centro_N cuyo núcleo_N es un nombre común y su AGX_G expandido al artículo determinado y también al demostrativo pospuesto en la periferia.
- (g) SN con un centro_N cuyo núcleo_N es un nombre común y su AGX_G expandido al artículo definido/indefinido o al determinante y al adjetivo en la periferia.
- (h) SN de anáfora cero con un centro_N cuyo núcleo_N es un AGX_G (proveniente del nombre elidido) expandido al artículo definido y al adjetivo en la periferia.
- (i) SN de anáfora cero con un centro_N cuyo núcleo_N es un pronombre y su AGX_G (proveniente del nombre elidido) expandido al adjetivo en la periferia.
- (j) SN con un centro_N cuyo núcleo_N es un nombre común y su AGX_G expandido al artículo definido/indefinido o al determinante, junto a un Sintagma Preposicional (SP) en la periferia³⁰.
- (k) SN de anáfora cero con un centro_N cuyo núcleo_N es un AGX_G (proveniente del nombre elidido) expandido al artículo definido, junto a un SP en la periferia.
- (l) SN de anáfora cero con un centro_N cuyo núcleo_N es un pronombre y su AGX_G (proveniente del nombre elidido), junto a un SP en la periferia.

³⁰ La proyección de constituyentes en el SP propuesta para el español se ha basado en el análisis del SP predicativo en la RRG (cfr. Van Valin, 2005, p. 23). En el anexo 1 se incluye la estructura para el SP predicativo y no predicativo en inglés y en español.

1.3.1.3. Representación informativa

En la RRG la distribución de la información de una oración se refleja en la **estructura focal**, en la cual se muestra el contraste entre el **tópico** y el **foco**; pues se separan los ámbitos del **dominio potencial de foco** o *potential focus domain* (PFD) y del **dominio real de foco** o *actual focus domain* (AFD)³³. La RRG integra en la proyección de la estructura de constituyentes y de operadores también la **proyección de la estructura focal**:

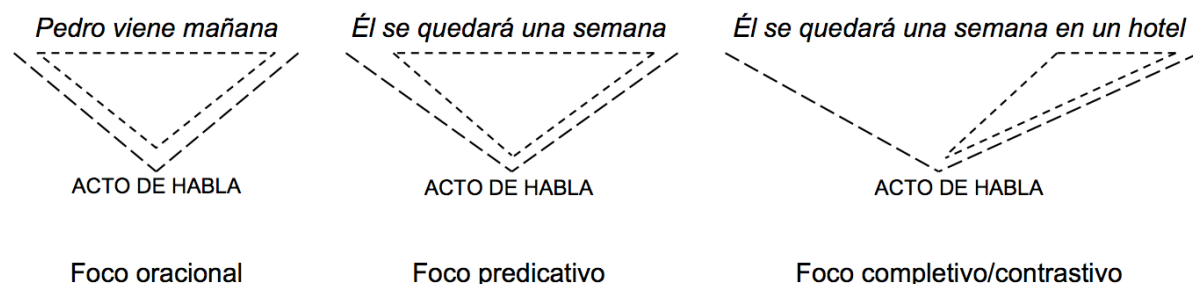


Figura 9. Representación del PFD y del AFD en las estructuras focales
Adaptado de González Vergara (2006a, p. 129)

A menudo los elementos anafóricos aparecen fuera del AFD. Van Valin y LaPolla (1997, p. 428) afirman que “the primary function of focus structure in the linking from syntax to semantics is anaphora resolution, that is, in helping to determine possible coreference relations with respect to pronouns”; pues, para que el pronombre sea coreferente con su antecedente, este debe estar fuera del AFD (véase Van Valin y LaPolla, 1997, p. 227). Además, la RRG incluye una proyección de la estructura del foco diferenciando la información nueva de la ya aportada, con los referentes de las variables entre paréntesis:

Foco oracional	Foco predicativo		Foco estrecho completivo		Foco estrecho contrastivo		
- MARY ARRIVED	- Mary KISSED SAM		- Who kissed Sam? - Mary KISSED SAM		- Sally kissed Sam - No, MARY kissed Sam		
Assertion	Presupposition	Assertion	Presupposition	Assertion	Presupposition	Assertion	
v	v	x, y	x, y	x, y	x, y	y, z	y, z
Mary(v) v arrived	Mary(x) x P	Mary(x) Sam(y) x kiss y	X(x) Sam (y) x kiss y	Mary(x) Sam(y) x kiss y	Sally(x) Sam(y) x kiss y	X(z) Sam (y) z kiss y	-Sally(z) Mary(z) Sam (y) z kiss y

Figura 10. Proyección de la estructura de los diferentes tipos de focos
Adaptado de Van Valin (2005, p. 172)

³³ Además se clasifican como estructuras de foco estrecho completivas o contrastivas (un único constituyente), de foco ancho predicativas (el predicado) o de foco ancho oracionales (toda la oración) (cfr. Van Valin, 2005, p. 69-73).

De igual modo, se ofrece también una representación discursiva para varias oraciones, donde se aprecia cómo aumenta la información:

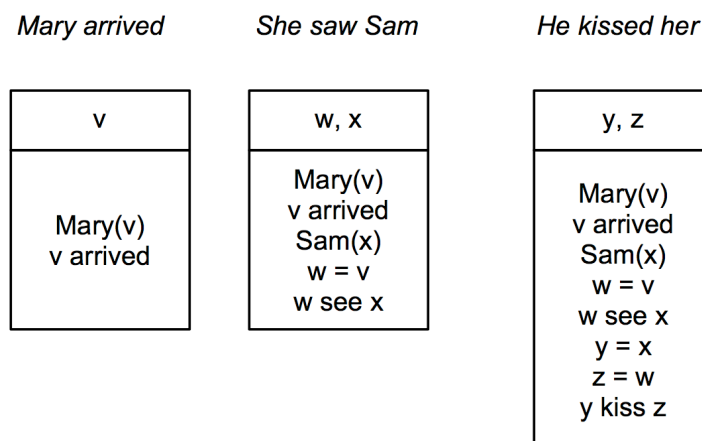


Figura 11. Estructura de representación discursiva
Tomado de Van Valin (2005, p. 172)

Por último, cabe mencionar el concepto de **relevancia** o *salience*. Consiste en que en el discurso se mantenga la alusión a un referente importante; siendo medible a través de la **frecuencia referencial** o *referential frequency* (RP-f) y la **referencia ininterrumpida** o *uninterrupted reference* (RP-d), pues cuanto mayores sean estas últimas, mayor será la relevancia (cfr. Shimojo, 2009).

Por otro lado, para nuestro tema tiene mucha relevancia la clasificación en la RRG de los diferentes niveles de activación del referente del SN según su capacidad de ser identificable:

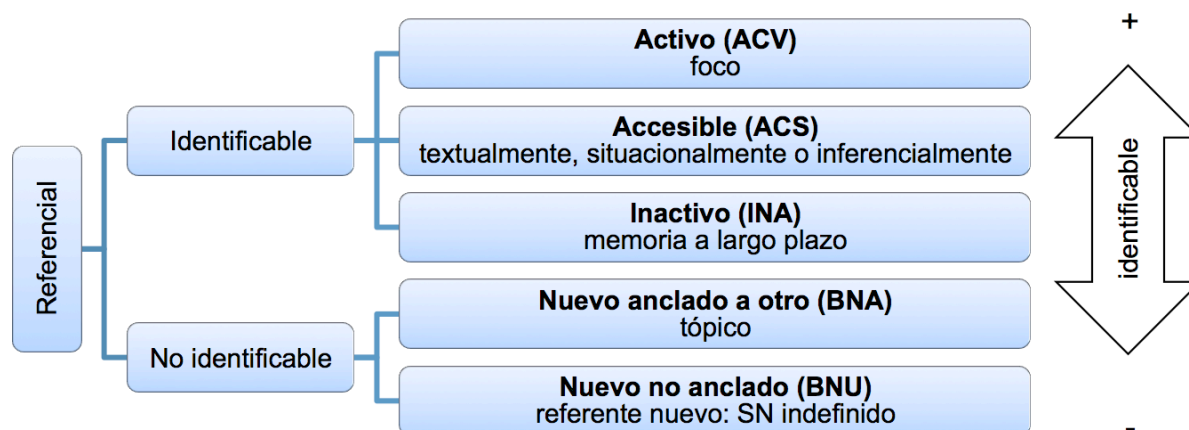


Figura 12. Niveles de activación del referente del SN
Basado en Van Valin y LaPolla (1997, pp. 200-201) y Van Valin (2005, p. 79)

La información del nivel de activación del referente del SN se integra en la LS:

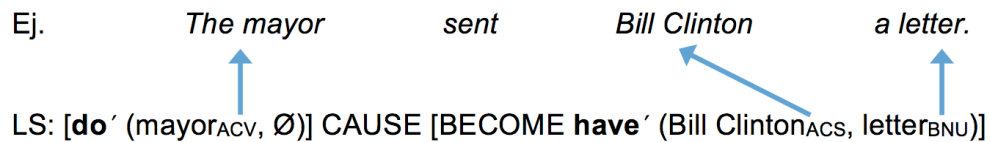


Figura 13. Ejemplo de la información sobre el nivel de activación integrada en la LS
Basado en Van Valin (2005, p. 80)

Respecto a la resolución inferencial automática del referente ACS, en el subapartado 1.3.3.3 veremos lo valioso que es el conocimiento ontológico, procedural y cultural de la base de conocimiento FunGramKB.

Por otro lado, la RRG propone una escala con la gradación de tópico a foco incluyendo elipsis, pronombres y determinantes, que debe tenerse en cuenta a la hora de resolver anáforas porque el elemento anafórico elidido y el empleo de pronombres implica que el antecedente es activo y accesible; además el uso del SN definido asegura que el referente es identificable, al contrario que el SN indefinido:

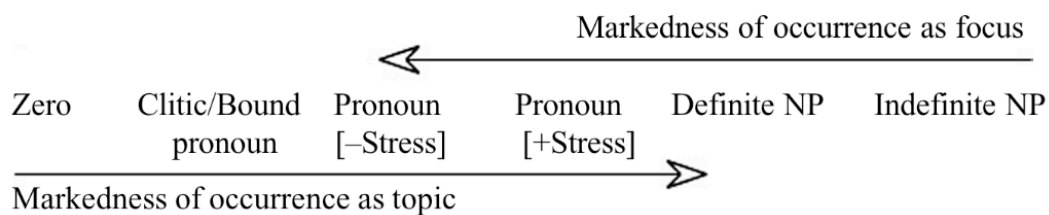


Figura 14. Escala de codificación de referentes en términos de sus posibles funciones
Tomado de Van Valin y LaPolla (1997, p. 205)

Por último, acerca del dativo en español, Belloro (2009, pp. 508-512) recuerda que el uso únicamente de pronombres clíticos indica el estado activo de su antecedente, el fenómeno de la reduplicación de pronombres clíticos marca su carácter accesible; mientras que emplear solo el SN implica su esencia inactiva. Esta misma autora (2007, p. 185) advierte de que los argumentos semánticos pueden corresponder en el plano sintáctico no solo a los nodos argumentales, sino también al nodo AGX_P, de modo que se complete la restricción de integridad o *Completeness Constraint* (véase Van Valin, 2005, pp. 129-130) en el español. Proponemos lo mismo sobre el AGX_G en el caso de la elisión en la anáfora cero.

1.3.2. El Modelo Léxico Construccional

El **Modelo Léxico Construccional** o *Lexical Constructional Model* (LCM) (Mairal Usón y Ruiz de Mendoza Ibáñez, 2007, 2008, 2009), influido por modelos

funcionalistas, cognitivistas y construccionistas³⁴, es una teoría sobre la construcción del significado (*i. e.* interpretación semántica) en base a la interacción de diferentes niveles lingüísticos, incluyendo implicaturas, fuerza ilocutiva y coherencia del discurso (Mairal Usón y Ruiz de Mendoza Ibáñez, 2009, p. 153). En el LCM, la **interpretación semántica** tiene lugar con el proceso de unión de la plantilla léxica y de la plantilla construccional en la que aparece; posibilitado por usar ambas el mismo metalenguaje y regulado en la gramática por restricciones internas y externas:

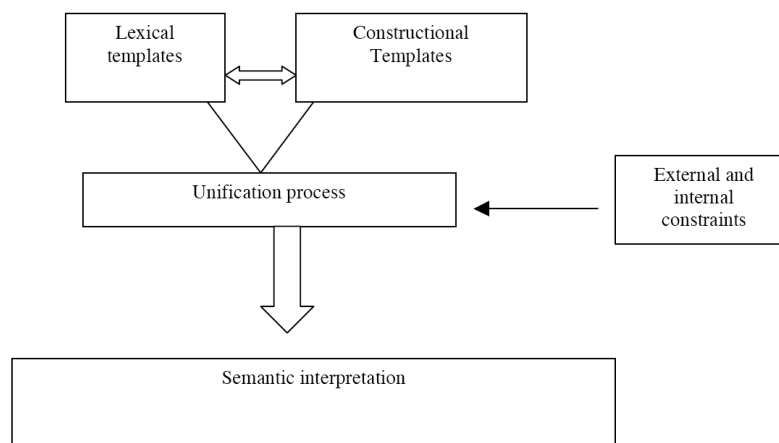


Figura 15. Estructura del LCM

Tomado de Ruiz de Mendoza Ibáñez y Mairal Usón (2007, p. 28)

La **plantilla léxica** ofrece información en tres módulos: <información pragmática> + [representación semántica] + [representación sintáctica]. En la representación pragmática se recoge información sobre el tipo de discurso, la connotación y el registro. En la representación semántica se aclara el significado del predicado para poder así diferenciar el significado de los distintos verbos. En la representación sintáctica se da información sobre el predicado verbal y sus argumentos a través de una LS de la RRG. Deben destacarse las relaciones de hiponimia dadas en la jerarquía semántica con la que se organizan los verbos³⁵, algo que resulta útil a la hora de resolver la correferencia y la anáfora indirecta.

La **plantilla construccional** es una representación no léxica con un impacto gramatical (Ruiz de Mendoza Ibáñez y Mairal Usón, 2008, p. 363) y está formada por

³⁴ Más información en <<http://www.lexicom.es/drupal>>. Véase Butler (2009) para un informe más detallado sobre los modelos que han influido en el LCM.

³⁵ En Jiménez Briones (2009, pp. 398-399) se encuentran ejemplos de la estructura jerárquica entre verbos de causar enfado, reflejándose así las diferencias y similitudes de sus significados.

argumentos relacionados a través de predicados abstractos como CAUSE, BECOME o MOVE, además de un conjunto de operadores³⁶.

Las **restricciones internas** se basan en rasgos semánticos e indican las condiciones para que las plantillas léxicas modifiquen su configuración interna y puedan unirse a las plantillas construccionales. Las **restricciones externas** se basan en mecanismos de asociación cognitivos y conceptuales como la metáfora y la metonimia³⁷ y apuntan las posibilidades de que estos mecanismos tengan lugar en los elementos que se fusionan de un nivel a otro.

Este proceso de construcción del significado ocurre en los cuatro niveles que conforman la **arquitectura del LCM** (véase la figura 16).

El **nivel 1 argumental** de la gramática nuclear o *core grammar* trabaja con plantillas léxicas y su subsunción (*i. e.* fusión) con las plantillas de construcciones argumentales. Aquí se llevan a cabo inferencias como *I will [marry you]* (Ruiz de Mendoza Ibáñez y Mairal Usón, 2008, p. 357). Esto contribuiría a resolver anáforas como:

(56) Me lo pidió y dije “sí, quiero _____”.

Sería posible inferir que el hecho de casarse con la persona que hizo la pregunta es el correcto antecedente del pronombre *lo* y de la anáfora cero.

El **nivel 2 implicativo** de ámbito pragmático bajo alberga construcciones implicativas que fueron subsumidas del primer al segundo nivel o surgieron de procesos inferenciales de bajo nivel. Aquí debe tenerse en cuenta la anáfora indirecta:

(57) ¿Qué haces escuchando ruido?

El interlocutor infiere que al hablante no le gusta que realice esa actividad e interpreta con los rasgos peyorativos de *ruido* que ese tipo de música no es del agrado del hablante.

El **nivel 3 ilocutivo** de ámbito pragmático alto contiene las construcciones ilocutivas que fueron fusionadas y/o aquellas producidas por inferencias de alto nivel relacionadas con convenciones sociales y relevantes para la resolución de la fuerza

³⁶ Por ejemplo, [LS1] CAUSE [LS2] para representar la construcción causativa (Ruiz de Mendoza Ibáñez y Mairal Usón, 2007, p. 36).

³⁷ En Ruiz de Mendoza Ibáñez y Mairal Usón (2007) se ofrecen ejemplos de ambas.

ilocutiva. Veamos un ejemplo propuesto por Ruiz de Mendoza Ibáñez y Mairal Usón (2009, p. 175):

(58) OK, I'll *buy you the bicycle you wanted*, I promise.

Yes, please, do!

Como mencionan estos autores, sabemos que el antecedente correcto para el elemento anafórico *do* es *comprar la bicicleta* (no *promise*) gracias al conocimiento de que, cuando se expresa aceptación tras una promesa, la respuesta no se refiere al hecho de prometer, sino a la futura acción realizada al cumplir la promesa.

El **nivel 4 discursivo** comprende las construcciones de nivel anterior anexionadas, junto a las ocasionadas por los marcos no situacionales relacionados implícitamente con conexiones lógicas (p. ej. causa/efecto), temporales (p. ej. simultaneidad) o conceptuales (p. ej. similitud)³⁸. Veamos un ejemplo con 'X let alone Y` de Ruiz de Mendoza Ibáñez y Marial Usón (2008, p. 369):

(59) Would any one buy this garbage, let alone eat it?

La anáfora indirecta *garbage*→*comida* es resuelta gracias al segundo verbo *eat*. Aunque alude a *garbage*, el elemento anafórico pronominal *it* mantiene la anáfora indirecta.

Al plantearnos cómo diferenciar a través de la coindización la anáfora indirecta de la directa, plasmando que el antecedente de *it* no es literalmente 'basura', sino 'comida', en este trabajo proponemos este procedimiento:

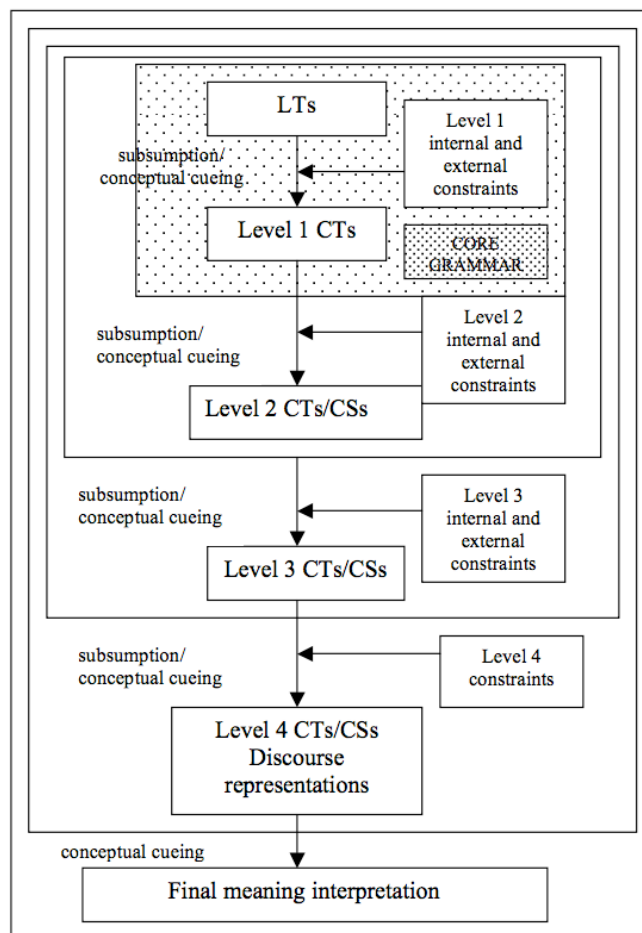
1. Señalizar con la misma letra latina (minúscula o mayúscula) los dos elementos de la anáfora: $i \dots i, j \dots j, k \dots k$, o de la anáfora indirecta: $I \dots i, J \dots j, K \dots k$.
2. Emplear las letras latinas mayúsculas I, J, K para indicar que se trata de un antecedente puente en una anáfora indirecta y por ello aún se debe localizar el antecedente. De este modo, se logra precisar si la anáfora ya está resuelta (antecedente_{minúscula}, p. ej. *this food_i*) o si aún no está resuelta (antecedente_{MAYÚSCULA}, p. ej. *this garbage_I*).
3. Precisar el antecedente al que en realidad se refiere la anáfora indirecta (p. ej. $I \rightarrow food$), de manera que quede resuelta la anáfora indirecta:

(60) Just because he cooked this food_i doesn't mean you have to eat it_i.

³⁸ Al tener lugar los fenómenos relacionados con la cohesión textual en el nivel 4, según se vaya desarrollando este nivel en el futuro, podrá arrojar más luz sobre los mecanismos que ayudan a la resolución anafórica y correferencial.

(61) Just because he cooked this garbage_i doesn't mean you have to eat it_j.
(I → food).

Tras los mecanismos del nivel 4, la representación de significado final puede requerir aún otros valores ilocutivos, pragmáticos o discursivos:



LT = lexical template; CT = constructional template; CS = Conceptual Structure

Figura 16. Arquitectura del LCM

Tomado de Ruiz de Mendoza Ibáñez y Mairal Usón (2008, p. 358)

Una de las virtudes del LCM es el papel otorgado a la metáfora y la metonimia en las restricciones externas a los procesos de subsunción (Butler, 2009, p. 143); presentando ambos fenómenos en el siguiente subapartado 1.3.2.1.

1.3.2.1. Metáforas y metonimias

En la Lingüística Cognitiva una **metáfora** es entendida como una proyección conceptual (*i. e.* conjunto de correspondencias) externa (*i. e.* entre dos dominios conceptuales o áreas semánticas), de modo que el dominio fuente nos permite razonar sobre el dominio meta (Ruiz de Mendoza Ibáñez y Mairal Usón, 2008, p. 372), p. ej. *you're getting nowhere that way* (LA VIDA ES UN VIAJE). En el LCM estas metáforas se rigen por varias normas: el **principio de invariabilidad**

extendida o *Extended Invariance Principle*, el **principio de correlación** o *Correlation Principle* y el **principio de aplicación de la proyección conceptual** o *Mapping Enforcement Principle*.

Por otro lado, la **metonimia** en la Lingüística Cognitiva es una proyección conceptual interna (*i. e.* dentro de un mismo ámbito semántico) donde un dominio fuente nos da acceso al dominio meta, usando el dominio fuente en representación del dominio meta (Ruiz de Mendoza Ibáñez y Mairal Usón, 2008, p. 374).

En la figura 17 se han simbolizado las diferencias entre la metáfora y la metonimia: el dominio fuente es más claro que el dominio meta y las flechas representan los procesos de reemplazo, expansión o reducción conceptual. En Ruiz de Mendoza Ibáñez (2001) y Ruiz de Mendoza Ibáñez y Otal Campo (2002) se diferencian dos tipos de metonimia: en el tipo **fuentes-en-meta** o *source in target* el dominio fuente es un subdominio del dominio meta, p. ej. *The ham sandwich is waiting for his check* (el sándwich de jamón es un subdominio del cliente que pidió un sándwich de jamón). Por el contrario, en el tipo **meta-en-fuente** o *target in source* el dominio meta es un subdominio del dominio fuente, como en *Chrysler has laid off a hundred workers* (el trabajador que realizó el despido es un subdominio de la empresa Chrysler donde está empleado):

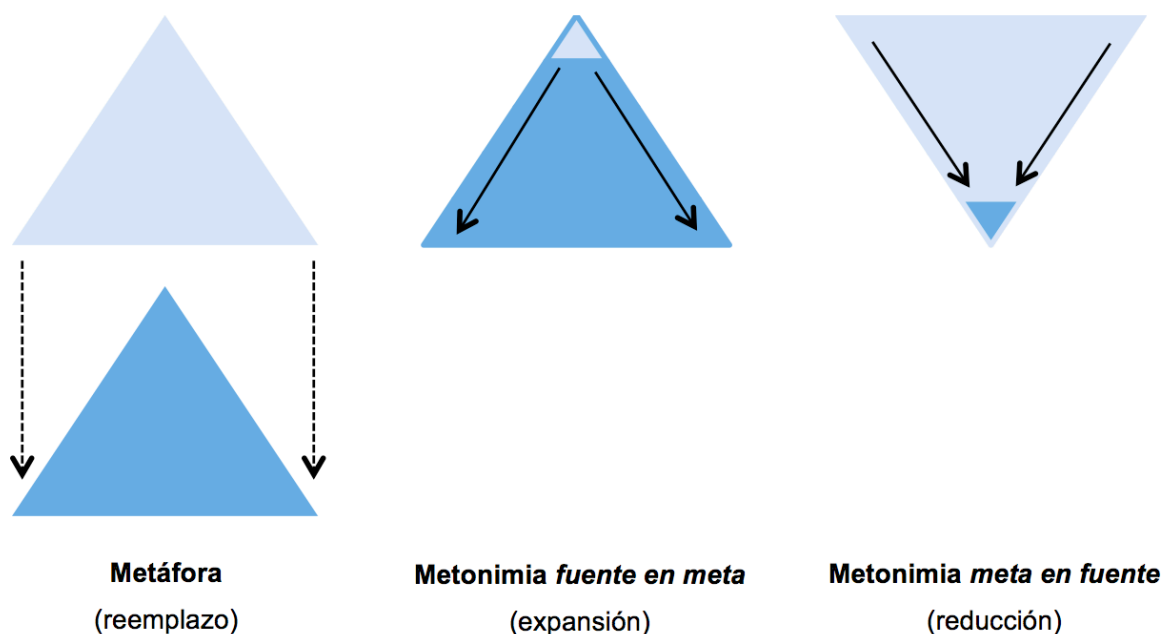


Figura 17. Metáfora y metonimias

Adaptado de Ruiz de Mendoza Ibáñez y Otal Campo (2002, p. 53)

La **anáfora metonímica** (Ruiz de Mendoza Ibáñez y Díez Velasco, 2004) se refiere a la selección de un pronombre anafórico cuando el antecedente es un

elemento metonímico, como en la ampliación del ejemplo anterior: *The ham sandwich is waiting for his check and he is getting restless*, donde el pronombre anafórico *he* se refiere a la metonimia *the ham sandwich*. Es importante señalar que con el elemento anafórico se está aludiendo al dominio meta (el cliente que pidió un sándwich), no al dominio fuente (el sándwich), como demuestra la imposibilidad de formular **The ham sandwich is waiting for his check and it is getting restless*. Con este ejemplo explican Ruiz de Mendoza Ibáñez y Díez Velasco (2004) que, en ambos tipos de metonimia, lo que se selecciona para la referencia es el **dominio matriz** de la metonimia, siendo este el dominio que más abarca de los dos implicados en la proyección conceptual, razón por la que normalmente es el dominio al que se alude en las referencias. Así, estos autores formulan en el **principio de disponibilidad de dominio** o *Domain Availability Principle* que para la referencia anafórica solamente está disponible el dominio matriz de la proyección conceptual metonímica.

Respecto a los casos de metonimias dobles³⁹, estos autores proponen dos principios más: primero, el **principio de combinación de dominio** o *Domain Combinability Principle* advierte que, habiendo dos dominios disponibles para la referencia anafórica a un nombre metonímico, seleccionamos el dominio semánticamente más compatible con el predicado de la frase donde aparece el elemento anafórico pronominal⁴⁰. Segundo, el **principio de precedencia de dominio** o *Domain Precedence Principle* señala que en las proyecciones metonímicas dobles se hace referencia al dominio matriz inicial, a no ser que el predicado combine mejor con el dominio matriz final (véase el principio anterior de precedencia de dominio).

Finalmente, para los casos de anáfora metonímica como el ejemplo anterior (*The ham sandwich is waiting for his check and he is getting restless*), los autores proponen la **restricción de anáfora metonímica** o *Constraint on Metonymic Anaphora* anunciando que, si se hace una referencia anafórica a un SN metonímico, el elemento anafórico pronominal no tiene una interpretación metonímica independiente diferente de aquella asignada a su antecedente, *i. e.* no puede ser metonímico.

³⁹ Las metonimias dobles tienen dos dominios matriz, por tanto con dos asociaciones. Puede consultarse un ejemplo en Ruiz de Mendoza Ibáñez y Díez Velasco (2004, p. 311).

⁴⁰ Véase un ejemplo en Ruiz de Mendoza Ibáñez y Díez Velasco (2004, p. 307).

En último lugar, en relación con nuestra investigación llevada a cabo con el corpus y analizada en el capítulo 4, es necesario resaltar la diferencia entre el **carácter referencial o predicativo** tanto de la metonimia como de la metáfora:

(62) We need one more brain in our team (metonimia referencial).

(63) John is a real brain (metonimia predicativa).

(64) John is a stinky rat (metáfora predicativa).

(65) The old cow wants to fire me (metáfora referencial).

Ruiz de Mendoza Ibáñez y Díez Velasco (2004, p. 296), tras proponer estos ejemplos, aclaran que lo más usual es el uso de metonimias referenciales (62) y de metáforas predicativas (64). Por esta razón creemos que debe tenerse muy en cuenta la metonimia a la hora de resolver anáforas, especialmente indirectas; de manera que se aumente así el éxito a la hora de reconocer la correferencia.

1.3.3. La base de conocimiento léxico-conceptual FunGramKB

FunGramKB es una base de conocimiento que almacena información conceptual y léxica para varias lenguas con un objetivo multifuncional, pues puede usarse en varias aplicaciones de comprensión del lenguaje natural⁴¹. Al contrario que otras bases de datos de semántica superficial (p. ej. WordNet) o de significado relacional (p. ej. EuroWordNet), FunGramKB está basada en una semántica profunda con un núcleo conceptual⁴², con lo cual no se la puede considerar como una base de datos sino de conocimiento.

Como se ve en la figura 18, la **arquitectura en FunGramKB** tiene tres niveles: léxico, gramatical y conceptual. La información léxica y morfológica para cada idioma está recogida en el Lexicón y en el Morficón, mientras que en el Gramaticón están las construcciones gramaticales de cada lengua. El conocimiento conceptual es compartido por todas las lenguas debido a que ahí se recoge información cognitiva que luego cada lengua lexicaliza de forma distinta. El modelo conceptual dispone de tres secciones donde se almacena diferente información: la semántica en la

⁴¹ Más información sobre FunGramKB en <www.fungramkb.com>. Jiménez Briones y Luzondo Oyón (2011, p. 12) aclaran que FunGramKB es el acrónimo de *Functional Grammar Knowledge Base*.

⁴² Más información sobre WordNet en <<http://wordnet.princeton.edu>> y acerca de EuroWordNet en <<http://www.illc.uva.nl/EuroWordNet>>.

Ontología por medio de postulados de significado, la procedimental en el Cognición a través de guiones y la cultural en el Onomasticón con retratos e historias:

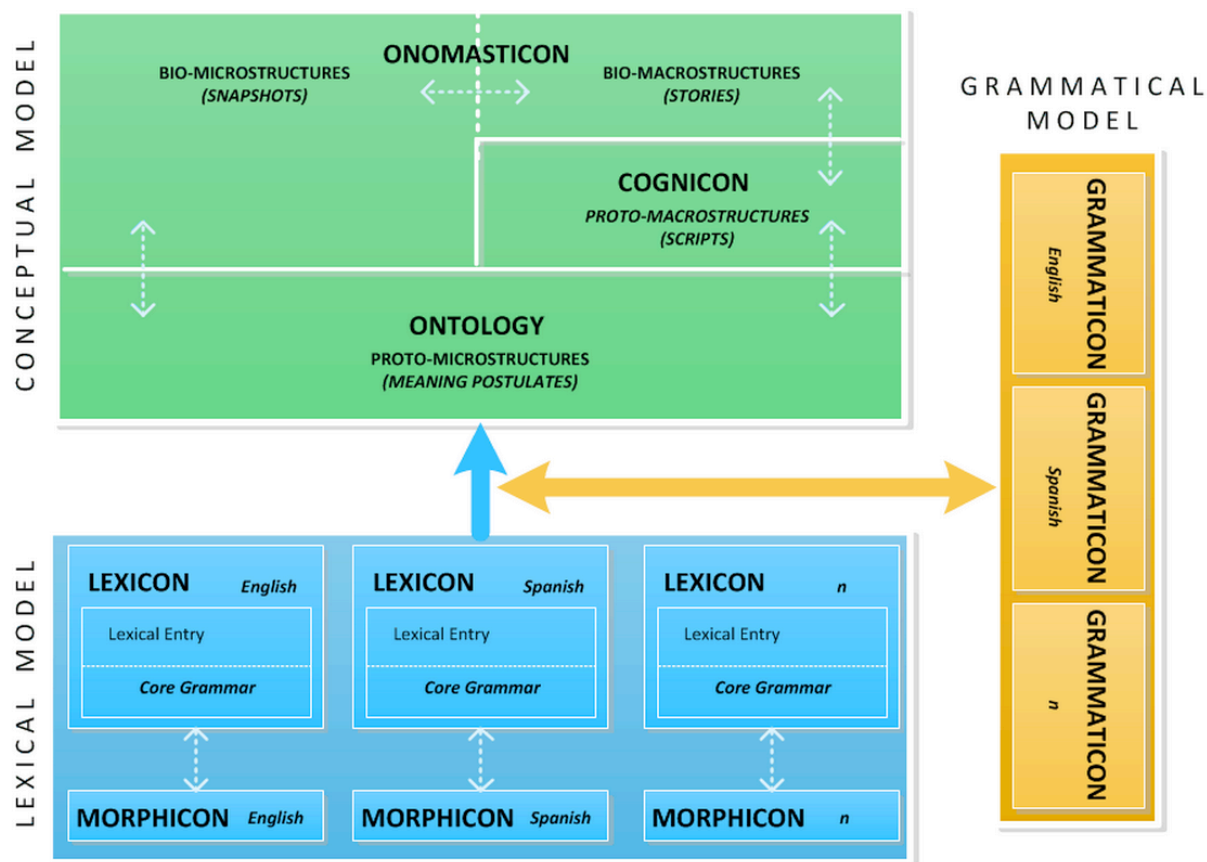


Figura 18. Arquitectura de FunGramKB

Tomado de <<http://www.fungramkb.com/architecture.htm>> [Fecha de consulta: 29.08.12]

Actualmente las lenguas más pobladas son el inglés y el español; aunque también se está completando toda la información referente a otras lenguas como el alemán, el búlgaro, el catalán, el francés y el italiano.

Las flechas que conectan todo en la figura representan la integración absoluta de la información almacenada por estar formalizada en el mismo metalenguaje COREL⁴³.

1.3.3.1. Información léxica y morfológica

Debido a la importancia para la resolución de la anáfora y de la correferencia, es necesario nombrar los beneficios que puede aportar el **Lexicón**, pues acumula la información morfosintáctica, pragmática y colocacional de las unidades léxicas.

⁴³ COREL es el acrónimo de *C*Onceptual *R*Epresentation *L*anguage, el metalenguaje empleado en todos los módulos y sobre el cual puede encontrarse información detallada en Perrián Pascual y Mairal Usón (2010b).

En Guerra García y Sacramento Lechado (2011, p. 55) se enumera la **información morfosintáctica** que se puede incluir en cada elemento léxico. Para la resolución de la correferencia, destacan las variantes gráficas o abreviaturas, el género de los sustantivos y, respecto a los verbos, su pronominalización, las variables, los macropapeles y las colocaciones. También es útil la **información miscelánea**: el dialecto, el estilo, el dominio (*i. e.* área de conocimiento) y la traducción de una unidad léxica a otras lenguas. Por otro lado, en el **Morficón** se encuentra la información sobre morfología flexiva.

1.3.3.2. Información gramatical

En el **Gramaticón**, actualmente en proceso de desarrollo por el grupo Lexicom⁴⁴, hay una gran influencia del LCM y de la RRG: aquí están los esquemas construccionales para cada uno de los cuatro niveles del LCM que ayudan a aplicar el algoritmo de enlace semántica-sintaxis para cada una de las lenguas de la base de conocimiento (Periñán Pascual, 2012). Su objetivo es: “to capture the properties that are specific to the most relevant constructional families [...]” (Jiménez Briones y Pérez Cabello de Alba, 2011, p. 104).

1.3.3.3. Información conceptual

FunGramKB tiene un **enfoque conceptualista y mentalista** debido a que en el centro de su arquitectura está la Ontología, donde se ordena la memoria semántica de los conceptos, entendidos como una representación mental estructurada (Periñán Pascual, 2010; Periñán Pascual y Mairal Usón, 2010a). Se basa en los **esquemas** (*i. e.* representación mental de una entidad o evento) de la Psicología Cognitiva. En este nivel conceptual de FunGramKB, la información se almacena en unos esquemas conceptuales organizados en torno a su prototipicidad y temporalidad:

⁴⁴ Más información sobre el grupo Lexicom en <<http://www.lexicom.es/drupal>>.

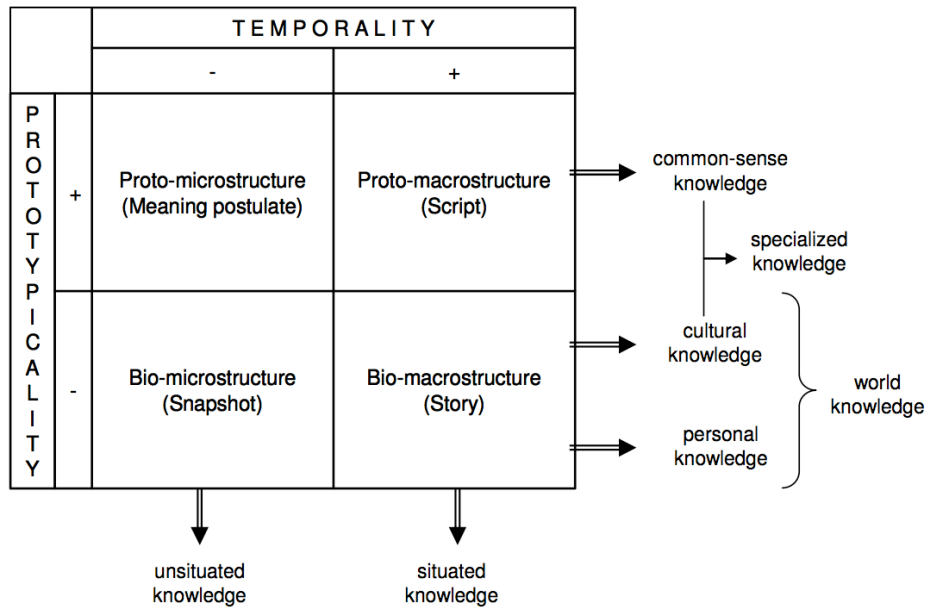


Figura 19. Esquemas conceptuales en FunGramKB
Tomado de Carlos Perrián Pascual (2012, p. 190)

Las protoestructuras (p. ej. *novela*) se diferencian de las bioestructuras (p. ej. *Gringo viejo*) en que aquellas reflejan el conocimiento del sentido común, mientras que estas representan el conocimiento del mundo. Por otro lado, las microestructuras (p. ej. *escritor*) se diferencian de las macroestructuras (p. ej. la biografía de Carlos Fuentes) en que las primeras son atemporales y las segundas temporales. Combinando estas posibilidades, obtenemos diferente conocimiento almacenado por separado en FunGramKB: las protomicroestructuras expresadas en los Postulados de Significado o *Meaning Postulates* (MMPP) de la Ontología, las protomacroestructuras recogidas con los guiones del Cognición, las biomicroestructuras reflejadas en los retratos del Onomasticón y las biomacroestructuras plasmadas con historias también del Onomasticón. El gran potencial de FunGramKB es que todo este conocimiento interactúa durante la comprensión textual gracias al uso siempre del mismo metalenguaje *COREL*:

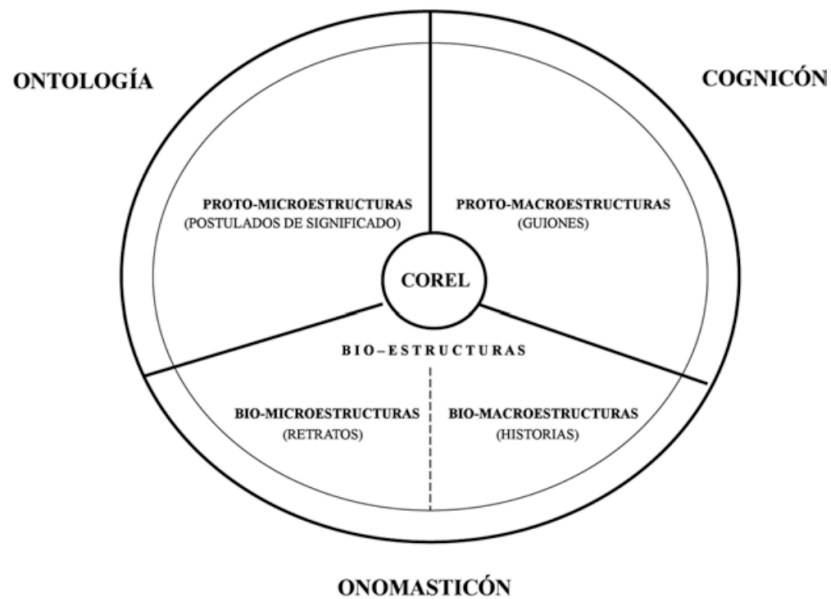


Figura 20. El planeta cognitivo de FunGramKB
Tomado de Perrián Pascual y Mairal Usón (2010b, p. 11)

Todo esto es extremadamente ventajoso para la resolución de la anáfora y la correferencia, ya que las anáforas que no se pueden resolver con la información ontológica quizás sí se puedan solventar con la procedimental o la cultural.

1.3.3.3.1. Ontología

La Ontología presenta los diferentes conceptos de los que dispone una persona al hablar de situaciones cotidianas⁴⁵. Los conceptos están ordenados en una estructura jerárquica de subsunción que posibilita la herencia múltiple monotónica y no monotónica. En la estructura de este modelo ontológico hay tres niveles: los metaconceptos (#) agrupan a los conceptos básicos (+), los cuales comprenden a su vez los conceptos terminales (\$). Tanto los básicos como los terminales pueden especificarse aún más con los subconceptos⁴⁶. Es importante recalcar que dentro de esta estructura cada concepto es lexicalizado en las diferentes lenguas con un significante distinto, p. ej. +BAG_00: *bag*, *bolsa*; algo que facilita enormemente la

⁴⁵ Se distingue entre *ontología nuclear* (más general) y *ontologías satélites* (con terminología de ámbitos más especializados); véase un ejemplo en Felices Lago, Ureña Gómez-Moreno y Alameda Hernández (2011).

⁴⁶ Como indican Jiménez Briones, Luzondo Oyón y Pérez Cabello de Alba (2011, pp. 19-20), en total hay 42 metaconceptos y unos 1.300 conceptos básicos. Además, debe tenerse en cuenta que en la Ontología los conceptos están agrupados según sean #ENTITIES, #EVENTS o #QUALITIES. De los 1657 conceptos actuales, 953 son entidades (nombres), 402 son eventos (verbos) y 320 son cualidades (adjetivos/adverbios).

detección de correferencia inter-lingüística en un mismo texto o un conjunto de textos.

Veamos ahora cómo se presenta la información dentro de cada concepto. El **marco temático** o *Thematic Frame* (TF) especifica cuántos y de qué tipo son los participantes de una situación cognitiva y el **postulado de significado** o *Meaning Postulate* (MP) define el concepto a través de sus rasgos prototípicos descritos en una o más predicaciones, en las cuales no se repite la información del TF.

Veamos un ejemplo con el MP del concepto básico +BAG_00:

The screenshot shows the FunGramKB Editor interface. On the left, there is a list of 935 entities, with +BAG_00 highlighted. Below the list are radio buttons for 'Entities', 'Events', and 'Qualities', and a checkbox for 'Non-Monotonic Inheritance'. The main area displays the 'Conceptual Information' for +BAG_00. This information is organized into a table-like structure:

CONCEPT:	+BAG_00
SUPERORDINATE(S):	+CONTAINER_00
SEMANTIC TYPE:	
MEANING POSTULATE:	<pre> +(e1: +BE_00 (x1: +BAG_00)Theme (x2: +CONTAINER_00 & +ARTIFICIAL_OBJECT_00)Referent) *(e2: +BE_01 (x1)Theme (x3: +PLASTIC_00 +PAPER_00)Attribute) *(e3: +TAKE_01 (x4)Agent (x1)Theme (x5)Location (x6)Origin (x7)Goal) *(e4: +COMPRISE_00 (x1)Theme (x8: 2 +HANDLE_00)Referent) </pre>
DESCRIPTION:	a container made of paper, cloth, or thin plastic which usually opens at the top

At the bottom of the interface, there are three columns for translations: English (bag), Spanish (bolsa), and Italian (empty).

Figura 21. MP de +BAG_00 en la Ontología

Este MP define los rasgos prototípicos de la entidad +BAG_00: una bolsa siempre es un contenedor y un objeto artificial, generalmente es de plástico o papel, normalmente es usada para trasportarla de lugar y está compuesta por dos asas (cfr. ejemplo 66).

Si recordamos el ejemplo 34, observaremos que junto al pronombre *it*, se hacía referencia a *this bag* con estos dos elementos anafóricos: the handle, the leather. Veamos cómo se asocian ambos al referente +BAG_00 en FunGramKB.

La entidad +HANDLE_00 se asocia a varios conceptos debido a que está presente en sus MMPP como preferencia de selección: p. ej. en aquellos conceptos que normalmente tienen un mango como +RACKET_00, +SWORD_00 o +KNIFE_00, pero también en los que por lo general tienen dos asas como +BAG_00 (véase la cuarta predicación del ejemplo 66):

(66) MP de la entidad +BAG_00

+(e1: +BE_00 (x1: +BAG_00)Theme (x2: +CONTAINER_00 & +ARTIFICIAL_OBJECT_00)Referent)

*(e2:+BE_01 (x1)Theme (x3: +PLASTIC_00 | +PAPER_00)Attribute)

*(e3: +TAKE_01 (x4)Agent (x1)Theme (x5)Location (x6)Origin (x7)Goal)

*(e4: +COMPRISE_00 (x1)Theme (x8: 2 +HANDLE_00)Referent)

Por otro lado, *the leather* es asociado al referente *this bag* también por una relación de inferencia dentro del MicroKnowing (véase el subapartado 1.3.3.4), pero esta vez en el sentido inverso al recién comentado, ya que en este caso +BAG_00 es la preferencia de selección dentro del MP de +LEATHER_00, en cuya quinta predicación se informa de que por lo general el cuero lo usa el ser humano para ropa o bolsas:

(67) MP de la entidad +LEATHER_00

(((e1: +BE_00 (x1: +LEATHER_00)Theme (x2: +SKIN_00)Referent)(e2: +COMPRISE_00 (x3: +ANIMAL_00)Theme (x2)Referent))

+(e3: n +EXIST_00 (x4: +HAIR_01)Theme (f1: x1)Location)

*(e4: +BE_01 (x1)Theme (x5: +SMOOTH_00)Attribute)

*(e5: +USE_00 (x6: +HUMAN_00)Theme (x1)Referent (f2: +CLOTHING_00 ^ +BAG_00)Purpose)

Este hecho tiene una importancia enorme para la correcta resolución de la anáfora indirecta. Por ejemplo en la frase *Mente brillante!*, ¿cómo funciona la metonimia *mente-persona*? La palabra *mente* está asociada al concepto +INTELLIGENCE_00, en cuyo MP aparece como preferencia de selección el concepto +COGNITIVE_ATT_00, ya que la primera predicación dice que es una habilidad cognitiva:

(68) MP de la entidad +INTELLIGENCE_00

+(e1: +BE_00 (x1: +INTELLIGENCE_00)Theme (x2: +COGNITIVE_ATT_00)Referent

*(e2: +THINK_00 (x3)Theme (x4)Referent (f1:x1)Means)

A su vez, el concepto +HUMAN_00 está asociado al concepto +COGNITIVE_ATT_00 porque aquel aparece como preferencia de selección en la segunda predicación del MP de este, donde se especifica que es una habilidad

cognitiva existente en el ser humano (véase la segunda predicación del siguiente ejemplo 69):

(69) MP de la entidad +COGNITIVE_ATT_00
 +(e1: +BE_00 (x1: +COGNITIVE_ATT_00)Theme (x2: +ABILITY_00)Referent)
 +(e2: +BE_01 (x3: **+HUMAN_00**)Theme (x1)Attribute)

Por último +HUMAN_00 está lexicalizado como *person*. Veamos la ruta conceptual en la figura 22, donde las flechas simbolizan el hecho de que un concepto aparece como preferencia de selección en el MP de otro concepto:

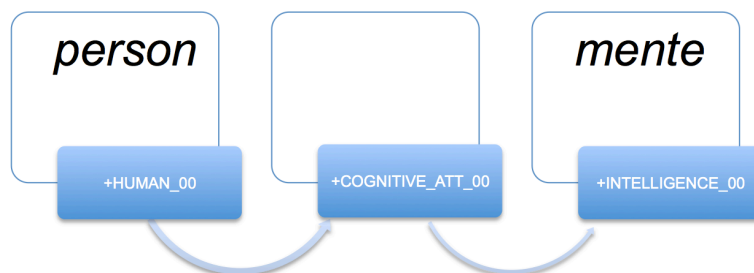


Figura 22. Camino conceptual entre *mente*-*person*

Otra característica importante para la resolución de anáforas es el hecho de que una misma palabra pueda estar asociada a más de un concepto, como ocurre con *brain*, lexicalizando tanto el concepto +BRAIN_00 como +INTELLIGENCE_00:

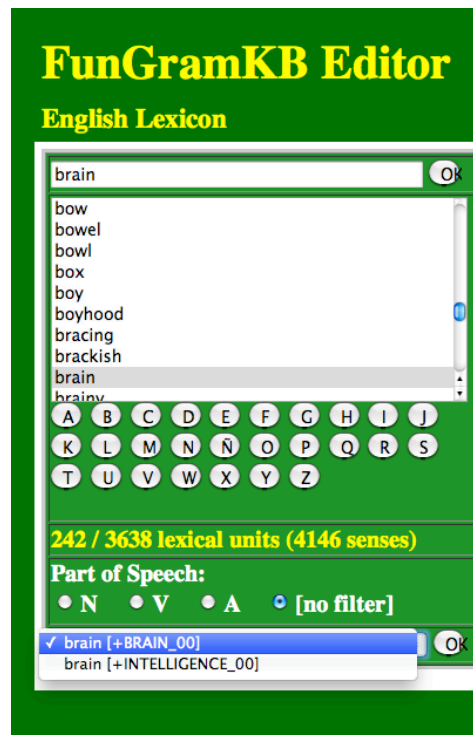


Figura 23. Búsqueda de *brain* en el Lexicón

Esto es algo muy beneficioso para la resolución de la anáfora y la correferencia a la hora de tratar con elementos anafóricos o antecedentes polisémicos, ya que con la ayuda del resto de conceptos presentes en el texto, sería posible seleccionar automáticamente el concepto más compatible con el sentido del texto.

1.3.3.3.2. Cognición

El Cognición almacena información procedimental a través de **guiones** compuestos de predicaciones ordenadas cronológicamente⁴⁷.

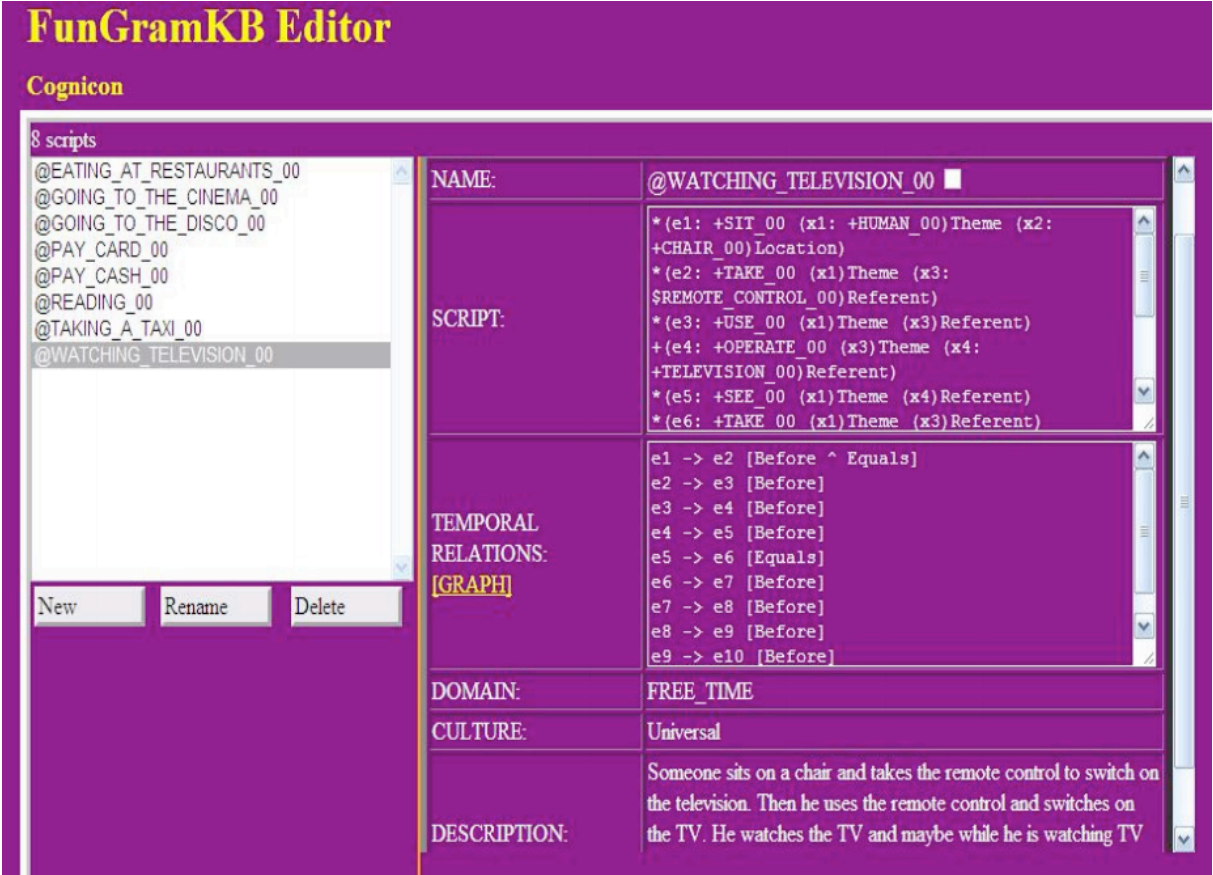


Figura 24. El guión @WATCHING_TELEVISION_00 en el Cognición Tomado de Garrido García y Ruiz de Mendoza Ibáñez (2011, p. 111)

Cada predicación es un evento y cada evento es un espacio de tiempo entre un comienzo y un final. En el guión se indican las relaciones temporales entre eventos:



Figura 25. Eventos e intervalos en el guión @WATCHING_TELEVISION_00 (E simboliza un evento descrito en la predicación, b -before-, antes, e -equals-, a la vez)

⁴⁷ Garrido García y Ruiz de Mendoza Ibáñez (2011) hacen referencia a las similitudes entre el nivel 2 del LCM y el Cognición en FunGramKB.

Como puede observarse en la figura 24, el guión @WATCHING_TELEVISION_00 ofrece información de lo que ocurre durante el acto de ver la televisión⁴⁸: un humano, por lo general sentado en una silla, toma el mando de la televisión y lo usa para encenderla. Normalmente ve la televisión, toma el mando y lo usa para cambiar el canal o para apagar la televisión.

La información archivada en el Cognición puede ser muy valiosa. En el ejemplo 32 vimos que se hacía referencia a *television* y a *remote control*. Una manera de asociar ambos conceptos es a través del guión simple @WATCHING_TELEVISION_00, ya que en todos los eventos excepto el primero y el quinto está presente el mando a distancia: para encender la televisión, cambiar de canal y apagarla.

El gran potencial del Cognición radica en la posibilidad de insertar unos guiones en otros, de manera que se plantean tres tipos: **simples** (p. ej. @WATCHING_TELEVISION_00), **complejos** (p. ej. el subguión @PAY_CASH_00 dentro de @GOING_TO_THE_DISCO_00) y **compuestos** (p. ej. @GOING_OUT_00 está compuesto de varias opciones como @GOING_TO_THE_DISCO_00, @GOING_TO_THE_CINEMA_00 y @GOING_TO_A_RESTAURANT_00).

Además, igual que en la Ontología podíamos asociar unos conceptos a otros (véase la figura 22), en el Cognición ocurre lo mismo: algunos guiones están alojados en otros y, al integrar ese conocimiento, se llevan a cabo proyecciones de salida y de entrada entre los guiones anfitriones y los guiones huéspedes como se refleja en el activador (véase *host script - guest script, outcoming mapping – incoming mapping e internal script activator* en Perrián Pascual, 2012, pp. 206-207), gracias a lo cual se pueden realizar numerosas inferencias, como se muestra en esta figura:

⁴⁸ El guión completo está disponible en Garrido García y Ruiz de Mendoza Ibáñez (2011, pp. 109-110).

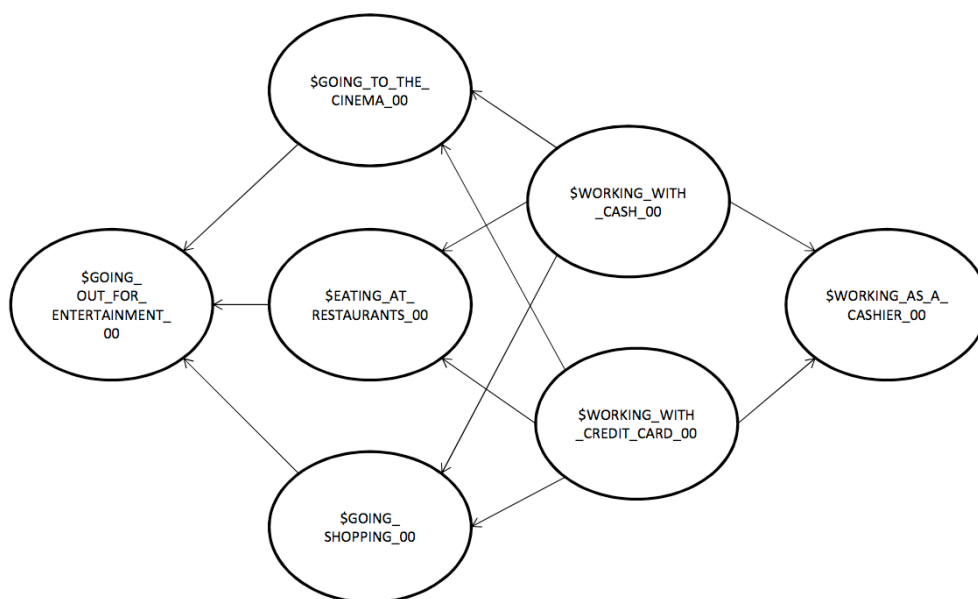


Figura 26. Relaciones establecidas entre guiones asociados
Tomado de Perrián Pascual (2012, p. 209)

Otro aspecto valioso del Cognición es que se reflejan las diferencias culturales por dos razones: en cada guión se señala el ámbito geográfico (continente, país y demás) en el que se procede generalmente de ese modo. Además, se pueden alternar las maneras de actuar a través del operador de exclusión ^, tal y como refleja el siguiente ejemplo propuesto por Perrián Pascual y Arcas Túnez (2010, p. 2670), donde se puede aclarar que en los restaurantes algunas sociedades suelen pagar en efectivo y otras lo hacen típicamente con tarjeta:

(70) *(e22: @PAY_CASH_00 (x1: x2, x2: f1) ^ @PAY_CARD_00 (x1: x1, x2: x2))

Estos fenómenos dentro del Cognición (aparecer unos guiones en/con otros, conectarlos y registrar diferencias culturales) son muy valiosos para ejemplos como los siguientes⁴⁹, donde se podría incluso dar cuenta de los distintos procedimientos en México en el caso del fallecimiento de una persona:

(71) [...] el cuerpo será velado en la casa del escritor [...].

(72) [...] personal de la agencia funeraria sacó el cuerpo del escritor en una camilla envuelta en una funda plástica de color rojo.

(73) [...] el cuerpo sería cremado durante la tarde de hoy [...].

1.3.3.3.3. Onomasticón

⁴⁹ Ejemplos del obituario de Carlos Fuentes: Piñón, A. (16 de mayo de 2012). Murió el novelista, vive la obra. *El Universal*. Recuperado de <<http://www.eluniversal.com.mx/espectaculos/113812.html>>.

Como ya se ha comentado al comienzo del apartado 1.3.3, FunGramKB almacena información cultural en forma de biomicroestructuras (**retratos**) y de biomacroestructuras (**historias**). Por ejemplo, en %TAH_MAHAL_00 encontramos los retratos atemporales del ejemplo 74 (con información como su localización, material y partes de las que se compone) y las historias temporales del ejemplo 75 (como año de construcción o año en el que fue declarado patrimonio de la humanidad)⁵⁰:

(74) Retrato de %TAH_MAHAL_00

+ (e1: +BE_02 (x1: %TAH_MAHAL_00)Theme (x2: %INDIA_00)Location)

* (e2: +BE_01 (x1)Theme (x3: +WHITE_00 & \$MARBLE_00)Attribute)

* (e3: +COMPRISE_00 (x1)Theme (x4: 1 \$DOME_00 & 4 +TOWER_00)Referent)

(75) Historia de %TAH_MAHAL_00

+ (e1: past +BUILD_00 (x1)Theme (x2: %TAH_MAHAL_00)Referent (f1: 1633)Time)

+ (e2: past +BE_00 (x2)Theme (x3: %WORLD_HERITAGE_SITE)Referent (f2:1983)Time)

Con esta información se pueden resolver correferencias como las del ejemplo 33. Perriñán Pascual y Carrión Varela (2011) resaltan el papel de las **entidades nombradas** en PLN⁵¹, p. ej. en la resolución de la correferencia. Para el reconocimiento de entidades nombradas no basta con una lista de nombres propios, sino que es necesario incluir conocimiento cultural en ellas. Por lo tanto, FunGramKB se plantea la meta de trasladar al Onomasticón dicho conocimiento cultural acumulado en bases de datos como DBpedia.

DBpedia es una base de datos que actualmente ofrece versiones en diferentes idiomas con información estructurada de forma jerárquica en su ontología⁵². DBpedia obtiene información de varias fuentes de datos, entre otras Wikipedia, YAGO, Project Gutenberg o EuroStat. Respecto a Wikipedia, para DBpedia la principal fuente de información es el **infobox** de cada artículo donde se resumen los datos

⁵⁰ Ejemplos de Perriñán Pascual y Arcas Túnez (2011, p. 5).

⁵¹ *Entidad nombrada* se entiende como una (agrupación de) palabra(s) identificada con un nombre de persona, lugar u organización.

⁵² Más información sobre DBpedia puede encontrarse en <http://dbpedia.org/>.

más importantes de forma estructurada mediante relaciones atributo-valor, como puede observarse en el ejemplo de la siguiente figura. Resaltado en negrita aparece la información de la página de edición que cualquier usuario puede ver en la infobox a la derecha:

```

{{Infobox writer <!-- for more information see [[:Template:Infobox writer/doc]] -->
| image      = Carlos Fuentes.jpg
| alt       = Head and shoulders photo of a greying man with a small moustache,
              wearing a suit, arms folded.
| imagesize = 200px
| caption   =
| pseudonym =
| birth_name = Carlos Fuentes Macías
| birth_date = {{birth date|1928|11|11|mf=y}}
| birth_place = [[Panama City]], Panama
| death_date = {{Death date and age|2012|5|15|1928|11|1|mf=yes}}
| death_place = [[Mexico City]], Mexico
| occupation = {{Flatlist|
* Novelist
* Writer
}}
| nationality = Mexican
| period     = 1954–2012
| genre     =
| subject   =
| movement  = [[Latin American Boom]]
| notableworks = {{Plainlist|
* "[[The Death of Artemio Cruz]]" (1962)
* "[[The Old Gringo]]" (1985)
}}
| spouse    = {{Plainlist|
* Rita Macedo (1959–1973)
* Silvia Lemus (1976–2012, his death)
}}
| partner   =
| children  = {{Plainlist|
* Cecilia Fuentes Macedo (1962–)
* [[Carlos Fuentes Lemus]] (1973–1999)
* Natasha Fuentes Lemus (1976–2005)
}}
| relatives =
| influences = [[Stendhal]], [[Marcel Proust]], [[James Joyce]], [[Luis Buñuel]]
| influenced =
| signature =
| website   =
| http://www.clubcultura.com/clubliteratura/clubescritores/carlosfuentes/index.htm
| www.clubcultura.com
}}

```

Carlos Fuentes



Born	Carlos Fuentes Macías November 11, 1928 Panama City, Panama
Died	May 15, 2012 (aged 83) Mexico City, Mexico
Occupation	Novelist · Writer
Nationality	Mexican
Period	1954–2012
Literary movement	Latin American Boom
Notable work(s)	The Death of Artemio Cruz (1962) The Old Gringo (1985)
Spouse(s)	Rita Macedo (1959–1973) Silvia Lemus (1976–2012, his death)
Children	Cecilia Fuentes Macedo (1962–) Carlos Fuentes Lemus (1973–1999) Natasha Fuentes Lemus (1976–2005)
Influences	[show]

[\[www.clubcultura.com\]](#) 
[www.clubcultura.com](#) 

Figura 27. Infobox de Carlos Fuentes en Wikipedia

Antes de describir cómo se añade información al Onomasticón, hay que aclarar que en el Onomasticón se podrá incluir información de DBpedia solamente en el caso de que los conceptos necesarios existan en la Ontología, en caso contrario deben ser creados primero en la Ontología.

El proceso de importación de conocimiento cultural al Onomasticón comienza con la **fase de construcción** en la que se crea una plantilla en el *DBpedia mapper* que pueda después ser rellanada con datos concretos de DBpedia durante la **fase de proyección**. Además, en la **fase de actualización** se renuevan los datos que hayan cambiado:

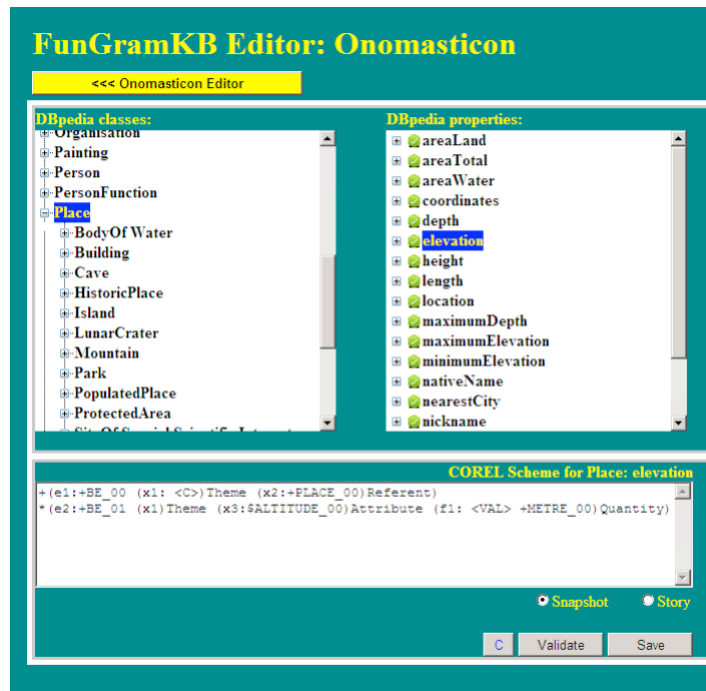


Figura 28. Plantilla *Place: elevation* del Onomasticon
Tomado de Perifán Pascual y Carrión Varela (2011, p. 95)

Así, con las plantillas y los datos de DBpedia, se consigue que los retratos e historias sean creados y actualizados en COREL de forma semi-automática. De esta manera, el conocimiento cultural y los procesos de expansión conceptual (véase el subapartado 1.3.3.4 sobre el razonamiento en FunGramKB) ayudarían a interpretar correctamente los siguientes ejemplos:

- (76) El _____ de naranja siempre será mi *Minute Maid* favorito.
- (77) Yo creo que el _____ naranja es mucho más feo que tu anterior *BMW*.
- (78) I had issues with *Orange*. They sent me a phone bill saying I was owing £500 for data usage.
- (79) Great *Halloween* songs for kids evoke the spirit of the orange holiday.
- (80) El resultado de la Naranja en 2010: España 1 – *Países Bajos* 0.
- (81) El resultado de la Naranja en 2010: *Syracuse* 71 - *Cincinnati* 54.

Gracias al conocimiento del Onomasticon sobre Minute Maid, BMW o la compañía Orange, y también a otros elementos léxicos presentes como *phone bill*, *holiday*, *Países Bajos* o *Syracuse*, es posible interpretar los antecedentes correctos de los elementos anafóricos con *naranja/orange*⁵³ y, por tanto, también su correferencia con un zumo, un coche, una compañía telefónica, la festividad de Halloween, la

⁵³ Rodríguez (2000) señala que las palabras, fuera de su contexto, tienen una categoría gramatical que puede ser ambigua; como el caso de *orange* ('fruta') asociado a la entidad +ORANGE_01 del ejemplo 76 u *orange* ('color') asociado a la cualidad +ORANGE_00 del ejemplo 77. El autor explica que la categoría de una palabra deja de ser ambigua al aparecer en un contexto.

selección de fútbol de los Países Bajos o el equipo de baloncesto de la Universidad estadounidense de Siracusa. En el siguiente subapartado 1.3.3.4 veremos cómo y dónde se llevan a cabo estos procesos.

1.3.3.4. Razonamiento y procesamiento

En el **significado relacional** (*i. e.* asociaciones entre unidades, véase Perrián Pascual, 2005) de FunGramKB se diferencian dos procesos para el **razonamiento** (Perrián Pascual y Arcas Túnez, 2005): por una parte, el **proceso de expansión micro-conceptual** o *Microconceptual-Knowledge Spreading* (MicroKnowing) que acude a las protomicroestructuras en forma de MMPP de la Ontología y, por otra parte, el **proceso de expansión macro-conceptual** o *Macroconceptual-Knowledge Spreading* (MacroKnowing) que se basa no solo en la Ontología, sino también en las protomacroestructuras en forma de guiones del Cognición. Recientemente se ha apuntado la opción de que “podemos diseñar un mismo motor de razonamiento que sirva para los conocimientos semántico, procedimental y episódico, facilitando así la integración del conocimiento” (Perrián Pascual y Carrión Varela, 2011, p. 99). En este trabajo se enfatiza la gran utilidad que tendría tal posibilidad a la hora de reconocer correferencias y resolver anáforas con éxito, por ello se propone añadir otro proceso de razonamiento: el **proceso de expansión bio-conceptual** o *Bioconceptual-Knowledge Spreading* (BioKnowing) que incluya todos los ámbitos de conocimiento conceptual en FunGramKB: la Ontología, el Cognición y las bioestructuras en forma de eventos e historias del Onomasticón.

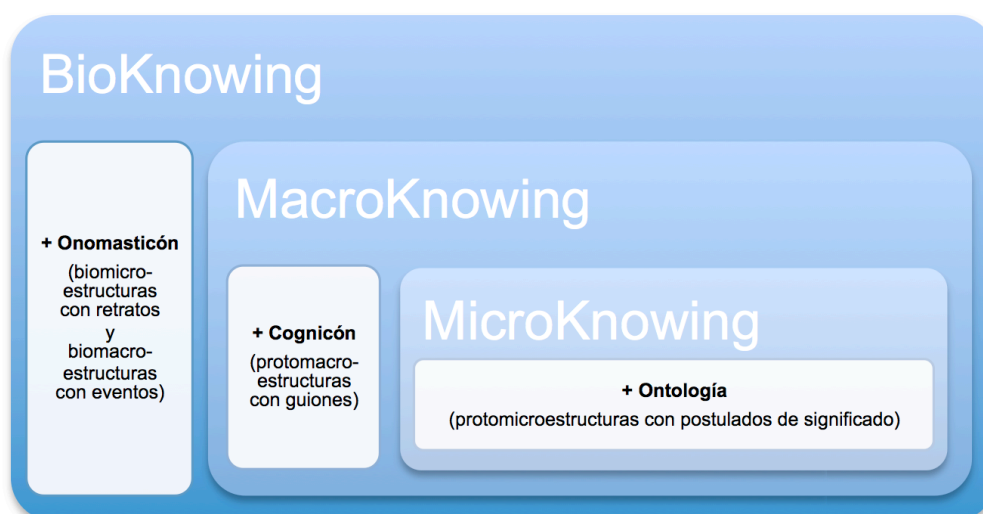


Figura 29. Propuesta del tercer proceso de expansión conceptual en FunGramKB

En todos estos procesos de razonamiento se producen dos posibles mecanismos de razonamiento basados en los dos tipos de relaciones que pueden darse entre los conceptos: herencia o inferencia (Periñán Pascual y Arcas Túnez, 2007a). Todo concepto subordinado tiene relaciones de **herencia**; la **inferencia** se da entre conceptos no subsumidos:

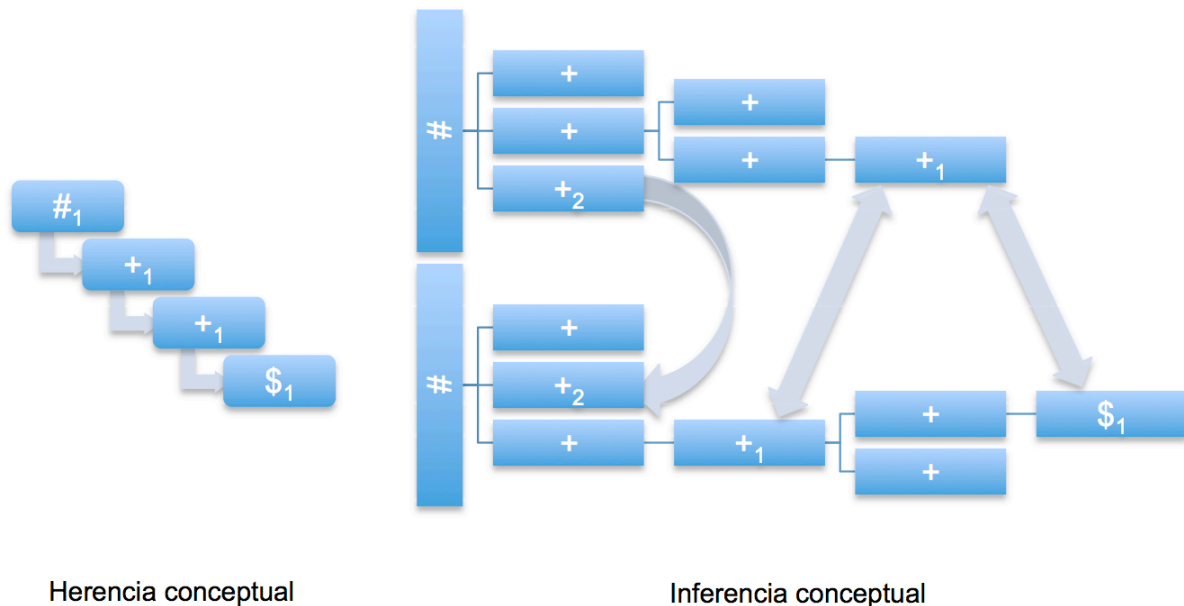


Figura 30. Relaciones de herencia y de inferencia

En último lugar, de las fases llevadas a cabo en FunGramKB en el **procesamiento de textos** (véase Periñán Pascual y Mairal Usón, 2010a; Periñán Pascual y Arcas Túnez, 2011), *i. e.* tokenización, análisis morfológico, procesamiento sintáctico-semántico, construcción de la LCS y razonamiento de lo proyectado a COREL, nos interesa especialmente esta última pues en el presente trabajo proponemos un algoritmo de resolución de la anáfora y la correferencia que puede incluirse en esta última fase donde se accede a todo el módulo conceptual (Ontología, Cognición y Onomasticón). Todo ello se puede esquematizar en este diagrama de secuencias de la figura 31:

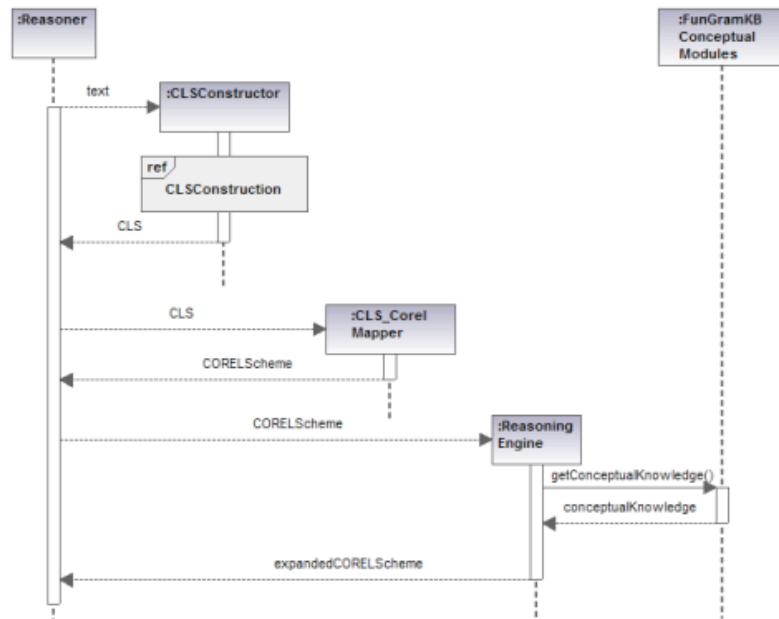
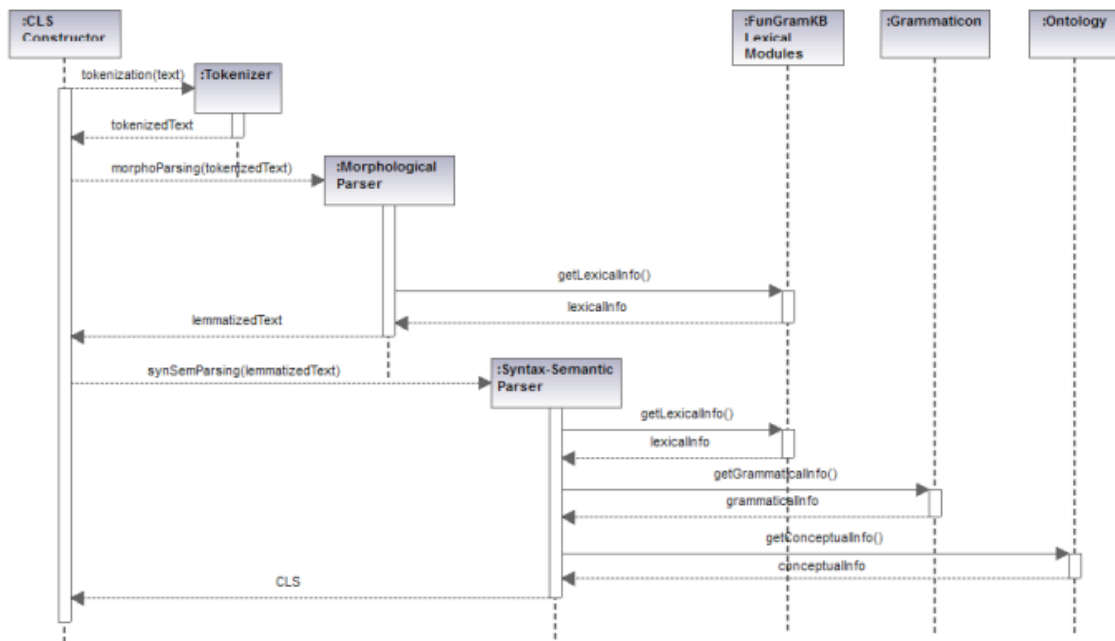


Figura 31. Procesamiento textual en FunGramKB
 Tomado de Perifán Pascual y Arcas Túnez (2011, p. 15)

2. OBJETIVOS

2.1. Objetivo principal

Mitkov (2002), tras mencionar la necesidad de sistemas multilingües de resolución de anáfora, señala la necesidad de incluir conocimiento del mundo para resolver con éxito las anáforas aunque, por otro lado, este autor muestra el pesimismo de encontrar soluciones para ello: “In fact this privileged factor shows even more clearly the limitations of automatic anaphora resolution given the unrealistic task of representing and acquiring real-world or common-sense knowledge” (Mitkov, 2002, p. 194).

Por lo tanto, partimos de una situación que se enfrenta al problema de falta de integración de conocimiento del mundo o sentido común a los sistemas automáticos de resolución de la anáfora y la correferencia. La pregunta planteada en esta investigación es qué ventajas puede aportar en este ámbito FunGramKB para resolver esa cuestión. Precisamente, el objetivo principal de este estudio es proponer la base de conocimiento multilingüe **FunGramKB como fuente de información** ontológica, procedimental y cultural para la resolución de anáforas correferenciales.

2.2. Objetivos secundarios

En primer lugar, se quiere comprobar qué beneficios aporta al objetivo principal la **información léxica, morfológica y gramatical** del módulo Lexicón de FunGramKB, pues muchos sistemas de conocimiento pobre (véase el apartado 1.2.1) se centran en el género y el número, la misma categoría gramatical y/o función sintáctica, y sinónimos basados en semántica superficial. Sin embargo, el Lexicón de nuestra base de conocimiento incluye además datos sobre los

macropapeles, las variantes gráficas, las abreviaturas, los colocados o los nombres propios y alias. Las asociaciones léxicas, junto a las semánticas, de sentido común y de conocimiento extralingüístico que veremos a continuación, cobran un papel protagonista.

Al usar FunGramKB una semántica profunda que no se limita a relacionar elementos léxicos, sino que aporta una rica descripción conceptual; esta base de conocimiento logra potenciar enormemente las redes de sinónimos, hipónimos e hiperónimos, los cuales son recursos fundamentales en la correferencia textual (véase el subapartado 1.3.2.1 sobre la metonimia referencial reductora y expansiva). Por tanto, en segundo lugar queremos comprobar cómo de útil puede llegar a ser para el objetivo principal este **conocimiento ontológico** sobre los conceptos a los que se refieren las unidades léxicas.

El conocimiento del mundo y del sentido común presente en la manera de actuar dentro de una sociedad determinada puede ser crucial para resolver anáforas. Por ello, en tercer lugar nos gustaría confirmar si este **conocimiento procedimental** ayuda a localizar el antecedente correcto y en qué medida se beneficia, tanto de la combinatoria de guiones (especialmente los complejos), como de las proyecciones de salida y entrada entre los guiones anfitriones y los guiones huéspedes. Sería también interesante ver hasta qué punto es práctico para el objetivo principal la diferenciación cultural según el ámbito geográfico.

En las correferencias, con frecuencia se presupone cierto conocimiento extralingüístico para poder resolver con éxito la anáfora. Por ello, en cuarto lugar queremos analizar la importancia de disponer de este **conocimiento del mundo**: no solo por la utilidad de la información aportada sobre un elemento determinado, sino también por los datos culturales aportados a través de las conexiones con otros elementos presentes en el discurso.

Todos los datos almacenados en FunGramKB son muy valiosos. Pero máspreciado aún es el hecho de poder acceder a todos ellos gracias al uso del mismo lenguaje formal o COREL, adquiriendo una potente capacidad de que un concepto nuclear expanda totalmente su MP al obtener conocimiento de sus conceptos superordinados, de los conceptos presentes en su MP y de los conceptos en cuyo MP está presente como preferencia de selección. Esta es la razón por la cual, en quinto lugar, queremos examinar las consecuencias para la resolución de

correferencias de que todo el conocimiento sea accesible por **herencia**, y especialmente por **inferencia**, en procesos como **Micro-**, **Macro-** y **BioKnowing**.

3. METODOLOGÍA

3.1. Hipótesis

En el capítulo anterior expusimos el objetivo de este trabajo: proponer FunGramKB como fuente de información para resolver anáforas correferenciales. Al ser necesario examinar su capacidad para lograr resolver con éxito y de forma autónoma dichas anáforas, se ha formulado esta hipótesis:

(H₁) Si se usa el conocimiento almacenado en FunGramKB, se puede resolver la anáfora correferencial.

Debido a que hasta la fecha aún no se ha propuesto ningún algoritmo que describa los pasos a dar en esta base de conocimiento para resolver anáforas, no contamos con datos empíricos con los que poder confirmar o refutar la anterior hipótesis. Esta es la razón por la que en este trabajo la efectividad de FunGramKB para resolver con éxito correferencias es valorada únicamente desde el ámbito **teórico**, planteando qué elementos pueden contribuir a conseguirlo.

A continuación veremos con qué herramientas y métodos se ha desarrollado el presente estudio.

3.2. Herramientas

Los materiales empleados en esta investigación se componen, primero, de los textos que conforman el **corpus**, los cuales comprenden dos artículos periodísticos en su versión online junto a los comentarios realizados por los lectores. Este corpus alcanza 7.582 palabras y en él se emplean dos idiomas: el inglés y el español (además de una oración en turco). Los rasgos que ambos textos tienen en común son varios: consisten en dos obituarios que tratan el mismo tema: el fallecimiento del escritor Carlos Fuentes; por lo cual comparten también la fecha de impresión: el 16 de mayo del año 2012. Las características que los diferencian se basan en la

extensión del artículo y la cantidad de comentarios: el artículo de *The New York Times* cuenta con 1.800 palabras, mientras que el texto del periódico *El Universal* tiene 1.187. En el artículo estadounidense aparecieron 60 comentarios con un total de 4.345 palabras; por otro lado, en el texto mexicano contribuyeron 6 personas extendiéndose en total 250 palabras.

El hecho de usar un corpus responde a varios motivos. En primer lugar, al deseo de trabajar con ejemplos de correferencias reales que reflejen un uso actual de ambas lenguas. En segundo lugar, el interés por abarcar distintas variedades diatópicas, pues aunque el ámbito geográfico esté determinado por tratarse de un periódico estadounidense y otro mexicano, en los comentarios de *The New York Times* se incluyen opiniones del continente americano, Europa y Asia. En tercer y último lugar, debido a que en este estudio la percepción de la investigadora se acepta como conocimiento válido, las conclusiones están condicionadas por su interpretación subjetiva, por lo que fue considerado necesario introducir una evidencia de carácter objetivo que diera cuenta de la frecuencia con la que ocurren los diferentes tipos de anáforas.

En segundo lugar, contamos con la información almacenada en la base de conocimiento FunGramKB accesible a través de **FunGramKB Suite**, en concreto el Lexicón y la Ontología para el inglés. Al no tener acceso al Cognición ni al Onomasticón, todavía en un estadio incipiente en su elaboración, hemos especulado con la creación de guiones y la información transferida de DBpedia mapper.

3.3. Procedimiento

El objetivo se logra analizando, respecto a la hipótesis planteada, cómo podrían diferentes aspectos de la variable independiente (si se usa el conocimiento almacenado en FunGramKB) llegar a afectar la variable dependiente (resolver la anáfora correferencial). Esto se pretende alcanzar a través de un estudio bimetódico con procedimientos de investigación **cuantitativos** y **cualitativos**.

Por otro lado, debe tenerse en cuenta que la validez externa de los datos está condicionada porque los resultados no son generalizables debido al tamaño limitado del corpus: únicamente se trata el género periodístico con dos obituarios y las reacciones de los lectores.

Para poder valorar la resolución correferencial en los textos, en ellos se ha realizado un análisis previo marcando y contabilizando las correferencias con el tópic del tema: el escritor Carlos Fuentes. Tras calcular la frecuencia con la que determinadas expresiones correferenciales aparecen en los textos, se ha comprobado qué conocimiento y mecanismos están disponibles en FunGramKB para resolverlas simulando en lo posible el razonamiento humano. Después se ha especulado sobre qué información y tipo de razonamiento en dicha base de conocimiento léxico-conceptual podría ser aprovechado por un sistema automático de resolución de correferencias. En último lugar, se ha propuesto un algoritmo donde se detallan los pasos a seguir basándose en las conclusiones de los resultados obtenidos, donde queda respondida la pregunta realizada al principio de esta investigación: qué ventajas puede aportar FunGramKB en la resolución automática de la anáfora correferencial.

4. DATOS

4.1. Descripción de datos obtenidos

En ambos textos incluidos en el anexo 2 se encuentran numerosos elementos correferenciales con función predicativa o referencial. En nuestro análisis no son tenidos en cuenta los SSNN con función predicativa (ya sea en forma de atributo o de aposición)⁵⁴, centrándonos en los **SSNN** que son **referenciales con el tópico** del texto: el escritor Carlos Fuentes⁵⁵. Por otro lado, además de los SSNN que establecen correferencia con él, se ha trabajado también con aquellos que aluden a este autor junto a otro(s) referente(s), p. ej. *both [writers]*.

En los textos de ambos periódicos se encuentran diferentes tipos de anáforas correferenciales con Carlos Fuentes: el elemento correferencial puede ser anáfora cero, pronombre, nombre propio o SN (in)definido:

	<i>The New York Times</i>	<i>El Universal</i>	Total
Pronombre	150 (48,38%)	19 (17,92%)	169
Anáfora cero	45 (14,51%)	59 (55,66%)	104
Nombre propio	91 (29,35%)	13 (12,26%)	104
SN definido / indefinido	24 (7,74%)	15 (14,15%)	39
Total	310	106	416

Tabla 2. Distribución de elementos correferenciales en el corpus según periódico

Tras realizar la prueba de χ^2 para descartar que las diferencias observadas se deben al azar, hemos comprobado que la manifestación de estas categorías correferenciales es significativamente ($p < 0,001$) dependiente de estos idiomas. Ello

⁵⁴ Véase van Deemter y Kibble (2001) para una explicación sobre los motivos por los que no deben tenerse en cuenta los SSNN no referenciales ni los SSNN atributivos.

⁵⁵ La posibilidad de sustituir el SN por el nombre propio *Carlos Fuentes* ha sido el criterio seguido para distinguir los usos referenciales y los usos predicativos. En el caso de aparecer ambos juntos, si uno de ellos estaba separado mediante comas en forma de aposición, ha sido considerado predicativo.

confirma que estos datos no son casuales, de hecho reflejan los fenómenos esperados: la anáfora cero tiene mayor presencia en español que en inglés, habiendo más pronombres en esta lengua que en aquella.

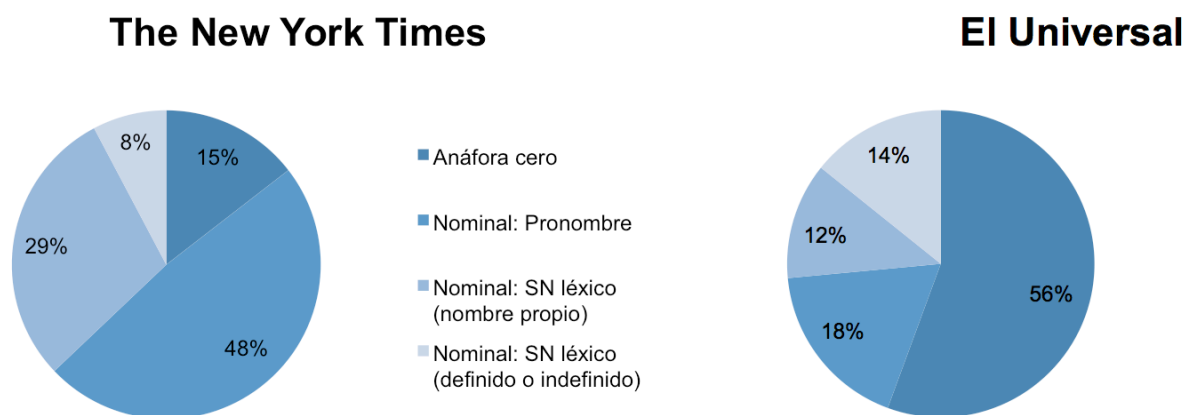


Figura 32. Porcentaje de diferentes tipos de anáfora correferencial con el tópic en artículos y comentarios

En este corpus de dos obituarios periodísticos se puede observar que la anáfora nominal de SN léxico con nombre propio es más frecuente en inglés, habiendo más anáforas nominales de SN léxico definido o indefinido en español.

Respecto a los nombres propios en *The New York Times*, por un lado llama la atención que, de las 31 ocasiones en las que se usa el nombre propio en el artículo, *Mr. Fuentes* es empleado con bastante frecuencia (26)⁵⁶. Por otro lado, esa denominación solamente se usa un cuarto de las veces (15 de 60) en los comentarios⁵⁷.

Hemos visto cómo el elemento correferencial con Carlos Fuentes puede ser, o bien elidido, o bien nominal (con un pronombre, un nombre propio o un SN definido o indefinido). Para realizar una clasificación tipológica completa, en segundo lugar debe añadirse el hecho de que todos los elementos correferenciales de ambos textos establecen relaciones anafóricas con el nombre propio *Carlos Fuentes* mencionado al comienzo; excepto el SN del título mexicano *la última gran conciencia* que es una catáfora. Tercero, aunque se encuentran algunas anáforas intra-oracionales, la mayoría son inter-oracionales. Por último, especialmente interesantes resultan las anáforas indirectas por el desafío que suponen para su resolución automática, tema tratado con mayor detenimiento en el apartado 4.2.2.

⁵⁶ En el artículo mexicano, el equivalente a *Mr. Fuentes* no se utiliza nunca. En su lugar se emplea el nombre, el apellido o ambos.

⁵⁷ Incluyendo todas las variantes formales (*Mr., Mr, Sr., Sr, Señor, Senor, Don, Maestro*).

4.2. Análisis de resultados

Como ya se explicó en el apartado 1.1.3, todo sistema de resolución automática de anáforas sigue tres pasos: identificación de los elementos anafóricos, localización de los candidatos de los antecedentes y selección del antecedente. En este análisis se expone qué puede aportar FunGramKB en cada una de esas fases. Sin embargo, antes es necesario aclarar lo siguiente: si después de haber comprobado la existencia de determinados conceptos en la Ontología se confirma que estos no existen, especularemos con el funcionamiento tras la **creación** de dichos conceptos. En relación con la información almacenada en el Onomasticón, igualmente especularemos con su **inclusión**; teniendo siempre en cuenta que cada dato añadido de DBpedia al Onomasticón debe tener un concepto en la Ontología, de modo que todo sea interpretable por la semántica profunda de FunGramKB.

En este análisis nos centramos en las correferencias con el tópico del texto: el escritor Carlos Fuentes. Se dedica especial atención a las **anáforas nominales de SN léxico definido e indefinido** del corpus, en las cuales se emplean palabras que ya tienen conceptos creados en FunGramKB (p. ej. +HUMAN_00, +MAN_00, +FRIEND_00, +ORGANISM_00, +VOICE_00, +SOUL_00, +SPIRIT_00, +FATHER_00), habiendo también palabras como *ganador*, *enamorado*, *amante*, *conciencia*, *presence*, *pillar*, *intellectual* y demás, para las cuales habría que **crear conceptos** a los que ligarlas. Por ejemplo, debido a que hasta la fecha no existe un concepto para la palabra *escritor* o *writer*, en este trabajo proponemos el siguiente concepto terminal, donde se aclara en su MP que un escritor es un trabajador y que, por lo general, un escritor escribe algo para publicarlo:

(82) Propuesta de MP para la entidad \$WRITER_00

+ (e1: +BE_00 (x1: \$WRITER_00)Theme (x2: +WORKER_00)Referent)

* (e2: +WRITE_00 (x1)Theme (x3)Referent (f1: (e3: \$PUBLISH_00 (x1)Theme (x3)Referent)Purpose))

Por otro lado, pueden añadirse las unidades léxicas *autor*, *biógrafo*, *dramaturgo*, *ensayista*, *guionista*, *narrador*, *novelista*, *poeta* y *redactor* especificando su traducción a otras lenguas.

Por último, cabe destacar el hecho de que en FunGramKB se puedan interpretar elementos léxicos en **diferentes idiomas**, lo que contribuye a reconocer

anáforas y correferencias de aquellas lenguas pobladas en la base de conocimiento léxico-conceptual.

4.2.1. Identificación de elementos correferenciales

Debemos ser conscientes de las consecuencias que tienen los **errores lingüísticos** de los textos como faltas de ortografía (*government, marvellous, generations*) o erratas al teclear (*didn;t, will, strats*), pues provocaría su no reconocimiento léxico; al igual que los casos donde falta alguna letra (*devenr, amrgura, hapy*), el espacio (*customs.His*) o algún carácter propio de una lengua (*Senor, pais*). El uso incorrecto de la tilde además provoca antecedentes erróneos como en esta anáfora cero, donde el AGX_P aparece como primera persona del singular en vez de la tercera:

- (83) [...] amante fiel de Silvia Lemus_i, a quien denomino_j “la galaxia misma” [...]. (Líneas 82 y 83 del corpus en el anexo 2).

Otro problema surge cuando se nombra el **referente erróneo**, el cual no debería ir coindizado con el antecedente correcto para así mantener el sentido de la corrección posterior:

- (84) While I didn;t always agree with Cruz's political choices [...]. (Comentario 8 del corpus en el anexo 2).
- (85) Sorry. Oops. Not Cruz_i, Fuentes_j. (Comentario 9 del corpus en el anexo 2).

Los algoritmos de resolución de correferencias suelen interpretar un **SN indefinido** como un elemento no anafórico que introduce una nueva entidad en el discurso. Esto puede confirmarse en el texto de *The New York Times*, donde encontramos un SN no referencial encabezado por un artículo indefinido, p. ej. *a Mexican novelist*. Este SN tendría un nivel de activación nuevo no anclado (véase figura 12) y estaría en el extremo más alejado del tópico (véase figura 14) debido a que no se refiere al tópico *Carlos Fuentes*, sino que se trata de una anáfora de sentido (no de referente) y alude por tanto a cualquier novelista mexicano, siendo por ello un SN no referencial.

Sin embargo, más adelante en los comentarios también encontramos varios SSNN indefinidos que sí se refieren al tópico y, aunque la mayoría tiene rasgos predicativos, pueden ser sustituidos por el nombre *Carlos Fuentes*; siendo por tanto

referenciales: *a man, a real intellectual, a most-gifted and elegant writer as well as a beautiful human being, a brilliant author, a brave and creative voice, a great man y a great writer*. En este trabajo interpretamos este hecho como una característica propia de los obituarios: referirse al fallecido a través de un SN indefinido en el que se destacan sus cualidades. Por ello debe resaltarse la necesidad de tener en cuenta las **particularidades del género textual** a la hora de resolver correferencias en este caso en los obituarios.

4.2.2. Localización de los candidatos a antecedente

Por un lado, en el caso de que el elemento correferencial sea un SN léxico con nombre propio⁵⁸, se buscará en el texto el mismo SN; proponiendo también como antecedentes aquellos SSNN con el mismo nombre o apellido (precedidos o no de fórmulas de tratamiento formal), así como los hipocorísticos, los diminutivos, las abreviaturas, los apodos y los alias, p. ej. *Carlos Fuentes, Fuentes, Mr. Fuentes, Señor Fuentes, Carlos, Don Carlos, Carlitos*.

Por otro lado, en el caso de que el elemento correferencial sea un SN léxico definido o indefinido, en primer lugar se tendrá en cuenta la **compatibilidad de género**. Para ello hay que detectar primero el género del elemento correferencial, algo que en FunGramKB es posible gracias a la información léxica, morfológica, semántica y cultural: el Morficon puede señalar el género masculino presente en el AGX_G de SSNN como *amigo, un grande más, ese enamorado o uno de los últimos grandes*. En el Lexicón español estaría registrado el género de *escritor, narrador, autor o padre*. En la Ontología también se indica el género de palabras como *man* a través del MP del concepto básico al que está asociada:

(86) MP de la entidad +MAN_00
+(e1: +BE_00 (x1: +MAN_00)Theme (x2: +ADULT_00)Referent)
+(e2: +BE_01 (x1)Theme (x3: +MALE_00)Attribute)

Segundo, debe detectarse el género de los SSNN que puedan incluirse en la lista de candidatos a antecedente. Esto en FunGramKB es posible, no solo con la información comentada anteriormente, sino también con la valiosa información almacenada en el Onomasticón. Si allí se recoge la información sobre los hombres y

⁵⁸ Sobre la resolución de la correferencia de nombres propios, véase García y Gamallo (2011).

mujeres nombrados en el texto (p. ej. *Carlos Fuentes, Octavio Paz* o *Jane Fonda*), o si respecto a sus nombres aparece el dato de que son usados para hombres (p. ej. *Arturo, Jorge* o *Alfonso*) o mujeres (p. ej. *Silvia, Cecilia* o *Rita*); pueden ser añadidos a la lista de candidatos a antecedente los SSNN que sean compatibles con el género masculino.

En segundo lugar, con la información proporcionada en el Lexicón y en el Morficon, debe tenerse en cuenta la **compatibilidad de número** entre el elemento correferencial y el referente al que este alude; proponiendo en el siguiente ejemplo candidatos donde se mencionen dos antecedentes, o bien cada antecedente por separado:

(87) The two became friends in 1950 [...]. (Línea 236 del corpus en el anexo 2).

En tercer lugar, hay que comprobar la **compatibilidad semántica del elemento correferencial** con los posibles SSNN que se quieren proponer como antecedentes. En el análisis realizado sobre nuestro corpus, hemos confirmado que un rasgo muy productivo es, por un lado, el carácter humano y, por otro lado, la profesión; como puede verse en estos ejemplos:

(88) +HUMAN_00: *amigo, a man, a real intellectual, a beautiful human being.*

(89) \$WRITER_00: *a great writer, a brilliant author, el novelista, el narrador.*

Gracias a las unidades léxicas ligadas a \$WRITER_00, a través de las relaciones de **herencia** dentro del **MicroKnowing** sería posible resolver casos de **sinonimia** como *el narrador, el novelista, y another great author*, independientemente de si esos sinónimos en el Onomasticón aparecen o no relacionados con Carlos Fuentes.

Por otro lado, en este análisis se ha dedicado especial atención a los SSNN correferenciales con **anáforas indirectas** por el reto que suponen en FunGramKB. Hemos podido observar cómo en nuestros textos se trataba siempre de metonimias del tipo *fuerza en meta* (véase el subapartado 1.3.2.1) como las siguientes:

(90) A brave and creative voice [...]. (Comentario 23 del corpus en el anexo 2).

(91) [...] this great and generous soul. (Comentario 28 del corpus en el anexo 2).

(92) [...] su espíritu[sic] [...]. (Comentario 17 del corpus en el anexo 2).

Para su posterior resolución, en primer lugar deben aclararse los rasgos semánticos aportados, de modo que estén presentes en los candidatos a antecedente. En este análisis proponemos el **MicroKnowing** como forma de obtener el rasgo semántico del referente al cual aluden estas anáforas indirectas: +HUMAN_00. Veamos cada ejemplo más detalladamente. En el caso de *voice*, se trazaría una relación de **inferencia**, ya que +HUMAN_00 no es un concepto superordinado de +VOICE_00, no obstante en la segunda predicación de su MP se especifica que es un sonido creado por un humano:

(93) MP de la entidad +VOICE_00

+ (e1: +BE_00 (x1: +VOICE_00)Theme (x2: +SOUND_00)Referent)
 + (e2: +CREATE_00 (x3: **+HUMAN_00**)Theme (x1)Referent (f1:
 +MOUTH_00 & +THROAT_00)Instrument)

El caso de *soul* es realmente interesante ya que está ligado a dos conceptos: o bien +SOUL_00 (en cuya segunda predicación del MP se indica que por lo general el ser humano está compuesto de alma, véase el ejemplo 94), o bien +HUMAN_00 (véase el ejemplo 95); con lo cual en FunGramKB esta anáfora no es tan indirecta porque el elemento léxico *soul* ya aparece asociado al concepto +HUMAN_00:

(94) MP de la entidad +SOUL_00

+ (e1: +BE_00 (x1: +SOUL_00)Theme (x2: **+SPIRIT_00**)Referent)
 * (e2: +COMPRISE_00 (x3: **+HUMAN_00**)Theme (x1)Referent)

(95) MP de la entidad +HUMAN_00

+ (e1: +BE_00 (x1: +HUMAN_00)Theme (x2:
 +ORGANISM_00)Referent)

Por último, *espíritu* se relaciona con +SPIRIT_00, concepto que carece de definición por ser un primitivo semántico, pero que sí aparece como preferencia de selección en el concepto +SOUL_00 (véase la primera predicación del ejemplo 94).

Volvamos de nuevo al ejemplo 87. ¿Por qué inferimos que el antecedente son dos personas? En el discurso, la información contextual puede llegar a ser muy valiosa: en este caso, *friends* está informando de que el elemento correferencial alude a dos personas por el MP del concepto +FRIEND_00 (además de ser subordinado de +HUMAN_00 en la Ontología):

(96) MP de la entidad +FRIEND_00

+ (e1: +BE_00 (x1: +FRIEND_00)Theme (x2: +HUMAN_00)Referent)

Así, vemos cómo la **compatibilidad semántica de otros elementos intraoracionales** puede ser muy útil a la hora de seleccionar los rasgos que deben poseer los candidatos a antecedentes.

Respecto al Cognición, debe señalarse la utilidad del **MacroKnowing**. Si recordamos los ejemplos 71, 72 y 73 sobre Carlos Fuentes, observamos que se puede detectar como elemento correferencial *el cuerpo*, pues está ligado al concepto +BODY_00 en cuyo MP se especifica que está en un ser humano o en un animal:

(97) MP de la entidad +BODY_00
 +(e1: +BE_00 (x1: +BODY_00)Theme (x2: +NATURAL_OBJECT_00 & +CORPUSCULAR_00 & +SOLID_00)Referent)
 +(e2: +BE_02 (x1)Theme (x3: +HUMAN_00 ^ +ANIMAL_00)Location)
 *(e3: +COMPRISE_00 (x1)Theme (x4: 1 +HEAD_00 & s +LIMB_00 & 1 +TRUNK_00)Referent)

En nuestro corpus hay un comentario que también podría aprovechar dicha información del Cognición sobre el luto en caso de un fallecimiento:

(98) Luto para las letras latinoamericanas, un grande mas[sic] que se ausenta... (Comentario 6 del corpus en el anexo 2).

Aunque el SP *para las letras latinoamericanas* aparece fuera del SN referencial, es elidido en la correferencia para evitar su repetición: (*un grande [de los escritores latinoamericanos] más que se ausenta*). Este caso es interesante por la **doble metonimia** que presenta: (*letra por obra y obra por escritor*). Con *letra* podría relacionarse +LETTER_01 cuyo superordinado es +MARK_00, concepto hermano de +WRITING_00, cuyo subordinado es +BOOK_00, en cuyo MP se especifica que por lo general un libro es escrito por un humano. Por último, se relacionaría +WRITE_00 con \$WRITER_00. Vemos por tanto las relaciones:

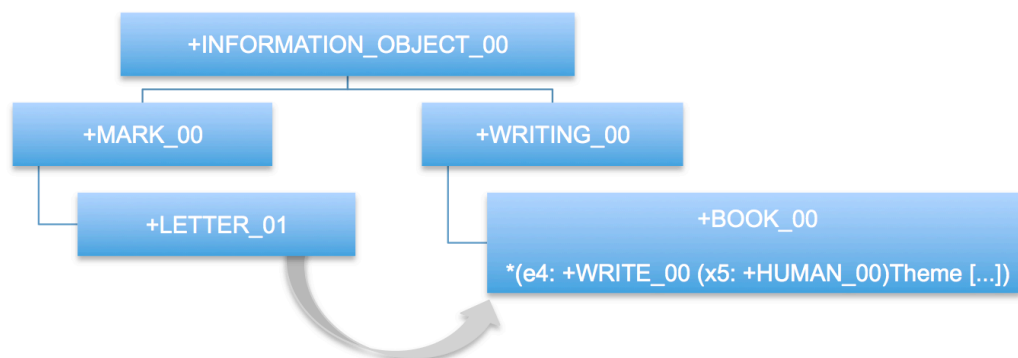


Figura 33. Relaciones entre +LETTER_01 y +BOOK_00

4.2.3. Selección del antecedente

Si en el Onomasticón están almacenados **datos** sobre la profesión, nacionalidad, fecha y lugar de nacimiento, obras emblemáticas y demás de Carlos Fuentes; este podría ser seleccionado de la lista de candidatos como el antecedente correcto:

- (99) El escritor mexicano nacido en Panamá [...], el 11 de noviembre de 1928 [...]. (Líneas 12 y 13 del corpus en el anexo 2).
- (100) El autor de obras emblemáticas como *La muerte de Artemio Cruz*, *Cristóbal Nonato*, *Aura* y *Las buenas conciencias* [...]. (Líneas 67 y 68 del corpus en el anexo 2).

La **cercanía con un posible antecedente** ayuda también a seleccionar el referente correcto:

- (101) Neither man apologized, diminishig[sic] the reputation of both. (Líneas 243 y 244 del corpus en el anexo 2).

Con los ejemplos 87 y 96 inferimos que los referentes de *the two* deben tener el rasgo +HUMAN_00. Los únicos dos referentes humanos aparecen precisamente en la oración precedente (Mr. Fuentes) y en la anterior (Mr. Paz), pudiéndose registrar en el Onomasticón también la información sobre el poeta de manera que sea reconocido. En el núcleo del ejemplo 101 tenemos una ayuda importante para resolver la correferencia: *man*, cuyo concepto +MAN_00 es subordinado de +ADULT_00 y este a la vez de +HUMAN_00, aclarando en su MP que se refiere a hombres (véase el ejemplo 86). Gracias a la información semántica y a las relaciones de herencia puede localizarse el referente correcto, siendo Mr. Paz y Mr. Fuentes nombrados de nuevo en las oraciones anteriores. Por último, *both* compartiría el mismo referente que *neither man* en la oración anterior: los escritores Octavio Paz y Carlos Fuentes.

Para terminar, en el corpus se han encontrado dos metáforas referidas a Carlos Fuentes: *A pillar of civilization* y *Nuestro Virgilio*. Proponemos que la primera sea resuelta asociando en la Ontología *pillar* y *pilar* al concepto +IMPORTANT_00, además de señalar en la información miscelánea del Lexicón su pertenencia a un registro literario. La segunda metáfora podría resolverse con el BioKnowing: en el Onomasticón habría información sobre Virgilio y en el Cognición se crearía un guión complejo @PRAISING con un subguión @COMPARING WITH CELEBRITIES

donde se refleje que en determinados ámbitos culturales las comparaciones con autores grecorromanos es una manera de alabar o ensalzar a una persona.

4.2.4. Propuesta de algoritmo para resolver la correferencia

Basándonos en el conocimiento adquirido a través del marco teórico y en las observaciones realizadas durante el análisis de la información obtenida con el corpus, proponemos el siguiente algoritmo para la resolución automática de correferencias de SN léxico no pronominal usando FunGramKB e implementándolo en su fase quinta de razonamiento (véase la figura 31) en la cual se tiene acceso a todo el conocimiento almacenado:

1. Identificación de elementos correferenciales:
 - 1.1. Busca SSNN con nombre propio.
 - 1.2. Busca SSNN definidos.
 - 1.3. En el caso de obituarios, busca también SSNN indefinidos.
2. Localización de los candidatos a antecedente:
 - 2.1. En el caso de 1.1:
 - 2.1.1. Localiza en el texto otros nombres propios con el mismo nombre o apellido (precedido o no de fórmulas de tratamiento formal), hipocorísticos, diminutivos, abreviaturas, apodos y alias.
 - 2.2. En el caso de 1.2 y 1.3:
 - 2.2.1. Localiza los candidatos con el mismo género que el elemento correferencial.
 - 2.2.2. Localiza los candidatos con el mismo número que el elemento correferencial.
 - 2.2.3. Localiza los candidatos con el mismo rasgo semántico que el elemento correferencial.
 - 2.2.4. Localiza los candidatos relacionados por herencia o inferencia con el elemento correferencial:
 - 2.2.4.1. Localiza los candidatos sinónimos al elemento correferencial.
 - 2.2.4.2. Localiza los candidatos hiperónimos al elemento correferencial.
 - 2.2.4.3. Si el rasgo +HUMAN_00 aparece en el concepto del núcleo del elemento correferencial, localiza antecedentes con ese mismo rasgo.

2.2.5. Si 2.2.3 y 2.2.4 no aportan ningún rasgo decisivo, localiza los candidatos a antecedente con el mismo rasgo semántico que otros elementos intra-oracionales del elemento correferencial.

3. Selección del antecedente:

3.1. Integra la información procedural y cultural para confirmar los datos presentes en la correferencia.

3.2. Activa el razonamiento para convertir medidas.

3.3. En caso de haber varios candidatos:

3.3.1. Selecciona el candidato que con mayor frecuencia aparezca en el texto.

3.3.2. Selecciona el candidato a antecedente más cercano al elemento correferencial.

5. CONCLUSIONES

En este trabajo se han realizado varias propuestas. En primer lugar, hemos planteado la existencia, dentro del SN español, de un índice de concordancia de género AGX_G ; resaltando su importancia como controlador para el pivote, hecho que queda reflejado en las plantillas de estructura estratificada para el SN que hemos desarrollado para explicar en términos de la RRG el fenómeno de la anáfora cero dentro del SN. En segundo lugar, hemos presentado una coindización diferenciada para la anáfora indirecta de modo que quede reflejada la asociación conceptual de la metonimia o la metáfora. En tercer lugar, se ha sugerido la denominación *BioKnowing* para el proceso de expansión informativa entre las micro, macro y bioestructuras de FunGramKB. Por último, se ha diseñado un algoritmo para la resolución de la anáfora correferencial que puede ser implementado para el procesamiento de textos con FunGramKB durante su fase final de razonamiento. Con ello hemos formalizado la consecución del objetivo principal en este estudio: proponer la base de conocimiento FunGramKB para la resolución de la anáfora correferencial. Los objetivos secundarios han sido alcanzados siguiendo el procedimiento expuesto en la metodología: hemos examinado en qué medida, tanto la información almacenada en FunGramKB (léxica, ontológica, procedural y / o cultural), como los procesos en dicha base de conocimiento, contribuyen a resolver las correferencias de los textos que conforman el corpus.

Con todo ello, podemos concluir que a nivel teórico la hipótesis de este estudio queda confirmada, siempre que FunGramKB disponga de la información necesaria en lo que se refiere al léxico, los conceptos, los guiones, los retratos y las historias. En este sentido, consideramos imprescindible enfocar los esfuerzos en poblar el Lexicón español y otros idiomas, desarrollar el Gramaticón de modo que esta base tenga mayor capacidad inferencial, crear más guiones en el Cognicón que describan el comportamiento de las diversas sociedades y crear plantillas para el Onomasticón de modo que se pueda importar más conocimiento cultural. Sin

embargo, la condición para poder realizar todo lo expuesto es que existan los conceptos en la Ontología, puerta que da acceso a toda la información almacenada en esta base de conocimiento. Por ello, la prioridad sin duda es aumentar lo máximo posible los conceptos en la Ontología. Debido a que FunGramKB actualmente se encuentra en fase de desarrollo, no consideramos que la información ya almacenada fuera suficiente para proponerla como única fuente de información en un sistema de resolución automática de anáfora correferencial; por lo que en estos momentos recomendamos su empleo junto a otras fuentes de información complementarias.

Respecto a FunGramKB Suite, sugerimos que exista la posibilidad de realizar búsquedas no solo en la Ontología, el Lexicón, los TTFF y los MMPP de otros conceptos (MicroKnowing), sino de forma simultánea en el Cognición (MacroKnowing) y también en el Onomasticón (BioKnowing). Algo muy práctico sería un buscador dentro de esta base de conocimiento que informara de las relaciones establecidas entre varias unidades léxicas en forma de red semántica, de manera que fuera posible añadir dos o más términos a una búsqueda y así se pudiese seleccionar cuál es el camino que puede unir los conceptos a los que están ligadas las unidades léxicas como p. ej. *banco – sentarse*; *banco – financiero*; *banco – órganos*; *banco – peces*.

Este tema de la polisemia podría abordarse en una misma variedad diatópica (p. ej. *banco*), en las diferentes variedades de una misma lengua (p. ej. *fregar*) o incluso en idiomas distintos (p. ej. *subir*), siendo esto último relevante en los textos plurilingües, así como en el terreno de la correferencia trans-documental multilingüe.

5.1. Futuras investigaciones

No descartamos la posibilidad de implementar FunGramKB con objetivos didácticos para la adquisición de lenguas extranjeras. En FunGramKB se encuentra perfectamente reflejado el empleo de *hay/ser/estar*, algo que en Español como Lengua Extranjera se confunde con frecuencia por el hecho de no ser consciente de a qué evento está ligado cada verbo: *hay* a +EXIST_00, pues expresa existencia; *estar* a +BE_02, ya que con ello se localiza; por último *ser* a +BE_01 cuando se describe o, cuando se define, a +BE_00. Sería positivo comprobar qué contenido del *Plan Curricular del Instituto Cervantes* está presente en FunGramKB, siendo para ello especialmente relevantes, primero, el inventario de nociones generales y

específicas para la Ontología junto al Lexicón; segundo, el inventario de saberes y comportamientos socioculturales para el Cognición y, tercero, el inventario de referentes culturales para el Onomasticón. Por supuesto, algo que potenciaría las aplicaciones didácticas sería añadir una casilla en la cual indicar a qué nivel del *Marco Común Europeo de Referencia para las Lenguas* (i. e. A1, A2, B1, B2, C1 o C2) pertenece el contenido en FunGramKB; de manera que se pudiera conocer el porcentaje de cada nivel presente en un texto dado, opción que ayudaría a la hora de evaluar la producción escrita de un estudiante de lengua extranjera en un test de nivel por ejemplo.

Si se decide desarrollar la propuesta preliminar del algoritmo para resolver la anáfora correferencial con FunGramKB, es necesario ponerlo en práctica con un sistema automático de resolución de anáforas de manera que se pueda comprobar qué no se resolvió con éxito, plantear posibles razones y proponer soluciones para mejorar el algoritmo. Por otro lado, sería recomendable llevar a cabo investigaciones basadas en corpus de otros géneros textuales, así como tratar de incluir la lengua oral. Según se vayan poblando otras lenguas en FunGramKB será positivo realizar estudios contrastivos que permitan reconocer peculiaridades de la anáfora correferencial a tener en cuenta en cada idioma.

En el caso de que en un futuro se obtengan datos con un sistema de resolución automática de anáforas basado en FunGramKB, será recomendable medir su grado de éxito llevando a cabo una investigación experimental en la que se aplique un test para comparar esos resultados con los conseguidos por otras herramientas actuales como ARKref o dcoref en Stanford CoreNLP⁵⁹.

Para seguir avanzando en el ámbito de la resolución de la anáfora consideramos necesarias futuras investigaciones sobre las posibilidades del Micro-, Macro- y BioKnowing y, de este modo, profundizar especialmente en el papel de la herencia e inferencia respecto a la anáfora indirecta y la importancia de desambiguar las unidades léxicas polisémicas de forma adecuada para encontrar el antecedente correcto. Nos gustaría terminar expresando nuestro convencimiento de la importancia que tiene seguir investigando y desarrollando la inclusión de conocimiento extralingüístico para la resolución automática de la anáfora correferencial. Ya vimos cómo Mitkov (2002) consideraba que la resolución

⁵⁹ Más información sobre *ARKref* en <<http://www.ark.cs.cmu.edu/ARKref/>> y acerca de *dcoref* en <<http://nlp.stanford.edu/software/dcoref.shtml>>.

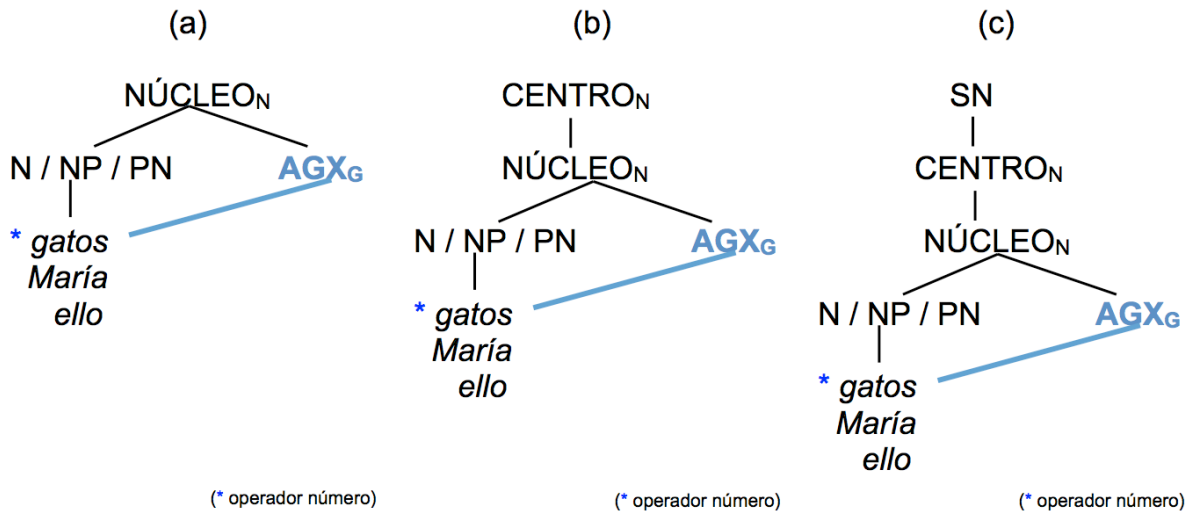
automática de la anáfora está limitada porque la representación y adquisición de conocimiento del mundo o del sentido común es una *tarea poco realista*. A lo largo de este trabajo hemos comprobado que FunGramKB sí es capaz de formalizar y obtener tal conocimiento, pero también de incluirlo y razonarlo de modo que se pueda emplear en la resolución de la anáfora; aumentando así el potencial de realizar esta tarea de forma automática. Como anunciamos en la introducción, son numerosas las aplicaciones de PLN que pueden beneficiarse de ello a la hora de traducir, resumir o clasificar textos, además de los sistemas de extracción de información, pregunta-respuesta y minería de datos.

ANEXOS

Anexo 1

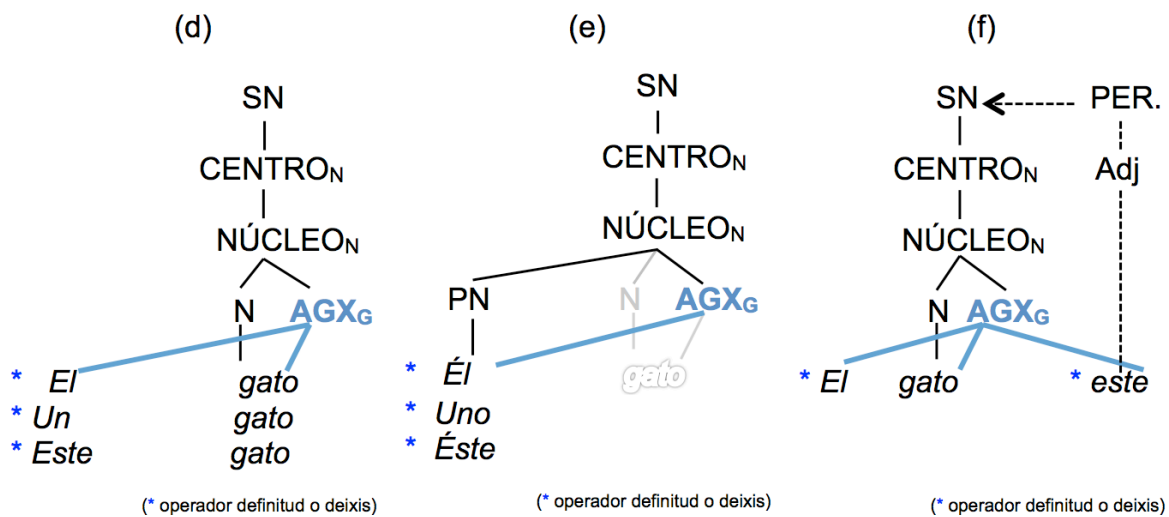
Propuesta de estructura estratificada para el SN en español siguiendo la RRG

Plantillas sintácticas nominales con nodo AGX_G en español para el SN a nivel de núcleo, centro y SN
(preliminar)



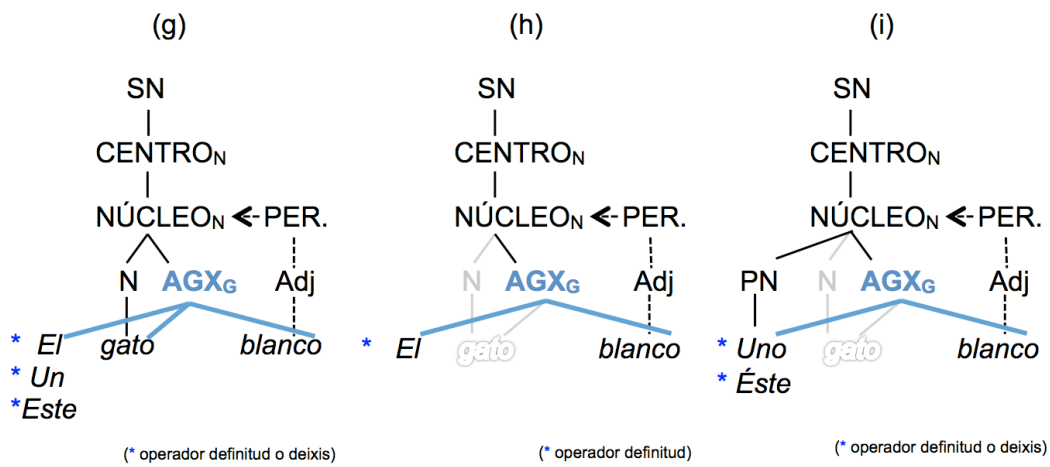
(a) a nivel de núcleo_N, (b) centro_N y (c) SN
 N = nombre común, NP = nombre propio, PN = pronombre, AGX_G = índice de concordancia de género

Plantillas sintácticas nominales con nodo AGX_G para el SN español (preliminar)



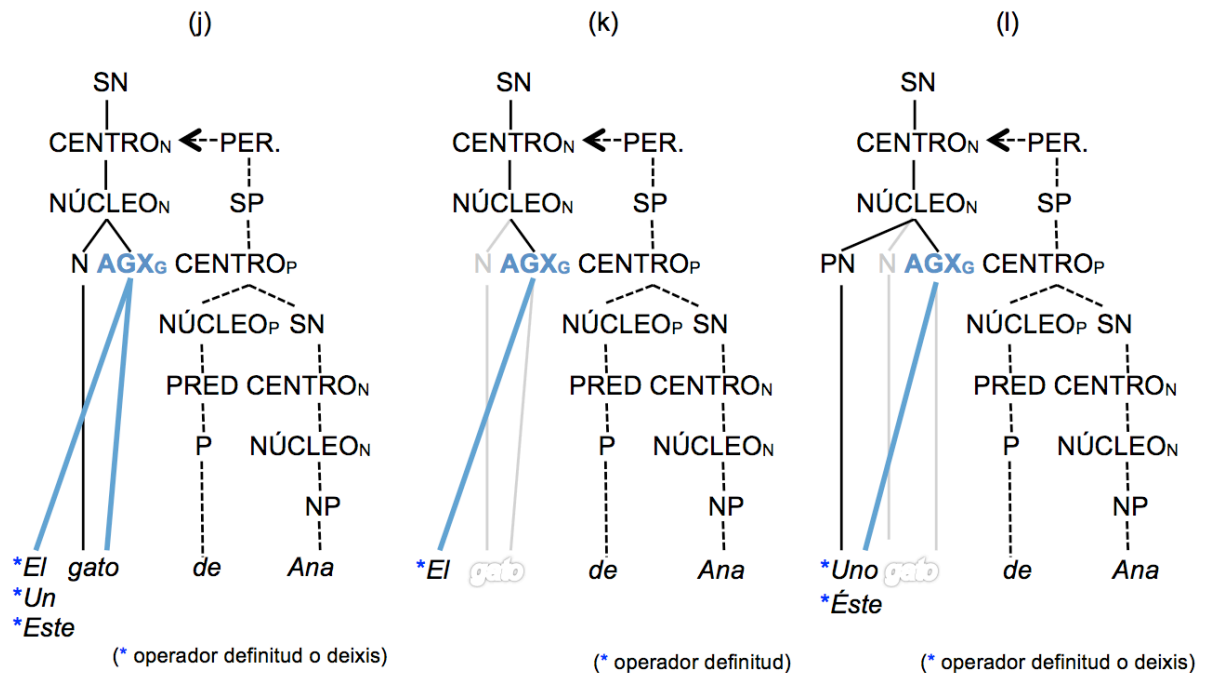
(d) SN definido, (e) SN con anáfora cero, (f) SN definido con demostrativo pospuesto
 Adj = Adjetivo, PER. = Periferia

Plantillas sintácticas nominales adjetivadas con nodo AGX_G para el SN español (preliminar)



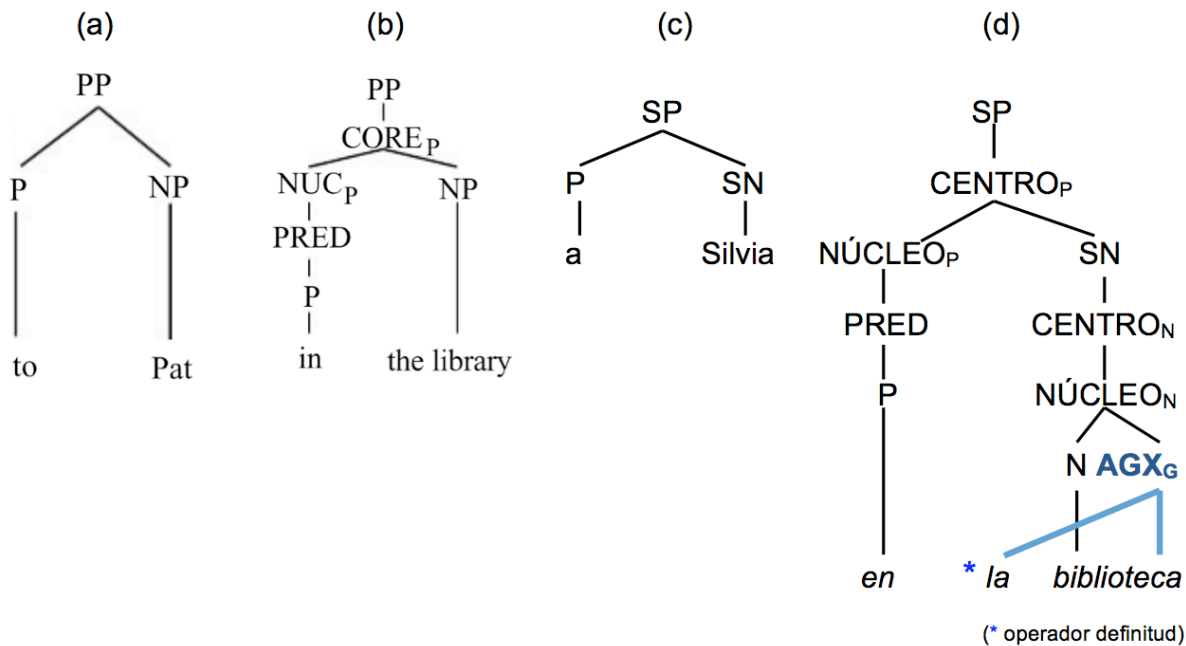
(g) SN adjetivado, (h) e (i) SN adjetivado con anáfora cero

Plantillas sintácticas nominales con SP y nodo AGX_G para el SN español (preliminar)



(i) SN con SP, (j) y (k) SN con SP y anáfora cero
SP = Sintagma Preposicional, P = Preposición

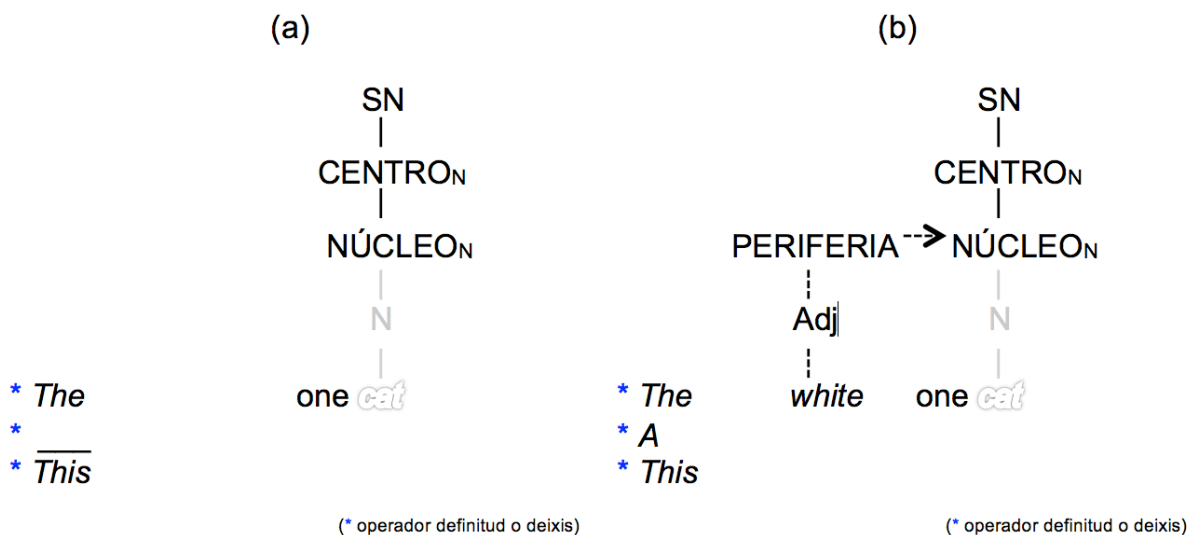
Plantillas sintácticas para el SP inglés y español (preliminar)



Tomado de y basado en Van Valin (2005: 23)

(a) SP no predicativo inglés, (b) SP predicativo inglés; (c) SP no predicativo español, (d) SP predicativo español

Plantillas sintácticas para el SN con anáfora cero en inglés (preliminar)



(a) SN con anáfora cero, (b) SN adjetivado con anáfora cero

Anexo 2

Corpus de dos artículos periodísticos junto a los comentarios de los lectores

1 **El Aura de México, adiós a la última gran conciencia.**

2 Carlos Fuentes murió ayer consolidado como una referencia de las letras hispánicas.
3 Deja en su vasta obra literaria un legado monumental.

4 Miércoles 16 de mayo de 2012
5 Yanet Aguilar Sosa | El Universal
6 Comenta la Nota
7 yanet.aguilar@eluniversal.com.mx

8 La vida de Carlos Fuentes fue agitada y productiva. No sólo era considerado uno de
9 los escritores más importantes de habla hispana y eterno candidato al Premio Nobel
10 de Literatura, sino que fue autor de una escritura ambiciosa que abarcó el cuento, la
11 novela, el ensayo y las colaboraciones periodísticas.

12 El escritor mexicano nacido en Panamá -por la movilidad de su padre, el diplomático
13 Rafael Fuentes Boettiger-, el 11 de noviembre de 1928, murió ayer en su patria,
14 México, a la que dedicó no sólo tres de sus novelas -la última La voluntad y la
15 fortuna- sino también su discurso de recepción del Premio Cervantes. Allí dijo:
16 “Comparto el Premio Cervantes, en primer lugar, con mi patria, México, patria de mi
17 sangre pero también de mi imaginación, a menudo conflictiva, a menudo
18 contradictoria, pero siempre apasionada. México es mi herencia, pero no mi
19 indiferencia”.

20 El narrador, guionista y diplomático murió a los 83 años por una hemorragia grave
21 del tubo digestivo que le provocó la pérdida del estado de conciencia unas horas
22 después de haber empezado e insuficiencia respiratoria, por lo que fue trasladado a
23 un hospital de la ciudad para ser atendido. Allí murió a las 12:15 horas, una hora
24 después de haber llegado.

25 Arturo Ballesteros, su médico desde hace 10 años, informó que “tuvo la hemorragia
26 por demás asintomática, fue súbita y le causó un malestar que lo despertó a las 5 de
27 la mañana, pero todavía tomó líquidos, se bañó y después se desmayó, fue
28 entonces que Silvia (Lemus) -su esposa- me llamó. Lo mas probable es que haya
29 sido una hemorragia por úlcera”.

30 Carlos Fuentes murió consolidado como una figura indispensable de las letras del
31 orbe, gran escritor e intelectual, voz crítica indispensable para comprender los
32 tiempos pasados, presentes y futuros. Hace apenas dos días dijo a el diario español
33 El País que no tenía ningún miedo literario y subrayó que escribía, desde el lunes
34 pasado, un nuevo libro en México.

35 “Miedos literarios no tengo ninguno. Siempre he sabido muy bien lo que quiero hacer
36 y me levanto y lo hago. Me levanto por la mañana y a las siete y ocho estoy
37 escribiendo. Ya tengo mis notas y ya empiezo. Así que entre mis libros, mi mujer,
38 mis amigos y mis amores, ya tengo bastantes razones para seguir viviendo”, dijo.

39 A Carlos Fuentes se le fue la vida dedicado a la literatura. Demostró que se podía
40 ser escritor de tiempo completo en un México en el que los que escribían hacían
41 miles de suertes más. Hablaba más de la vida que de la muerte; su narrativa habla

42 por él, en Gringo viejo dijo: “Nadie me verá decrepito. Siempre seré joven porque
43 hoy me atrevo a volver a ser joven. Siempre seré recordado como fui”.

44 **... Y el novelista nació**

45 Carlos Fuentes, el hombre gallardo y galante llamado por muchos dandy guerrillero y
46 por otros un hombre que vivía entre el cosmopolitismo gringo y la mexicanidad,
47 siempre se enorgulleció de estar rodeado de muchos protagonistas de la “República
48 de las Letras”; entre sus amigos estaban Carlos Monsiváis, Octavio Paz y José
49 Emilio Pacheco, pero también Sergio Ramírez, Gabriel García Márquez y Pablo
50 Neruda. Sin embargo, no se quedó allí, el autor fue mentor de generaciones más
51 jóvenes: Jorge Volpi, Juan Villoro, Cristina Rivera Garza y Juan Gabriel Vásquez.

52 Su orgullo también estaba en el fondo, en cómo confirmó que debía ser escritor, la
53 anécdota la contaba a diestra y siniestra, así lo hizo en la UNAM durante un coloquio
54 realizado en 2008 para analizar su obra y celebrar su vida:

55 “Una noche de verano conocí a Thomas Mann en un café. Era un hombre alto,
56 elegante y sabía manejar a la perfección los cubiertos con los que comía. En ese
57 instante se me reveló un deseo, y dije: ‘Si no puedo ser él, ni escribir como él, al
58 menos quiero hacer lo que él’, y al regresar a México con esa idea sentí claramente
59 que esta ciudad me dijo al oído: ‘Carlos, aquí te tocó, qué le vas a hacer, en la
60 región más transparente’. Y el novelista nació”, afirmó ante admiradores y
61 estudiosos de su obra.

62 Allí mismo rememoró cómo aprendió literatura “en las rodillas de Alfonso Reyes” y
63 cómo lo hizo también a través de las enseñanzas de sus de sus maestros y amigos.
64 Ese día, recordó que en Santiago de Chile, en la gaceta del Instituto Chileno,
65 escribió su primera novela: Estampas mexicanas “muy patrioter, pero eso sí, llena
66 de las mejores intenciones”.

67 El autor de obras emblemáticas como La muerte de Artemio Cruz, Cristóbal Nonato,
68 Aura y Las buenas conciencias entró con paso firme al concierto de las letras, en
69 1958, con una novela emblemática La región más transparente, que se insertó en
70 las obras del llamado “boom latinoamericano”, a partir de allí la vida le sonrió, tuvo
71 voluntad y tuvo fortuna.

72 A pesar de la tragedia que circundó a su familia, sus dos hijos con Silvia Lemus:
73 Carlos y Natasha, murieron jóvenes en situaciones dolorosas -de ellos habló en su
74 libro más autobiográfico En esto creo-, Carlos Fuentes y su esposa salieron avante,
75 allí señala que compartieron “la alegría de tener hijos. La pena de perderlos”.

76 **Nuestro Virgilio**

77 Fuentes es sin duda figura clave de las letras en español, prolífico hacedor de
78 tramas e historias que se han traducido a más de 20 idiomas como francés, inglés,
79 polaco, chino, italiano y portugués; muchas de esas traducciones las donó a la
80 Biblioteca Vasconcelos, su colección de revistas las entregó a la Biblioteca Nacional
81 y su archivo personal lo vendió a la Universidad de Princeton.

82 De ese enamorado de las mujeres, amante fiel de Silvia Lemus, a quien denomino
83 “la galaxia misma”, padre de Cecilia, la hija que le sobrevive y procreó con la actriz
84 Rita Macedo, escritores, editores y amigos celebraron su vitalidad. Jorge Volpi lo

85 llamó “nuestro Virgilio” y en un artículo en El País dijo que es un “crítico tan agudo
86 como feroz, tan profundo como descarnado. Un guía generoso”. Cristina Rivera-
87 Garza asegura: “Fuentes nos enseñó a pensar críticamente la posición de México en
88 la historia y en el contexto de este mundo contemporáneo”.

89 Distintas instituciones culturales y educativas expresaron su pesar por la pérdida del
90 novelista mexicano.

91 Del ganador del Premio Rómulo Gallegos y del Príncipe de Asturias queda su
92 literatura y los libros póstumos que publicará Alfaguara. Marcela González, su
93 editora, dijo que la próxima semana saldrá Personas donde perfila retratos de líderes
94 latinoamericanos, y para noviembre esperan la salida de la novela Federico en su
95 balcón; contó también que Fuentes hacía investigaciones para su libro El baile del
96 Centenario.

97 Quedan pendientes los festejos por el 50 aniversario de Aura y La muerte de
98 Artemio Cruz que se realizarán en junio, y la entrega del tomo cuatro de Obras
99 Reunidas que publica el FCE.

100 Compartir

Comenta la Nota

1	<p style="text-align: center;">EmMa NV · Comentarista destacado</p> <p>Más allá de un homenaje, deseaba que hubiese un México diferente en los próximos días, su deseo puede ser cumplido y sabemos cual es, y ese deseo depende del millón de personas que como sociedad andamos día a día, no permitamos que un pobre diablo llegue al poder, hagase SU DESEO REALIDAD del Maestro Fuentes!</p> <p style="text-align: right;">Responder · 2 Me gusta · El 16 de mayo a la(s) 6:58</p>
2	<p style="text-align: center;">Bernardo Morales · Comentarista destacado · Abogado Litigante en Freelancer</p> <p>Dueños de la noche, porque en ella soñamos... eternamente Don Carlos.</p> <p style="text-align: right;">Responder · 2 Me gusta · El 16 de mayo a la(s) 2:25</p>
3	<p style="text-align: center;">Albenis Andrade</p> <p>paz a uno de los últimos grandes de la literatura latinoamericana</p> <p style="text-align: right;">Responder · Me gusta · El 16 de mayo a la(s) 2:51</p>
4	<p style="text-align: center;">Alma Escarcega Pedregon · Juárez, Chihuahua, Mexico</p> <p>Mente brillante!</p> <p style="text-align: right;">Responder · Me gusta · El 16 de mayo a la(s) 5:46</p>
5	<p style="text-align: center;">Jesus Gamiño Garcia · Comentarista destacado · UNiversidad de Guadalajara; GUadalajara Jaisco</p> <p>CARLOS: Es la Noche del Comienzo, duerme la nada desnuda, más ya el intento del buda, será el devenr inmenso, por cada micra de ensueño, miriámetros de amrgura; eras el verbo, la palabra encaranada y nos regalabas esa esperanza en que un día todo será diferente y mejor; aún cuando en este pais sus conductores se empeñen en hacer lo contrario, adíos amigo.</p> <p style="text-align: right;">Responder · Me gusta · El 16 de mayo a la(s) 7:55</p>
6	<p style="text-align: center;">Alex Draven · Comentarista destacado</p> <p>Luto para las letras latinoamericanas, un grande mas que se ausenta...</p> <p style="text-align: right;">Responder · Me gusta · El 16 de mayo a la(s) 8:56</p>

6 comentarios

Carlos Fuentes, Mexican Man of Letters, Dies at 83

By ANTHONY DePALMA

Published: May 15, 2012

Carlos Fuentes, Mexico's elegant public intellectual and grand man of letters, whose panoramic novels captured the complicated essence of his country's history for readers around the world, died on Tuesday in Mexico City. He was 83.

His death was confirmed by Julio Ortega, his biographer and a professor of Hispanic studies at Brown University, where Mr. Fuentes taught for several years. He died at the Angeles del Pedregal hospital after his doctor, Arturo Ballesteros, found him in shock in his Mexico City home, The Associated Press reported. The doctor told reporters that Mr. Fuentes had had an internal hemorrhage.

Mr. Fuentes was one of the most admired writers in the Spanish-speaking world, a catalyst, along with Gabriel García Márquez, Mario Vargas Llosa and Julio Cortázar, of the explosion of Latin American literature in the 1960s and '70s, known as El Boom. He wrote plays, short stories, political nonfiction and novels, many of them chronicles of tangled love.

Mr. Fuentes received wide recognition in the United States in 1985 with his novel "The Old Gringo," a convoluted tale about the American writer Ambrose Bierce, who disappeared during the Mexican Revolution. It was the first book by a Mexican novelist to become a best seller north of the border, and it was made into a 1989 film starring Gregory Peck and Jane Fonda.

In the tradition of Latin American writers, Mr. Fuentes was politically engaged, writing magazine, newspaper and journal articles that criticized the Mexican government during the long period of sometimes repressive single-party rule that ended in 2000 with the election of an opposition candidate, Vicente Fox Quesada.

Mr. Fuentes was more ideological than political. He tended to embrace justice and basic human rights regardless of political labels. He supported Fidel Castro's revolution in Cuba, but turned against it as Mr. Castro became increasingly authoritarian. He sympathized with Indian rebels in the southern Mexican state of Chiapas and skewered the administration of George W. Bush over its antiterrorism tactics and immigration policies, calling them unduly harsh.

He was also critical of Venezuela's leftist leader, Hugo Chávez, however, calling him a "tropical Mussolini," and of his own country's failure to stem its rampant drug violence. On the day he died the newspaper Reforma published a hopeful essay by him on the change of power in France.

Mr. Fuentes was appointed the Mexican ambassador to France in 1975, but he resigned two years later to protest the appointment of Gustavo Díaz Ordaz as ambassador to Spain. Mr. Díaz Ordaz had been president of Mexico in 1968 when Mexican troops opened fire on student protesters in Mexico City.

142 But it was mainly through his literature, Mr. Fuentes believed, that he could make his
143 voice heard, and he did so prolifically and inventively, tracing the history of modern
144 Mexico in layered stories that also explored universal themes of love, memory and
145 death. In “The Death of Artemio Cruz,” a 1962 novel that many call his masterpiece,
146 his title character, an ailing newspaper baron confined to his bed, looks back at his
147 climb out of poverty and his heroic exploits in the Mexican Revolution, concluding
148 that it had failed in its promise of a more egalitarian society.

149 His novels remained ambitious and topical. His last, “Destiny and Desire” (2011), is a
150 sprawling work that Michael Wood, writing in The New York Times Book Review,
151 described as “not exactly a parody of ‘War and Peace,’ but certainly a spectral,
152 playful revision of the idea of a novel that competes with history.”

153 He added, “It offers lavish quantities of comedy, satire, allegory, fantasy and brilliant
154 political commentary; makes coded allusions to recognizable celebrities like the
155 communications magnate Carlos Slim; evokes the work of Spinoza and Machiavelli;
156 includes ghosts, graves, murders, a voluble flying prophet and a talking severed
157 head.”

158 The severed head had fallen victim to Mexico’s drug-gang wars, which Mr. Fuentes
159 believed pose an ever-graver threat to Mexican society. The head speaks in darkly
160 comic tones.

161 “I speak of my body because I’ve lost it,” the character says, then adds: “I am a 27-
162 year-old man, one meter seventy-eight centimeters tall. Every morning I look at
163 myself naked in my bathroom mirror and caress my cheeks in anticipation of the daily
164 ceremony: Shave my beard and upper lip, provoke a strong response with Jean-
165 Marie Farina cologne on my face, resign myself to combing black, thick, untamable
166 hair. Close my eyes. Deny to my face and head the central role my death will be
167 certain to give them. Concentrate instead on my body. The trunk that is going to be
168 separated from my head. The body that occupies me from my neck to my
169 extremities, covered in skin the color of pale cinnamon and tipped with nails that will
170 continue to grow for hours and days after death, as if they wanted to scratch at the lid
171 of the coffin and shout I’m here, I’m still alive, you made a mistake when you buried
172 me.”

173

174 Though Mr. Fuentes wrote in just about every genre, including opera (a 2008 work
175 inspired by the life of Gen. Antonio López de Santa Anna, the wooden-legged
176 president of Mexico during the mid 19th-century), he declined to write an
177 autobiography.

178 “One puts off the biography like you put off death,” he once said. “To write an
179 autobiography is to etch the words on your own gravestone.”

180 Carlos Fuentes was born on Nov. 11, 1928, in Panama, the son of Berta Macías and
181 Rafael Fuentes, a member of Mexico’s diplomatic corps. As his father moved among
182 Mexican embassies, Mr. Fuentes spent his early childhood in several South
183 American countries. Then, in 1936, the family was transferred to Washington, where
184 Mr. Fuentes learned to speak English fluently while enrolled in a public school.

185 In 1940 the family was transferred again, this time to Santiago, Chile, where he
186 began to experiment with writing. In an interview with *The Times* in 1985, Mr.
187 Fuentes said he first had to decide “whether to write in the language of my father or
188 the language of my teachers.” He chose Spanish, he said, because he believed that
189 it offered more flexibility than English. There was also a practical reason. English, he
190 said, “with a long and uninterrupted literary tradition, did not need one more writer.”

191 He was 16 when his family finally returned to Mexico. He knew his homeland through
192 the stories that his grandmothers had told during the summers he spent with them.

193 “I think I became a writer because I heard those stories,” he said in 2006 in an
194 interview with the Academy of Achievement, a nonprofit organization in Washington.
195 His grandmothers fascinated him with their tales of bandits, revolution and reckless
196 love. “They had the whole storehouse of the past in their heads and hearts,” Mr.
197 Fuentes said. “So this was, for me, very fascinating, this relationship with my two
198 grannies — the two authors of my books, really.”

199 When he told his family that he wanted to be a writer, his father was encouraging, but
200 insisted that he also study law, which he did in Mexico and Switzerland.

201 After completing his degree, Mr. Fuentes entered Mexico’s diplomatic service, while
202 also carving out time for his fiction. His first novel, “Where the Air Is Clear,” was
203 published in 1958 when he turned 30. It was a literary sensation, mixing biting social
204 commentary with interior monologues and portrayals of the subconscious. His
205 reputation established, Mr. Fuentes left government service to devote all his energies
206 to writing.

207 As an author, he said, he did not spend much time rewriting and never suffered from
208 writer’s block. He liked to write on the right-hand pages of lined notebooks, making
209 changes and corrections on the left-hand pages before sending a manuscript to be
210 typed.

211 Professor Ortega called Mr. Fuentes “an unleashed cultural force” who avoided some
212 of the trappings of literary celebrity. In a retrospective book that he wrote about Mr.
213 Fuentes’s life when the writer turned 80 in 2008, Mr. Ortega wrote, “Fuentes detests
214 the literary life, its obligations and commitments.”

215 “He hasn’t created his own group, and he belongs neither to parties nor ideologies,”
216 Mr. Ortega added. “He isn’t controlled by either the power of the state nor the power
217 of the market.”

218 Mr. Fuentes’s independent thought and reputation for supporting leftist causes led to
219 his being denied visas to enter the United States in the early 1960s. When he was
220 refused permission to come to New York in 1963 for a presentation of an English
221 translation of one of his books, he reacted angrily, saying, “The real bombs are my
222 books, not me.”

223 Congress intervened in 1967, and the restrictions against him were lifted. Later he
224 traveled to the United States frequently, teaching at several Ivy League universities.

225 Mr. Fuentes is survived by his wife, Silvia Lemus, and a daughter, Cecilia, by a
226 previous marriage to the actress Rita Macedo, who died in 1993. Two children from

227 his marriage to Ms. Lemus, Carlos and Natasha, both died of illness before they were
228 30.

229 For much of his career Mr. Fuentes competed for recognition and influence in Mexico
230 and abroad with another titan of Mexican letters, the poet Octavio Paz. Mr. Fuentes
231 received the National Order of Merit, France's highest civilian award given to a
232 foreigner; Spain's Prince of Asturias Award for literature in 1994; and, in 1987, the
233 Cervantes Prize, the Spanish-speaking world's highest literary honor. Mr. Paz,
234 however, won the Nobel Prize in Literature in 1990. Mr. Fuentes, a perennial on the
235 shortlist for the honor, never did.

236 The two became friends in 1950, when Mr. Paz published his landmark work on
237 Mexican identity, "The Labyrinth of Solitude." They worked together on several
238 literary projects. But by the mid-1980s their political opinions had started to differ. Mr.
239 Fuentes supported the Sandinistas, the leftist rebel group in Nicaragua, but Mr. Paz,
240 who had more conservative views, condemned them. Then, in 1988, the literary
241 magazine *Vuelta*, which Mr. Paz directed, published an article fiercely critical of Mr.
242 Fuentes, accusing him of lacking true Mexican identity. That set off an often public
243 feud that lasted until Mr. Paz died in 1998. Neither man apologized, diminishing the
244 reputations of both.

245 Still, in his later years, Mr. Fuentes became an elder statesman of international
246 letters. On his 80th birthday hundreds gathered at the Metropolitan Museum of Art in
247 New York to celebrate his life and work. He was introduced by Rubén Beltrán, the
248 consul general of Mexico in New York at the time.

249 "To speak about Carlos Fuentes is to engage inexorably in Mexican history and
250 culture," Mr. Beltrán said. "We cannot fathom a debate on Mexican literary and
251 humanistic traditions in which his name and work are absent."

7	<p style="text-align: center;">Bonehead Lancaster, Pa</p> <p>Mexico and the world has lost a great writer, the soul of Mexico, You will be missed but your writing will live on forever.</p> <p style="text-align: right;">May 17, 2012 at 7:21 a.m. Recommended5</p>
8	<p style="text-align: center;">JF Singapore</p> <p>Reading "The Death of Artemio Cruz" and "The Old Gringo" in my mid-80s Latin American Literature course in prep school had a huge impact on how I have viewed the world ever since. While I didn;t always agree with Cruz's political choices, I always respected the fact that he rarely let ideology get in the way of honest, principled argument. Carlos Fuentes is a man to be emulated.</p> <p style="text-align: right;">May 17, 2012 at 3:27 a.m. Recommended3</p>
9	<p style="text-align: center;">JF Singapore</p> <p>Sorry. Oops. Not Cruz, Fuentes. That's what happens when one writes a comment immediately after waking.</p> <p style="text-align: right;">May 17, 2012 at 1:51 p.m. Recommended1</p>
10	<p style="text-align: center;">Bob Horan Menomonie, WI</p> <p>As a reader, I will always be grateful for Fuentes' books. However, I got acquainted with Carlos when I studied creative writing with him at Pennsylvania 1979-81, and I will miss him as a man. It's hard to imagine meeting him and not being struck by his extraordinary vitality; he was a man who never yawned. Carlos was a generous teacher—offering encouragement every class period and professional help when he could. He read student work carefully and commented on it with a startling hopefulness about what the student work wanted to do. His remarks on student work and our reading took his classes on a whirlwind tour of world literature from Homer to Joyce Carol Oates. For Carlos, literature was a way to investigate and know the world—it wasn't ancillary to other, more "serious" methods. To make this point, he taught us to read his way—emphasizing both structure and form and the great seriousness about the writer and reader's shared adventure. Carlos was a physically striking man, and deeply serious and playful. My own life has been different, better, because I knew him; I have tried to pay forward what he gave me. I join his readers and friends in mourning the end of Carlos Fuentes' life and wishing him safe travels to the</p>

next world, where I hope he is reunited with all his loved ones who have preceded him, gets all the time he'd like with Cervantes, and doesn't have to wait too long for a long chat with God. To Silvia, I send my deepest sympathies and my heartfelt gratitude.

May 17, 2012 at 3:27 a.m.
Recommended6

11

alberto angli
Camepche, Mexico

beyond the sadness that is among us for his great books, the lost in this moment in elections times is a big shock for people that don't want PRI to come back to the government because he was one of the main opositors

May 17, 2012 at 12:11 a.m.
Recommended3

12

Hugo
San Diego, CA

Ever since I was little, I used to hear his name being mentioned as one of the great Mexican writers of our time. Many years later I read Aura which left me spell bound. I had the priviledge of listening to a lecture at UCSD a few years ago. He had a book signing. As my turn was up he gave me a strong hand shake, had a picture taken with him and he kindly signed La Region Mas Transparente-Where the Air is Clear. His voice has something of the friend you wish you had, the philosopher you wish you could talk to, or the inquisitive mind that never lets go and takes you to higher states of comprehension without understanding. To read Carlos Fuentes is to have a conversation with him and all the authors he read. I will read him often. Descanse En Paz.

May 16, 2012 at 10:52 p.m.
Recommended10

13

Dane Walters
Dallas, Texas

You can see a video interview from 1992 where Fuentes is interviewed by Lee Cullum for KERA's Conversations series here:

<http://artandseek.net/2012/05/16/video-a-conversation-with-carlos-fuentes/>

Vigorous and animated, Fuentes talks about growing up in Washington D.C., connecting with his Mexican roots in a movie theater, realizing how that identity alienated him from his classmates in the U.S., and falling for the Spanish language. He also touches on Mexico's culture of Catholicism and obsession with death, and the role of myth and dreams in his work. Enjoy!

May 16, 2012 at 10:52 p.m.
Recommended5

14

Guillermo Peimbert
Mexico City

	<p>He gave sense to what is to be a cosmopolitan writer. We will missed his books, articles and political opinions, in a time where Mexico is among a major dichotomy, between returning to the past or continue in a path of new challenges. .</p> <p style="text-align: right;">May 16, 2012 at 10:50 p.m. Recommended3</p>
15	<p style="text-align: center;">Laura Iriarte Florida</p> <p>I really feel so sad for this... he is one of my favorite writers and he was a marvellous human . Descanse en Paz.</p> <p style="text-align: right;">May 16, 2012 at 10:50 p.m. Recommended2</p>
16	<p style="text-align: center;">Alfred D. Los Angeles, California</p> <p>I majored in History at UC Santa Barbara. My emphasis of study was Mexican History. "Aura" & "The Death of Artemio Cruz" changed me forever. Fuentes's words looked to the past as a means to decipher Mexico's future. Sometimes the correlation with the past and future had an atmosphere of darkness.....but Fuentes' characters were portrayed in a way that let them radiate from this darkness. Sometimes the characters flaws and perfections were one in the same. He left it up to us to decide the moral meaning. We will always have his work as way of reminding us that we still have hope among the chaos. Descanse en Paz, Nuestro Amigo.</p> <p style="text-align: right;">May 16, 2012 at 10:49 p.m. Recommended5</p>
17	<p style="text-align: center;">EM Tempe,AZ</p> <p>Carlos Fuentes was a great ambassador of literature to the world at large, and a magnificent writer and human being. Que descanse en paz, y muchisimas gracias por haber compartido su espiritu con nosotros.</p> <p style="text-align: right;">May 16, 2012 at 10:49 p.m. Recommended4</p>
18	<p style="text-align: center;">Paul Dresman Eugene, Oregon</p> <p>Fuentes was a native of Xalapa, the small city-capital of the Mexican state of Veracruz. When we visited that city a few years ago, we discovered that the library had been built by Fuentes. It is a wonderful building with glass walls, a good collection, especially of literature, and a coffee shop on the rooftop that looks out over the city. It was his gift to his native place, just as his inspired fiction, essays and histories were gifts to the world at large.</p> <p>There were tears in my Chilean friend's voice when he called yesterday with the news of Fuentes' death. An entire hemisphere mourns this passing.</p>

		May 16, 2012 at 6:50 p.m. Recommended14
19	<p style="text-align: center;">Ashley Chaney New York</p> <p>Que Descanse en Paz. I am deeply saddened by his passing. My love for Latin American Literature, Mexico, and short stories was brought on by Constanca and Other Stories for Virgins. Since that moment, I have loved a man without ever meeting him.</p>	May 16, 2012 at 6:50 p.m. Recommended8
20	<p style="text-align: center;">Miguelito Mexico City</p> <p>Beautiful thought. Thank you.</p>	May 17, 2012 at 1:51 p.m.
21	<p style="text-align: center;">Atakan Izmir, Turkey</p> <p>A great man of letters, a real intellectual passed away and as one of his devoted readers I felt deeply sad. "The Death of Artemio Cruz" is among the five best novels I have ever read. I have taken immense pleasure from reading "Cambio del Piel". Fuentes was one of the most inspiring and exciting literary men of our times. I have always regretted not being able to read him in Spanish. However, I could get deep esthetic joy even from Turkish translations, because he has written perfect books. A distinguished member of "the invisible college" of humanity which extends beyond national and mental borders and enriches the soul of humanity in a humble, slow yet powerful way, Senor Fuentes will be missed much across the world. Huzur icinde yat ustad!</p>	May 16, 2012 at 6:50 p.m. Recommended8
22	<p style="text-align: center;">Myriam Yvonne United Kingdom</p> <p>A lovely obituary to a most-gifted and elegant writer as well as a beautiful human being. You will be missed, dear Carlos. World literature has lost a brilliant author in your passing. We will not see the likes of you again....a sad farewell and a heart-felt thanks for what you gave us all.</p>	May 16, 2012 at 6:50 p.m. Recommended7
23	<p style="text-align: center;">Rachelle Chinnery Hornby Island</p> <p>Terra Nostra inspired me to learn Spanish. Carlos Fuentes led me to Mexico, he changed how I read and why I read. A brave and creative voice</p>	

now watches from another place. You will be loved forever. Que en paz descanse.

May 16, 2012 at 6:50 p.m.
Recommended5

24

David Ramirez-Moreno

Cali-Colombia

Carlos Fuentes is a powerful presence on the world intelectual arena. We'll miss him. His thoughts and ideas will be around for decades, even centuries. Descanse en paz Sr. Fuentes.

May 16, 2012 at 6:50 p.m.
Recommended4

25

ATOM

NY, NY

I feel a profound sadness. When reading La Muerte de Artemio Cruz for the first time, I realized that even though my parents were Latin American immigrants and I am fluent in Spanish, that I knew nothing about identity, culture, politics, and history of Latin America. That book changed my life and made me who I am today.

Carlos, sus obras perduraran por una eternidad y seran una gran inspiracion para muchas generaciones. [Your works will survive and eternity and will be a great inspiration for many generations.]

May 16, 2012 at 6:49 p.m.
Recommended7

26

LM

Germany

Q.E.P.D.

May 16, 2012 at 6:23 p.m.
Recommended3

27

VdO

NY - Paris

Carlos Fuentes will remain famous for centuries for his masterpieces and by the proverb he created: "Sex without sin is like an egg without salt". Just for this proverb, he should have been awarded the Nobel Prize in Literature

May 16, 2012 at 4:20 p.m.
Recommended8

28

Maggie

New Orleans

I have no words to express my sadness at the passing of this great and generous soul. I had the great pleasure of meeting him but once. It was the dream of my life. He was not only the epitome of true class and elegance, but

a brilliant mind and kind human being. I have treasured his words for many years. His books have challenged and enlightened all thinking people. I write this with tears streaming down my face because of the great loss. I have had his name on my google news page for a long time, so that I could track his brilliant and dogged pursuits of truth and justice in the cause of his country and the world.

Now, his voice has been silenced, but hopefully his words will be heard for many generations. Adios, Senor Fuentes.

May 16, 2012 at 4:20 p.m.
Recommended7

29

Ted Morgan
Baton Rouge

I do not speak Spanish nor do I read it well, but I deeply admire Spanish language intellectuals including Unamuno, Borges, Neruda, Paz, and Fuentes. I especially appreciate writers who voice great depth, wisdom, balance, and fairness. The range of outlooks from these writes is great.

Mr. Fuentes is one whom I have not read as fully as some of those I mention but that is my loss—one that maybe I have time to mend. Such people leave us very much in their debt because they enrich us individually and collectively.

I very much appreciate the quality of this obituary. Thank you.

May 16, 2012 at 4:20 p.m.
Recommended13

30

Luca Formenton
Milan, Italy

Carlos Fuentes, a long time friend, will be missed by anyone who cares of Literature and Culture in their best meaning. I'd like to quote a statement by him regarding the mission of a writer "You have to create readers, not just give them what they want".

My deep sympathy and affection to Silvia, remembering our happy times here in Italy

May 16, 2012 at 4:19 p.m.
Recommended7

31

S.A.Gonzalez
Los Angeles

JJ you must read and finish Terra Nostra. Truly, i tell you, it is a breathtaking marvel.

You will be immeasurably missed, maestro.

I send you all of my love and admiration.

May 16, 2012 at 4:19 p.m.
Recommended3

32

JJ
atlanta

Dear S.A.

After many previous attempts and false strats, to celebrate the man and mourn his passing, I will certainly read and complete Terre Nostra for Carlos and for me.....

JJ

May 16, 2012 at 10:49 p.m.
Recommended2

33

el chinche
Libra de Congress

he will continue to communicate ...

May 16, 2012 at 4:19 p.m.
Recommended2

34

ProfWombat
Andover MA

To all of the above I'd add my thanks to Sr Fuentes for 'The Buried Mirror', which taught me more about Latin America, Spain and the world than I've seen anywhere else

May 16, 2012 at 4:19 p.m.
Recommended5

35

Peter Samis
Berkeley, CA

What a brilliant Renaissance man, virtuoso writer, inspiring and kind teacher. His sweeping vision encompassing both cyclic and linear time inspired me as no other; his novel "The Death of Artemio Cruz" a precision-wrought jewel perfectly merging form and content, Citizen Kane and Faulkner fused in a top-to-bottom x-ray of Mexican society and history in the first half of the twentieth century, lit by lights cosmic and visceral; his love of Proust, Tristram Shandy, and Cervantes; of language; of women in all their forms from Aura to Malinche to the wizened crones who have seen and known life's circus from inside out, who have lived to bear fruit and watch its corrupting cycles of decay (were these perhaps inspired by his grandmothers, the ones whom he called "the two authors of my books, really"?). He was generous, supportive, a force of nature, vital source (as his name connotes), a gift to us all. I salute you, Fuentes, my one-time friend and mentor and perpetual inspiration! And my sympathy to Silvia, who knew him as no other.

May 16, 2012 at 7:11 a.m.
Recommended24

36

jessiekitty

Evanston, IL

Carlos Fuentes lived across the street from my family for a year in Princeton. I remember him as a very elegant, serious person, but also a nice neighbor. My sister spent a lot of time with his children (at the request of Sr. Fuentes) so they would learn to speak English a bit better. Of course, it is his remarkable writing which will continue his memory.

May 16, 2012 at 7:11 a.m.
Recommended11

37

Sleater
Chicago

Like several posters here, I took an undergraduate class with Fuentes, in which he delivered magisterial lectures, and at that time he represented for me what a man of letters, an international novelist, an engaged writer, was supposed to be. I vividly remember reading and finishing *Where the Air is Clear,* which he published just at the end of his 20s, and being entranced by the even better *The Death of Artemio Cruz,* which led me to study carefully how he constructed the narrative. In that novel as in all great works of literature, a world comes into being through language. Carlos Fuentes was a great writer, perhaps not as well appreciated as he should have been, but I always attributed that to his not living in Mexico full time. I, like many, will miss his voice and presence in the republic of letters. RIP, Carlos Fuentes.

May 16, 2012 at 7:11 a.m.
Recommended7

38

rainyday
earth

A pillar of civilization has fallen.

May 16, 2012 at 7:11 a.m.
Recommended7

39

Browniana
CS, TX

It is a shame.. Mexico lost a great man. Just 45 days before the Mexican election... He will be missed!!

May 16, 2012 at 7:10 a.m.
Recommended5

40

JJ
atlanta

WOW..NO...NO...NO... this really really hurts.... have followed, been obsessed with and dreamed about Carlos for over 40 years... and over this past year I have been trying to read Terra Nostra but couldn't get through its density... I will try harder now.... the book has been on by bed for years...

One day in the mid-80s while walking down nassau street in Princeton I

ran into him on his way home.... he was carrying two bags of groceries... I was probably the only one to recognized... I offered and he allowed me to help him with the bags... when we reached his house we sat on the steps and talked for a while... mainly about Mexico where I had lived for three years and about Harvard (which he was very interedted in... after princeton he went to teach at Harvard the following term)...

I never forgot that day in princeton and I never forget him...

May 16, 2012 at 5:52 a.m.
Recommended7

41

Hartmut Wiesner
Peru

I just had the great opportunity to see and listen to Carlos Fuentes in the Bookfair in Buenos Aires not even 10 days ago. I thought he looked so handsome and in good shape. How cruel can nature be to take away someone who is offering us so much insight into where Latin America is and where it should be going.

May 16, 2012 at 5:22 a.m.
Recommended10

42

Dario Gonzalez
Washington DC

That is very true, especially as Mexico comes close to elections. We needed his voice during this tough years.

May 16, 2012 at 7:10 a.m.
Recommended4

43

OE
Boulder, CO

One of the most important moments of my education at University of Pennsylvania was taking a graduate seminar with Carlos Fuentes (this, despite my being an undergraduate and a biology major at that!). He had such a depth of understanding about literature, and such passion. He shaped forever how I would read and understand literature. A great writer, and a great teacher.

May 16, 2012 at 5:22 a.m.
Recommended7

44

frugalfish
rio de janeiro

Yet another great author who should have won the Nobel, but didn't, like Jorge Amado or John Updike and Philip Roth, who were, usually, too true to their native culture to attract the votes of the Nobelians. The last Nobel Prize winner for literature who really deserved it was José Saramago, another Latino.

May 16, 2012 at 5:21 a.m.

	Recommended6
45	<p style="text-align: center;">Old Gringo San Luis Obispo, California</p> <p>Carlos, I've taught Latin American Culture for 11 years at a State University, in Chile, Mexico, Peru, Spain and the U.S. Finding your "Buried Mirror" offered me an insight that is personal and political, patriotic and global, sincere and sensitive. With your spirit at my side, I share with students more than a balanced view. Your inclusive and cross-cultural perspectives offer us a dawn to dusk sun to see ourselves and the Other. Beyond borders then, beyond death now, Carlitos, your words continue to guide us look long, reflect deep and share love. Eres un espejo eterno. An eternal mirror.</p> <p style="text-align: right;">May 16, 2012 at 4:24 a.m. Recommended17</p>
46	<p style="text-align: center;">Helena Dallas, Texas</p> <p>Rest in peace, maestro Carlos Fuentes. Your mastery of letters, your nuanced insight, and your grand chronicles of and about Mexico--of hopelessness sometimes, of hopefulness on others--will always inform my conscience of the native, childhood land I left behind so long ago.</p> <p style="text-align: right;">May 16, 2012 at 4:14 a.m. Recommended12</p>
47	<p style="text-align: center;">Ted Morgan Baton Rouge</p> <p>Along with Octavio Paz, a great man and a fine writer. He blessed us with his life and work.</p> <p style="text-align: right;">May 16, 2012 at 3:52 a.m. Recommended6</p>
48	<p style="text-align: center;">David Keys Las Cruces, NM</p> <p>Fuentes had the courage to produce prose that was truly artful and abstract. He trusted his readers more than anything else and he was successful because he did so. I first read Fuentes in a college spanish class (Death of Artemio Cruz) and his work opened my mind to a way of thinking and expressing emotions that would otherwise never have happened. I thank you for that Señor Fuentes. Vaya con Dios.</p> <p style="text-align: right;">May 16, 2012 at 3:21 a.m. Recommended18</p>
49	<p style="text-align: center;">auto-de-fe Los Angeles</p> <p>Did we come to laugh or cry? Are we dying or being born? Is it the beginning or the end, cause or solution? What are we living through?"</p>

-Terra Nostra, Carlos Fuentes
'Carnal-val'

In my own readings of Don Carlos the inspiration carry over into my own artworks! 'Hasta la proxima Carlos!'

<http://www.artbreak.com/work/show/4255-carnal-val-sunofman>

May 16, 2012 at 3:21 a.m.

Recommended4

50

Paty Chenge

Xalapa, Veracruz, México

Carlos Fuentes represented and identified half a century of our history. Cervantes Prize in 1987 and Prince of Asturias Award in 1994, Fuentes remained only get the novel in literature. Just last March 2 attended the third issue of the Chair Carlos Fuentes in Xalapa, Veracruz, Mexico, organized by the Universidad Veracruzana.

His words: "The past is alive in memory, the future present in desire."

May 16, 2012 at 3:21 a.m.

Recommended12

51

Richard Young

Bayfield, CO

I am saddened to hear of the death of Carlos Fuentes, my favorite author not only in the Spanish language but in any language. Fortunately Mr. Fuentes lives on in his approximately two dozen novels. To those readers who are not familiar with his work, you cannot make a mistake by reading any of his invariably challenging and brilliant books. In lieu of reading the new Fuentes novel which will never arrive, I now look forward with pleasure to rereading his life work from start to finish. Gracias por la vida Fuentes.

May 16, 2012 at 3:21 a.m.

Recommended12

52

Sergio Camacho

Cholula, Puebla

I just heard as I was driving that he had finished one more novel, which is already with the publishing house. As I was driving, I din not catch more details. Something to look forward to.

May 16, 2012 at 4:24 a.m.

Recommended2

53

Gerardo Mondragon

Chicago

He was perhaps, the greatest Mexican writer after Octavio Paz, whose work left a deep impression in the Latin America letters and a void difficult to fill in the Mexican conscience. RIP Don Carlos!

May 16, 2012 at 2:58 a.m.

Recommended10

54

SteveZ
NYC

My wife and I saw him deliver a lecture in the Great Hall at Cooper Union in 1993. He was introduced by his friend, William Styron. One of the most memorable evenings of my life. What a brilliant soul. What a loss to us all.

May 16, 2012 at 2:58 a.m.

Recommended7

55

Enrique
Miami, Fl

The world has lost a great writer today. A man with a clear vision of our time who loves his country. His legacy will remain with us forever. Descansa en paz, querido Carlos

May 16, 2012 at 2:58 a.m.

Recommended4

56

LEFTY
North Carolina

He was the great one! Farewell, Mr. Fuentes!

May 16, 2012 at 2:58 a.m.

Recommended1

57

DR Hyatt
Carefree

When reading the works of Carlos Fuentes, Pablo Neruda and Octavio Paz, I struggle with fleeting moments of reality and loose touch with my surroundings therefore with these great writers: Carlos Fuentes, Neruda and Paz I journey into the sublime of day dreams. For Carlos Fuentes he allowed me to explore a mélange of dreams, with Neruda just looking at the ocean—imaginary flights of words and imagination and with Paz the hyperbolization of various sensory impressions. Blessed is Carlos Fuentes to allow me to be an immobile traveler to witness strange ancestors and strange lands.

May 16, 2012 at 2:26 a.m.

Recommended7

58

Miguel Rubiano
Texas

United States baned his traveling to USA because he was considered a dangerous person in 1963, he could travel in 1967 to USA. He became professor in Ivy League Universities later. Few years ago, the secretary of Government in Mexico, Mr Abascal required to ban the reading Mr Fuentes book in a private High school because he considered against the righth customs.His daugther was student in such school. Now, his name appears on first page in all the newspapers in Latin America and also in NY TIMES.

May 16, 2012 at 2:26 a.m.

Recommended7

59	<p style="text-align: center;">Lorenzo Moreno Princeton, NJ</p> <p>Carlos Fuentes created the voices of a city, a political era, and a moral standard...Ixca Cienfuegos, Artemio Cruz, and Felipe Montero. Only he could do it.</p> <p style="text-align: right;">May 16, 2012 at 2:26 a.m. Recommended3</p>
60	<p style="text-align: center;">Carolivia Herron Washington, DC</p> <p>He was my friend and teacher. When I read his Terra Nostra I quit my government job and went off to write my first novel. I sought him out to be my teacher at Penn. He got me my first book contract and first literary agent. I was the president of his fan club, the Fuentettes. I spent a year as a Fulbright fellow in Mexico, following his images. He taught me at Penn, and we taught together at Harvard. Oh, my friend, I believed you to be the greatest living novelist in Terra Nostra, in all your fictions. Farewell.</p> <p style="text-align: right;">May 16, 2012 at 2:26 a.m. Recommended22</p>
61	<p style="text-align: center;">Kwame Okoampa-Ahoofe Bronx, New York</p> <p>The gradual and several departures of the great ones bring to a profound and peaceful understanding of the self amidst the cosmos; we are, each, a fragment of one another.</p> <p style="text-align: right;">May 16, 2012 at 1:40 a.m. Recommended16</p>
62	<p style="text-align: center;">Suzanne Wheat North Carolina</p> <p>Mr. Fuentes always had something incisive to say about current events. His voice will be missed.</p> <p style="text-align: right;">May 16, 2012 at 1:40 a.m. Recommended9</p>
63	<p style="text-align: center;">edwin claremont, ca</p> <p>A wonderful person as well as write whom i had the excelent fortune to meet twice and converse with on one of those occasions. Rest in peace, Carlos.</p> <p style="text-align: right;">May 16, 2012 at 1:29 a.m. Recommended2</p>
64	<p style="text-align: center;">faust Nowhere</p>

I knew him a little, I liked him a lot; he accepted things from me. My friend Saliva, in the 1980's, said he was the most important public man in Western culture--I told him. I had to drive him and his wife to the airport once, and I watched as we waited for the short ride to get going: he sat in a living room chair, reflecting, his fingers joined as if in prayer, touching his upper lip--silent. I remember this pose all the time. It taught me what to do before I start writing.

May 16, 2012 at 1:06 a.m.
Recommended17

65

ZihuaRob

Zihuatanejo, Guerrero, Mexico

Carlos Fuentes will long remain a giant in Mexican culture, a soothing voice guiding the consciousness of millions of Mexicans and Latin Americans. His wisdom and passion will be missed. Que en paz descance.

May 16, 2012 at 12:48 a.m.
Recommended19

66

Douglas

Sacramento

"The Crystal Frontier" taught me more about Mexican-American relations than I ever could have learned in any other context. The world has lost a great writer today. I am grateful that we still have Mr. Fuentes with us, dispersed throughout his books.

May 16, 2012 at 12:48 a.m.
Recommended28

Anexo 3

Glosario de la terminología empleada en este trabajo⁶⁰

⁶⁰ En las definiciones se han marcado con cursiva aquellos términos presentes en el glosario. La realización de este glosario está basada en las definiciones encontradas en las siguientes publicaciones: Cole, Mariani, Uszkoreit, Zaenen y Zue (1997); Garrido García y Ruiz de Mendoza Ibáñez (2011); Gómez Guinovart (2000); González Vergara (2006a); Guerra García y Sacramento Lechado (2011); Jiménez Briones, Luzondo Oyón y Pérez Cabello de Alba (2011); Liddy (2001); Llisterri (2003); Mairal Usón y Cortés Rodríguez (2006); Mitkov (2003); Moure y Llisterri (1996); Perrián Pascual (2005, 2012); Perrián Pascual y Arcas Túnez (2005, 2007b, 2010, 2011); Perrián Pascual y Carrión Varela (2011); Perrián Pascual y Mairal Usón (2009, 2010a, 2010b, 2011); Ruiz de Mendoza y Díez Velasco (2004); Ruiz de Mendoza Ibáñez y Marial Usón (2008); Ruiz de Mendoza y Otal Campo (2002); Salem y Nolan (2008), Shimojo (2009); Van Valin (2005).

Activador de guión (*script activator*): en el *Cognición de FunGramKB*, un *activador de guión* indica la *expansión conceptual saliente y entrante de guiones huéspedes y anfitriones*, de manera que los *guiones* estén interconectados entre sí.

Aktionsart (modo de acción): en la *Gramática del Papel y la Referencia*, clasificación que según su acción atribuye a los verbos una clase (estado, actividad, logro, semelfactivo y realización junto a sus causativos).

Algoritmo de resolución de anáfora (*anaphora resolution algorithm*): algoritmo en el que se especifican los pasos a seguir en las fases de la resolución automática de anáfora: identificación del elemento anafórico, localización de los candidatos a antecedente y selección del antecedente correcto.

Algoritmo de enlace (*linking algorithm*): algoritmo que relaciona la sintaxis y la semántica en la *Gramática del Papel y la Referencia*, especificando cuáles son los pasos a seguir en el proceso de comprensión y producción del lenguaje.

Anáfora metonímica (*metonymic anaphora*): anáfora en la cual el elemento anafórico es un pronombre que se refiere a un antecedente caracterizado por ser metonímico.

Análisis morfológico (*morphological parsing*): segundo paso del procesamiento textual en *FunGramKB* en el cual se despoja a los tokens léxicos de los afijos flexivos.

Analizador sintáctico-semántico (*parser*): analizador que otorga a una frase una estructura sintáctica y una forma lógica basándose en el lexicon y la gramática.

Analizador sintáctico-semántico ligero (*shallow parser*): *analizador sintáctico-semántico* que reconoce solo determinadas estructuras.

Árbol de decisión (*decision tree*): estructura de datos en forma de diagrama que categoriza ejemplos adquiridos induciéndolos de un conjunto de datos etiquetados.

Argumento (*argument*): en *FunGramKB* cada argumento aparece como una *variable* x_n , en algunos casos junto a una *preferencia de selección* y por último su *papel temático*, p. ej. (x_1 :+HUMANO_00)Theme. El tipo de *papel temático* que tiene un argumento depende de a qué *metaconcepto* está ligado.

Argumento Sintáctico Privilegiado (*Privileged Syntactic Argument, PSA*): relación sintáctica en una construcción sintáctica dada con neutralización restringida de los papeles semánticos y las funciones pragmáticas. En la *Gramática del Papel y la Referencia* es la única relación sintáctica reconocida por darse en todas las lenguas.

Base de conocimiento (*knowledge base*): base de datos con representaciones de conocimiento que aportan información a una aplicación de *procesamiento del lenguaje natural*.

Bio-estructuras (*bio-structures*): *entidades* o *eventos* dentro de la clasificación de los *esquemas* conceptuales en *FunGramKB* (véase *retratos* e *historias*).

Bio-macroestructura (*bio-macrostructure*): véase *historia*.

Bio-microestructura (*bio-microstructure*): véase *retrato*.

Centro (*core*): unidad sintáctica presente en todas las lenguas y compuesta por los argumentos y el predicado (*i. e. núcleo*) de la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Clasificación de textos (*text categorization*): proceso de decidir a qué categoría pertenecen los documentos.

Cláusula (*clause*): elemento presente en todas las lenguas formado por el *centro* y la *periferia* en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Cognición (*Cognicon*): almacenamiento del *conocimiento procedural* a través de *guiones* en el nivel conceptual de *FunGramKB*.

Concepto (*concept*): representación mental estructurada lexicalizada en alguna manera y definida por un *marco temático* y un *postulado de significado*.

Concepto básico (*basic concept*): nivel conceptual de la *ontología nuclear* de *FunGramKB* representado por +.

Concepto terminal (*terminal*): nivel conceptual de la *ontología nuclear* de *FunGramKB* simbolizado por \$.

Construcción de la estructura lógico-conceptual (*Conceptual Logical Structure construction*): cuarto paso del procesamiento textual en *FunGramKB* en el que se construye la *estructura lógica conceptual*.

COREL: lenguaje formal de representación conceptual común a los tres tipos de conocimiento, postulado por *FunGramKB* e independiente de cualquier lengua.

Conocimiento cultural / del mundo / sobre entidades nombradas y eventos (*episodic knowledge / knowledge about named entities and events*): conocimiento sobre *entidades* y *eventos* almacenado en el *Onomasticón* de *FunGramKB*.

Conocimiento discursivo (*discursive knowledge*): información creada con un conjunto de oraciones producidas por una o más personas para transmitir

información.

Conocimiento ontológico (*ontological knowledge*): información almacenada en la *Ontología* de *FunGramKB* donde los *conceptos*, por medio de relaciones IS-A (*i. e.* siendo hipónimos unos de otros), siguen la jerarquía *metaconceptos* > *conceptos básicos* > *conceptos terminales* y donde se permite la *herencia no monotónica*.

Conocimiento procedural (*procedural knowledge*): información sobre determinados comportamientos y actuaciones estereotípicas en situaciones concretas almacenado en el *Cognición* de *FunGramKB*.

Controlador (*controller*): en la *Gramática del Papel y la Referencia*, tipo de *Argumento Sintáctico Privilegiado* que ayuda a interpretar el *pivote*.

Corpus anotado / etiquetado (*tagged corpus*): corpus en el que las palabras tienen etiquetas con información como clases de palabras, relaciones, estructuras y demás.

DBpedia: base de datos que extrae información automáticamente de varias fuentes como Wikipedia o YAGO, la almacena de manera jerárquica y la hace accesible a la web.

DBpedia mapper: herramienta de *FunGramKB Suite* que asiste en la fase de construcción del proceso de población del *Onomasticón* durante la cual se importa el *conocimiento cultural* de *DBpedia* a *FunGramKB*.

Dominio fuente (*source domain*): origen del ámbito semántico que se importa.

Dominio matriz (*matrix domain*): ámbito semántico mayor de los que forman parte de una proyección conceptual.

Dominio meta (*target domain*): destino del ámbito semántico importado.

Dominio Potencial de Foco (*Potential Focus Domain*, PFD): parte de la oración que puede constituir el *foco*.

Dominio Real de Foco (*Actual Focus Domain*, AFD): parte de la oración que efectivamente constituye el *foco*.

Entidades (*entities*): subontología en *FunGramKB* compuesta por nombres.

Entidad nombrada (*named entity*): palabra(s) que se identifican con un nombre de persona, organización, lugar, etc.

Espacio pre-central (*PreCore Slot*, *PrCS*): unidad sintáctica presente en algunas lenguas que precede al *centro* de la *estructura estratificada de la cláusula* en la *Gramática del Papel y la Referencia*.

Espacio post-central (*PostCore Slot, PoCS*): unidad sintáctica presente en algunas lenguas que sigue al *centro* de la *estructura estratificada de la cláusula* en la *Gramática del Papel y la Referencia*.

Esquema (*schemata*): representación mental de una *entidad* o *evento*.

Estructura lógica (*Logical Structure, LS*): representación formal de cada clase léxica que en el lexicón forma el núcleo de la representación semántica de unidades léxicas.

Estructura lógico-conceptual (*Conceptual Logical Structure, CLS*): formalismo en *FunGramKB* de la *estructura lógica*.

Estructura estratificada de la cláusula (*Layered Clause Structure, LCS*): representación sintáctica de la *cláusula* en la *Gramática del Papel y la Referencia* con elementos universales (*núcleo, centro, periferia*) y otros presentes en algunas lenguas (espacio pre-/post-central, elemento dislocado a la derecha/izquierda).

Estructura estratificada del SN (*Layered NP Structure, LNPS*): representación sintáctica del SN en la *Gramática del Papel y la Referencia* con elementos universales (*núcleo, centro, periferia*) y otros presentes en algunas lenguas (*posición de comienzo/final del SN*).

Estructura de foco (*focus structure*): estructura de *foco* que puede ser de foco estrecho completiva o contrastiva, de *foco* ancho predicativa o de *foco* ancho oracional.

Etiquetador (*tagger*): elemento que asigna etiquetas con información gramatical, semántica y demás a (conjuntos de) palabras en un texto.

EuroWordNet: base de datos léxica multilingüe.

Eventos (*events*): subontología en *FunGramKB* que reúne los verbos. En los *guiones* del *Cognición*, un *evento* es una predicación desarrollada en un espacio de tiempo entre un comienzo y un final.

Expansión conceptual entrante (*incoming mapping*): expansión conceptual de algunos participantes en el *guión huésped* a sus participantes correspondientes en el *guión anfitrión* dentro del *Cognición* de *FunGramKB*.

Expansión conceptual saliente (*outcoming mapping*): expansión conceptual de algunos participantes en el *guión anfitrión* a sus participantes correspondientes en el *guión huésped* dentro del *Cognición* de *FunGramKB*.

Extracción automática de términos (*term recognition*): reconocimiento de términos en las diversas apariciones de un conjunto de textos.

Extracción de información (*information extraction*): proceso de hallar unos datos

concretos dentro de un grupo de documentos presentándolos de forma organizada.

Foco (*focus*): información nueva que indica la diferencia entre la presuposición y la aserción pragmática.

Frecuencia referencial (*referential frequency, RP-f*): frecuencia de la referencia a una entidad dentro de diez oraciones.

FunGramKB: base de conocimiento léxico conceptual basada en gramática funcionalista con varios usos y para diferentes lenguas diseñada para tareas del *procesamiento del lenguaje natural*.

FunGramKB Suite: entorno computacional para construir semiautomáticamente una base de sistemas automatizados del *procesamiento del lenguaje natural*.

Gramática del Papel y la Referencia (*Role and Reference Grammar, RRG*): teoría funcional con la que se representa un texto de entrada en términos de su *estructura lógica*.

Gramaticón (*Grammaticon*): acumulación de esquemas estructurales para que la *Gramática del Papel y la Referencia* construya la conexión entre semántica y sintaxis.

Guión (*script*): esquemas conceptuales que en el *Cognicón* de *FunGramKB* describen eventos estereotípicos en un marco temporal.

Guión anfitrión (*host script*): nodo desde el cual se origina la *expansión conceptual* en los *guiones* del *Cognicón* de *FunGramKB*.

Guiones simples (*simple script*): *guiones* que constan únicamente de ellos mismos con varios eventos.

Guiones complejos (*complex script*): *guiones* que integran en ellos sub-guiones.

Guiones compuestos (*composite script*): *guiones* formados por dos o más *guiones* sin ser dependientes unos de otros, es decir, ninguno es necesario para comprender otro.

Guión huésped (*guest script*): nodo al cual se dirige la *expansión conceptual* en los *guiones* del *Cognicón* de *FunGramKB*.

Herencia (*inheritance*): mecanismo del *proceso de expansión microconceptual* que transfiere predicaciones del *concepto* superordinado al subordinado en *FunGramKB*. Puede ser *monotónica* o *no monotónica*.

Herencia múltiple (*multiple inheritance*): quiere decir que un *concepto* puede ser *subsumido* por más de un *concepto*.

Historia (*story*): entidad o evento representado cronológicamente en *FunGramKB* (i. e. *bio-macroestructura*).

Índice de concordancia de género (*agreement index-genus*, AGX_G): nodo dependiente del *núcleo* en la *estructura estratificada del SN* de la *Gramática del Papel y la Referencia* que contiene los rasgos de concordancia de género.

Índice de concordancia de persona (*agreement index-person*, AGX_P): nodo dependiente del *núcleo* en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia* que contiene los rasgos de concordancia de persona de todos los argumentos centrales.

Inferencia (*inference*): mecanismo del *proceso de expansión microconceptual* que transfiere información entre *conceptos* sin *subsunción* en *FunGramKB*.

Ingeniería lingüística / industrias de la lengua (*linguistic engineering*): explicación de productos comerciales con técnicas del *procesamiento del lenguaje natural*.

Inteligencia Artificial (*Artificial Intelligence*): rama de las ciencias computacionales que pretende alcanzar los resultados del razonamiento humano.

Interfaz sintáctico-semántica (*syntax-semantics interface*): niveles de estructura sintáctica en los que se aplican principios de interpretación semántica para generar una representación semántica.

Lexicón (*Lexicon*): acumulación de la información morfosintáctica, semántica y pragmática de cómo se comportan las unidades léxicas en una lengua en concreto.

Lingüística computacional (*Computational linguistics*): teorías y procedimientos lingüísticos para crear sistemas computacionales que analicen y/o produzcan textos en lenguaje natural.

Macroestructura (*macrostructure*): representación del conocimiento situado temporalmente que se dividen en *proto-macroestructuras* y *bio-macroestructuras*.

Macropapel (*macrorole*): generalizaciones de diferentes tipos de argumento que predicen su comportamiento sintáctico (transitividad del predicado). Según el número de macropapeles, se distingue entre verbos atransitivos (sin macropapeles), intransitivos (con un macropapel) o transitivos (con dos macropapeles).

Macropapel de Actor (*actor*): argumento como el agente, similar al sujeto lógico.

Macropapel de Padecedor (*undergoer*): argumento como el paciente, similar al objeto lógico.

Marco temático (*thematic frame*, *TF*): construcción conceptual con los participantes

de la situación cognitiva prototípica del *concepto* formada por *variables* x_n , *preferencias de selección* y *papeles temáticos*.

Metaconcepto (*metaconcept*): nivel conceptual de la *ontología nuclear* de *FunGramKB* representado por #.

Metáfora (*metaphor*): proyección conceptual externa entre dos áreas semánticas. Puede ser referencial o predicativa.

Metonimia (*metonymy*): proyección conceptual interna dentro de un mismo área semántica. Puede ser referencial o predicativa.

Metonimia fuente-en-meta (*source in target metonymy*): *metonimia* con una proyección conceptual expansiva, en la cual el *dominio fuente* es un subdominio del *dominio meta*.

Metonimia meta-en-fuente (*target in source metonymy*): *metonimia* con una proyección conceptual reductora, en la cual el *dominio meta* es un subdominio del *dominio fuente*.

Microestructuras (*micro-structure*): representación del conocimiento sin unas coordenadas temporales que se divide en *proto-microestructuras* y *bio-microestructuras*.

Minería de opiniones (*opinion mining*): análisis de opiniones y sentimientos expresados en un conjunto de textos pudiendo observar si son positivos, neutros o negativos.

Modelo Léxico Construccional (*Lexical Constructional Model, LCM*): modelo de construcción del significado basado en modelos funcionalistas como la *Gramática del Papel y la Referencia* y la *Lingüística Cognitiva*.

Módulo conceptual (*conceptual module*): módulo del nivel conceptual en *FunGramKB*, *i. e. Ontología, Cognición y Onomasticón*.

Monotónico (*monotonic*): relación según la cual los *conceptos* subordinados tendrán todos los rasgos del *concepto* superior más los que ellos mismos añadan.

Morficón (*Morphicon*): módulo en *FunGramKB* que se encarga de los casos de morfología flexiva en el nivel léxico. Está compuesto por dos submódulos: datos de formas regulares (morforegla) e irregulares (morfoDB).

Nivel 1 argumental (*level 1, argumental layer*): nivel 1 de construcción del significado en el *Modelo Léxico Conceptual* con modelos no-situacionales (*i. e.* no relacionados con el contexto discursivo), donde los elementos léxicos son *subsumidos* a las construcciones.

Nivel 2 de implicatura (*level 2, implicational layer*): nivel 2 de construcción del significado en el *Modelo Léxico Conceptual* con modelos situacionales de bajo nivel (*i. e.* específicos), donde se tratan fenómenos relacionados con la teoría de implicaturas.

Nivel 3 ilocutivo (*level 3, illocutionary layer*): nivel 3 de construcción del significado en el *Modelo Léxico Conceptual* con modelos situacionales de alto nivel (*i. e.* generales), donde se trata la fuerza ilocutiva tradicional.

Nivel 4 del discurso (*level 4, discourse layer*): nivel 4 de construcción del significado en el *Modelo Léxico Conceptual* con modelos situacionales de alto nivel (*i. e.* generales) y en pares relacionales, donde se trata la construcción discursiva: cohesión/coherencia.

No monotónico (*non-monotonic*): relación en la que algunos rasgos del *concepto* superodenado no serán heredados por los *conceptos* inferiores.

Núcleo (*nucleus*): unidad sintáctica presente en todas las lenguas que contiene el predicado en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Onomasticón (*Onomasticon*): acumulación de información de *entidades* y *eventos* en el nivel conceptual de *FunGramKB* con 2 tipos de *esquemas* (*bioestructuras*): *retratos* e *historias*.

Ontología (*Ontology*): catálogo jerarquizado de *conceptos* manejado por las personas al hablar que refleja el modelo del mundo compartido por una comunidad.

Ontología nuclear (*Core Ontology*): módulo de propósito general de la *ontología* que constituye el modelo ontológico junto a las *ontologías satélites*.

Ontologías satélites (*Satellite Ontologies*): módulos terminológicos específicos de dominio que constituye el modelo ontológico junto a la *ontología nuclear*.

Operador (*operator*): elementos representados por mayúsculas que modifican las constantes y representan diferentes aspectos semánticos. En la *Gramática del Papel y la Referencia* se distinguen varios operadores para diferentes niveles en la *estructura estratificada de la cláusula* y del *SN*.

Papel temático (*semantic role*): papel semántico directamente vinculado con la clase de predicado en el que aparece.

Periferia (*periphery*): unidad sintáctica presente en todas las lenguas que comprende los adjuntos (no argumentos) en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Pivote (*pivot*): en la *Gramática del Papel y la Referencia*, tipo de *Argumento*

Sintáctico Privilegiado que es elidido en una construcción determinada.

Plantilla léxica (*lexical template*): plantilla que aporta información pragmática, semántica y sintáctica en el *Modelo Léxico Conceptual*.

Plantilla construccional (*constructional template*): plantilla que agrupa argumentos interrelacionados a través de predicados abstractos en el *Modelo Léxico Conceptual*.

Posición de comienzo del SN (*NP-initial position, NPIP*): constituyente presente en algunas lenguas que ocupa el primer lugar dentro de la estructura estratificada del SN en la *Gramática del Papel y la Referencia*.

Posición del final del SN (*NP-final position, NPFP*): constituyente presente en algunas lenguas que ocupa el último lugar dentro de la estructura estratificada del SN en la *Gramática del Papel y la Referencia*.

Posición dislocada a la izquierda (*Left Detached Position, LDP*): constituyente presente en algunas lenguas que aparece a la izquierda de la cláusula en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Posición dislocada a la derecha (*Right Detached Position, RDP*): constituyente presente en algunas lenguas que aparece a la derecha de la cláusula en la *estructura estratificada de la cláusula* de la *Gramática del Papel y la Referencia*.

Postulado de Significado (*Meaning Postulate, MP*): constructo que describe el conocimiento semántico y el sentido común de los *conceptos*. Se compone de predicaciones e_n estrictas o rebatibles, con una descripción donde se integran los *argumentos* coindizados del *marco temático* x_n y con satélites f_n que introducen una *predicación* ligada o un *concepto*.

Predicación (*predication*): elemento precedido por el operador de razonamiento + si es estricta o * si es rebatible, representado por una *variable*, indizado y con dos puntos.

Preferencia de selección (*selectional preferences*): restricción conceptual en *marco temático* y *postulado de significado* de *conceptos básicos*, *conceptos terminales* y *subconceptos* asociada prototípicamente a una situación cognitiva concreta.

Procesamiento del lenguaje natural (PLN, *Natural Language Processing, NLP*): análisis de textos con la finalidad de conseguir procesar un lenguaje como el humano para un grupo de aplicaciones o tareas o aplicaciones.

Procesamiento sintáctico-semántico (*syntax-semantics processing*): tercer paso del procesamiento textual en *FunGramKB* en el que tiene lugar la desambiguación y se decide la estructura de la frase de entrada.

Proceso de expansión bio-conceptual (*BioKnowing*): capacidad de razonamiento en *FunGramKB* que permite extender semánticamente los *postulados de significado*, los *guiones*, los *retratos* y las *historias* mediante los mecanismos de *herencia e inferencia*, los *activadores de guiones* y la información del *Onomasticón* almacenada en el mismo lenguaje *COREL*.

Proceso de expansión macro-conceptual (*MacroKnowing*): capacidad de razonamiento en *FunGramKB* que permite extender semánticamente los *postulados de significado* y los *guiones* mediante los mecanismos de *herencia e inferencia* y los *activadores de guiones*.

Proceso de expansión micro-conceptual (*MicroKnowing*): capacidad de razonamiento en *FunGramKB* que permite extender semánticamente los *postulados de significado* mediante los mecanismos de *herencia e inferencia*.

Proto-estructura (*proto-structure*): conocimiento prototípico en el parámetro de la prototipicidad que clasifica los esquemas conceptuales en *proto-macroestructuras* y *proto-microestructuras* dentro de *FunGramKB*.

Proto-macroestructura (*proto-macrostructure*): véase *guión*.

Proto-microestructura (*proto-microstructure*): véase *postulado de significado*.

Proyección de la estructura de foco (*focus structure projection*): proyección integrada en la *Gramática del Papel y la Referencia* a la proyección de la estructura de constituyentes y de *operadores*. En ella se distinguen dos ángulos de líneas discontinuas para el *dominio potencial de foco* y el *dominio real de foco* que parten del nodo acto de habla.

Razonamiento (*reasoning*): quinto paso opcional en el procesamiento textual en *FunGramKB* que proyecta la *estructura estratificada de la cláusula* a un esquema *COREL* para inferir el conocimiento relevante.

Recuperación de información (*information retrieval*): selección, dentro de un grupo de textos, de aquellos que tengan la información que busca la persona que hace la consulta.

Referencia ininterrumpida (*uninterrupted reference, RP-d*): duración de una referencia ininterrumpida en un número de oraciones dado, dependiendo de la presencia de esa entidad en diez oraciones seguidas. Cuanto mayor sea la duración, mayor será la *relevancia*.

Referente identificable activo (*activated, ACV*): referente que, como *foco*, es resuelto activamente por medio de una referencia directa.

Referente identificable accesible (*accessible, ACS*): referente no resuelto activamente sino que el destinatario lo reconoce gracias al conocimiento del mundo o a la aparición en el contexto (extra)lingüístico.

Referente identificable inactivo (*inactivated, INA*): referente que en el discurso ya fue mencionado previamente pero no es resuelto activamente y, además, el hablante no asume que sea reconocido por el destinatario.

Referente no identificable nuevo anclado a otro más identificable (*brand new anchored, BNA*): referente que, como *tópico*, es la primera vez que se menciona, aunque está relacionado con otro elemento ya dicho o accesible.

Referente no identificable nuevo no anclado (*brand new unanchored, BNU*): referente que aparece por primera vez en el discurso sin estar relacionado con nada mencionado antes.

Relevancia (*saliency*): importancia de la información. La alusión a un referente importante suele ser continua en un discurso dado. Es medible a través de la *frecuencia referencial* y la *referencia ininterrumpida*.

Restricción interna (*internal constraints*): condiciones para que las *plantillas léxicas* se unan con las *plantillas constructivas*.

Restricción externa (*external constraints*): condiciones para que determinados mecanismos de asociación cognitivos y conceptuales se den en los elementos que son *subsumidos* de un nivel a otro en el *Modelo Léxico Conceptual*.

Retrato (*snapshot*): *entidad* o *evento* sin unas coordenadas temporales en *FunGramKB* (i. e. *bio-microestructura*).

Semántica profunda (*deep semantics*): conocimiento conceptual con gran poder expresivo como la base de conocimiento *FunGramKB*.

Semántica superficial (*surface semantics*): conocimiento de significado *relacional* con poder expresivo restringido como *EuroWordNet*.

Significado relacional (*relational meaning*): descripción del uso de un elemento léxico a través de las relaciones de significado que establece con otros elementos léxicos.

Síntesis automática (*summarization*): aplicación para resumir un texto con la información requerida por un usuario u otra aplicación.

Sistema de búsqueda de respuestas (*Question Answering system*): sistema que busca respuestas a preguntas realizadas usando técnicas de *procesamiento del lenguaje natural*.

Sistema de comprensión del lenguaje natural (*natural language comprehension system*): aplicación para entender el significado de las oraciones de entrada.

Sistema de resolución de anáfora (*anaphora resolution system*): sistema que procesa textos a varios niveles (morfológico, sintáctico, semántico, pragmático, etc.) aportando los textos analizados al *algoritmo de resolución de anáfora*.

Sistemas de resolución de anáfora alternativo: sistema que recurre a conocimiento alternativo (aprendizaje automático basado en corpus, análisis probabilístico, colocados y demás) para resolver la anáfora.

Sistemas de resolución de anáfora de conocimiento enriquecido: (*knowledge-rich anaphora resolution system*): sistema que recurre a varios niveles de conocimiento lingüístico (p. ej. léxico, sintáctico, semántico, pragmático) y extralingüístico (conocimiento del mundo) para resolver la anáfora.

Sistema de resolución de anáfora de conocimiento pobre o limitado: (*knowledge-poor/limited anaphora resolution system*): sistema que recurre a pocos niveles de conocimiento lingüístico (p. ej. léxico, sintáctico) para resolver la anáfora.

Sistema de diálogo (*dialogue system*): sistema que controla el diálogo persona – máquina con reconocedor de habla, conversor de texto a habla, analizador, *base de conocimiento* y generador lingüístico.

Subconcepto (*subconcept*): unidad ontológica con el mismo *postulado de significado* que el *concepto básico* o *concepto terminal* al que se asocia pero con una/s diferente/s *preferencia/s de selección* en el *marco temático*.

Subsunción (*subsumption*): en la *Ontología* de *FunGramKB*, relación entre dos *conceptos* de manera que el primero subsume al segundo si todas las propiedades de aquel son también ciertas en este, habiendo un conjunto de rasgos del segundo que no tienen por qué ser ciertos en el primero. En el *Modelo Léxico Conceptual*, la subsunción construccional consiste en incorporar construcciones de nivel inferior en las de nivel superior.

Synset: conjunto de palabras consideradas sinónimos en uno o todos los contextos.

Tema (*Theme*): *papel temático* presente en los ámbitos metacognitivos y que tiene un valor definitorio diferente en cada uno de ellos.

Tokenización (*tokenization*): primer paso del procesamiento textual en *FunGramKB* en que el texto de entrada se segmenta en oraciones y en palabras (unidades básicas del análisis).

Tópico (*topic*): información conocida que forma parte de la presuposición pragmática.

Traducción automática (*machine translation*): ámbito de la *lingüística computacional* que estudia el uso de aplicaciones informáticas para traducir texto de una lengua natural a otra.

Variable (*variable*): elemento que simboliza los *argumentos* prototípicos de la unidad léxica (x, y, z).

WordNet: base de datos con una red semántica que relaciona *synsets* entre sí.

BIBLIOGRAFÍA

- Aguilar Sosa, Y. (16 de mayo de 2012). El Aura de México, adiós a la última gran conciencia. *El Universal*. Recuperado de <http://www.eluniversal.com.mx/cultura/68749.html>.
- Belloro, V. A. (2004). *A Role and Reference Grammar account of third-person clitic cluster in Spanish*. (Trabajo de Fin de Máster, Universidad de Buffalo, Nueva York, Estados Unidos). Recuperado de <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/A%20Role%20and%20Reference%20Grammar%20Account%20of%20Third-Person%20Clitics%20Clusters%20in%20Spanish.pdf>.
- . (2007). *Spanish Clitic Doubling: A Study of the Syntax-Pragmatics Interface*. (Tesis Doctoral, Universidad de Buffalo, Nueva York, Estados Unidos). Recuperado de http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/Belloro-Spanish_Clitic_Doubling.pdf.
- . (2009). Spanish datives: Remarks on the Information-Structure side of the story. En Guerrero, L., Ibáñez Cerda, S. y Belloro, V. A. (eds.), *Studies in Role and Reference Grammar*, (pp. 491-516). México: UNAM.
- Bisbal Asensi, E., Molina Marco, A., Moreno Boronat, L., Pla Santamaría, F., Saiz Noeda, M. y Sanchis Arnal, E. (2003). 3LB-SAT: una herramienta de anotación semántica. *Procesamiento del lenguaje natural*, 31, 193-200. Recuperado de http://rua.ua.es/dspace/bitstream/10045/1510/1/PLN_31_23.pdf.
- Boldrini, E., Puchol-Blasco, M., Navarro, B., Martínez-Barco, P. y Vargas-Sierra, C. (2009). AQA: a multilingual anaphora annotation scheme for question answering. *Procesamiento del lenguaje natural*, 42, 97-104. Recuperado de http://rua.ua.es/dspace/bitstream/10045/10550/1/PLN_42_12.pdf.
- Branco, A. (2007). Null Subjects are Reflexives, not Pronouns. En Branco, A. (ed.), *Anaphora: Analysis, Algorithms and Applications. DAARC 2007. LNCS (LNAI)*,

- 4410, (pp. 59-76). Berlín: Springer-Verlag. Recuperado de <http://www.di.fc.ul.pt/~ahb/BrancoLNAI2007.pdf>.
- Butler, C. S. (2009). The Lexical Constructional Model: Genesis, strengths and challenges. En Butler, C. S. y Martín Arista, J. (eds.), *Deconstructing Constructions*, (pp. 117-152). Ámsterdam/Filadelfia: John Benjamins. Recuperado de http://www.lexicom.es/drupal/files/Butler_2009.pdf.
- Caselli, T. (2009). Using a generative lexicon resource to compute bridging anaphora in Italian. *Procesamiento del lenguaje natural*, 42, 71-78. Recuperado de http://rua.ua.es/dspace/bitstream/10045/10547/1/PLN_42_09.pdf.
- Cole, R. A., Mariani, J., Uszkoreit, H., Zaenen, A. y Zue, V. (1997). *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press.
- Consejo de Europa. 2002. *Marco Común Europeo de Referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Ministerio de Educación, Cultura y Deporte. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf.
- Davies, M. *BYU-BNC*. (Basado en el *British National Corpus* de Oxford University Press). [En línea] <http://corpus.byu.edu/bnc/> [19.07.12].
- . *The Corpus of Contemporary American English: 450 million words, 1990-present*. [En línea] <http://corpus.byu.edu/coca/> [19.07.12].
- Deemter, K. y Kibble, R. (2001). On Coreferring: Coreference in MUC and Related Annotation Schemes. *Journal of Computational Linguistics*, 26(4), 629–637. Recuperado de <http://acl.ldc.upenn.edu/J/J00/J00-4005.pdf>.
- Delmonte, R. (2003). The Semantic Web Needs Anaphora Resolution. *International Symposium on Reference Resolution and Its Applications to Question Answering and Summarization*. Recuperado de <http://project.cgm.unive.it/events/ARQAS/ARQAS/pdf/delmonte.pdf>.

- DePalma, A. (15 de mayo de 2012). Carlos Fuentes, Mexican Man of Letters, Dies at 83. *The New York Times*. Recuperado de http://www.nytimes.com/2012/05/16/books/carlos-fuentes-mexican-novelist-dies-at-83.html?_r=3&pagewanted=all#comments.
- Devi, S. L. (2008). Anaphora Resolution Using Named Entity and Ontology. En Johansson, C. (ed.), *Proceedings of the Second Workshop on Anaphora Resolution (WAR II)*, (pp. 91-96). Tartu: NEALT. Recuperado de <http://dspace.utlib.ee/dspace/bitstream/handle/10062/7369/Sobha.pdf?sequence=1>.
- Felices Lago, A., Ureña Gómez-Moreno, P. y Alameda Hernández, A. (2011). FunGramKB y la Adquisición Terminológica. *ANGLOGERMANICA ONLINE*, 8, 66-86. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb05.pdf>.
- García, M. y Gamallo, P. (2011). Resolución de Correferencia de Nombres de Persona para Extracción de Información Biográfica. *Procesamiento del Lenguaje Natural*, 47, 47-55. Recuperado de <http://sinai.ujaen.es/sepln/ojs/ojs/index.php/pln/article/download/965/718>.
- Garrido García, N. y Ruiz de Mendoza Ibáñez, F. J. (2011). La modelación del conocimiento procedimental en el Cognición de FunGramKB. *ANGLOGERMANICA ONLINE*, 8, 106-120. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb07.pdf>.
- Gómez Guinovart, X. (2000). Perspectivas de la lingüística computacional. *Novática: Revista de la Asociación de Técnicos de Informática*, 145, 85-87. Recuperado de <http://www.ati.es/novatica/2000/145/javgom-145.pdf>.
- González Vergara, C. (2006a). La Gramática del Papel y la Referencia: una aproximación al modelo. *Onomázein*, 14(2), 101-140. Recuperado de <http://www.onomazein.net/14/14-3.pdf>.
- . (2006b). *Las construcciones no reflexivas con “se”: una propuesta desde la*

Gramática del Papel y la Referencia. (Tesis doctoral, Universidad Complutense de Madrid, Madrid, España). Recuperada de <http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/Las%20construcciones%20no%20reflexivas%20con%20se.pdf>.

———. (2009). One rule to rule them all: logical structures for Spanish non reflexive se sentences. En Guerrero, L., Ibáñez Cerda, S. y Belloro, V. A. (eds.), *Studies in Role and Reference Grammar*, (pp. 361-380). México: UNAM.

Greenbaum, S. y Randolph, Q. (1990). *A Student's Grammar of the English Language*. Londres: Longman.

Guerra García, F. y Sacramento Lechado, E. (2011). El módulo léxico de FunGramKB. *ANGLOGERMANICA ONLINE*, 8, 52-65. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb04.pdf>.

Halliday, M. A. K. y Hasan, R. (1976). *Cohesion in English*. Londres: Longman.

Hendrickx, I., De Clercq, O. y Hoste, V. (2011). Analysis and Reference Resolution of Bridge Anaphora across Different Text Genres. En Hendrickx, I., Devi, S. L., Branco, A. y Mitkov, R. (eds.), *Anaphora Processing and Applications. DAARC 2011. LNCS (LNAI), 7099*, (pp. 1-11). Berlín: Springer-Verlag. Recuperado de <https://biblio.ugent.be/input/download?func=downloadFile&recordId=2000619&fileId=2129088>.

Hendrickx, I., Hoste, V. y Daelemans, W. (2007). Evaluating hybrid versus data-driven coreference resolution. En Branco, A. (ed.), *Anaphora: Analysis, Algorithms and Applications. DAARC 2007. LNCS (LNAI), 4410*, (pp. 137-150). Berlín: Springer-Verlag. Recuperado de <http://www.cnts.ua.ac.be/~iris/corea/publications/daarc2007.pdf>.

Hirst, G. (1981). *Anaphora in Natural Language Understanding*. Berlín: Springer-Verlag.

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

- Instituto Cervantes. 2006. *Plan curricular del Instituto Cervantes. Niveles de referencia para el español*. Madrid: Biblioteca nueva. Recuperado de http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/default.htm.
- Jegatha-Deborah, L., Baskaran, R. y Kannan, A. (2011). Visualizing Domain Ontology using Enhanced Anaphora Resolution Algorithm. *International Journal of Database Management Systems*, 3(3), 110-123. Recuperado de <http://airccse.org/journal/ijdms/papers/3311ijdms10.pdf>.
- Jiménez Briones, R. (2009). Lexical templates for the Spanish verbs of feeling: A further elaboration of RRG logical structure. En Guerrero, L., Ibáñez Cerda, S. y Bello, V. A. (eds.), *Studies in Role and Reference Grammar*, (pp. 381-403). México: UNAM.
- Jiménez Briones, R. y Luzondo Oyón, A. (2011). Building ontological meaning in a lexico-conceptual knowledge base. *Onomázein*, 23(1), 11-40. Recuperado de <http://www.onomazein.net/23/23-01.pdf>.
- Jiménez Briones, R., Luzondo Oyón, A. y Pérez Cabello de Alba, M. B. (2011). FunGramKB y la organización ontológica. *ANGLOGERMANICA ONLINE*, 8, 16-36. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb02.pdf>.
- Jiménez Briones, R. y Pérez Cabello de Alba, M. B. (2011). An Account of Selection Restrictions in Role and Reference Grammar. *Revista Canaria de Estudios Ingleses*, 62, 99-122. Recuperado de <http://www.fungramkb.com/resources/papers/015.pdf>.
- Kimm, N. y Horn, C. (2011). Nominal Associative Anaphors – A Text-Based Analysis at the Semantics-Pragmatics Interface. En Hendrickx, I., Devi, S. L., Branco, A. y Mitkov, R. (eds.), *Anaphora Processing and Applications. DAARC 2011. LNCS (LNAI), 7099*, (pp. 108-118). Berlín: Springer-Verlag.
- Klenner, M., Tuggener, D., Fahrni A. y Sennrich R. (2010). Anaphora Resolution with Real Preprocessing. En Loftsson, H., Rögnvaldsson, E. y Helgadóttir, S. (eds.),

- Advances in Natural Language Processing. LNCS (LNAI), 6233*, (pp. 215-225). Berlín: Springer-Verlag.
- Kuno, S. (1987). *Functional Syntax. Anaphora, Discourse and Empathy*. Chicago: The University of Chicago Press.
- Lassalle, E. y Denis, P. (2011). Leveraging different meronym discovery methods for bridging resolution in French. En Hendrickx, I., Devi, S. L., Branco, A. y Mitkov, R. (eds.), *Anaphora Processing and Applications. DAARC 2011. LNCS (LNAI), 7099*, (pp. 35-46). Berlín: Springer-Verlag. Recuperado de <http://researchers.lille.inria.fr/~pdenis/papers/daarc11.pdf>.
- Lech, T. C. y de Smedt, K. (2007). Ontology extraction for coreference chaining. En Johansson, C. (ed.), *Proceedings from the first Bergen Workshop on Anaphora Resolution WAR I*, (pp. 26-38). Newcastle upon Tyne: Cambridge Scholars Publishing. Recuperado de <http://folk.uib.no/hfosm/papers/war.pdf>.
- Liddy, E. D. (2001). Natural language processing. En *Encyclopedia of Library and Information Science*. Nueva York: Marcel Decker.
- Llisterri, J. (2003). Lingüística y tecnologías del lenguaje. *Lynx. Panorámica de Estudios Lingüísticos* 2, 9-71. Recuperado de http://liceu.uab.es/~joaquim/publicacions/Llisterri_03_Linguistica_Tecnologias_Lenguaje.pdf.
- Mairal Usón, R. y Cortés Rodríguez, F. (2006). An overview of Role and Reference Grammar. En Mairal Usón, R., Escobar Álvarez, M. A., Peña Cervel, M. S., Samaniego Fernández, E. y Cortés Rodríguez, F. (eds.), *Current Trends in Linguistic Theory*, (pp. 97-176). Madrid: UNED.
- Martínez Barco, P. (1999). Algoritmo de resolución de la anáfora pronominal en diálogos. *Procesamiento del lenguaje natural*, 24, 75-84. Recuperado de <http://www.sepln.org/revistaSEPLN/revista/24/24-articulo8.pdf>.
- Mitkov, R. (2002). *Anaphora resolution*. Londres: Longman.

- . (2003). Anaphora resolution. En Mitkov, R. (ed.), *The Oxford Handbook of Computational Linguistics*, (pp. 266-283). Oxford: Oxford University Press.
- . (2010). Discourse Processing. En Clark, A., Fox, C. y Lappin, S. (eds.), *Computational Linguistics and Natural Language Processing*, (pp. 599-629). Singapur: Wiley-Blackwell.
- Moreno Cabrera, J. C. (1998). Retórica, texto y lingüística contemporánea: de la anáfora a la elipsis y de Chomsky a Montague. En Albaladejo, T., Caballero, J. A., y del Río, E. (eds.), *Quintiliano, historia y actualidad de la retórica, I*, (pp. 358-366). Logroño: Ediciones Instituto de Estudios Riojanos.
- Moure, T. y Llisterri, J. (1996). Lenguaje y nuevas tecnologías. El campo de la lingüística computacional. En Fernández Pérez, M. (ed.), *Avances en lingüística aplicada*, pp. 147-228. Santiago de Compostela: Universidade de Santiago de Compostela.
- Novák, M. y Žabokrtský, Z. (2011). Resolving Noun Phrase Coreference in Czech. En Devi, S. L., Branco, A. y Mitkov R. (eds.), *Anaphora Processing and Applications. DAARC 2009. LNCS (LNAI)*, 5847. (pp. 24-34). Berlín: Springer-Verlag. Recuperado de http://ufal.mff.cuni.cz/~mnovak/papers/DAARC_2011_paper.pdf.
- Palomar, M., Moreno, L., Peral, J., Muñoz, R., Ferrández, A., Martínez-Barco, P. y Saiz-Noeda, M. (2001). An algorithm for anaphora resolution in Spanish texts. *Computational Linguistics*, 27(4), 545-567. Recuperado de <http://www.mitpressjournals.org/doi/pdf/10.1162/089120101753342662>.
- Pavey, E. (2008). Predication and reference in specificational sentences – functions of noun phrases. En Van Valin, R. D. (ed.), *Investigations of the Syntax–Semantics–Pragmatics Interface. Studies in Language Companion Series*, 105. (pp. 305–317). Ámsterdam: John Benjamins.
- Peral, J., Palomar, M. y Ferrández, A. (1999). Coreference-oriented Interlingual Slot Structure & Machine Translation. En *Proceedings of the ACL Workshop Coreference and its Applications (ACL99 Workshop)*, (pp. 69-76). Maryland:

College Park. Recuperado de <http://acl.ldc.upenn.edu/W/W99/W99-0210.pdf>.

Periñán Pascual, C. (2005). Procesamiento del lenguaje natural: de lingüista a ingeniero del conocimiento. En Brady, I. K., Navarro Coy, M. y Periñán Pascual, C. (eds.), *Nuevas Tendencias en Lingüística Aplicada*, (pp. 293-317). Murcia: Quaderna.

———. (2012). The situated common-sense knowledge in FunGramKB. *Review of Cognitive Linguistics*, 10(1), 184-214. Recuperado de <http://www.fungramkb.com/resources/papers/020.pdf>.

Periñán Pascual, C. y Arcas Túnez, F. (2005). Microconceptual-Knowledge Spreading in FunGramKB. *En 9th IASTED International Conference on Artificial Intelligence and Soft Computing*, (pp. 239-244). Anaheim-Calgary-Zurich: ACTA Press. Recuperado de <http://www.fungramkb.com/resources/papers/002.pdf>.

———. (2007a). Cognitive Modules of an NLP Knowledge Base for Language Understanding. *Procesamiento del Lenguaje Natural*, 39, 197-204. Recuperado de <http://www.sepln.org/revistaSEPLN/revista/39/24.pdf>.

———. (2007b). Deep semantics in an NLP knowledge base. En Borrajo, D., Castillo, L. y Corchado, J. M. (eds.), *12th Conference of the Spanish Association for Artificial Intelligence*, (pp. 279-288). Salamanca: Universidad de Salamanca. Recuperado de <http://www.fungramkb.com/resources/papers/005.pdf>.

———. (2010). The Architecture of FunGramKB. En Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odjik, J., Piperidis, S., Rosner, M. y Tapias, D. (eds.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, (pp. 2667-2674). Valletta, Malta: European Language Resources Association (ELRA). Recuperado de <http://www.fungramkb.com/resources/papers/011.pdf>.

———. (2011). Introducción a FunGramKB. *ANGLOGERMANICA ONLINE*, 8, 1-15. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb01.pdf>.

- Periñán Pascual, C. y Carrión Varela, M. L. (2011). FungramKB y el Conocimiento Cultural. *ANGLOGERMANICA ONLINE*, 8, 87-105. Recuperado de <http://www.fungramkb.com/resources/papers/fgkb06.pdf>.
- Periñán Pascual, C. y Mairal Usón, R. (2009). Bringing Role and Reference Grammar to natural language understanding. *Procesamiento del Lenguaje Natural*, 43, 265-273. Recuperado de <http://www.sepln.org/revistaSEPLN/revista/43/articulos/art29.pdf>.
- . (2010a). Enhancing UniArab with FunGramKB. *Procesamiento del Lenguaje Natural*, 44, 19-26. Recuperado de <http://www.lexicom.es/drupal/files/010.pdf>.
- . (2010b). La gramática de COREL: un lenguaje de representación conceptual. *Onomázein*, 21, 11-45. Recuperado de <http://www.fungramkb.com/resources/papers/012.pdf>.
- . (2011). The COHERENT Methodology in FunGramKB. *Onomázein*, 24, 13-33. Recuperado de <http://www.onomazein.net/24/01.pdf>.
- Piñón, A. (16 de mayo de 2012). Murió el novelista, vive la obra. *El Universal*. Recuperado de <http://www.eluniversal.com.mx/espectaculos/113812.html>.
- Poesio, M., Uryupina, O., Vieira, R., Alexandrov-Kabadjov, M. y Goulart, R. (2004). Discourse-New Detectors for Definite Description Resolution: A Survey and a Preliminary Proposal. En Harabagiu, S. y Farwell, D. (eds.), *ACL 2004: Workshop on Reference Resolution and its Applications*, (p. 47-54). Barcelona: Association for Computational Linguistics. Recuperado de <http://acl.ldc.upenn.edu/W/W04/W04-0707.pdf>.
- Poesio, M. y Vieira, R. (1998). A Corpus-based Investigation of Definite Description Use. *Computational Linguistics*, 24(2), 183-216. Recuperado de <http://cswww.essex.ac.uk/staff/poesio/publications/CL98.pdf>.
- Poesio, M., Vieira, R. y Teufel, S. (1997). Resolving bridging references in unrestricted text. En Mitkov, R. (ed.), *Proceedings of the ACL'97/EACL'97 workshop on Operational factors in practical, robust anaphora resolution*, (p.

- 1-6). Madrid. Recuperado de <http://cswww.essex.ac.uk/staff/poesio/publications/ACL97-anaphora.pdf>.
- Real Academia Española. Banco de datos (CREA). *Corpus de referencia del español actual*. [En línea] <http://www.rae.es> [19.07.12].
- . (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Recasens, M. y Hovy, E. (2009). A Deeper Look into Features for Coreference Resolution. En Devi, L. S., Branco, A. y Mitkov R. (eds.), *Anaphora Processing and Applications. DAARC 2009. LNCS (LNAI)*, 5847. (pp. 29-42). Berlín: Springer-Verlag. Recuperado de <http://www.isi.edu/natural-language/people/hovy/papers/09DAARC-coref-final.pdf>.
- Recasens, M., Martí, M. A. y Taulé, M. (2007). Text as scene: discourse deixis and bridging relations. *Procesamiento del lenguaje natural*, 39, 205-212. Recuperado de http://rua.ua.es/dspace/bitstream/10045/3014/1/PLN_39_25.pdf.
- Rodríguez, H. (2000). Técnicas básicas en el tratamiento informático de la lengua. *Quark Ciencia, Medicina, Comunicación y Cultura*, 19, 26-34. Recuperado de <http://quark.prbb.org/19/019026.htm>.
- Ruiz de Mendoza Ibáñez, F. J. (2001). Lingüística cognitiva: semántica, pragmática y construcciones. *Círculo de lingüística aplicada a la comunicación*, 8, 1-32. Recuperado de <http://www.ucm.es/info/circulo/no8/ruiz.htm>.
- Ruiz de Mendoza Ibáñez, F. J. y Díez Velasco, O. (2004). Metonymic Motivation in Anaphoric Reference. En Günter R. y Panther, K. U. (eds), *Studies in Linguistic Motivation (Cognitive Linguistics Research)*, (pp. 293-320). Berlín y Nueva York: Mouton de Gruyter.
- Ruiz de Mendoza Ibáñez, F. J. y Mairal Usón, R. (2007). Levels of semantic representation: where lexicon and grammar meet. *Interlingüística*, 17, 26-47. Recuperado de

http://dialnet.unirioja.es/servlet/fichero_articulo?codigo=2316363&orden=84250

- . (2008). Levels of description and constraining factors in meaning construction: an introduction to the Lexical Constructional Model. *Folia Linguistica* 42(2), 355-400. Recuperado de http://www.lexicom.es/drupal/files/RM_Mairal_2008_Folia_Linguistica.pdf.
- . (2009). Levels of description and explanation in meaning construction. En Butler, C. S. y Martín Arista, J. (eds.), *Deconstructing Constructions*, (pp. 153-198). Ámsterdam/Filadelfia: John Benjamins. Recuperado de <http://www.fungramkb.com/resources/papers/013.pdf>.
- Ruiz de Mendoza Ibáñez, F. J. y Otal Campo, J. L. (2002). *Metonymy, grammar and communication*. Granada: Comares.
- Saiz Noeda, M. (2002). Influencia y aplicación de papeles sintácticos e información semántica en la resolución de la anáfora pronominal en español. *Procesamiento del Lenguaje Natural*, 28, 113-114. Recuperado de http://gplsi.dlsi.ua.es/gplsi11/sites/default/files/tesis_maximiliano.pdf.
- Salem, Y., Hensman, A. y Nolan, B. (2008). Towards Arabic to English machine translation. *ITB Journal* 17, 20-31. Recuperado de http://members.upc.ie/y.salon/pdf/Towards%20Arabic%20to%20English%20_ITB%20Journal-May-2008.pdf.
- Sasano, R., Kawahara, D., Kurohashi, S. (2007). Improving coreference resolution using bridging reference resolution and automatically acquired synonyms. En Branco, A. (ed.), *Anaphora: Analysis, Algorithms and Applications. DAARC 2007. LNCS (LNAI), 4410*, (pp. 125-136). Berlín: Springer-Verlag. Recuperado de <http://www.lr.pi.titech.ac.jp/~sasano/paper/DAARC07.pdf>.
- Shimojo, M. (2009). Focus structure and beyond: discourse-pragmatics in RRG. En Guerrero, L., Ibáñez Cerda, S. y Belloro, V. A. (eds.), *Studies in Role and Reference Grammar*, (pp. 113-142). México: UNAM.

- Uryupina, O. (2009). Detecting Anaphoricity and Antecedenthood for Coreference Resolution. *Procesamiento del Lenguaje Natural*, 42, 113-120. Recuperado de <http://www.sepln.org/revistaSEPLN/revista/42/03Articulos-p22-113a120.pdf>.
- Van Valin, R. D. (1990). Functionalism, anaphora and syntax. *Studies in Language*, 14(1), 169-219. Recuperado de http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg/Studies_in_Language_Scans.pdf.
- . (2001). *An introduction to Syntax*. Cambridge: Cambridge University Press.
- . (2005). *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.
- . (2009). Privileged syntactic arguments, pivots, and controllers. En Guerrero, L., Ibáñez Cerda, S. y Belloro, V. A. (eds.), *Studies in Role and Reference Grammar*, (pp. 45-68). México: UNAM.
- Van Valin, R. D. y R. J. LaPolla. (1997). *Syntax: Structure, Meaning and Function*. Cambridge: Cambridge University Press.
- Versley, Y. (2007). Using the web to resolve coreferent bridging in German newspaper text. En Rehm, G., (ed.), *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of GLDV-Frühjahrstagung 2007*, (pp. 253-262). Tübinga: Narr. Recuperado de <http://www.versley.de/gldv07.pdf>.