

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN



**“Llenado de una Base de Patrones para
la Valencia de los Verbos del Idioma Español”**

**PARA OBTENER EL GRADO DE:
MAESTRO EN CIENCIAS EN
CIENCIAS DE LA COMPUTACIÓN**

P R E S E N T A:

I.S.C. ALBERTO CASTRO HERNÁNDEZ

D I R E C T O R:

DR. JOSÉ ANTONIO MARTÍNEZ FLORES

C O D I R E C T O R:

DR. RODOLFO ABRAHAM PAZOS RANGEL

J U R A D O:

Presidente: **DR. JOSÉ ANTONIO MARTÍNEZ FLORES**
Secretario: **M.C. CLAUDIA GUADALUPE GÓMEZ SANTILLÁN**
Vocal: **DR. JUAN JAVIER GONZÁLEZ BARBOSA**
Suplente: **DR. HÉCTOR JOAQUÍN FRAIRE HUACUJA**



"2006, Año del Bicentenario del Natalicio del Benemerito de las Américas, Don Benito Juárez García"

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

COORDINACIÓN DE POSGRADO EN
COMPUTACIÓN
U5.498/06
AUTORIZACIÓN DE IMPRESIÓN
DE TESIS DE GRADO

2006-NOV-27

C. ING. ALBERTO CASTRO HERNÁNDEZ
Presente.

Me es grato comunicarle que después de la revisión realizada por el Jurado designado para su examen de grado de Maestro en Ciencias en Ciencias de la Computación, se acordó autorizar la impresión de su tesis titulada:

**"LLENADO DE UNA BASE DE PATRONES PARA LA VALENCIA DE LOS VERBOS
DEL IDIOMA ESPAÑOL"**

Es muy satisfactorio para la División de Estudios de Posgrado e Investigación compartir con Usted el logro de esta meta. Espero que continúe con éxito su desarrollo profesional y dedique su experiencia e inteligencia en beneficio de México.

ATENTAMENTE
"POR MI PATRIA Y POR MI BIEN"

M. Yolanda Chávez Cinco
M.P. MARÍA YOLANDA CHÁVEZ CINCO
JEFA DE LA DIVISIÓN



MYCHC ' NLCO ' cerc*

Av. 1° de Mayo s/n Pte. Col. Los Mangos
Cd. Madero, Tam. C. P. 89440
Internet : www.itcm.edu.mx
Correo Electrónico : itcm@itcm.edu.mx

Teléfono : 3 - 57 - 48 - 20 al 29
Fax : 3 - 57 - 48 - 20 ext. 1002

DEDICATORIA

Dedico este trabajo a mis cinco ángeles guardianes: mi padre Rafael Castro Rodríguez, mi madre Ana María Hernández Ramírez, mi hermano Rafael Ignacio Castro Hernández, mi tía Luz Hernández Ramírez y mi tío Francisco Hernández Ramírez.

AGRADECIMIENTOS

Agradezco a Dios por haberme dado la vida, por haberme mostrado el camino hacia él, y por la oportunidad de lograr esta meta.

Un agradecimiento especial a la Santísima Virgen María, a Jesucristo y a San Judas Tadeo, por brindarme fortaleza en los momentos difíciles.

Agradezco a mi familia, por todo el amor, apoyo, comprensión, sabiduría y paciencia que me han brindado en todo momento.

Mi agradecimiento a los miembros del comité tutorial de esta tesis: Dr. José Antonio Martínez Flores, M.C. Claudia Guadalupe Gómez Santillán, Dr. Juan Javier González Barbosa y Dr. Héctor Joaquín Fraire Huacuja.

Agradezco al Dr. José Antonio Martínez Flores, al Dr. Rodolfo Abraham Pazos y al Dr. Juan Javier González Barbosa por su apoyo brindado en el desarrollo de esta tesis.

Mi cariño y agradecimiento a mis hermanos de maestría: Carlos Camacho, Pedro Torres, Cristina Cruz, Georgina Castillo, Víctor Álvarez y Arturo Herrera, por su gran amistad y apoyo en esta etapa tan importante de mi vida.

Mi sincero agradecimiento a todos mis amigos, por brindarme una palabra, un momento, un consejo cuando más lo necesité. En especial, agradezco a Luis Omar del Angel, Eder Hernández, Fernando Balderas y Gabriela Cardel, por ofrecerme su invaluable amistad.

Agradezco el apoyo brindado por el Consejo Nacional de Ciencia y Tecnología (CONACYT), y la Dirección General de Educación Superior Tecnológica (DGEST).

RESUMEN

En el Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) se desarrolló un analizador sintáctico para una interfaz de lenguaje natural para consulta a bases de datos en español. Dicho analizador usa una base de patrones para la valencia de los verbos, la cual no contiene suficiente información lingüística para el análisis sintáctico. Sin embargo, la población manual de una base de patrones es una tarea ardua que requiere mucho tiempo.

En este trabajo se aborda el tema del llenado de una base de patrones para la valencia de los verbos del idioma español; específicamente, dado un conjunto de verbos del idioma español y sus respectivos sentidos se desea obtener sus patrones de manejo. En esta tesis se propone un método semiautomático de llenado.

El método utiliza los patrones de manejo de un conjunto de verbos base para adquirir los patrones de manejo de otros. La relación entre los verbos base y otros verbos se obtuvo de la ontología Spanish WordNet, la cual contiene relaciones semánticas entre las palabras. Además, se aprovechó la relación que existe entre el patrón de manejo y la clase funcional de los verbos. Esta información se utilizó para generar los patrones de manejo. Para evaluar el método se elaboraron manualmente y validaron 239 patrones de manejo. Este conocimiento lingüístico puede ser utilizado en otras investigaciones relacionadas a los patrones de manejo.

Los resultados obtenidos al generar los patrones de manejo de verbos, utilizando el método anterior, tienen un porcentaje de acierto de 89.38%.

Este trabajo aporta un nuevo método para la generación de los patrones de manejo de verbos, el cual utiliza las relaciones semánticas que existen entre ellos y su clase funcional.

Debido al éxito de los resultados obtenidos, se propone realizar un método semejante para encontrar las valencias de otras clases de palabras, es decir, sustantivos y adjetivos.

SUMMARY

At the Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), a parser for a natural language interface to databases in Spanish was developed. This parser uses a pattern database for verbs valence, which does not contain enough linguistic information for parsing. However, the manual population of a pattern database is an arduous task which requires a great amount of time.

This work focuses on the topic of populating a pattern database for the valence of Spanish verbs; specifically, given a set of Spanish verbs and its respective senses, one wishes to obtain its government patterns. In this thesis a semiautomatic population method is proposed.

The method uses the government patterns of a set of base verbs in order to acquire the government patterns of other verbs. The relationship among base verbs and the other verbs was obtained from the ontology Spanish WordNet, which contains semantic relations among words. Additionally, the relation between the government pattern and the functional class of verbs was exploited. This information was used for generating the government patterns. For evaluating the method, 239 government patterns were manually developed and validated. This linguistic knowledge can be used in other investigations on government patterns.

The results obtained when generating the verbal government patterns, using the method described above, had an accuracy percentage of 89.38%.

The contribution of this work consists of a new method for generating verbal government patterns, which uses the semantic relations among verbs and their functional class.

Due to the success of the results obtained, a similar method is proposed for generating the valence of other parts of speech, i.e., nouns and adjectives.

TABLA DE CONTENIDO

TABLA DE CONTENIDO	i
LISTA DE FIGURAS	iii
LISTA DE TABLAS	iv
Capítulo 1 Introducción	1
1.1. Antecedentes.....	1
1.2. Descripción del Problema.....	3
1.3. Objetivo de la Tesis	3
1.4. Justificación.....	4
1.5. Alcances y Limitaciones de la Investigación	4
1.6. Estructura de la tesis.....	5
Capítulo 2 Marco Teórico	6
2.1. Procesamiento de Lenguaje Natural	6
2.1.1. Análisis Sintáctico	9
2.1.2. Modelos de Análisis Sintáctico	10
2.1.3. Teoría Texto-Significado.....	12
2.2. Subcategorización.....	16
2.2.1. Valencia	16
2.2.2. Valencias Existentes para los Verbos.....	18
2.2.3. Valencias con Respecto a la Clase Funcional Verbal.....	27
2.2.4. Representación Formal de las Valencias	29
2.3. WordNet	32
2.3.1. Ontología	32
2.3.2. WordNet	32
2.3.3. Spanish WordNet.....	36
2.4. Base de Patrones de CENIDET	38
2.4.1. Tablas	38
2.4.2. Relaciones.....	39
2.5. Trabajos Relacionados.....	39

Capítulo 3 Método de Solución	45
3.1. Definición de la Estructura del Patrón de Manejo Verbal	45
3.2. Análisis de la Ontología Spanish WordNet 1.0	47
3.2.1. Conjunto de Palabras como Elemento de un <i>Synset</i>	49
3.2.2. Obtención de los Significados para cada <i>Synset</i>	49
3.2.3. <i>Synset</i> sin ninguna palabra	51
3.2.4. Hipónimos Relacionados con más de un Hiperónimo	51
3.3. La Animidad en el Complemento Directo	51
3.4. Elaboración de los Patrones de Manejo Verbal Base	52
3.5. Modificaciones en la Base de Patrones	53
3.6. Algoritmo del Módulo de Llenado	54
Capítulo 4 Experimentación	60
4.1. Casos de Prueba	60
4.2. Experimentación	61
4.2.1. Resultados de Experimento 1	62
4.2.2. Resultados del Experimento 2	65
4.2.3. Resultados del Experimento 3	66
4.2.4. Análisis de los Resultados	67
Capítulo 5 Conclusiones	70
5.1. Conclusiones	70
5.2. Trabajos Futuros	71
Referencias	73
Anexo A	77
Anexo B	81

LISTA DE FIGURAS

Figura 1. Transferencia del conocimiento mediante un lenguaje.....	7
Figura 2. Ejemplo de una oración con el modelo de constituyentes	11
Figura 3. Ejemplo de una oración con el modelo de dependencias.....	12
Figura 4. Niveles de la Teoría Texto-Significado.	13
Figura 5. Ejemplo de entrada en un ECD.....	15
Figura 6. Ejemplo de patrón de manejo sintáctico del verbo <i>solicitar</i>	31
Figura 7. Base de patrones de CENIDET.....	39
Figura 8. Zona semántica del PMV del verbo inhalar con el sentido 1.....	46
Figura 9. Zona sintáctica del PMV del verbo/sentido inhalar ₁	46
Figura 10. PMV del verbo/sentido inhalar ₁	47
Figura 11. Ejemplo de relación semántica de hiponimia verbal en Spanish WordNet 1.0	48
Figura 12. Ejemplo de un conjunto de palabras como elemento de un synset.....	49
Figura 13. Significado de los synsets	50
Figura 14. Ejemplo de un hipónimo que tiene más de un hiperónimo.....	51
Figura 15. Submódulo de llenado automático total.....	55
Figura 16. Submódulo de llenado automático por niveles.	56
Figura 17. Módulo de consulta de verbo con sentido.....	56
Figura 18. Módulo de consulta, asignación y elaboración de patrones - parte 1.....	57
Figura 19. Módulo de consulta, asignación y elaboración de patrones - parte 2.....	58
Figura 20. Casos de prueba.	61

LISTA DE TABLAS

Tabla 1. Descripción de los niveles de análisis en el PLN.....	8
Tabla 2. Información de subcategorización obtenida mediante la clase funcional	29
Tabla 3. Ejemplo de la tabla spWN1.0_variants	36
Tabla 4. Ejemplo de la tabla spWN1.0_synsets	37
Tabla 5. Ejemplo de la tabla spWN1.0 relations	38
Tabla 6. Características de los trabajos relacionados y este trabajo.....	44
Tabla 7. Relación de profundidad y palabras en Spanish WordNet.....	49
Tabla 8. Resultados obtenidos del experimento 1	68
Tabla 9. Resultados obtenidos del experimento 2	68
Tabla 10. Resultados obtenidos del experimento 3	69
Tabla 11. Porcentaje de acierto de los experimentos.....	69

Capítulo 1

Introducción

En este capítulo se presenta una perspectiva general de este trabajo de investigación, en el cual se menciona la descripción del problema abordado, los objetivos planteados y los alcances de la investigación. El capítulo finaliza con una descripción del contenido de cada capítulo de la tesis.

1.1. Antecedentes

Uno de los grandes patrimonios de los seres humanos es el conocimiento, que es el resultado del aprendizaje. La mayoría de éste lo vemos representado en diferentes medios a través del lenguaje propio, porque es la forma de expresarnos. Uno de estos medios es la comunicación escrita.

En la actualidad, el uso de computadoras es importante ya que facilita en gran medida las tareas cotidianas; además, de que son almacenes de grandes cantidades de datos que representan nuestro conocimiento.

Desafortunadamente, estas máquinas necesitan recibir instrucciones en un lenguaje formal y diferente al común utilizado por las personas. Por tal motivo se requiere la capacitación de personal en un lenguaje formal para que puedan interactuar con las computadoras.

Un ejemplo de lenguaje en sistemas de cómputo es SQL (Structured Query Language) que permite la extracción de información.

La manera ideal es que las personas pudieran comunicarse con las computadoras a través de su lenguaje natural.

Este tema ha sido un asunto de interés para los investigadores de la computación del área denominada procesamiento de lenguaje natural (PLN), disciplina que busca utilizar el lenguaje común para la interacción humano-máquina. Esta disciplina ha dado origen a las interfaces de lenguaje natural (ILN), que son sistemas cuya función es interpretar una solicitud en lenguaje natural realizada por un usuario.

El procesamiento de lenguaje natural es utilizado para diferentes aplicaciones, algunas de ellas son: filtrado, clasificación, recuperación de información, extracción de información y resúmenes, reconocimiento de voz y traducción automática.

El procesamiento de lenguaje natural requiere un análisis amplio, de manera que se ha dividido en varios, algunos son: análisis léxico, análisis morfológico, análisis sintáctico y análisis semántico; cada uno dependiente del anterior. Aunque existen situaciones en las cuales comparten herramientas dentro de sus procesos.

Una de estas herramientas es la base de patrones, la cual guarda la valencia de las palabras, que es el número de elementos regidos por una palabra; y el patrón de manejo de las palabras, el cual es una secuencia de elementos sintácticos basada en la valencia de la palabra. Esta información ayuda a interpretar las consultas formuladas por el usuario.

El Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) desarrolló un analizador sintáctico para una interfaz de lenguaje natural, el cual utiliza una base de patrones.

Desafortunadamente, la base de patrones para la valencia de los verbos que se usa en dicho analizador, no contiene suficiente información lingüística para el análisis sintáctico.

El presente trabajo trata precisamente sobre el tema de una base de patrones, en particular la implementación de un módulo para llenar una base de este tipo para la valencia de los verbos en el idioma español.

1.2. Descripción del Problema

El procesamiento de Lenguaje Natural utiliza herramientas de análisis; una de ellas es el análisis sintáctico. Para este análisis existen diferentes modelos. Los más utilizados son el modelo de constituyentes y el modelo de dependencias. Dentro de los modelos de dependencias se encuentra la Teoría de Texto-Significado (en inglés MTT, Meaning-Text Theory). El uso de la MTT permite describir algunas características del español, tal como el establecimiento de relaciones entre valencias sintácticas y semánticas.

En este método se presenta la necesidad de crear una base de conocimiento lingüístico que contenga los patrones de manejo de las valencias de los verbos, que permitan conocer cómo se relacionan las palabras en una oración.

Como se mencionó anteriormente, el CENIDET desarrolló un analizador sintáctico para una interfaz de lenguaje natural, el cual utiliza el modelo de dependencias para el análisis sintáctico, haciendo uso de la MTT. Dicho analizador usa una base de patrones para la valencia de los verbos, la cual no contiene suficiente información lingüística para el análisis sintáctico.

1.3. Objetivo de la Tesis

Desarrollar un método para llenar una base de patrones para la valencia de los verbos del idioma español, que sirva como conocimiento lingüístico en el análisis sintáctico de una oración.

Para lograr este fin se propuso alcanzar los siguiente objetivos específicos:

- Elaborar los patrones de manejo de los verbos *base* de manera manual.
- Desarrollar un procedimiento para obtener los patrones de nuevos verbos, a partir de verbos que ya existan en la base de patrones y que sean afines en significado.

1.4. Justificación

La extracción de patrones es una herramienta importante para el análisis sintáctico, además de proporcionar información semántica útil para análisis posteriores [1].

Como se muestra en [2], se señala que la predicción de la preferencia de subcategorización de un verbo puede obtenerse de mejor manera, si se toma en cuenta el sentido del verbo. Por ejemplo, en (1), (2) y (3) se presentan tres oraciones con el verbo *empezar*, cada oración está relacionada con un sentido diferente en la ontología WordNet.

- (1) Empezar [CD a trabajar]
- (2) Empezar [CD una batalla]
- (3) El día empezó.

En (1), la acepción del verbo es *iniciar una acción o actividad* y lleva un Complemento Directo (CD), que está formado por la preposición *a* y el verbo en infinitivo *trabajar*. Por otra parte en (2), la acepción del verbo es *poner en marcha o provocar que inicie* y lleva un CD que es un sintagma nominal (*una batalla*). Mientras que en (3), la acepción del verbo es *tener un inicio* y no lleva un CD.

Por lo que se concluye que existe correlación entre la acepción del verbo y su subcategorización.

Mediante este trabajo se espera contribuir con la elaboración de un procedimiento que permita extraer información sobre los patrones de los verbos de manera semiautomática.

1.5. Alcances y Limitaciones de la Investigación

- Debido al gran número de verbos que existen en el idioma español, se decidió identificar la valencia de 44 verbos *base* que sirvan como base del conocimiento lingüístico y la valencia de verbos que sirvan en la evaluación del módulo.

- La valencia semántica es establecida con base en las acepciones contenidas en MultiWordNet.
- Las acepciones contenidas en MultiWordNet se encuentran en el idioma inglés.
- Las oraciones analizadas para conocer los patrones de manejo de los verbos son en el idioma español.
- Las relaciones semánticas utilizadas en este trabajo se obtuvieron de Spanish WordNet 1.0.
- La información obtenida por el *módulo de llenado* se creó en base al formato utilizado en la base de patrones del Centro Nacional de Investigación y Desarrollo Tecnológico.

1.6. Estructura de la tesis

En el capítulo 2 se presenta el marco teórico que fundamenta esta investigación, el cual incluye los fundamentos sobre procesamiento de lenguaje natural, la subcategorización verbal, la ontología WordNet, la base de patrones diseñada en el CENIDET y los trabajos relacionados con este trabajo. En el capítulo 3 se describe el procedimiento de solución propuesto, el cual se basa en un método semiautomático de adquisición de patrones de manejo de verbos. En el capítulo 4 se describen los experimentos realizados y el análisis de los resultados obtenidos para determinar el porcentaje de acierto del enfoque propuesto.. Finalmente, en el capítulo 5 se presentan las contribuciones de esta investigación, y se describen los posibles trabajos futuros identificados en el desarrollo del trabajo realizado.

Capítulo 2

Marco Teórico

En este capítulo se presenta el marco teórico que fundamenta esta investigación. El marco teórico incluye los fundamentos sobre procesamiento de lenguaje natural, la subcategorización verbal, la ontología WordNet, la base de patrones diseñada en el CENIDET y los trabajos relacionados con este trabajo.

2.1. Procesamiento de Lenguaje Natural

Un lenguaje es un conjunto de señales y sonidos que utilizan los seres y las máquinas con la intención de comunicarse. Existen dos clases de lenguajes: los lenguajes naturales, por ejemplo: español, inglés, francés, etcétera; y los lenguajes formales, por ejemplo lógico, matemático, SQL, etcétera [3].

El lenguaje natural es el medio que utilizamos para establecer una comunicación con las demás personas. Este lenguaje permite designar las cosas y razonar sobre ellas. Fue desarrollado y organizado a partir de la experiencia humana y puede ser utilizado para analizar situaciones complejas. Una propiedad de estos lenguajes es la posibilidad de que una palabra u oración tenga diferentes significados [3].

Por otro lado, el lenguaje formal es el que el hombre ha desarrollado para expresar las situaciones específicas en un área de conocimiento científico o tecnológico.

Los lenguajes formales están exentos de cualquier componente semántico fuera de sus operadores y relaciones. Pueden ser utilizados para modelar en una teoría de cualquier ciencia, por su ventaja de que no tienen ambigüedad [3].

El propósito de un lenguaje natural es transferir conocimientos de una persona a otra. El conocimiento es una estructura compleja, multidimensional, que usualmente se representa como una red, o grafo, de conceptos. Pero el modo que usamos para transferir el conocimiento es unidimensional, esto es, un sonido o una letra. Entonces, un objetivo del lenguaje natural es codificar el conocimiento en una cadena de símbolos. Posteriormente, en el cerebro del escuchante o el lector, esta secuencia se decodifica en el conocimiento original, como se muestra en la Figura 1 [4].

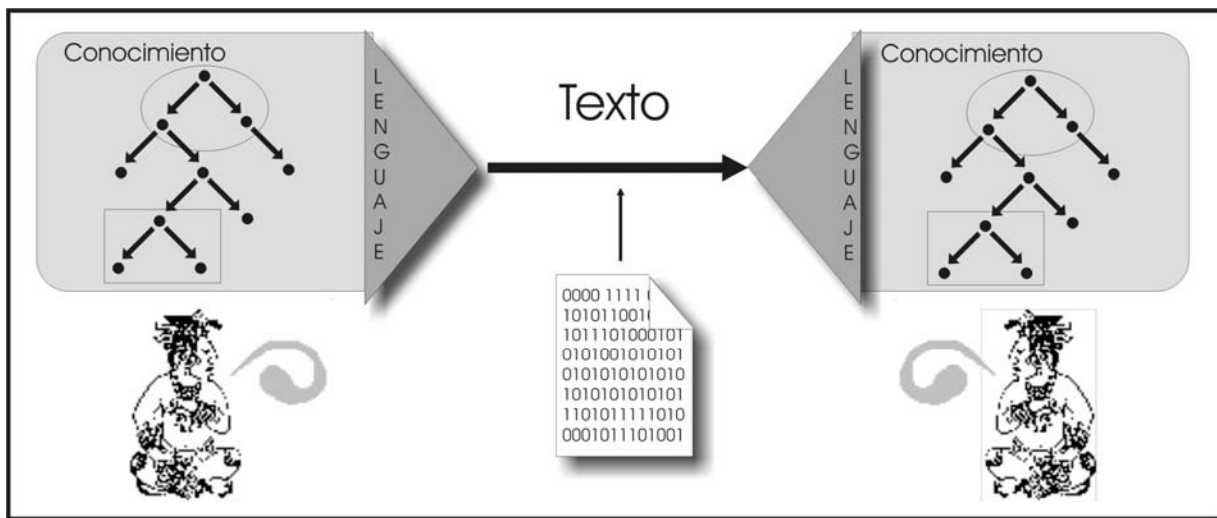


Figura 1. Transferencia del conocimiento mediante un lenguaje.

Si el escuchante o el lector es una computadora, ésta debe tener el conocimiento suficiente sobre el lenguaje para entender el mensaje que le sea proporcionado.

Lograr que una computadora ‘comprenda’ lo que una persona dice o escribe es algo complejo. Esta problemática se aborda en el Procesamiento del Lenguaje Natural (PLN), el cual es una parte de la Inteligencia Artificial, que investiga y formula mecanismos computacionales efectivos, cuyo objetivo es facilitar la interrelación hombre-máquina y permitir una comunicación mucho más fluida y menos rígida que los lenguajes formales [5].

Para lograr este objetivo es necesario utilizar conocimiento sobre el lenguaje, el cual puede ser separado en seis niveles [6].

- Fonética y fonología — El estudio de sonidos lingüísticos.
- Morfología — El estudio de los componentes significativos de las palabras.
- Sintaxis — El estudio de las relaciones estructurales entre las palabras.
- Semántica — El estudio del significado.
- Pragmática — El estudio de cómo el lenguaje es usado para alcanzar objetivos.
- Discurso — El estudio de las unidades lingüísticas más grandes que una única expresión.

De la misma forma, en un sistema computacional existen diferentes niveles de análisis, donde cada nivel tiene un formato de entrada y salida. Los tipos de estructura mostrados en la Tabla 1 son los que se utilizan comúnmente [7].

Tabla 1. Descripción de los niveles de análisis en el PLN.

Niveles de análisis para una oración en un sistema computacional				
Análisis	Entrada	Salida	Estructura utilizada	Cómo lo reconoce
Morfológico-Léxico	Caracteres	Unidades con significado	Diccionario y reglas morfológicas	Las convierte de cadenas de letras a números en el diccionario y a etiquetas de tiempo, género, número, etcétera.
Sintáctico	Unidades léxicas o representación morfológica	Representación de su estructura	Árboles y autómatas	Las convierte de cadenas de palabras etiquetadas en estructuras gráficas de oraciones, sujeto, objeto, etcétera.
Semántico	Estructuras sintácticas	Forma lógica	Redes semánticas	Lo convierte a una red semántica. Resuelve las relaciones entre los pronombres y sus antecedentes.

El objetivo del análisis morfológico consiste en detectar la relación que se establece entre las unidades mínimas que forman una palabra. Este nivel de análisis mantiene una estrecha relación con el léxico.

Igualmente, el léxico es el conjunto de información sobre cada palabra que el sistema utiliza para el procesamiento. La información incluye la información morfológica, la categoría gramatical, irregularidades sintácticas y representación del significado.

Por otra parte, el análisis sintáctico tiene como función identificar cada uno de los componentes sintácticos que aparecen en la oración y analizar cómo las palabras se combinan para formar construcciones gramaticalmente correctas. El resultado de este proceso consiste en generar la estructura correspondiente a las categorías sintácticas formadas por cada una de las unidades léxicas que aparecen en la oración.

El análisis semántico consiste en examinar el significado de las oraciones. Primero, se hace referencia al significado que las palabras tienen por sí mismas, sin considerar el significado adquirido según el uso en una determinada circunstancia; y se genera una estructura semántica. Después, se interpreta la estructura anterior en su contexto para obtener una representación final y examinar su significado [8].

2.1.1. Análisis Sintáctico

Como se mencionó anteriormente, el análisis sintáctico se basa en las estructuras de las expresiones de un lenguaje, es decir, busca identificar cómo las palabras se relacionan entre sí. La información que se examina puede ser:

- Cómo se organizan las palabras en grupos o sintagmas.
- Qué palabras o grupos modifican a otras palabras o grupos.
- Qué palabras o grupos son los más importantes en la oración.

Los métodos empleados en el análisis sintáctico se clasifican en descendentes o ascendentes.

Análisis sintáctico descendente. Este método empieza a partir de la representación de la frase, para descomponerla en subconstituyentes hasta que se deriva cada palabra específica en la frase. En esta técnica se crea un árbol de análisis sintáctico desde la raíz hacia abajo.

Análisis sintáctico ascendente. En este tipo de análisis se inicia con las palabras individuales, las cuales se remplazan con su categoría sintáctica, creando un árbol de análisis sintáctico desde las hojas hacia arriba [9].

2.1.2. Modelos de Análisis Sintáctico

El análisis sintáctico establece métodos que determinen únicamente las secuencias gramaticales en el procesamiento de textos. Esto ha sido el objetivo de diferentes formalismos gramaticales, en los cuales se han considerado dos enfoques principalmente para describir formalmente la gramática de las oraciones: el modelo de constituyentes y el modelo de dependencias.

Modelo de constituyentes

Los constituyentes y la suposición de la estructura de frase, es el enfoque donde las oraciones se analizan mediante un proceso de segmentación y clasificación. Se segmenta la oración en sus partes constituyentes, se clasifican estas partes como categorías gramaticales, después se repite el proceso para cada parte dividiéndola en subconstituyentes, y así sucesivamente hasta que las partes sean las partes de la palabra indivisibles dentro de la gramática (morfemas) [10].

En la Figura 2 se presenta la oración (O) *Los jóvenes contemplan el bello amanecer*, la cual está compuesta por dos constituyentes: un sintagma nominal (SN) y un sintagma verbal (SV). El SN está, a su vez, compuesto por un determinante (d) y un sustantivo (s). Por otro lado, el SV está compuesto por un verbo (v) y un SN, que a su vez, está compuesto por un determinante, un adjetivo (a) y un sustantivo.

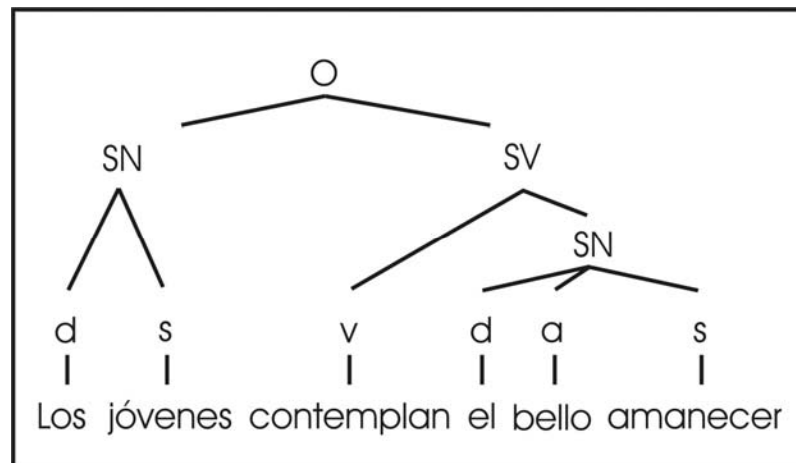


Figura 2. Ejemplo de una oración con el modelo de constituyentes

Modelo de dependencias

En este modelo, se establecen las dependencias entre pares de palabras, donde una es principal o rectora y la otra está subordinada a (o dependiente de) la primera. Si cada palabra de la oración tiene una palabra propia rectora, la oración entera se ve como una estructura jerárquica de diferentes niveles, como un árbol de dependencias. La única palabra que no está subordinada a otra es la raíz del árbol [10].

En el enfoque de dependencias, una de las formalizaciones más importante es la Teoría de Texto-Significado (Meaning-Text Theory, MTT).

En la Figura 3 se presenta la misma oración de la sección anterior *Los jóvenes contemplan el bello amanecer*, la cual está regida por el verbo *contemplan*, el cual tiene una relación predicativa con el sustantivo *jóvenes* y una relación circunstancial con el sustantivo *amanecer*. El sustantivo *jóvenes* tiene una relación determinativa con el determinante *Los*. Por otro lado, el sustantivo *amanecer* tiene una relación determinativa con el determinante *el* y una relación modificadora con el adjetivo *bello*.

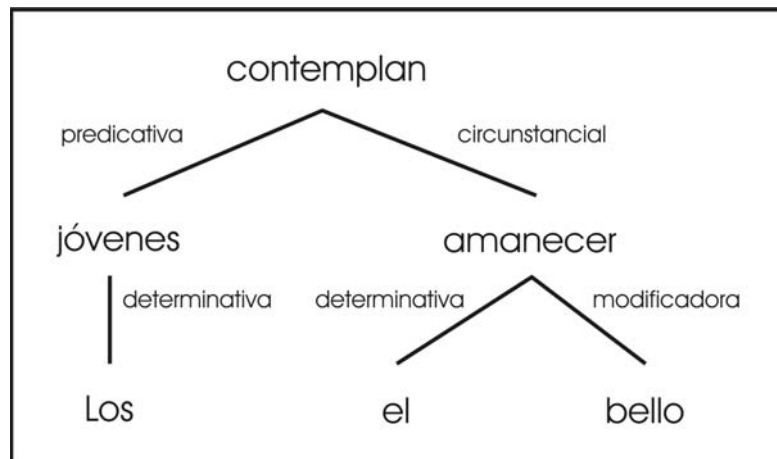


Figura 3. Ejemplo de una oración con el modelo de dependencias

2.1.3. Teoría Texto-Significado

La Teoría Texto-Significado es un formalismo en el enfoque de dependencias que alude a la concepción de cómo los significados léxicos interactúan con las reglas sintácticas.

La meta de esta teoría es modelar la comprensión del lenguaje como un mecanismo que convierta los significados en los textos correspondientes y los textos en los significados correspondientes. Aunque no existe una correspondencia de uno a uno, ya que el mismo significado puede expresarse mediante diferentes textos, y un mismo texto puede tener diferentes significados.

La MTT emplea un mayor número de niveles de representación; así, tanto la sintaxis como la morfología y la fonología se dividen en dos niveles: profundo (D) y superficial (S). Bajo estos términos, la morfología profunda (DMorR) es más superficial que la sintaxis superficial (SSintR). Las nociones de profundo y superficial significan que conforme progresa la representación de la semántica a la fonología superficial (SFonR), ésta se vuelve más y más detallada y específica del lenguaje.

La MTT es un sistema estratificado. Cada oración se caracteriza simultáneamente por siete diferentes representaciones, cada una especifica la oración desde la perspectiva del nivel correspondiente.

Cada nivel de representación se traslada a la adyacente mediante una de las seis componentes de la MTT. En la Figura 4 se muestran los siete niveles [11].

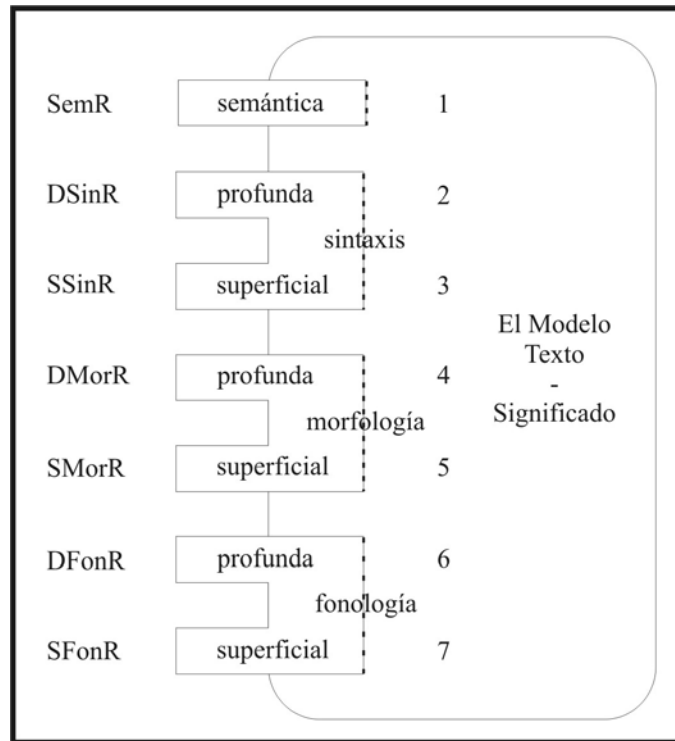


Figura 4. Niveles de la Teoría Texto-Significado.

Cada nivel de representación se considera como un lenguaje separado en el sentido de que tiene su propio vocabulario diferente y reglas distintas de combinación. La transición de un nivel a otro es un proceso de tipo traducción, que involucra el cambio tanto de los elementos como de las relaciones entre ellos, pero que no cambia el contenido informativo de la representación.

Tres conjuntos de conceptos y términos son esenciales para la MTT en su enfoque de la sintaxis:

- Una situación y sus participantes (actuantes). Al número de actuantes se le denomina valencia, la cual será descrita a detalle en la sección 2.2.1.

- Una palabra y sus actuantes semánticos (valencia semántica¹ de la palabra).
- Una palabra y sus actuantes sintácticos (valencia sintáctica² de la palabra).

La *situación*, en esta teoría, significa una parte de la realidad expresada mediante el léxico de un lenguaje dado. Los actuantes semánticos de una situación corresponden a esas entidades cuya existencia está implicada por su significado léxico. Los actuantes sintácticos corresponden a la representación sintáctica de los actuantes semánticos. Por ejemplo, en la oración de la Figura 3, la situación es *contemplar*, los actuantes semánticos son: *quién contempla* y *qué contempla*; y los actuantes sintácticos son: un sintagma nominal que corresponde a *quién contempla* y un sintagma nominal que corresponde a *qué contempla*.

Dentro de la MTT, se utiliza un lexicón el cual es llamado Diccionario Explicativo Combinatorio (Explanatory Combinatorial Dictionary, ECD). Una entrada en este diccionario cubre sólo una unidad léxica [12].

Se desarrolló un ECD para el ruso en [13]. También un ECD en francés está siendo desarrollado en la Universidad de Montreal [14]. Cabe hacer mención que algunos otros ECDs para el idioma inglés han sido desarrollados en otros trabajos [12, 15, 16].

En la Figura 5 se presenta un ejemplo de entrada en un ECD en inglés. Una entrada de un ECD está dividida en tres zonas:

- Una zona semántica (Figura 5.A), donde se presenta la definición lexicográfica o la descomposición semántica del vocablo.
- Una zona sintáctica (Figura 5.B), donde se presenta el patrón de manejo (información de subcategorización), el cual especifica cada actuante semántico (X, Y, etc.) la correspondiente relación sintáctica profunda (I, II, etc.) y una lista de todos los medios de superficie para expresarlos en el texto (N's, against N, etc.); además, algunas condiciones particulares, tales como: Si Y está expresada por *toward N*, N debe

¹ Definición semántica de los actuantes.

² Definición sintáctica de los actuantes.

referirse a la gente ($C_{II,4}$: *N denotes people*, donde $C_{II,4}$ significa Columna II, línea 4); y también, algunos ejemplos que muestran qué es posible y qué no es posible.

- Una zona de coocurrencia (Figura 5.C) que consta de funciones léxicas, las cuales describen la coocurrencia léxica del vocablo.

REVULSION

Semantic definition

(A)

X's revulsion for Y \equiv X's (strong) negative emotion about Y similar to what people normally experience when they are in contact with something that makes them sick and such that it causes that X wants to avoid any contact with Y.

Government Pattern

X = I	Y = II
1. N's	1. against N
2. A_{poss}	2. at N
	3. for N
	4. toward N

(B)

$C_{II,2}$: N denotes something that happens and can be seen or felt.

$C_{II,4}$: N denotes people.

Lexical Functions

Syn_{\cap}	: repugnance; repulsion; disgust; loathing; distaste
$Anti_{\cap}$: attraction
$Conv_{21}Anti_{\cap}$: appeal
A_1	: revulsed
$Able_2$: revulsive
$Magn+Able_2$: of utmost ~ Y = SCENE, SIGHT
$Magn$: deep < extreme < utmost
$AntiMagn$: slight
$Propt$: in, out of [~]
Adv_1Manif	: with [~]
$Oper_1$: experience, feel [~]
$Magn+Labor_{12}$: fill [N=X with ~]
$IncepLabor_{12}$: drive [N=X to ~]

(C)

Figura 5. Ejemplo de entrada en un ECD.

El patrón de manejo de la zona sintáctica (*government pattern*) se utiliza para conocer la relación entre la valencia sintáctica y la valencia semántica, así como los elementos sintácticos que representan a cada actuante [17].

2.2. Subcategorización

La subcategorización identifica los argumentos que son unidos por una de las partes de la oración. Esta parte puede ser un verbo, un sustantivo o un adjetivo [18].

A la información de subcategorización sobre una parte de la oración, en el enfoque de dependencias, se le conoce como valencia.

2.2.1. Valencia

La valencia es conocida en química como la capacidad de un elemento químico que hace posible la combinación de un número fijo de átomos con algún otro elemento. Por ejemplo, el oxígeno es divalente porque tiene la capacidad de vincular 2 átomos. Por ejemplo, combinando 1 átomo de oxígeno con 2 de hidrógeno se forma agua.

De la misma manera, algunas clases de palabras tienen la capacidad de demandar argumentos. Si su demanda es satisfecha, resultan oraciones o frases completamente entendibles [19].

Por lo tanto, la valencia es el número de argumentos que requiere una palabra, esta palabra puede ser un verbo, un sustantivo o adjetivo.

Por ejemplo, en (1), el verbo *comió* es la principal parte de la oración que gobierna dos frases: *La novia de tu hermano* y *ensalada*.

(1) La novia de tu hermano comió ensalada.

La oración (1) puede ser ampliada como en (2). Sin embargo, la valencia del verbo *comer* exige solamente dos argumentos (llamados actuantes en la MTT), los cuales están

representados por alguien (*La novia de tu hermano*) y algo (*ensalada*). Con sólo estos dos actuantes, la oración es completamente entendible.

(2) La novia de tu hermano comió ensalada ayer en mi casa.

Los complementos adicionales en la oración ampliada anteriormente que marcan el tiempo (*ayer*) y el lugar de la acción (*en mi casa*) no son gobernados por la valencia. Ambos complementos son opcionales y se denominan *adjuntos*. Los nombres de los *adjuntos* (es decir, adjuntos temporales, causales, modales, condicionales, etc.) son tomados de la terminología de la gramática tradicional. Generalmente, todas las cláusulas relativas o adverbiales pertenecen al grupo de *adjuntos* [19].

En la MTT se menciona la valencia semántica que se refiere a la definición semántica de los actuantes requeridos por el verbo. En el ejemplo anterior, la valencia semántica está dada por *alguien que come* y *lo que come*.

También, se menciona la valencia sintáctica que es la realización sintáctica de la valencia semántica. En el ejemplo anterior, la realización sintáctica de *alguien que come* es un sintagma nominal, y la realización sintáctica de *lo que come* es, también, un sintagma nominal.

Cabe mencionar que, aunque los actuantes semánticos son requeridos por el verbo, su realización sintáctica puede no ser obligatoria. Por ejemplo, en (3) para el verbo *comprar*, sintáctica y semánticamente, existen cuatro actuantes: *quién compra* (*Juan*), *qué compra* (*un carro*), *a quién lo compra* (*a Luis*) y *en cuánto lo compra* (*en 9000 pesos*).

(3) Juan compró un carro a Luis en 9000 pesos.

Sin embargo, en (4) se usa el mismo verbo *comprar*, pero sólo se presentan, sintácticamente, dos actuantes: *quién compra* (*Anita*) y *qué compra* (*un helado*). Aunque, semánticamente, se supone que a alguien le compró el helado y que tuvo un costo la compra del helado.

(4) Anita compró un helado.

Por lo tanto, no es necesario que todas las valencias semánticas tengan una realización sintáctica (valencia sintáctica) para que una oración sea comprensible.

2.2.2. Valencias Existentes para los Verbos

Los argumentos relacionados a una palabra son elementos de una oración, por lo que a continuación se describen.

Una oración es una unidad, dentro del discurso, que expresa un sentido completo y está constituida por sujeto y predicado. El sujeto (Suj) es de quien se habla en la oración, y el predicado (Pred) es lo que se dice del sujeto [20, 21].

(5) [Suj Los astronautas] [Pred llegarán a la tierra el próximo martes].

(6) [Suj Los avestruces] [Pred corren a gran velocidad].

Otra unidad en el discurso es el sintagma que está conformado por una palabra que es la más importante y que funciona como núcleo. Existen diferentes tipos de sintagmas según la categoría gramatical del núcleo.

Sintagma Nominal (SN). Tiene como núcleo un nombre o sustantivo (Sus, parte variable de una oración que sirve para designar realidades tanto contables como incontables); también puede ser un pronombre.

(7) [SN El nido de las palomas].

(8) [SN Ella misma].

El núcleo de los sintagmas nominales puede tener artículos o determinantes (Det, clase de palabra que precede al sustantivo para determinarlo y concuerda en él en género y número) y adjetivos (Adj, parte de la oración que se refiere al sustantivo para calificarlo y determinarlo) funcionando como sus complementos o modificadores.

(9) [Sus Ríos] [Adj anchos y profundos].

(10) [Det La] [Sus tierra] [Adj estéril].

Sujeto

Como se mencionó anteriormente, el sujeto es un elemento de la oración. Éste es la palabra o frase que se refiere a una idea, un concepto, una persona, un animal o una cosa [20, 21, 22].

(11) [Suj La silueta de la muchacha] se reflejó en el espejo.

(12) [Suj Las hormigas] trabajaban afanosamente.

El sujeto puede encontrarse al principio, en medio o al final de la oración:

(13) [Suj Las primas de Silvestre] piensan viajar disfrazadas.

(14) Mañana [Suj nosotros] prepararemos una cena al estilo italiano.

(15) Comenzarán a trabajar [Suj todos los amigos de Leopoldo].

El sujeto puede estar constituido por: un pronombre (Pro, clase de palabras que se emplea para sustituir un sintagma nominal y evitar, en ocasiones, su repetición) y un sintagma nominal, o puede ser omitido (sujeto tácito).

(16) [Pro Tú] atenderás las llamadas.

(17) En ese momento [Pro todos] levantaron la mano.

(18) [SN La historia del mundo] registra muchas guerras.

Predicado

El predicado es la parte de la oración que expresa la acción que realiza el sujeto o los diferentes estados en los que éste puede encontrarse. Está formado por un verbo y sus complementos [20,21, 22].

(19) Los asistentes del médico [Pred guardaron silencio durante la operación].

(20) [Pred Por la mañana circuló] la noticia sobre el secuestro.

El verbo puede aparecer sin complementos y constituir un predicado.

(21) Unos desconocidos [Pred cantaban].

(22) [Pred Cociné].

(23) Ustedes [Pred van a descansar].

El núcleo del predicado siempre es un verbo (V, simple o perifrástico, es decir, perteneciente a la perífrasis verbal³); es la palabra más importante y concuerda en número y persona con el núcleo del sujeto.

(24) La familia de mis primos no [v asistió] al funeral.

(25) Sara nunca [v ha visto] un eclipse.

(26) El lunes yo [v voy a preparar] una comida tailandesa.

La estructura del predicado está conformada por el verbo que funciona como núcleo y por los complementos de éste. Los complementos del verbo son: complemento directo, complemento indirecto, complemento circunstancial, atributo, complemento preposicional y complemento agente.

Complemento Directo

El complemento directo (CD) se refiere a la persona, animal o cosa que recibe directamente la acción del verbo; se conoce también como paciente, dado que es el que resulta afectado o modificado por la acción del verbo [20, 21, 22,23].

(27) Mi hermano construyó [CD un helicóptero de madera].

(28) Todos asumieron [CD su destino trágico] con entereza.

El complemento directo puede estar formado por:

a) Un pronombre (me, te, se, la, los, las, nos, os, todo, algo, etc.) [24]:

(29) [CD Me] asaltaron anoche.

(30) La noticia [CD los] dejó satisfechos.

³ Es la unidad que consiste en la unión de un verbo con otro en forma no personal (infinitivo, participio o gerundio).

- (31) El economista negó [CD todo].
- b) Un sintagma nominal, constituido por un sustantivo y sus modificadores:
- (32) Esa canción transmitía [CD alegría].
- (33) Ella extendió [CD sus largos y blancos brazos].
- c) Un sintagma preposicional introducido por la preposición *a*. Esta forma sólo se usa cuando el objeto se refiere a personas o seres personificados o singularizados (a esto se le denomina la animidad en el Complemento Directo):
- (34) Visité [CD a mi prima].
- (35) Buscaban [CD a los estudiantes].

Complemento Indirecto

El complemento indirecto (CI) es la persona, animal o cosa que recibe indirectamente la acción del verbo; es el beneficiado o perjudicado por la acción. Si es un sintagma preposicional se une con la preposición *a* [20, 21, 22].

- (36) Armando les dio una sorpresa [CI a sus socios].
- (37) No le cuentes todo [CI a Raquel].

El complemento indirecto puede estar formado por:

- a) Un sintagma preposicional:
- (38) Carmela pidió las copias [CI a Rocío].
- b) Un pronombre (me, nos, te, os, se, le, les) [24]:
- (39) Rafael [CI me] entregó los billetes.
- (40) [CI Nos] dieron una esperanza.

Complemento Circunstancial

El Complemento Circunstancial (CC) expresa la manera, el tiempo, el lugar y demás circunstancias en las que se realiza la acción del verbo [20, 21, 25].

El complemento circunstancial puede estar formado por:

a) Un adverbio, un sintagma adverbial o una locución adverbial:

(41) El sol [CC nunca] girará alrededor de la tierra.

(42) Llovió [CC muy fuertemente].

(43) Hizo la maleta [CC a la ligera].

b) Un sintagma preposicional:

(44) Se comporta [CC de una forma extraña].

(45) Habla [CC con mucha seguridad].

c) Un sintagma nominal:

(46) Se paseaba [CC todos los fines de semana].

(47) [CC La próxima semana] podrás conocer las nuevas instalaciones.

Las múltiples circunstancias en las que se realiza la acción del verbo pueden ser de: modo, tiempo, lugar, cantidad, instrumento, compañía, tema, causa, finalidad o duda.

Complemento Preposicional

En lo que respecta a las estructuras verbales, el término de régimen (relación de dependencia entre elementos sintácticos, donde uno aparece como *núcleo* y otro como subordinado) se ha ampliado hasta hacerlo equivalente a diferentes posibilidades de complementación de un verbo en la oración, posibilidades habituales y ocasionales [22].

La denominación ‘complemento de régimen verbal’ ha ido restringiendo su aplicación a complementos como:

(48) No abuses [CP del alcohol].

(49) Esto carece [CP de sentido].

(50) Piensa [CP en la muerte].

(51) He soñado [CP contigo].

Estos complementos tienen una categoría distinta a la de los clásicos ‘objetos’ (directo e indirecto), pero distintos, también, a las ‘circunstancias’ de lugar, tiempo o manera.

Por lo que, la oración *Pienso en ti* no debería ser analizada igual que *Trabajo en mi casa* pese a su forma sintáctica idéntica, ni *Sueño contigo* igual que *Mañana comeré contigo*, Ni *Me acuerdo de ella* como *Se acercó de rodillas*. Los complementos de la primera frase de cada par denotan una relación mucho más estrecha con el verbo que la circunstancia designada por los de la segunda frase de cada par.

Se puede decir que el complemento de la primera frase de cada par es parte de la significación del verbo, y como tal es dependiente del verbo.

Como en el siguiente ejemplo, donde el verbo *poner* tiene el significado de colocar en un sitio o lugar a alguien o algo:

(52) [Suj Juan] [v puso] [CD los pollos] [CP en el corral] [CC por la mañana].

(53) [Suj Juan] [v puso] [CD los pollos] [CP en el corral].

(54) [Suj Juan] [v puso] [CD los pollos].

Donde se observa cómo el CC en (54) puede ser omitido en la oración sin perder el sentido, mientras que en (55) la ausencia del CP deja la oración sin el significado original. Por otro lado, utilizando el mismo verbo *poner*, podemos encontrar la oración (56) (oración presentada en [26]).

(55) [Suj Juan] [v puso] [CD la mesa].

Donde el significado de *poner* es disponer algo para un cierto fin.

Como se observa, la presencia de un complemento preposicional, puede determinar el significado del verbo. Y debido a que son parte de la significación verbal no pueden ser encasillados dentro de los complementos circunstanciales.

Por otra parte, la estructura sintáctica de los complementos preposicionales está definida por cada verbo, así que se considera que se requiere un estudio particular para obtener todas las estructuras posibles [22].

Complemento Predicativo o Atributo

El complemento predicativo o atributo (Atr) predica o informa sobre cualidades, atributos o peculiaridades del sujeto [20, 21, 22]. Aparece en las oraciones con verbos copulativos. Puede estar formado por:

a) Un sintagma nominal:

(56) El arquitecto era [Atr un hombre de avanzada edad].

b) Un sintagma adjetivo:

(57) Angelina estaba [Atr muy avergonzada].

c) Un pronombre:

(58) El problema es [Atr aquél].

d) Una oración:

(59) El novelista es [Atr el que sabe].

Después de conocer los elementos de una oración, se pueden identificar éstos en diferentes oraciones como las siguientes:

(60) [Suj Luis] [V corre].

(61) [Suj Luis] [V corre] [CC por las mañanas].

(62) [Suj Alicia] [V está] [Atr feliz].

(63) [Suj Alicia] [V estuvo] [Atr feliz] [CC durante el viaje].

(64) [Suj Sandra] [V compró] [CD un carro].

(65) [Suj Sandra] [V compró] [CD un carro] [CC ayer].

(66) [Suj Sandra] [V compró] [CD un carro] [CI a don Francisco].

- (67) [Suj Sandra] [v compró] [CD un carro] [CI a don Francisco] [CC ayer].
- (68) [Suj Sandra] [CI le] [v compró] [CD un carro].
- (69) [Suj Sandra] [CI le] [v compró] [CD un carro] [CC ayer].
- (70) [Suj Sandra] [CI se] [CD lo] [v compró].
- (71) [Suj Sandra] [CI se] [CD lo] [v compró] [CC ayer].
- (72) [Suj Brenda] [v habló] [CP de la pobreza] [CI a las personas].
- (73) [Suj Brenda] [CI les] [v habló] [CP de la pobreza].

Después de mencionar los diferentes complementos que pueden aparecer en una oración, cabe hacer mención que la valencia de un verbo, generalmente, requiere los siguientes tipos de argumentos: sujeto, atributo, complemento directo, complemento indirecto y complemento preposicional. Aunque existen excepciones donde un complemento circunstancial formado por un adverbio es necesario para completar el significado del verbo, por lo que también es considerado un argumento del verbo.

Por otra parte, los verbos en español presentan hasta cinco valencias. A continuación se presentan algunos ejemplos de verbos con su valencia y los actuantes correspondientes.

Verbos sin valencias

Los verbos en español que no presentan ninguna valencia (la inexistencia de algún actuante se representa con \emptyset) son los verbos que sólo se conjugan en tercera persona como: *llover, granizar, nevar, etc.*

Estructura	Ejemplos
\emptyset - Verbo - \emptyset	Llueve

Verbos con una valencia

Algunos verbos intransitivos en español tienen únicamente la valencia que corresponde al sujeto, por ejemplo el verbo *cojear*:

Estructura	Ejemplos
------------	----------

Suj - V - Ø	[_{Suj} El hombre] [_V cojea]
-------------	---

Verbos con dos valencias

Los verbos transitivos, por definición, tienen una segunda valencia semántica denominada en el nivel sintáctico como objeto directo o complemento directo, aunque en ocasiones también puede ser un complemento preposicional o atributo.

Estructura	Ejemplos
Suj - V - CD	[_{Suj} El padre] [_V quiere] [_{CD} a su hijo]
Suj - V - CP	[_{Suj} Ese saco] [_V huele] [_{CP} a mojado]
Suj - V - Atr	[_{Suj} Ese hombre] [_V es] [_{Atr} tu padre]

Verbos con tres valencias

Los verbos considerados en la gramática clásica como doble transitivos tienen tres valencias. La tercera valencia, en la mayoría de las ocasiones, se denomina objeto indirecto o complemento indirecto; aunque también existe la posibilidad de que un complemento preposicional sea el tercer actuante.

Estructura	Ejemplos
Suj - V - CD - CP	[_{Suj} Juan] [_V puso] [_{CD} el dinero] [_{CP} en la mesa]
Suj - V - CP - CI	[_{Suj} Dolores] [_V habló] [_{CP} de tí] [_{CI} a la gente]

Verbos con cuatro valencias

La mayoría de los verbos que usan cuatro valencias son transitivos, un ejemplo es el verbo *condenar*.

Estructura	Ejemplos
Suj - V - CD - CP - CP	[_{Suj} El juez] [_V condena] [_{CD} al acusado] [_{CP} a cadena perpetua] [_{CP} por asesinato]

Verbos con cinco valencias

Por último, presentamos un ejemplo, el verbo *rentar*, que tiene cinco valencias.

Estructura	Ejemplos
Suj - V - CD - CI - CP - CP	[_{Suj} María] [_V rentó] [_{CD} un departamento] [_{CI} a la compañía Zeta] [_{CP} en dos mil pesos]

[CP por mes]

2.2.3. Valencias con Respecto a la Clase Funcional Verbal

La valencia también puede identificarse con la clase funcional (tipo de verbo).

Según la RAE [26], existe la siguiente clasificación de verbos⁴:

- Transitivo (activo): el que se construye con complemento directo; p. ej. *comprar*.
- Adjetivo: tradicionalmente, cualquiera de los verbos; p. ej. *ser correcto*.
- Auxiliar: el que se emplea en la formación de la voz pasiva, de los tiempos compuestos y de las perífrasis verbales; p. ej. *haber*.
- Causativo: verbo, o perífrasis verbal, cuyo sujeto no ejecuta por sí mismo la acción, sino que la hace ejecutar por otro. *El abogado permitió que su cliente hablara*.
- Copulativo: aquél que, junto con el atributo, forma el predicado nominal de una oración; p. ej. *estar*.
- Defectivo: aquél que no se usa en todos los modos, tiempos o personas, p; ej. *abolir*.
- Frecuentativo (iterativo, reiterativo): el que expresa una acción que se compone de acciones repetidas; p. ej. *besuquear*.
- Impersonal (unipersonal): el que se emplea generalmente en la tercera persona de singular de todos los tiempos y modos, simples y compuestos, y en infinitivo y gerundio, sin referencia ninguna a sujeto léxico elíptico o expreso; p. ej. *llover*.
- Incoativo: el que indica el comienzo de una acción; p. ej. *florecer*.
- Intransitivo (neutro): el que se construye sin complemento directo; p. ej. *nacer*.

⁴ Se han omitido algunos verbos en la lista, porque son aplicables sólo a ciertos dialectos del lenguaje español o porque son considerados antiguos en su definición.

- Irregular: el que se conjuga alterando o combinando la raíz, el tema o las desinencias de la conjugación regular; p. ej. *cab*er.
- Pronominal (pronominado, reflejo): el que se construye en todas sus formas con un pronombre átono que concuerda con el sujeto y que no desempeña ninguna función sintáctica oracional. Algunos verbos son exclusivamente pronominales, como *arrepentirse*, y otros adoptan determinados matices significativos o expresivos en las formas reflexivas; p. ej., *caer o morir*.
- Recíproco: tradicionalmente, aquél que denota reciprocidad o cambio mutuo de acción entre dos o más personas, animales o cosas, llevando siempre por complemento un pronombre; p. ej. *Pedro y Juan se tutean*.
- Reflexivo: tradicionalmente, el que se construye con un pronombre reflexivo átono. *Tú te peinas*.
- Regular: el que se ajusta en su conjugación al modelo que se fija como propio de esa conjugación; p. ej. *amar*.
- Sustantivo: tradicionalmente, verbo copulativo ser, único con idea de esencia o sustancia, sin denotar, como los demás verbos, otros atributos o modos de ser.
- Terciopersonal: tradicionalmente, el que se construye sólo en tercera persona de singular y plural; p. ej. *ocurrir*.

Cabe mencionar que las clasificaciones mencionadas anteriormente se empalman. Por ejemplo puede que exista un verbo transitivo que sea, a su vez, causativo (permitir), o que un verbo auxiliar sea, también, un verbo impersonal (haber), etc.

Debido a que este trabajo se centra en los actuantes de la oración según el significado del verbo, se seleccionaron las clases funcionales que proporcionan información sobre la valencia. Las cuales son:

- Transitivo (activo): el que se construye con complemento directo; p. ej. *comprar, cortar, pegar, abrazar, patear, morder, llamar, abalanzar, abandonar, abocar,*

abochornar, abordar, aborrecer, abrasar, abrir, aburrir, cerrar, encender, apagar, borrar, romper, ensuciar.

- Intransitivo (neutro): el que se construye sin complemento directo; p. ej. *nacer, abogar, abusar, acaecer, abundar, caer.*
- Impersonal: el que se construye sin complemento directo; p. ej. *llover, nevar, granizar,* etc.
- Pronominal: el que se construye en todas sus formas con un pronombre átono que concuerda con el sujeto y que no desempeña ninguna función sintáctica oracional; p. ej. *hacerse, vestirse, comportarse, etc.*

En la Tabla 2 se presenta la información de subcategorización obtenida mediante estas clases funcionales.

Tabla 2. Información de subcategorización obtenida mediante la clase funcional

Clase funcional	Información de subcategorización	Ejemplo
Impersonal – intransitivo	Ningún actuante	Nieva
Impersonal – transitivo	Complemento Directo	Llueven bendiciones
Intransitivo	Sujeto	Beatriz corre
Transitivo	Sujeto y Complemento Directo	La niña escribe una carta
Pronominal	Sujeto y pronombre átono	El bebé se despertó

2.2.4. Representación Formal de las Valencias

Marcos de Subcategorización

En los métodos para caracterizar formalmente las valencias, los actuantes se describen en ellas desde un enfoque sintáctico y se denominan complementos. Las valencias se consideran características sintácticas. La información de la estructura de los complementos de un verbo se conoce como subcategorización. Cada subcategoría del verbo tiene su propio conjunto de complementos que habitualmente van en orden lineal predeterminado. Dentro de este enfoque, se considera que los verbos pueden pertenecer a más de un tipo de subcategorización.

A continuación se presentan los marcos de subcategorización para el verbo *acusar*. Donde el verbo está implícito en el marco, SN representa un sintagma nominal, V_INF representa un verbo en infinitivo y \emptyset representa vacío.

1. a SN
2. de SN
3. de V_INF
4. a SN de SN
5. a SN de V_INF
6. \emptyset
7. SN

Cabe resaltar, que cada marco muestra una secuencia lineal de constituyentes. Este tipo de descripción es útil para los lenguajes de orden fijo (como el inglés), ya que su orden de palabras es más estricto.

Patrones de Manejo

El patrón de manejo es una estructura que permite almacenar mayor información que los marcos de subcategorización. El patrón de manejo consta de cuatro secciones que se describen a continuación:

- *Primera sección:* La palabra encabezado que corresponde al verbo considerado con un significado específico.
- *Segunda sección:* La explicación semántica de la situación relacionada a cada verbo específico. En esta sección se definen las valencias, cuyo orden es arbitrario, aunque cada lexema normalmente impone un orden “natural” en las valencias.
- *Tercera sección:* La descripción de cada valencia sintáctica. La lista exhaustiva de todas las posibles formas de representación de cada valencia sintáctica en los textos. Dentro de esta sección se encuentra la condición de obligatoriedad en los textos. Si una valencia semántica no está presente en esta parte, significa que es opcional en su realización sintáctica.

- *Cuarta sección:* En esta sección se muestra la información acerca de los posibles combinaciones de valencias sintácticas.

En la Figura 6 se presenta un ejemplo de patrones de manejo sintáctico para el verbo *solicitar* [10]. Donde S denota un sustantivo; C, una cláusula subordinada; (an) denota que un sustantivo es animado y (na) denota que un sustantivo es inanimado.

← Primera Sección
← Segunda Sección

solicitar ←
X asks something Y from Z ←

Número	Patrón de manejo	Ejemplo
X = 1; who asks?		
1.1	S (an)	Juan / el gobierno ~
Y = 2; what?		
2.1	S (na)	~ una prórroga / un préstamo
2.2	que C	~ que este libro se le dé
Z = 3; from whom?		
3.1	A S (an)	~ a la secretaria
3.2	con S (an)	~ con el secretario
3.3	de S (an)	~ de usted
3.4	en S (na)	~ en urgencias
POSIBLE:		
(1) 0 2 3	(El partido) solicita una prórroga al gobierno.	
(1) 0 2	(Ella) solicita un préstamo.	
IMPOSIBLE:		
(1) 0	*(El partido) solicita.	
(1) 0 3	*(El partido) solicita al gobierno.	

} Tercera Sección

} Cuarta Sección

Figura 6. Ejemplo de patrón de manejo sintáctico del verbo *solicitar*

2.3. WordNet

2.3.1. Ontología

Una ontología es una representación formal del conocimiento donde los conceptos, las relaciones y las restricciones conceptuales son explicitadas mediante formalismos para un determinado dominio. Su función más frecuente en la lingüística es de apoyo para sistemas de Traducción Automática Basada en el Conocimiento y para la Terminografía (práctica de la Terminología). En ambos casos, la ontología es una representación formal y explícita de la estructura conceptual del campo sobre el que se trabaja. Este recurso lingüístico incluye como mecanismo de inferencia a la herencia, que implica una economía en la codificación de la información: los conceptos superiores transmiten sus características a los conceptos inferiores.

Por consiguiente, la ontología es uno de los módulos asociados a un sistema de conocimiento donde su función es la de apoyo semántico para las unidades léxicas, es decir que las unidades léxicas son descritas como objetos lingüísticos en una base de datos léxica y son relacionadas con una jerarquía conceptual localizada en una ontología [27].

2.3.2. WordNet

WordNet es una ontología para el lenguaje inglés, la cual está formada, principalmente, por conjuntos de sinónimos denominados *synsets*. Además, provee definiciones para las palabras y contiene varias relaciones semánticas entre los *synsets*.

Posterior a WordNet, se desarrolló EuroWordNet[28], la cual es una interfaz que contiene bases de datos (ontologías) de lenguajes europeos (alemán, italiano, español, francés, entre otros). Cada base de datos está estructurada de la misma manera que el original WordNet (en inglés). La base de datos que se utiliza en el presente trabajo es Spanish WordNet [29].

Synset

La información en WordNet está organizada alrededor de grupos lógicos llamados *synsets*. Cada *synset* consiste en una lista de palabras y punteros que describen la relación entre un *synset* y otros [30].

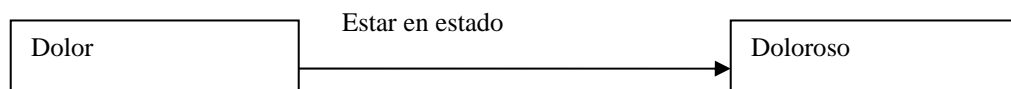
Una palabra puede aparecer en más de un *synset*, y como más de una clase de palabra (sustantivo, verbo, adjetivo, etc.). Las palabras en un *synset* están agrupadas de tal manera que son intercambiables en algún contexto, por ejemplo, el grupo de palabras *carro*, *automóvil* y *auto*.

Relaciones semánticas

Las relaciones semánticas que existen dentro de la estructura de Spanish WordNet, sus funciones, alcances y algunos ejemplos, se presentan a continuación.

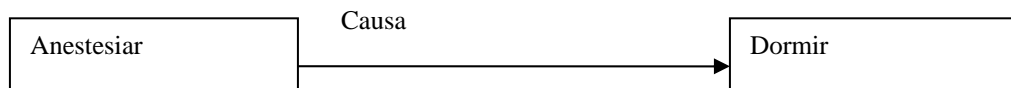
Be in state. A estar en estado B.

Aplica para sustantivos con adjetivos.



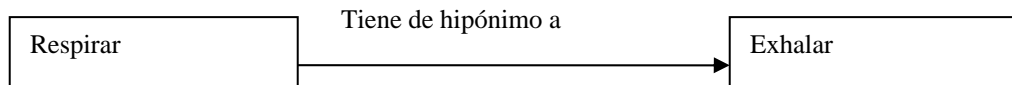
Causes. A causa B.

Aplica para verbos o sustantivos con sus homogéneos.



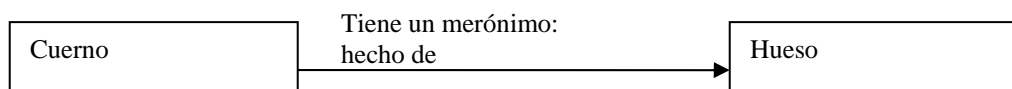
Has hyponym. A tiene de hipónimo a B.

Aplica para verbos o sustantivos con sus homogéneos.



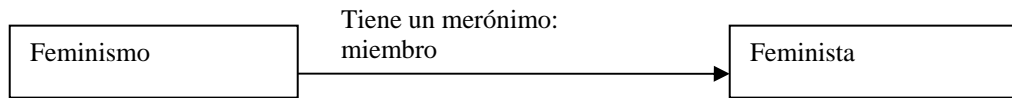
Has mero madeof. A tiene un merónimo: hecho de B

Aplica entre sustantivos.



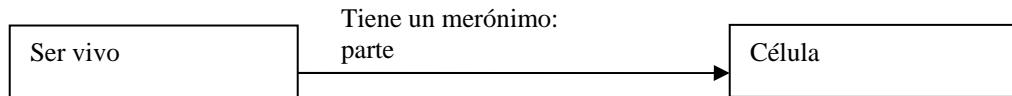
Has mero member. A tiene un merónimo: miembro B

Aplica entre sustantivos.



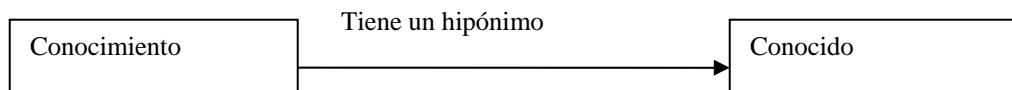
Has mero part. A tiene un merónimo: parte B.

Aplica entre sustantivos.



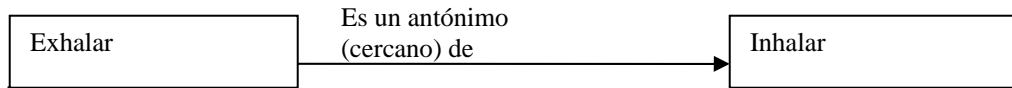
Has xpos hyponym. A tiene un hipónimo B.

Aplica para sustantivos con adjetivos. La diferencia con *has hyponym* es que este permite una relación entre sustantivos y adjetivos.



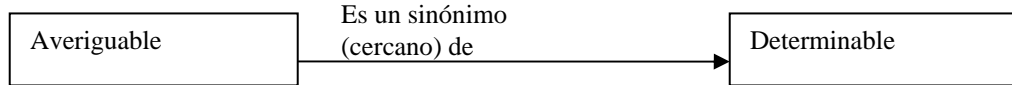
Near antonym. A es un antónimo (cercano) de B

Aplica para verbos o adjetivos con sus homogéneos.



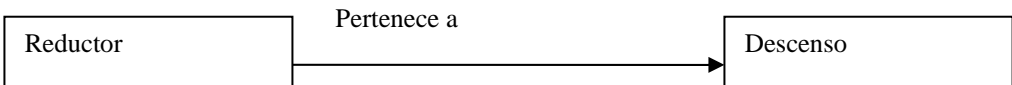
Near synonym. A es un sinónimo (cercano) de B.

Aplica entre adjetivos.



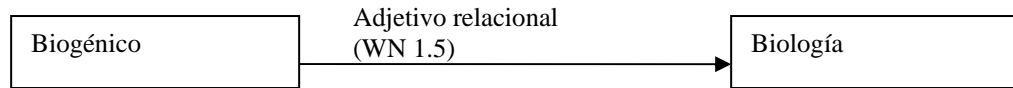
Pertains to. A pertenece a B.

Aplica para adjetivos con sustantivos.



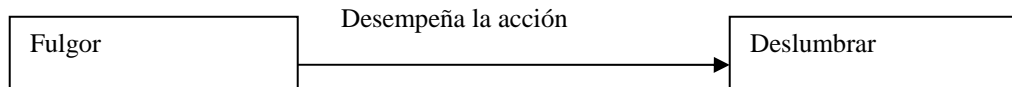
Relational adj wn15. A adjetivo relacional (wn15) B.

Aplica para adjetivos con sustantivos.



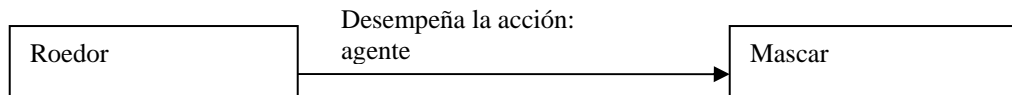
Role. A desempeña la acción B.

Aplica para sustantivos con verbos.



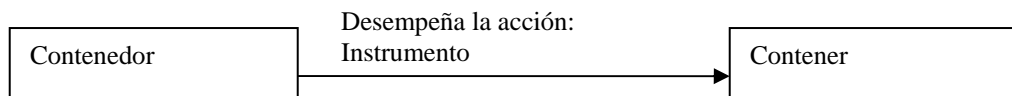
Role agent. A desempeña la acción: agente B.

Aplica para sustantivos con verbos.



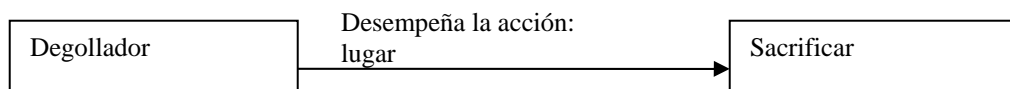
Role instrument. A desempeña la acción: instrumento B

Aplica para sustantivos con verbos.



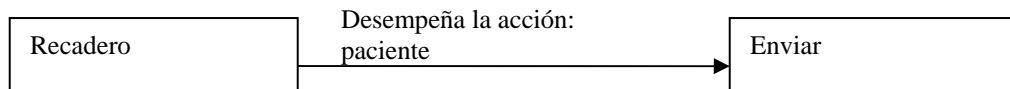
Role location. A desempeña la acción: lugar B

Aplica para sustantivos con verbos.



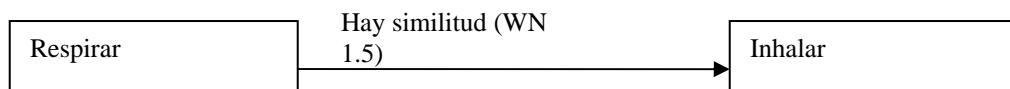
Role patient. A desempeña la acción: paciente B

Aplica para sustantivos con verbos.



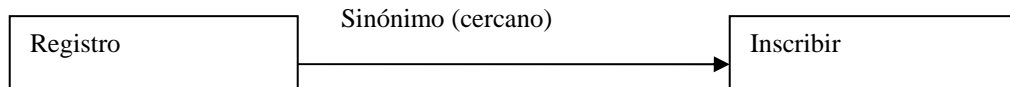
See also wn15. A hay similitud (wn15) B.

Aplica para verbos o adjetivos con sus homogéneos.



Xpos near synonym. A sinónimo (cercano) B

Aplica para sustantivos con verbos. La diferencia con *near synonym* es que éste permite una relación entre sustantivos y verbos.



2.3.3. Spanish WordNet

La base de datos de Spanish WordNet está implementada en tres tablas, que son las siguientes:

spWN1.0_variants

Esta tabla contiene las columnas POS, SYN, WORD, SENSE, CS y <>, donde:

- POS es la parte de la oración.
- SYN es el synset español en SpWN1.0.
- WORD es la palabra sinónima en español.
- SENSE es el número de sentido correspondiente.
- CS es el marcador de confianza.
- <> contiene guión (-) o nulo, en esta versión.

En la Tabla 3 se ve un ejemplo de esta tabla.

Tabla 3. Ejemplo de la tabla spWN1.0_variants

POS	SYN	WORD	SENSE	CS	<>
A	00320588	Empañado	1	99	-
A	01966111	Abacial	1	99	-
A	00788774	Abandonado	1	99	-

spWN1.0_synsets

Esta tabla contiene los campos SPOS, SSYN, ILI, NDES, STAT y SDEF, donde:

- SPOS es la parte de la oración del *synset* en español.
- SSYN es el identificador del *synset* en español en SpWN1.0.
- ILI es el índice interlengua. Este número corresponde principalmente a un *synset* de WN1.5, excepto aquellos *synsets* en Spanish WordNet que empiecen con 5.
- NDES es el número de *synsets* hipónimos.
- STAT es el estado del *synset*. Donde la letra 1 significa cerrado y la letra *n* significa ningún concepto lexicalizado.
- SDEF es la definición del *synset* en español.

En la Tabla 4 se ve un ejemplo de esta tabla.

Tabla 4. Ejemplo de la tabla spWN1.0_synsets

SPOS	SSYN	ILI	NDES	STAT	SDEF
n	09185137	09185137	0	1-	
n	09185279	09185279	0	1-	Período de mayor producción y prosperidad
n	09185628	09185628	0	1-	Tiempo en que transcurre algo de principio a fin

spWN1.0_relations

Esta tabla contiene los campos EWNR, SPOS, SSYN, TPOS, TSYN y CS, donde:

- EWNR es una de las relaciones de EuroWordNet.
- SPOS es la parte de la oración fuente.
- SSYN es el identificador del *synset* fuente en spWN1.0.
- TPOS es la parte de la oración objetivo.
- TSYN es el identificador del *synset* objetivo en spWN1.0.
- CS es el marcador de confianza.

En la Tabla 5 se ve un ejemplo de esta tabla.

Tabla 5. Ejemplo de la tabla spWN1.0 relations

EWNR	SPOS	SSYN	TPOS	TSYN	CS
has_hyponym	N	03892008	n	08677522	90
has_xpos_hyponym	N	03464624	a	50000826	99
has_xpos_hyponym	N	03464624	a	00296574	99

2.4. Base de Patrones de CENIDET

En el Centro Nacional de Investigación y Desarrollo Tecnológico se implementó un analizador léxico-sintáctico de lenguaje natural.

Dicho analizador utiliza una base de patrones de manejo como conocimiento lingüístico para el *filtrado* de estructuras sintácticas.

La base de datos del conocimiento lingüístico para el manejo de patrones está implementada en PostgreSQL 7.2.1 y consta de cuatro tablas.

2.4.1. Tablas

La tabla denominada *verbos* contiene tres columnas: la clave primaria (*claveverbo*), el verbo en infinitivo (*verbo*) y la clasificación del verbo (*clasefuncional*).

La tabla denominada *oraciones* contiene dos campos: *tipooracion*, permite manejar una clave primaria, y *estructuraoracion*, que permite definir si la oración está en forma activa, pasiva o imperativa.

La tabla *patrones* proporciona información acerca de la valencia sintáctica que puede tener cada verbo, al mismo tiempo indica si los actuantes de la valencia sintáctica necesitan o no una preposición para su representación, mostrando cuáles son las preposiciones usadas para cada actuante si es que procede.

La tabla *combinaciones* contiene campos para los diez tiempos que existen en el español; además, contiene los campos *claveverbo*, *tipooracion* y *clavepatron*.

2.4.2. Relaciones

En la Figura 7 se muestran las tablas de la base de patrones y sus relaciones.

La tabla *verbos* está relacionada con la tabla *combinaciones*. La relación que existe entre estas tablas es de uno a muchos (1-N).

La tabla *oraciones* está relacionada con la tabla *combinaciones*. El tipo de relación que existe entre estas tablas es de uno a muchos (1-N).

La tabla *patrones* está relacionada con la tabla *combinaciones*. El tipo de relación que existe entre estas tablas es de uno a muchos (1-N).

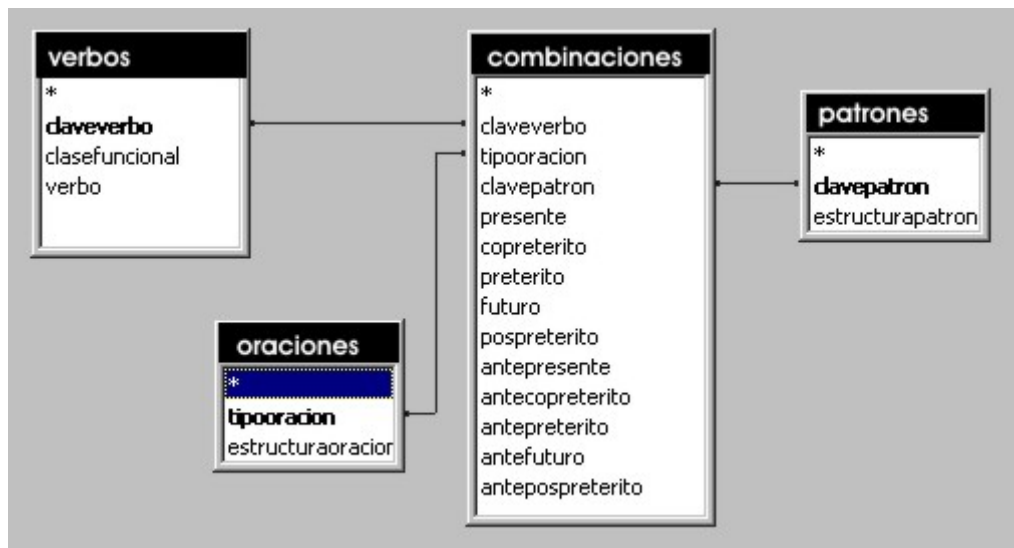


Figura 7. Base de patrones de CENIDET.

2.5. Trabajos Relacionados

Una base de patrones es una herramienta importante para el análisis sintáctico [1], además de proporcionar información semántica útil para análisis posteriores. Sin embargo, la mayoría de los diccionarios que contienen esta información han sido llenados de manera manual con ayuda de lingüistas. Actualmente, existen más esfuerzos enfocados a la

determinación automática de los patrones de subcategorización. Algunas técnicas desarrolladas al respecto se describen a continuación.

En [31] se describe la creación de un diccionario indexado gramaticalmente a partir de un diccionario base. El objetivo de este trabajo fue extraer los datos almacenados en el diccionario base, codificarlos y almacenarlos en un lexicón elaborado con una mejor estructura.

En esa investigación, para poder definir la subcategorización de un verbo, se utilizan cuatro clasificaciones de significado para los verbos. De esta manera, pueden asignar la valencia según el significado del verbo sin examinar otras propiedades de él.

Es menester decir que ese trabajo se desarrolló de manera semiautomática, utilizando un programa para la extracción de información, y la participación de lingüistas para afinar los posibles errores dentro de la codificación. El lexicón obtenido se encuentra en inglés y fue elaborado por las siguientes instituciones: University of Cambridge, University of Lancaster y University of Edinburgh.

Se describe en [1] un mecanismo para adquirir automáticamente los marcos de subcategorización de los verbos y sus frecuencias en un conjunto de textos. En tal proceso, un corpus etiquetado es analizado sintácticamente de manera parcial para identificar frases; posteriormente usan una gramática lineal para estimar el marco de subcategorización apropiado para cada token-verbo en el conjunto de textos.

En un experimento sobre la identificación de seis marcos de subcategorización, el sistema mostró un 80% de exactitud. Además, el enfoque estadístico mejora la exactitud de la estimación de frecuencia.

En [32] se describe el diseño de la sintaxis de COMLEX, un lexicón computacional que provee información sintáctica detallada para aproximadamente 38,000 palabras importantes en inglés. En él se aborda información detallada sobre subcategorización de los verbos, sustantivos y adjetivos que toman complementos. Sus resultados incluyen 92 características sobre subcategorización para verbos, 14 para adjetivos y 9 para sustantivos. Los datos lingüísticos anteriores fueron recopilados de manera manual.

En [33] se describe una técnica y la implementación de un sistema para la construcción de un diccionario de subcategorización a partir de un conjunto de textos. Cada entrada del diccionario cifra la frecuencia relativa de ocurrencia de un conjunto de clases de subcategorización para el inglés.

Dentro de la investigación se realizó un experimento inicial sobre una muestra de 14 verbos, los que mostraban múltiples patrones de complementación. Y se demostró que esta técnica logra una exactitud comparable con enfoques anteriores, los cuales son limitados altamente por un conjunto de clases de categorización. Además, se comparó el diccionario de subcategorización construido con el sistema contra un analizador sintáctico, y mejoró la exactitud en una cantidad apreciable.

Un trabajo desarrollado por el grupo COLE (Compiladores y Lenguajes) se expone en [34]. Este documento presenta un análisis automático del régimen verbal en español. Para lograr ese objetivo se utilizó una base de datos sintácticos informatizada, que los lingüistas de ese grupo utilizaron como punto de partida para la elaboración de un lexicón formal de verbos, en el que se incluyó la caracterización sintáctica y semántica de los argumentos. También, se describe el desarrollo de gramáticas parciales de las categorías que funcionan como constituyentes de la cláusula en español.

Esta investigación arrojó como resultado principal un lexicón con la información sintáctica pertinente para los verbos del español, completando y refinando así los datos contenidos en la base de datos sintácticos. Algunas de las aplicaciones de esta base son las siguientes:

- Programas de etiquetación morfosintáctica y lematización.
- Recuperación de información.
- Traducción automática.
- Correctores automáticos de la sintaxis y el estilo.
- Enseñanza de español como lengua extranjera.

En [10] se presenta un modelo para resolver el problema del análisis sintáctico relacionado a la gran cantidad de variantes generadas cuando se analizan textos sin restricciones. Para lograr esto, mencionan que analizaron las características del español que difieren de los lenguajes cuyo orden de palabras es más estricto, y se representaron bajo un enfoque generalizado de descripción de valencias sustentado en el formalismo de la Teoría de Texto-Significado.

Dentro de la misma investigación se expone una forma nueva de descripción de los Patrones de manejo, denominada patrones de manejo avanzados, con información cualitativa para el análisis sintáctico. Debido al conocimiento lingüístico que se requiere en dichos patrones, se plantea un método semiautomático de adquisición de esa información, a partir de un corpus de textos. Este método combina métodos lingüísticos que permiten extraer estructuras sintácticas, y métodos estadísticos para la selección de variantes de estructuras con la finalidad de obtener los complementos de palabras específicas (verbos, adjetivos y sustantivos).

En [35] se describe una Base de Datos Sintácticos del español actual (BDS), la cual contiene el resultado de analizar aproximadamente 160,000 cláusulas de forma manual. Éstas son parte del Archivo de Textos Hispánicos de la Universidad de Santiago (ARTHUS). Cada registro del archivo central consta de 63 campos organizados en cuatro grandes bloques, los cuales contienen la siguiente información:

- Datos referentes al verbo que actúa como predicado y su localización.
- Datos referentes a la cláusula como conjunto (tipo de cláusula, función que desempeña, voz, polaridad, modalidad, forma verbal empleada, forma verbal de la que depende, orden de los elementos, etc).
- Datos referentes a cada una de las funciones sintácticas que pueden existir en la cláusula (sujeto, complemento directo, complemento preposicional, etc.). Para cada una se indica el tipo de unidad que la realiza y otros factores de interés.

El documento [36] presenta un método para mejorar la calidad de los marcos de subcategorización (SCF, SubCategorization Frame) adquiridos de un conjunto de textos para aumentar un diccionario de gramática lexicalizada.

Primero, se estimó un valor de confianza para cada palabra sobre si puede tener un SCF, y se creó un vector de valores de confianza SCF para cada palabra. Tomando en cuenta que los vectores obtenidos involucran tendencia de concurrencia entre los SCFs para las palabras, se puede mejorar la calidad de los SCFs adquiridos a través de vectores de agrupación obtenidos del diccionario de SCFs adquirido y del diccionario de la gramática objetivo.

Se aplicó el método de adquirir SCFs de un corpus, usando un subconjunto del diccionario de marcos de subcategorización de la gramática del inglés XTAG (XTAG English Grammar, del grupo de investigación XTAG). Los resultados mostraron que se puede lograr una mejor precisión usando este método.

En [37] se presenta el diseño y la implementación de un analizador sintáctico usando el método de dependencias para oraciones en español. Su principal objetivo es desarrollar un analizador sintáctico que permita disminuir el número de variantes de estructura sintáctica.

Otro objetivo dentro de ese trabajo fue generar un analizador sintáctico flexible y dinámico; es decir, que permita la modificación de las reglas sintácticas, de acuerdo a las necesidades del usuario y ofrecer la posibilidad de que el número de reglas sintácticas del analizador se pueda ir incrementando. Este trabajo también utiliza una base de patrones para la valencia de los verbos, la cual fue llenada de manera manual por el investigador, y con pocos elementos dentro de ella.

En el trabajo SENSEM [38], se presenta la elaboración de una base de datos con información verbal del español. El objetivo de este proyecto es la creación de una base de datos con información sintáctica–semántica de cada verbo en español. La información que se almacena consta de los papeles semánticos, la estructura argumental y preposiciones. Una de las características importantes de este recurso es la lista de esquemas sintácticos que acepta cada verbo, asociados a su significado oracional.

En una investigación desarrollada en la Universidad de Barcelona [39], se muestra la elaboración de un corpus etiquetado sintácticamente y semánticamente. Tal corpus fue obtenido usando uno ya existente con información solamente sintáctica (3LB-LEX). Revisando manualmente cada oración del corpus se le agregó información sintáctica. Y toda esta información es almacenada en un diccionario, que servirá para etiquetar el corpus objetivo (3LB-SEM). Dentro del diccionario verbal se toma en cuenta el sentido del verbo para definir la estructura que se acepta como correcta.

En la Universidad de Cambridge se desarrolló una investigación [40] donde se trata la adquisición de marcos de subcategorización para el idioma inglés. En dicho trabajo se considera que la información sintáctica de los marcos de subcategorización verbales, obtenida a través de la revisión automática de un corpus, puede contener información ambigua. Por lo que proponen utilizar información semántica para el filtrado de marcos de subcategorización, usando un diccionario con estructuras sintácticas basado en las clases de Levin.

Tabla 6. Características de los trabajos relacionados y este trabajo.

Trabajos	Idioma		Procedimiento			Relaciones semánticas y la clase funcional
	Español	Inglés	Manual	Semiautomático	Automático	
Boguraev [31]		✓		✓		
Ushioda [1]		✓			✓	
Grishman [32]		✓	✓			
Briscoe [33]		✓			✓	
Graña [34]	✓		✓			
Galicia [10]	✓			✓		
Muñiz [35]	✓		✓			
Korhonen [40]		✓		✓		
Cervantes [37]	✓		✓			
Yoshinaga [36]		✓			✓	
Fernandez [38]	✓		✓			
Taulé [39]	✓			✓		
Este trabajo	✓			✓		✓

Capítulo 3

Método de Solución

En este capítulo se describe el procedimiento de solución propuesto, el cual se basa en un método semiautomático de adquisición de patrones de manejo de verbos. Éste consiste en la elaboración manual de algunos patrones de manejo de verbos y el diseño de un método para el llenado automático de la base de patrones.

3.1. Definición de la Estructura del Patrón de Manejo Verbal

Como se menciona en la sección 2.1.3, una entrada en un ECD tiene 3 zonas: la zona semántica, la zona sintáctica y la zona de co-ocurrencia.

En esta investigación, un patrón de manejo verbal (PMV) comprende la zona semántica (acepción) y la zona sintáctica (patrón de manejo) de una entrada en un ECD.

La Figura 8 muestra la zona semántica del PMV del verbo *inhalar* con el sentido número 1. Al inicio de esta zona, se presenta el identificador (número) del *synset* que contiene al verbo, este identificador sirve para relacionar el PMV elaborado con la base de datos Spanish WordNet. Posteriormente, se muestra el verbo (*inhalar*) con su número de sentido, el cual se presenta como un subíndice (*inhalar₁*).

Cabe hacer mención que en la base de datos Spanish WordNet cada verbo puede tener uno o más sentidos, y por consiguiente, un verbo puede tener uno o más patrones de manejo.

00004020 inhalar ₁ (Medicine) X draws in air.
--

Figura 8. Zona semántica del PMV del verbo inhalar con el sentido 1.

Por lo anterior, en esta investigación se considera necesaria la relación verbo-sentido. Por lo que se hará referencia a esta pareja como verbo/sentido.

Finalmente, se presenta la acepción del verbo/sentido *inhalar₁ ((Medicine) X draws in air)*. En ésta se muestran los actuantes relacionados con el verbo con una literal (en el ejemplo de la Figura 8, el actuante es X).

En la Figura 9.A se muestra el patrón de manejo, el cual describe, de manera sintáctica, cada actuante regido por el verbo. En la Figura 9.B se describe cada elemento de los actuantes (*C_{1.1}: SN denota un sintagma nominal con un sustantivo animado*). En la Figura 9.C, se presenta una tabla que muestra las combinaciones de aparición de los actuantes con el verbo, y un ejemplo correspondiente a la combinación.

X = I				
1. SN				
<i>C_{1.1}: SN denota un sintagma nominal con un sustantivo animado.</i>				
<table border="1"> <thead> <tr> <th style="text-align: center;">Combinaciones</th> <th style="text-align: center;">Ejemplo</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;"><i>C_{1.1}</i></td> <td style="text-align: center;">{Un buen hombre} inhala [, exhala y acomoda sus ideas.]</td> </tr> </tbody> </table>	Combinaciones	Ejemplo	<i>C_{1.1}</i>	{Un buen hombre} inhala [, exhala y acomoda sus ideas.]
Combinaciones	Ejemplo			
<i>C_{1.1}</i>	{Un buen hombre} inhala [, exhala y acomoda sus ideas.]			

Figura 9. Zona sintáctica del PMV del verbo/sentido inhalar₁.

Al unir estas dos zonas, finalmente, obtenemos nuestro PMV correspondiente a algún verbo/sentido, por ejemplo, en la Figura 10 se presenta el PMV del verbo/sentido *inhalar₁*.

00004020			
inhalar ₁			
(Medicine) X draws in air.			
<table border="1"> <tr> <td>X = I</td> </tr> <tr> <td>1. SN</td> </tr> </table>		X = I	1. SN
X = I			
1. SN			
C _{1,1} : SN denota un sintagma nominal con un sustantivo animado.			
Combinaciones	Ejemplo		
C _{1,1}	{Un buen hombre} inhala [, exhala y acomoda sus ideas.]		

Figura 10. PMV del verbo/sentido inhalar₁.

3.2. Análisis de la Ontología Spanish WordNet 1.0

Como se muestra en la sección 2.3.2, existen 18 relaciones semánticas en la ontología Spanish WordNet 1.0, de las cuales se seleccionó la relación semántica de hiponimia (*has_hyponym*). Esta relación permite enlazar los synsets de la ontología en una jerarquía, lo cual es importante para obtener los patrones de manejo.

Por ejemplo, en la Figura 11 se observa la relación semántica de hiponimia verbal del verbo/sentido *respirar*₁. De tal manera que si se conoce el patrón de manejo del verbo/sentido *respirar*₁, se puede conocer el patrón de manejo de los verbo/sentido *espirar*₁ y *exhalar*₁, debido a que estos verbo/sentido son hipónimos de *respirar*₁. Asimismo, si se conoce el patrón de manejo de estos dos verbo/sentido, se puede conocer el patrón de manejo del verbo/sentido *soplar*₁, por lo antes descrito. Y éste, a su vez, servirá para conocer el patrón de manejo del verbo/sentido *resoplar*₂, por la misma situación antes mencionada.

Por lo tanto, el método básico de solución propuesto en esta investigación consiste en:

- Elaborar los patrones de manejo de los verbo/sentido que se encuentren en el nivel 1 de la jerarquía.
- Desarrollar un procedimiento automático para obtener los patrones de manejo de los verbo/sentido restantes mediante la hiponimia.

En la ontología Spanish WordNet 1.0, mediante la relación de hiponimia, los *synsets* de sustantivos y de verbos están clasificados según su nivel.

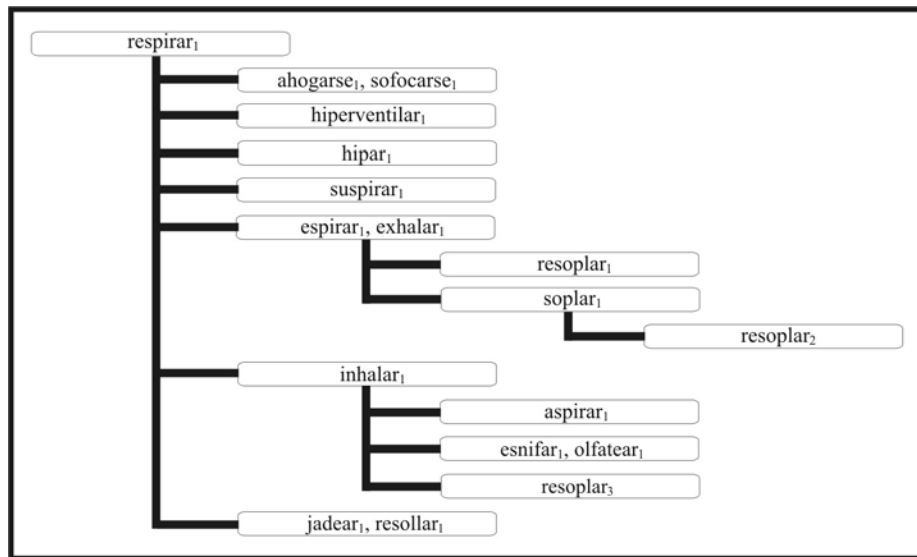


Figura 11. Ejemplo de relación semántica de hiponimia verbal en Spanish WordNet 1.0

Como se muestra en la Tabla 7, existen 366 *synsets* de verbos en el nivel 1 de Spanish WordNet [41]. Éstos se definieron como los *synsets base* utilizados para crear las valencias de los restantes. De tal manera que se hizo la asociación entre los patrones de manejo de los verbos con menos profundidad y los restantes, utilizando su relación de hiponimia.

Al analizar la relación semántica de hiponimia de los *synsets* en la ontología Spanish WordNet 1.0, se encontraron algunos inconvenientes, los cuales se describen a continuación.

Tabla 7. Relación de profundidad y palabras en Spanish WordNet

Profundidad	Sustantivos	Verbos
1	11	366
2	149	1165
3	1041	1174
4	2511	752
5	4354	375
6	5303	160
7	5016	66
8	3285	30
9	1856	3
10	967	2
11	347	
12	135	
13	76	
14	18	
15	4	

3.2.1. Conjunto de Palabras como Elemento de un *Synset*

Un *synset* está formado por diferentes elementos (generalmente, cada elemento es una sola palabra) que son sinónimos. En algunos casos, se encontraron conjuntos de palabras como elemento de un *synset*. Por ejemplo, en la Figura 12 se presenta el *synset* 00585450 que contiene 4 elementos; donde uno de los elementos no es sólo una palabra, sino un conjunto de palabras (*ponerse_de_acuerdo*₁).

avenirse₁, concordar₂, convenir₁, ponerse_de_acuerdo₁

Figura 12. Ejemplo de un conjunto de palabras como elemento de un *synset*.

Debido a que el objetivo de este trabajo es obtener patrones de manejo de verbos, los conjuntos de palabras como elementos de un *synset* no son tomados en consideración.

3.2.2. Obtención de los Significados para cada *Synset*

Una de las tablas (*sp_synsets*) de la ontología Spanish WordNet 1.0 contiene el significado de cada *synset* (ver Figura 13).

	SPOS	SSYN	ILI	NDES	STAT	SDEF
▶	v	00009090	00009090	12	-	
	v	00002403	00002403	20	--	
	v	00007021	00007021	32	-	
	v	00021823	00021823	3	--	
	v	00033668	00033668	2	-	
*						

Registro: 1 de 5

Figura 13. Significado de los synsets

Como se aprecia en la tabla *sp_synsets*, la columna *SDEF* no contiene el significado de los *synsets* de los verbo/sentido.

Debido a que el significado de un *synset* es la fuente de información de donde se obtienen los actuantes del verbo/sentido, fue necesario obtener de otra fuente los significados de los *synsets* (que fueron seleccionados en esta investigación).

Para esto, se consiguió una fuente de significados relacionada con la ontología que se utiliza (Spanish WordNet 1.0). La fuente de información es MultiWordNet [42], la cual es un conjunto de ontologías de diferentes idiomas con la estructura de WordNet. Los idiomas que se pueden consultar mediante la interfaz son italiano, inglés, español, hebreo y rumano. El acceso a la base de datos en español está disponible por cortesía del Centro de Tecnologías y Aplicaciones del Lenguaje y del Habla (Centro TALP) de la Universitat Politècnica de Catalunya (España).

Cabe mencionar que se encontraron algunos casos donde, al momento de obtener los sentidos de algún verbo, no coincidió el número de sentidos en Spanish WordNet para ese verbo con el número de sentidos en MultiWordNet para ese verbo. Por lo que aquellos verbos que se encontraron en este caso no fueron considerados dentro de la experimentación.

3.2.3. Synset sin ninguna palabra

Se encontraron casos donde un *synset* no está relacionado con alguna palabra, es decir, no existe representación léxica del significado que representa el *synset*. Estos *synsets* no fueron considerados en la experimentación realizada en esta investigación.

3.2.4. Hipónimos Relacionados con más de un Hiperónimo

Se encontraron casos donde un *synset* está relacionado con más de un *synset* hiperónimo. 25 *synsets* se encontraron en tal caso. En la Figura 14 se presenta el *synset* hipónimo 00971216 que está relacionado con los *synsets* hiperónimos 00965279 y 00968570.

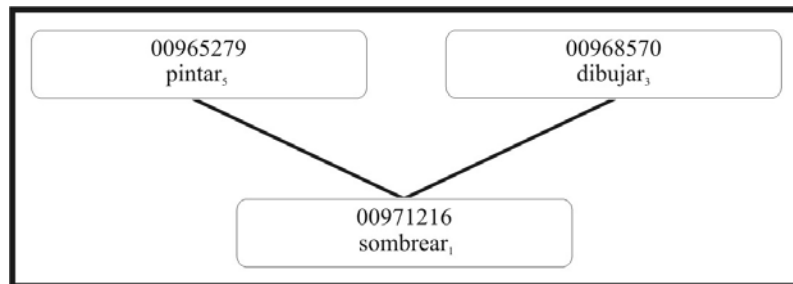


Figura 14. Ejemplo de un hipónimo que tiene más de un hiperónimo.

Este caso se tomó en cuenta en el desarrollo del módulo de patrones de manejo verbales.

3.3. La Animidad en el Complemento Directo

En este trabajo se considera que la preposición *a* en el Complemento Directo, la cual define la animidad del complemento, está determinada por la semántica del sustantivo que la acompaña.

Por lo tanto, **el verbo no determina si su Complemento Directo va acompañado de una preposición.** Por ejemplo, el verbo *animar* en la oración (1) permite la preposición *a* dado que su sustantivo es animado (*los aficionados del Espanyol*); por otra parte, el mismo

verbo en la oración (2) no permite la preposición *a* debido a que el sustantivo es inanimado (*el partido*). Lo mismo sucede con el verbo *exhalar* en (3) y (4).

- (1) “El gol acabó con los bostezos y animó **a** los aficionados del Espanyol...”⁵
- (2) “Una canasta triple de Manel Sánchez animó el partido...”⁶
- (3) “... debemos exhalar **a** Dios para impartir vida a otros.”⁷
- (4) “Cada una de ellas exhala unos veinte gramos de vapor de agua...”⁸

3.4. Elaboración de los Patrones de Manejo Verbal Base

Como se mencionó en la sección 3.2, existen 366 *synsets base* en la ontología Spanish WordNet. Debido a que la elaboración manual de más de 366 patrones de manejo⁹ de nivel 1 es una tarea ardua que requiere mucho tiempo, se trabajó con una muestra de 44 verbo/sentidos, la cual se obtuvo de 22 *synsets* de nivel 1.

Para la elaboración manual de los patrones de manejo de los verbo/sentido *base*, se decidió entablar relación con alguna persona experta en el tema. De tal manera que se contactó a la M. L. Gladis Miriam Galiana Bravo, profesora del Instituto Tecnológico de Ciudad Madero, quien es Maestra en Letras de Lingüística de la Université de Toulouse-le-Mirail (Francia), para participar en el desarrollo de esta investigación.

⁵ T.I.S.A (Barcelona), La Vanguardia, 03/04/1995 : DAGOBERTO ESCORCIA, España, 1995.

⁶ Control (La Coruña), La Voz de Galicia : Tikhonenko derrotó al Oar, mientras el Breogán ganó con apuros al Valvi, España, 1991.

⁷ Hechos 6:4

⁸ T.I.S.A (Barcelona), La Vanguardia, 02/10/1995 : La pirámide de Kefrén cierra para ser restaurada, España, 1995.

⁹ Un *synset* puede contener más de un verbo.

Los pasos realizados para elaborar manualmente los patrones de manejo de los verbo/sentido *base* fueron los siguientes:

1. Seleccionar un verbo/sentido de la muestra.
2. Establecer manualmente la valencia del verbo/sentido seleccionado, mediante su acepción.
3. Crear la estructura sintáctica de una oración con respecto al verbo, según los ejemplos presentados en la acepción (si existen).
4. Verificar el patrón de manejo del verbo/sentido con oraciones del corpus publicado por la RAE (Real Academia Española) [43].
5. Revisar el patrón de manejo por parte de un experto en lingüística.

3.5. Modificaciones en la Base de Patrones

Como se menciona en la sección 2.4, la base de patrones contiene una tabla denominada *combinaciones*; en la cual no aparece una columna que identifique uno de los varios sentidos que puede tener un verbo.

En (1) y (2) se presentan oraciones con el verbo *enajenar*. En (1) el verbo *enajenar* tiene la acepción *Sacar a alguien fuera de sí, entorpecerle o turbarle el uso de la razón o de los sentidos* y su patrón es Suj(SN) V CD(SN); mientras que en (2) el verbo *enajenar* tiene la acepción *Pasar o transmitir a alguien el dominio de algo o algún otro derecho sobre ello* y su patrón es Suj(SN) V CD(SN) CI(SN).

(1) [Suj El miedo] [v enajenó] [CD a Javier].

(2) [Suj Don Luis] [v enajenó] [CD el carro] [CI a su hijo].

Para diferenciar los sentidos de cada verbo se agregó una columna en la tabla *combinaciones* que representa el sentido del verbo. Este campo es denominado *synset*, que

hace referencia al número de *synset* al que pertenece el verbo en Spanish WordNet 1.0, según su sentido.

Debido a la relación que existe entre la clase funcional de un verbo y su valencia, se agregó una nueva tabla a la base de patrones, la cual es denominada *clases*.

Esta tabla contiene tres columnas: *claveverbo*, que representa la clave del verbo; *synset*, que representa el sentido del verbo en Spanish WordNet; y *clasefuncional*, que representa la clase funcional del verbo (intransitivo, transitivo, impersonal-intransitivo, impersonal-transitivo, pronominal). Esta información ayudará a obtener los patrones de manejo restantes.

3.6. Algoritmo del Módulo de Llenado

Para el desarrollo del módulo de llenado se diseñaron dos algoritmos: el *submódulo de llenado automático total* (Figura 15), el cual utiliza los verbo/sentido *base* para llenar automáticamente la base de patrones; y el *submódulo de llenado automático por niveles* (Figura 16) que utiliza un archivo donde se encuentran algunos verbo/sentido que toma como *base*, y obtiene los patrones de manejo de los verbo/sentido hipónimos.

El *submódulo de llenado automático total* realiza una búsqueda en anchura (tomando como punto de partida los *synsets base*) en la ontología Spanish WordNet para obtener los *synsets* hipónimos. En cada *synset* hipónimo, obtiene los verbo/sentido, los cuales se procesan en el *submódulo de consulta de verbo con sentido*, para crearles su PMV.

En la Figura 17 se muestra el *submódulo de consulta de verbo con sentido*, el cual verifica que exista el verbo/sentido en la ontología Spanish WordNet. Si lo encuentra, éste es procesado por el *submódulo de consulta, asignación y elaboración de patrones*, para obtener su patrón de manejo (PM).

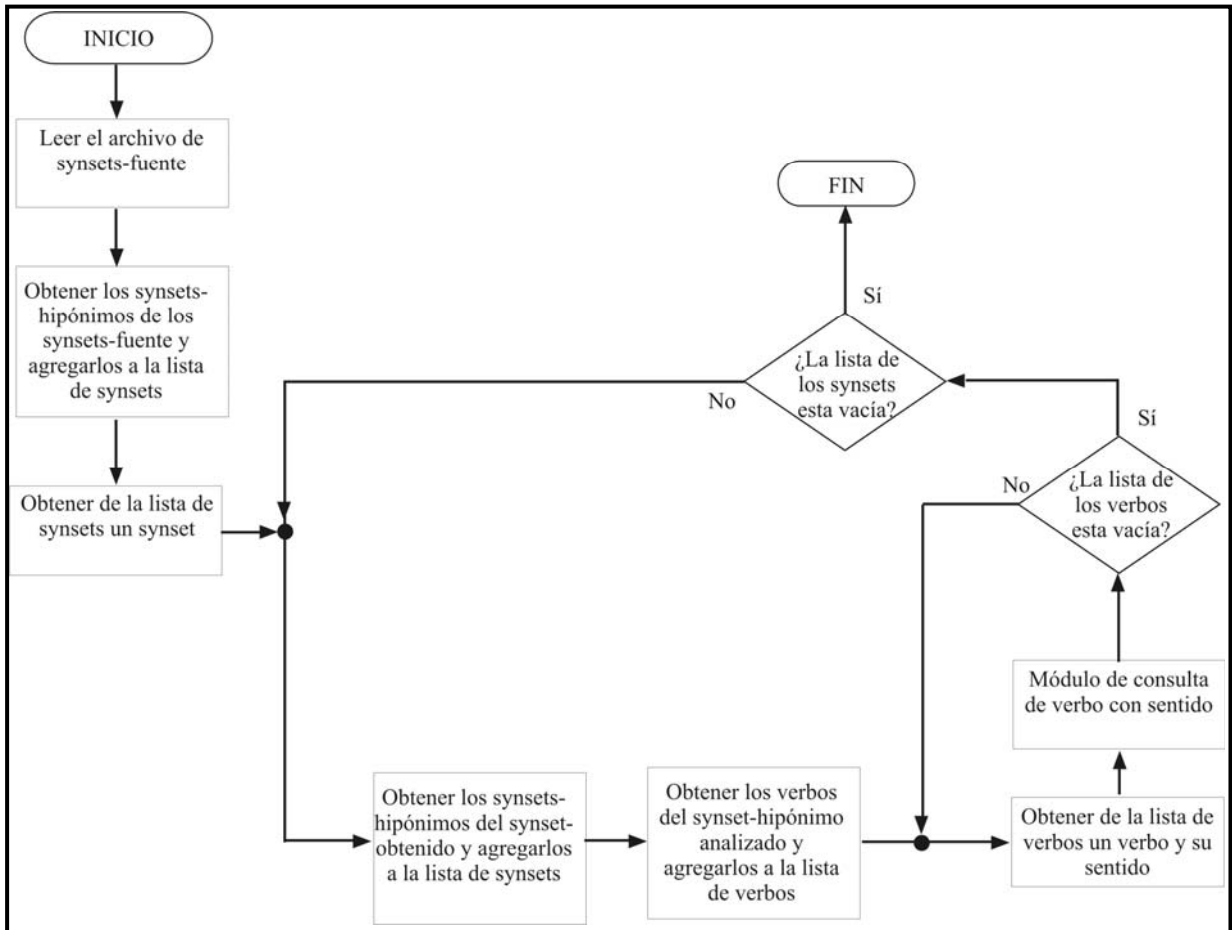


Figura 15. Submódulo de llenado automático total.

En el *submódulo de consulta, asignación y elaboración de patrones* (Figura 18 y Figura 19), dado un verbo/sentido, éste se consulta en la base de patrones. Si existe el PM en la base de patrones, se obtiene y se muestra su clase funcional y su PM.

En caso de que no exista el patrón de manejo en la base de patrones para el verbo/sentido analizado, se identifica cuáles son sus verbo/sentido-hiperónimos, así como las clases funcionales de éstos. Una vez obtenido lo anterior, se crea el patrón de manejo verbal al verbo/sentido analizado, en uno de los modos que se describen a continuación.

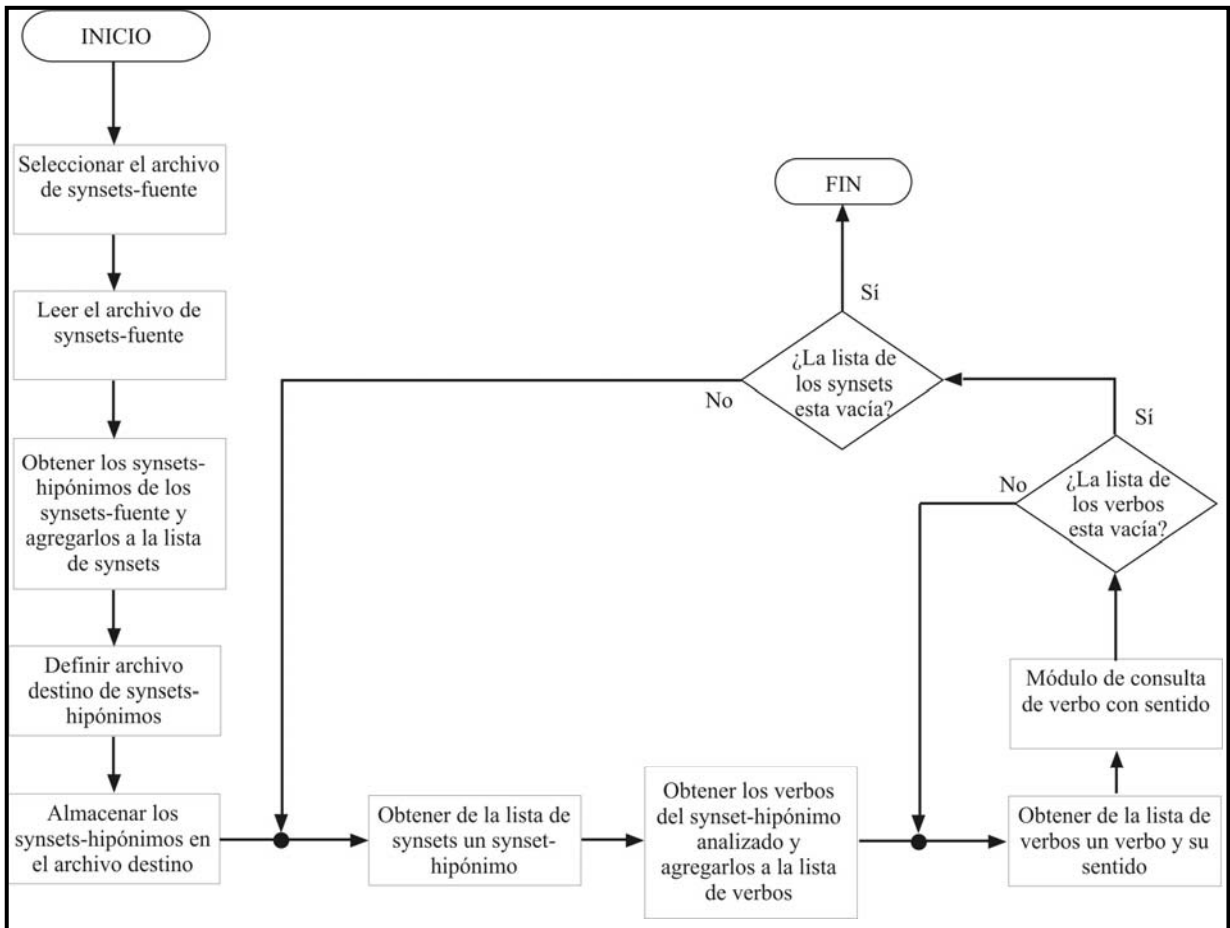


Figura 16. Submódulo de llenado automático por niveles.

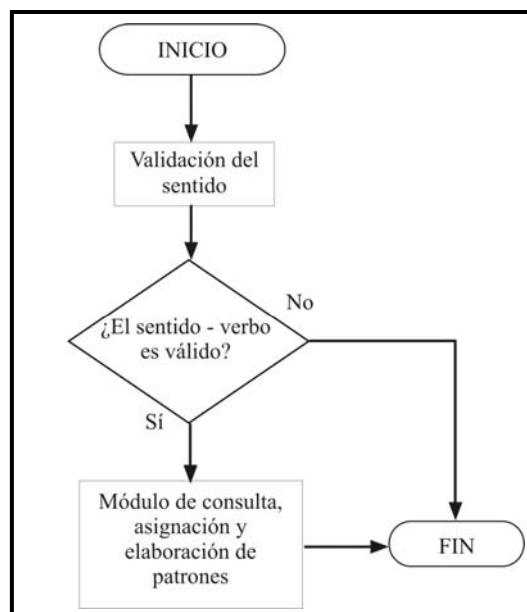


Figura 17. Módulo de consulta de verbo con sentido.

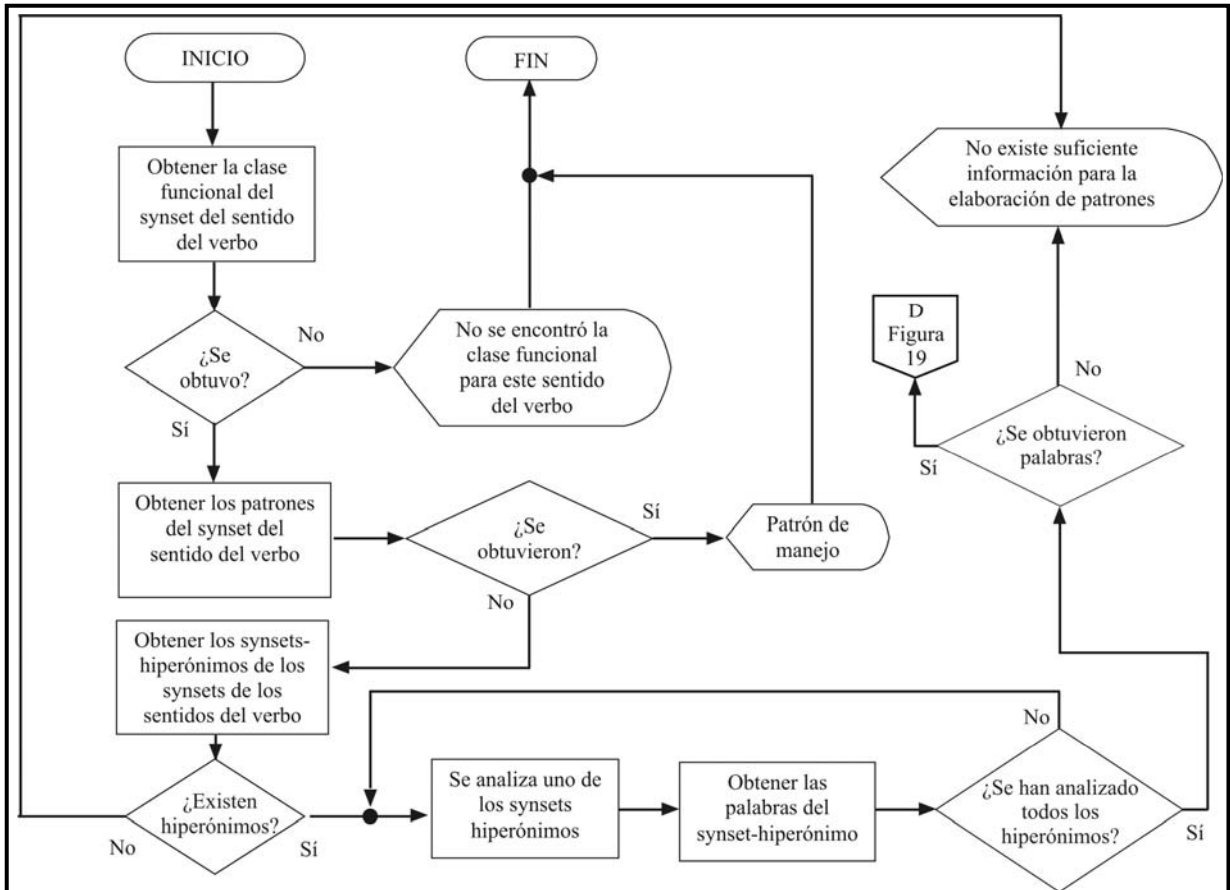


Figura 18. Módulo de consulta, asignación y elaboración de patrones - parte 1.

- Asignación: Cuando la clase funcional del verbo/sentido analizado es igual a la de algún verbo/sentido-hiperónimo, se le copia el PM de su verbo/sentido hiperónimo.

Por ejemplo, el verbo/sentido *inyectar₂* con clase funcional transitiva para crearle su PM, primeramente se le obtiene su *synset*-hiperónimo; el cual es: 00110396. Posteriormente se obtiene la lista de sus verbos/sentido con su clase funcional, los cuales son: *adicionar₁* (transitivo), *agregar₁* (transitivo), *añadir₁* (transitivo), *agregarse₁* (pronominal). De esta lista, se seleccionan los verbos/sentido con la misma clase funcional de *inyectar₂*, que son: *adicionar₁*, *agregar₁*, *añadir₁*.

De la lista resultante, se selecciona al azar alguno de los verbos/sentido y se obtiene su PM, para este ejemplo es *agregar₁* el cual tiene el PM [Suj(SN)] V CD(S|a S|Pro) CI(a S|Pro). Por último, este PM se le asigna al verbo/sentido analizado.

Por lo tanto, el PM de *inyectar*₂ es [Suj(SN)] V CD(S| a S| Pro) CI(a S|Pro).

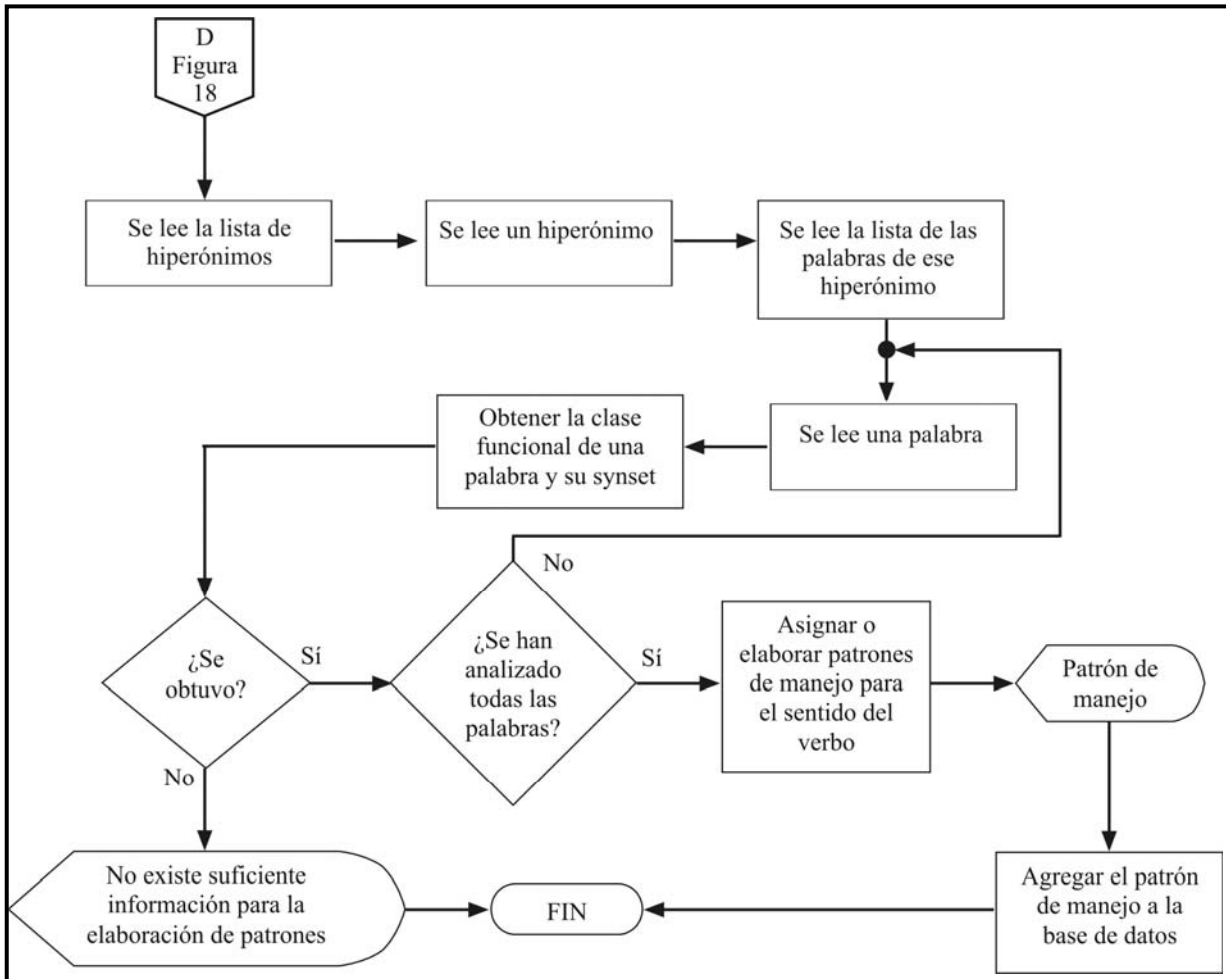


Figura 19. Módulo de consulta, asignación y elaboración de patrones - parte 2.

- **Elaboración:** Cuando la clase funcional del verbo/sentido analizado es diferente a la de sus verbo/sentido-hiperónimo, se elabora su PM a partir del PM de algún verbo/sentido-hiperónimo seleccionado al azar y su clase funcional.

Por ejemplo, dado el verbo/sentido *ahogarse*₁ con su clase funcional pronominal, se obtiene su *synset*-hiperónimo. De éste se obtiene una lista de los verbos/sentido y su clase funcional; para este *synset*-hiperónimo sólo se tiene un verbo/sentido, *respirar*₁ (intransitivo).

Como en la lista anterior no existe algún verbo/sentido con la misma clase funcional del verbo/sentido *ahogarse*₁, se elabora el PM con respecto a las reglas siguientes:

- Si la clase funcional es impersonal-intransitivo el PM es V.
- Si la clase funcional es impersonal-transitivo el PM es V CD(S|a S|Pro).
- Si la clase funcional es intransitivo el PM es [Suj(SN)] V.
- Si la clase funcional es transitivo el PM es [Suj(SN)] V CD(S|a S|Pro).
- Si la clase funcional es pronominal el PM es [Suj(SN)] Pro-V.

De la lista de los verbos/sentido anterior, se selecciona al azar alguno de los verbos/sentido y se obtiene su PM, para este ejemplo, la lista sólo contiene *respirar*₁, el cual tiene el PM [Suj(SN)] V. Por último, a este PM se le extraen aquellos complementos diferentes al CD, y se le copian al PM elaborado. Para este ejemplo, *respirar*₁ no tiene complementos adicionales, tal como CP o CI.

Por lo tanto, el PM de *ahogarse*₂ es [Suj(SN)] Pro-V.

El *submódulo de llenado automático por niveles* obtiene los *synsets*-hipónimos, en la ontología Spanish WordNet, de un conjunto de *synsets*-fuente¹⁰. Para cada *synset*, se obtienen los verbo/sentido y se procesan en el *submódulo de consulta de verbo con sentido*.

Como se mencionó en la sección 3.2.4, un hipónimo puede tener más de un hiperónimo. En estos casos se identificó que algunos hipónimos pueden tener un hiperónimo de un nivel de profundidad distinto al nivel de profundidad de otro de sus hiperónimos.

Por ejemplo, el hiperónimo A tiene el nivel de profundidad 3, y el hiperónimo B tiene el nivel de profundidad 4; si el *synset* C es hipónimo de ambos (A y B), entonces ¿cuál será el nivel de profundidad de C? Esta situación influye en el llenado automático total de la base de patrones. Por lo que, para solucionar esta situación, se propone la etiquetación de cada *synset* con su nivel de profundidad antes de iniciar el proceso de llenado, mediante un algoritmo del camino más largo que defina la profundidad desde el *synset* analizado hasta el *synset base*. La descripción de dicho algoritmo se puede encontrar en [44].

¹⁰ Un *synset*-fuente es un *synset* tomado como base para obtener sus hipónimos. El *synset*-fuente se obtiene de un archivo de texto que contiene una lista de *synsets*.

Capítulo 4

Experimentación

En este capítulo se describe la experimentación realizada para probar la eficacia del enfoque propuesto. Se presentan los casos de prueba, se describe la plataforma de implementación del software en el que se realizó la experimentación. Finalmente, se describen los experimentos realizados, se muestran los resultados obtenidos y se realiza el análisis de los mismos.

4.1. Casos de Prueba

Como se mostró en la Tabla 7 de la sección 3.4, en la ontología Spanish WordNet los *synsets* están clasificados por niveles. La muestra de 44 verbo/sentido-*base* que se les utilizó para las pruebas, corresponden al nivel 1 (verbo/sentido nivel 1).

A partir de éstos, se obtuvieron los primeros casos de prueba que corresponden a los verbo/sentido de los *synsets* de nivel 2 (verbo/sentido nivel 2), a los cuales se les elaboraron o asignaron los patrones de manejo.

De la misma forma, estos patrones de manejo sirvieron como apoyo para obtener los patrones de manejo de los verbo/sentido de los *synsets* de nivel 3 (verbo/sentido nivel 3).

Asimismo, con estos últimos se obtuvieron los patrones de manejo de los verbo/sentido de los *synsets* de nivel 4 (verbo/sentido nivel 4).

De la misma manera, este proceso puede continuar hasta llegar al último nivel.

Cabe mencionar que, cada vez que se inicia la obtención de los patrones de manejo de los verbo/sentido de un nivel, se verifica que los patrones de manejo de sus *synsets* hiperónimos estén correctos.

Como se aprecia en la Figura 20, los casos de prueba corresponden al área sombreada de la ontología Spanish WordNet.

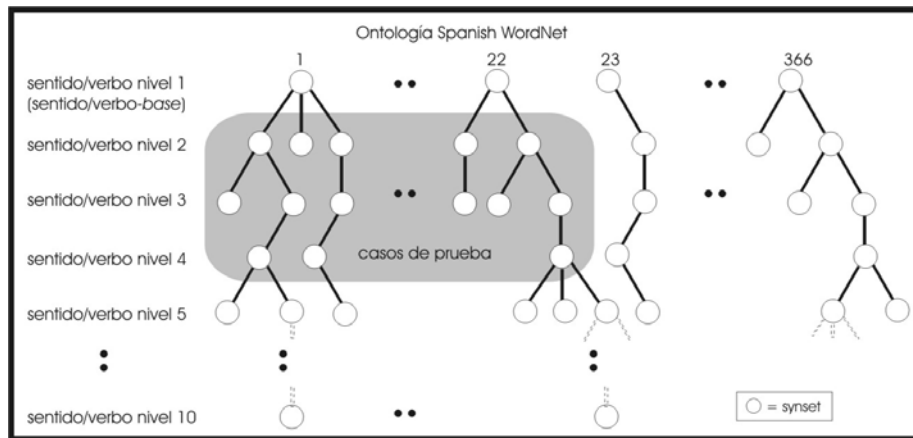


Figura 20. Casos de prueba.

4.2. Experimentación

El *módulo de llenado de la base de patrones* fue implementado (a excepción del proceso de etiquetación) en el lenguaje de programación orientado a objetos Java y en un equipo de cómputo con un procesador Pentium 4 a 1.6 GHz, 352 MB de RAM y 27.9 GB de disco duro.

La experimentación consistió en tres pruebas. Y para verificar que fueran correctos los 195 patrones de manejo obtenidos por el *módulo de llenado de la base de patrones*, éstos se elaboraron previamente de manera manual.

4.2.1. Resultados de Experimento 1

El primer experimento consistió en elaborar o asignar los PM de 100 verbo/sentido de nivel 2, que son los verbo/sentido-hipónimos de los verbo/sentido de nivel 1.

De los PMs obtenidos de los 100 verbo/sentido de nivel 2, 66 son correctos y 34 son incorrectos. Estos últimos presentan diferentes errores (9 en el modo de elaboración y 25 en el modo de asignación), los cuales son descritos a continuación.

Errores obtenidos en el modo de elaboración

Cuando un PM se elabora a partir del PM de su(s) hiperónimo(s) que tiene(n) diferente clase funcional a la del verbo/sentido analizado, en ocasiones, se llega a los siguientes errores:

Verbo/sentido pronominal sin complementos necesarios

Este error ocurrió 4 veces; al generar para verbos pronominales el PM siguiente:

- [Suj(S)] Pro-V

Dicho PM no acepta Complementos Preposicionales (CP) o Atributos (Atr), los cuales son necesarios para algunos verbos pronominales.

Por ejemplo, el PM generado para el verbo *convertirse*₁ permite la oración incorrecta (1).

(1) [Suj La plástica] [v se convirtió].

Mientras que el PM correcto acepta un complemento preposicional como se muestra en la oración (2).

(2) [Suj La plástica] [v se convirtió] [CP en un desastre].

El PM del verbo/sentido *convertirse*₁ se obtuvo del PM [Suj(SN)] V del verbo/sentido *cambiar*₂. Este PM no incluye el CP(en SN), motivo por el cual el PM que se le elabora al verbo/sentido *convertirse*₁ es incorrecto.

Verbo/sentido transitivo con complemento incorrecto

Este error ocurrió 5 veces al generar para verbos transitivos el PM siguiente:

- [Suj(S)] V CD(a S|Pro)

A dicho PM se le agrega algún(os) complemento(s) como Complemento Circunstancial (CC), obtenido de uno de sus hiperónimos seleccionado al azar, tal(es) complemento(s) es(son) incorrecto(s) para algunos verbos transitivos.

Por ejemplo, el PM obtenido para el verbo *hacer*₁ permite la oración incorrecta (3).

(3) [Suj Samuel] [v hizo] [CD a su novia][CC bien].

Mientras que el PM correcto acepta un atributo como se muestra en la oración (4).

(4) [Suj Samuel] [v hizo] [CD a su novia] [Atr feliz].

El PM del verbo/sentido *hacer*₁ se obtuvo del PM [Suj(SN)].Pro-V.CC(Adv) del verbo/sentido *cambiar*₂. Este PM incluye el CC(Adv) y no incluye el Atr(Adj), motivo por el cual el PM que se le elabora al verbo/sentido *hacer*₁ es incorrecto.

Errores obtenidos en el modo de asignación

Cuando un PM se asigna a partir del PM de su(s) hiperónimo(s) sin tomar en cuenta la clase funcional del verbo/sentido analizado, en ocasiones, se llega a los siguientes errores:

Verbo/sentido transitivo con Complemento Directo opcional

- Este error ocurrió 3 veces al copiar el PM a un verbo transitivo, cuyo CD es opcional.

Por ejemplo, el PM obtenido para el verbo *ensanchar*₄ es incorrecto debido a que permite la ausencia del CD como se muestra en la oración (5).

(5) [Suj Los trabajadores] [v ensancharon].

Cuando el Complemento Directo debe ser obligatorio, por ejemplo, la oración (6).

(6) [Suj Los trabajadores] [v ensancharon] [CD la carretera].

El PM del verbo/sentido *ensanchar*₄ se obtuvo del PM [Suj(SN)] V CD(SN) del verbo/sentido *cambiar*₃. Este PM incluye el CD opcional, motivo por el cual el PM que se le asigna al verbo/sentido *ensanchar*₄ es incorrecto.

Verbo/sentido con complementos incorrectos

Este error ocurrió 15 veces al copiar el PM a un verbo, cuyo complemento (indirecto, preposicional o atributo) es incorrecto.

Un ejemplo de este error es el verbo *fanfarronear*₁ en la oración (7).

(7) [Suj Juan] [v fanfarronea] [CC mal].

Mientras que la forma correcta de la oraciones anterior es (8).

(8) [Suj Juan] [v fanfarronea].

El PM del verbo/sentido *fanfarronear*₁ se obtuvo del PM [Suj(SN)] V CC(Adv) del verbo/sentido *actuar*₁. Este PM incluye el CC(Adv), motivo por el cual el PM que se le asigna al verbo/sentido *fanfarronear*₁ es incorrecto.

Verbo/sentido sin complementos necesarios

Este error ocurrió 7 veces al copiar el PM a un verbo, el cual no acepta Complementos Preposicionales (CP) o Complementos Circunstanciales (CC), los cuales son necesarios para algunos verbos.

Por ejemplo, el PM obtenido para el verbo *comportarse*₂ no permite la oración correcta (9).

(9) [Suj Luis] [v se comportó] [CC como loco].

El PM del verbo/sentido *comportarse*₂ se obtuvo del PM [Suj(SN)] V CC(Adv) del verbo/sentido *actuar*₁. Este PM no incluye el CC(como SN|Adv), motivo por el cual el PM que se le asigna al verbo/sentido *comportarse*₂ es incorrecto.

4.2.2. Resultados del Experimento 2

El segundo experimento consistió en elaborar o asignar los patrones de manejo de 60 verbo/sentido de nivel 3, que son los verbo/sentido-hipónimos de los verbo/sentido de nivel 2. Los verbo/sentido de nivel 2 fueron modificados con el patrón de manejo correcto en caso de que éste no fuera obtenido.

De los PMs obtenidos de los 60 verbo/sentido de nivel 2, 54 son correctos y 6 son incorrectos. Estos últimos presentan diferentes errores (los 6 en el modo de asignación), los cuales se describen a continuación.

Errores obtenidos en el modo de asignación

Cuando un PM se asigna a partir del PM de su(s) hiperónimo(s) sin tomar en cuenta la clase funcional del verbo/sentido analizado, en ocasiones, se llega a los siguientes errores:

Verbo/sentido sin complementos necesarios

Este error ocurrió 6 veces al asignar el PM a un verbo, el cual no permite que existan Complementos Preposicionales (CP) o Complementos Indirectos (CI), los cuales son necesarios junto al verbo.

Por ejemplo, el PM obtenido para el verbo *ajustar*₂ permite la oración incorrecta (10).

(10) [Suj La señorita] [v ajustó] [CD el vestido].

Mientras que la oración correcta es (11).

(11) [Suj La señorita] [v ajustó] [CD el vestido] [CI a su figura].

El PM del verbo/sentido *ajustar*₂ se obtuvo del PM [Suj(SN)] V CD(S|a S|Pro) del verbo/sentido *adaptar*₁. Este PM no incluye el CI(a SN|Pro), motivo por el cual el PM que se le asigna al verbo/sentido *ajustar*₂ es incorrecto.

4.2.3. Resultados del Experimento 3

El tercer experimento consistió en elaborar o asignar los patrones de manejo de 35 verbo/sentido de nivel 4, que son los verbo/sentido-hipónimos de los verbo/sentido de nivel 3. Los verbo/sentido de nivel 3 fueron modificados con el patrón de manejo correcto en caso de que éste no fuera obtenido.

De los patrones de manejo obtenidos de los 35 verbo/sentido de nivel 4, 34 son correctos y 1 es incorrecto. Este último se presentó en el modo de asignación, y se describe a continuación.

Error obtenido en el modo de asignación

Cuando un PM se asigna a partir del PM de su(s) hiperónimo(s) sin tomar en cuenta la clase funcional del verbo/sentido analizado, se llegó al siguiente error:

Verbo/sentido transitivo sin complementos necesarios

Este error ocurrió 1 vez al asignar el PM a un verbo transitivo, el cual no permite que existan Complementos Preposicionales (CP) o Complementos Indirectos (CI), los cuales son necesarios junto al verbo.

El PM obtenido para el verbo *administrar*₂ permite la oración incorrecta (12).

(12) [Suj El sacerdote] [v administró] [CD la extrema unción].

Mientras que la oración correcta es (13).

(13) [Suj El sacerdote] [v administró] [CD la extrema unción] [CI al moribundo].

El PM asignado se obtuvo del verbo/sentido *aplicar*₁, el cual tiene el PM [Suj(S)] V CD(S|a S|Pro). Dicho PM no tiene un CI(a SN|Pro), por lo cual aparece este error.

El PM del verbo/sentido *administrar*₂ se obtuvo del PM [Suj(S)] V CD(S|a S|Pro) del verbo/sentido *aplicar*₁. Este PM no incluye el CI(a SN|Pro), motivo por el cual el PM que se le asigna al verbo/sentido *administrar*₂ es incorrecto.

4.2.4. Análisis de los Resultados

Para conocer la eficacia del *módulo de llenado de la base de patrones*, se calculó el porcentaje de correctez de los patrones de manejo obtenidos.

El porcentaje de acierto (PA) de un verbo sentido se obtuvo comparando cada una de las combinaciones¹¹ posibles del patrón de manejo correcto con las combinaciones posibles del correspondiente patrón de manejo obtenido, y utilizando la siguiente fórmula:

$$PA = \frac{\text{Número de combinaciones obtenidas correctas}}{\text{Número de combinaciones posibles correctas}}$$

En la fórmula anterior se divide el número de combinaciones obtenidas correctas (NCOC) entre el número de combinaciones posibles correctas (NCPC). También, como parte de la evaluación del *módulo de llenado de la base de patrones*, se calculó el número de combinaciones obtenidas incorrectas (NCOI) y número de combinaciones sobregeneradas (NCS).

En la Tabla 8 se muestran los resultados obtenidos del experimento 1. Donde cada caso representa una combinación de NCPC, NCOC, NCOI y NCS. Además, se muestra el número de verbos/sentido (NVS) que corresponden a cada caso. Por ejemplo, el caso 4 en la Tabla 8 tiene la combinación 4-NCPC, 2-NCOC, 0-NCOI, 0-NCS; dicho caso tiene un PA de 50% (esto es, 2-NCOC / 4-NCPC). Además, tal caso ocurrió con un verbo/sentido (1-NVS).

En la Tabla 9 se muestran los resultados obtenidos del experimento 2, donde se aprecia que el promedio de PA por verbo/sentido aumenta en 9% con respecto al obtenido en el experimento 1 y , además, el promedio de NCOI por verbo/sentido disminuye en 1.35.

En la Tabla 10 se muestran los resultados obtenidos del experimento 3, donde se aprecia que el promedio de PA por verbo/sentido aumenta en 7.14% con respecto al obtenido en el experimento 1 y , además, el promedio de NCOI por verbo/sentido disminuye en 36.

¹¹ Ver Figura 6

Tabla 8. Resultados obtenidos del experimento 1

Caso	NCPC	NCOC	NCOI	NCS	PA	NVS
1	2	0	2	0	0%	9
2	2	2	0	0	100%	23
3	4	0	2	0	0%	1
4	4	2	0	0	50%	1
5	6	0	6	0	0%	5
6	6	0	6	12	0%	1
7	6	6	0	0	100%	30
8	6	6	0	2	100%	6
9	6	6	0	6	100%	11
10	6	6	0	12	100%	2
11	12	0	6	0	0%	2
12	12	6	6	0	50%	1
13	12	12	0	0	100%	4
14	18	18	0	0	100%	4
$\text{Porcentaje de NCS} = \frac{\sum_{Caso=1}^{14} NCS_{Caso}}{\sum_{Caso=1}^{14} (NCOC_{Caso} + NCOI_{Caso} + NCS_{Caso})} = 23.53\%$						
$\text{Promedio de PA} = \frac{\sum_{Caso=1}^{14} (PA_{Caso} * NVS_{Caso})}{\sum_{Caso=1}^{14} NVS_{Caso}} = 81.00\%$						

Tabla 9. Resultados obtenidos del experimento 2

Caso	NCPC	NCOC	NCOI	NCS	PA	NVS
1	2	0	2	0	0%	1
2	2	2	0	0	100%	18
3	6	0	6	0	0%	2
4	6	6	0	0	100%	36
5	12	0	6	0	0%	3
$\text{Porcentaje de NCS} = \frac{\sum_{Caso=1}^5 NCS_{Caso}}{\sum_{Caso=1}^5 (NCOC_{Caso} + NCOI_{Caso} + NCS_{Caso})} = 0.00\%$						
$\text{Promedio de PA} = \frac{\sum_{Caso=1}^5 (PA_{Caso} * NVS_{Caso})}{\sum_{Caso=1}^5 NVS_{Caso}} = 90.00\%$						

Tabla 10. Resultados obtenidos del experimento 3

Caso	NCPC	NCOC	NCOI	NCS	PA	NVS
1	2	2	0	0	100%	6
2	6	6	0	0	100%	28
3	12	0	6	0	0%	1
$\text{Porcentaje de NCS} = \frac{\sum_{Caso=1}^3 NCS_{Caso}}{\sum_{Caso=1}^3 (NCOC_{Caso} + NCOI_{Caso} + NCS_{Caso})} = 0.00\%$						
$\text{Promedio de PA} = \frac{\sum_{Caso=1}^3 (PA_{Caso} * NVS_{Caso})}{\sum_{Caso=1}^3 NVS_{Caso}} = 97.14\%$						

En la Tabla 11 se presentan los resultados obtenidos, donde se aprecia el promedio de PA y el porcentaje de NCOI por experimento, y un promedio total de todos los experimentos.

Tabla 11. Porcentaje de acierto de los experimentos.

Número de experimento	Promedio de PA	Porcentaje de NCS
Experimento 1	81.00%	23.53%
Experimento 2	90.00%	0.00%
Experimento 3	97.14%	0.00%
Promedio	89.38%	7.84%

Como se puede apreciar en la Tabla 11 el porcentaje de acierto se incrementó en cada experimento. Como cada experimento representa un nivel en la ontología, por lo tanto el porcentaje de acierto incrementa conforme se avanza en cada nivel de la ontología.

Capítulo 5

Conclusiones

En este capítulo se presentan las contribuciones de esta investigación, y se describen los posibles trabajos futuros identificados en el desarrollo del trabajo realizado.

5.1. Conclusiones

Las aportaciones más importantes de esta investigación son las siguientes:

- Se elaboraron manualmente y validaron 239 patrones de manejo de verbos en el idioma español. Este conocimiento lingüístico puede ser utilizado en otras investigaciones relacionadas a los patrones de manejo como referencia para evaluar su funcionamiento.
- Se modificó la base de patrones del Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET), de tal manera que dicha base de datos permitiera diferenciar los sentidos de los verbos.

- Se desarrolló un módulo de llenado de la base de patrones. Los resultados obtenidos al elaborar los patrones de manejo de verbo/sentido en tres niveles (nivel 2, nivel 3 y nivel 4) de la ontología Spanish WordNet tienen un promedio de PA del 89.38% y un promedio de porcentaje de NCS de 7.84%.
 - Se identificó que el porcentaje de acierto de los resultados obtenidos de la experimentación en la obtención de los patrones de manejo mejora conforme se incrementa el nivel de profundidad en la ontología. Posiblemente, esto es debido a que el significado de los verbos hace posible que se requieran menos valencias.
 - Se identificó que la animidad o inanimidad en el Complemento Directo, está determinada por la semántica de éste, y no está determinada por el significado del verbo. Por ejemplo, en la oración (1), el Sintagma Nominal del Complemento Directo es precedido por la preposición *a*, debido a que el Sintagma Nominal es animado (*Dios*); mientras que en la oración (2), el Sintagma Nominal del Complemento Directo no es precedido por la preposición *a*, debido a que el Sintagma Nominal es inanimado (*unos veinte gramos de vapor de agua*).
- (1) “... debemos exhalar a Dios para impartir vida a otros.”
- (2) “Cada una de ellas exhala unos veinte gramos de vapor de agua...”

5.2. Trabajos Futuros

Los resultados de esta investigación permiten identificar algunos nichos de oportunidad que sirven de base a futuras investigaciones. Algunos de esos nichos son los siguientes:

- Debido a que existe la necesidad de llenar la base de patrones del Centro Nacional de Investigación y Desarrollo Tecnológico para que sirva como conocimiento lingüístico en el análisis sintáctico, se propone generar los patrones de manejo del resto de los verbo/sentido-*base* en la ontología Spanish WordNet.
- La información de la clase funcional de los verbo/sentido fue obtenida del Diccionario de la Lengua Española, y capturada manualmente en la base de patrones. Se propone que esta información pueda ser obtenida automáticamente del mismo diccionario u otra fuente de información.
- Debido a que se obtuvo un promedio de PA de 89.38% y un promedio de porcentaje de NCS de 7.84 en los resultados obtenidos, se propone realizar un método semejante para encontrar las valencias de otras clases de palabras, es decir, sustantivos y adjetivos.

Referencias

- [1] Ushioda A., Evans D., Gibson T., Waibel A., “The Automatic Acquisition of Frequencies of Verb Subcategorization Frames from Tagged Corpora”, Computational Linguistics Program, Univ. Carnegie Mellon, 1993.
- [2] Andrew G., Grenager T., Manning C., “Verb Sense and Subcategorization: Using Joint Inference to Improve Performance on Complementary Tasks”, EMNLP, 2004.
- [3] Domínguez A., “Implementación de un Analizador Gramatical en Lenguaje Español”, tesis de maestría, Instituto Tecnológico de Ciudad Madero, 2002.
- [4] Gelbukh A., Bolshakov I., “Avances y Perspectivas de Procesamiento Automático de Lenguaje Natural”, J. IPN Ciencia, Arte: Cultura, N 31 Vol. II, mayo-junio 2000, ISSN 1405-2822. IPN, Mexico, pp. 10-18.
- [5] Palomar M., Ferrández A., "Introducción al Procesamiento del Lenguaje Natural", <http://www.dlsi.ua.es/~antonio/PLN.htm>, junio 2001
- [6] Jurafsky D., Martín J., “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistic, and Speech Recognition”, Prentice Hall, USA, 2000.
- [7] García R., “Modelado Gramatical de un Subconjunto del Lenguaje Español”, Centro Nacional de Investigación y Desarrollo Tecnológico, 2003.
- [8] Sosa E., “Procesamiento del lenguaje natural: revisión del estado actual, bases teóricas y aplicaciones (Parte I)”, http://www.elprofesionaldelainformacion.com/contenidos/1997/enero/procesamiento_del_lenguaje_natural_revisin_del_estado_actual_bases_tericas_y_aplicaciones_parte_i.html, 1997.
- [9] Aho V., Sethi R., Ullman J., “Compiladores principios, técnicas y herramientas”, Addison-Wesley Longman, 1998.
- [10] Galicia S., “Análisis Sintáctico Conducido por un Diccionario de Patrones de Manejo Sintáctico para Lenguaje Español”, tesis doctoral, Centro de Investigación en Computación, Instituto Politécnico Nacional, 2000.

-
- [11] Galicia S., Gelbukh A., “Investigaciones en Análisis Sintáctico para el Español”, Centro de Investigación en Computación, Instituto Politécnico Nacional, 2006.
- [12] Mel’cuk I., Polguère A., “A Formal Lexicon in the Meaning-Text Theory”, Computational Linguistics, Volume 13, Numbers 3-4, July-December 1987.
- [13] Mel’cuk I., Žolkovskij A., “Explanatory Combinatorial Dictionary of Modern Russian”, Vienna: Wiener Slawisticher Almanach, 1984.
- [14] Mel’cuk I., et al., “Dictionnaire explicatif et combinatoire du français contemporain” Vol. 1, 2, 3, 4, Montreal: Presses de l’Univ. Montréal, 1984, 1988, 1992, 1999.
- [15] Ilson R., Mel’cuk, I., “English BAKE revisited (BAKE-ing an ECD)”, International Journal of Lexicography, 2, 325-45, 1989.
- [16] Steele J., “Meaning-Text Theory: Linguistics, lexicography and Implications”, Ottawa, Univ. of Ottawa Press, 1990.
- [17] Kahane S., “The Meaning-Text Theory”, Dependency and Valence, An International Handbook of Contemporary Research, Berlín: De Gruyter.
- [18] Flores D., “Como crear un lenguaje”, http://www.pueblacity.com/ego-pdf/sp/lng/como/como_verbos.html
- [19] Dobrenov-Major M., “Reading Comprehension Enhancement in Foreign Language Learners with a Verb Valency based Reading-Strategy”, Applied Linguistics Association of Australia, 1998.
- [20] Munguía I., Murguía M., Rocha G., “LAROUSSE, Gramática, Lengua Española, Reglas y Ejercicios”, ediciones Larousse S.A. de C.V.
- [21] “Gramática y dudas del español (Guía de consulta rápida)”, segunda edición, editorial Everest S. A.
- [22] “Gramática descriptiva de la lengua española”, Editorial Real Academia Española / Espasa Calpe, Madrid, 1999.

-
- [23] Fernando L., “Los complementos verbales, cómo reconocerlos, qué categorías los desempeñan y sus peculiaridades”, *Lingüística por Fernando Liroz*, <http://www.amerschmad.org/spanish/gram/inicio.htm>, 03/06.
- [24] Tarradas Agea D., “Los Pronombres”, fichas útiles, *Amnesia: El juego*, http://amnesia.eljuego.free.fr/Fichas_gramatica/FG_pronombres.htm#23, 03/06
- [25] Fernández J., “Complemento circunstancial”, *Hispanoteca: Lengua y Cultura*, <http://culturitalia.uibk.ac.at/hispanoteca/Foro-preguntas/ARCHIVO-Foro/Complemento%20circunstancial.htm>, 03/06.
- [26] REAL ACADEMIA ESPAÑOLA: Diccionario de la lengua española (en línea), vigésima edición, <http://www.rae.es>, 03/06.
- [27] Arano S., “La ontología: una zona de interacción entre la Lingüística y la Documentación [on line]”, *Hipertext.net*, núm. 2, 2003. <http://www.hipertext.net>, ISSN 1695-5498
- [28] EuroWordNet: Buildin a multilingual database with wordnets for several European languages, <http://www.ilic.uva.nl/EuroWordNet/>
- [29] Atserias J., Climent S., Farreres X., Rigau G., Rodríguez H., Combining Multiple Methods for the Automatic Construction of Multilingual WordNets, In *Proceedings of RANLP'97*, pages 143-149, Bulgaria, 1997.
- [30] “Glossary of Terms used in WordNet System”, Manual Page, <http://wordnet.princeton.edu/man/wngloss.7WN>
- [31] Boguraev B., Briscoe T., Carroll J., Carter D. y Grover C., “The Derivation of a Grammatically Indexed Lexicon from the Longman Dictionary of Contemporary English”, Computer Laboratory, Univ. de Cambridge, 1987.
- [32] Grishman, R., Macleod C. y Meyers A., “Complex Syntax: Building a Computational Lexicon”, Computer Science Department, New York University, 1994.
- [33] Briscoe, T. Carrol J. “Automatic Extraction of Subcategorization from Corpora”, Univ. de Cambridge, Univ. de Sussex, 1997.

-
- [34] Graña J., “Automatic Analysis of Verbal Constructions in Spanish”, proyecto fundado por Xunta de Galicia (XUGA 20402B97), 1999.
- [35] Muñiz E., Rebolledo M., Rojo G., Santalla M. y Sotelo S., “Description and Exploitation of BDS: a Syntactic Database about Verb Government in Spanish”, Proceedings of Recent Advances in Natural Language Processing, Borovets, Bulgaria, septiembre, 2003.
- [36] Yoshinaga N. y Tsujii J., “Generalizing Subcategorization Frames Acquired from Corpora Using Lexicalized Grammars”, Univ. de Tokyo, CREST, JST, 2004.
- [37] Cervantes A. “Diseño e Implementación de un Analizador Sintáctico para las Oraciones en Español Usando el Método de Dependencias”, tesis de maestría, Centro Nacional de Investigación y Desarrollo Tecnológico, 2005.
- [38] Fernández A. et al. “SENSEM: base de datos verbal del español”. Universidad Autónoma de Barcelona. Taller de Herramientas y Recursos Lingüísticos para el español y el portugués, Tonantzintla, México, 2004.
- [39] Taulé M. et al. “Mapping syntactic functions into semantic roles”. Universidad de Barcelona, 2006.
- [40] Korhonen A. “Subcategorization Acquisition”, tesis doctoral, Universidad de Cambridge, 2002.
- [41] Verdejo, M., “Spanish EWN Database”, Universidad Nacional de Educación a Distancia (UNED), Madrid.
- [42] Ciravegna F., Magnini B., Pianta E., Strapparava C., "A Project for the Construction of an Italian Lexical Knowledge Base in the Framework of WordNet", IRST Technical Report # 9406-15, June 1994.
- [43] REAL ACADEMIA ESPAÑOLA, Banco de datos (CREA), Corpus de referencia del español actual, <http://www.rae.es>
- [44] Longest Path - Approach and Algorithms, http://users.csc.tntech.edu/~algviz/longest_path/algorithms.php

Anexo A

Glosario

- **Acepción**

Es cada uno de los significados de una palabra según los contextos en que aparece. Por ejemplo, el verbo *traspapelar* tiene 2 acepciones.

traspapelar.

1. Confundir, perder un papel entre otros, haciendo perder el lugar o colocación que tenía.
2. Perder o colocar en sitio equivocado cualquier otra cosa.

- **Animidad**

La animidad se considera como una personificación, por ejemplo *gobierno* en español es un sustantivo animado y al dirigirse a él se utiliza la preposición *a* (*al gobierno*). Además de personas, la animidad abarca grupos de personas, animales, países, entidades abstractas (organizaciones, partidos políticos), etc.

- **Divalente**

Se dice de un elemento o de un compuesto químico: Que funciona con dos valencias.

- **Hiperónimo**

Es aquel término general que puede ser utilizado para referirse a la realidad nombrada por un término más particular, es decir, es una palabra cuyo significado incluye al de otra u otras.

Por ejemplo: pájaro es un hiperónimo de jilguero; y también, es hiperónimo de gorrión.

- **Hipónimo**

Es aquella palabra que posee todos los rasgos semánticos de otra más general, pero que añade en su definición otros rasgos semánticos que la diferencian de la segunda, es decir, es una palabra cuyo significado está incluido en el de otra.

Por ejemplo: gorrión es hipónimo de pájaro.

- **Lexicón**

Lexicón es un diccionario o libro en el que se recogen y explican de forma ordenada voces de una o más lenguas, de una ciencia o de una materia determinada.

- **Sintagma**

El sintagma es un tipo de constituyente sintáctico formado por un grupo de palabras que forman otros sub-constituyentes, al menos uno de los cuales es un núcleo sintáctico. Las propiedades combinatorias de un sintagma se derivan de las propiedades de su núcleo sintáctico. Existen los siguientes tipos de sintagmas:

- Sintagma nominal

Donde el núcleo es un sustantivo o pronombre.

Por ejemplo: "aquellos chicos de azul".

En español, el sintagma nominal puede funcionar como sujeto agente o paciente, como atributo, como complemento directo de cosa, como complemento predicativo, como aposición, como vocativo y como complemento circunstancial. Si su núcleo es un pronombre, puede funcionar además como complemento indirecto y complemento directo de persona.

- Sintagma preposicional

Donde hay una preposición más un sintagma como complemento.

Por ejemplo: "entre los sauces llorones".

En español, el sintagma preposicional puede funcionar como sujeto excepcionalmente (Entre Pedro y Luis lo hicieron), como complemento directo de persona, como complemento indirecto, como complemento de régimen o suplemento, como complemento agente, como complemento predicativo, como complemento circunstancial, como atributo y como complemento preposicional de un nombre, de un adjetivo, de un adverbio o de una interjección.

- Sintagma verbal

El verbo es el núcleo, por ej: "come pescado".

En español el Sintagma verbal puede funcionar solamente como predicado, bien sea como predicado verbal, como predicado nominal o como predicado mixto.

- Sintagma adjetival

El adjetivo es el núcleo, por ej: "muy guapo".

En español el sintagma adjetival puede funcionar como adyacente, como atributo y como complemento predicativo.

- Sintagma adverbial

El adverbio es el núcleo, por ejemplo: "muy cuidadosamente".

En español el sintagma adverbial puede funcionar como complemento circunstancial, como complemento de otro adverbio, como complemento de un adjetivo y como grado o cuantificador de un adjetivo.

Anexo B

Muestra de Patrones de Manejo de Verbos

(1)

00002403

respirar₁

(Medicine) X draws air into, and expel out of, the lungs.

X = I
1. SN

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{I,1}	{ <i>El bebe</i> } <i>respira!</i>



(2)

00007021

actuar₁

(Psychology) X behaves in a certain manner Y; X show a certain behavior Y; X conduct or comport oneself on Y way.

X = I	Y = II
1. SN	1. Adv

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.C_{II,1}: Adv denota un adverbio de modo.

Combinaciones posibles	Ejemplos
C _{I,1} + C _{II,1}	{ <i>Jorge</i> } <i>actuó</i> { <i>mal</i> } [<i>con su novia</i>]"



(3)

00007021

comportarse₁

(Psychology) X behaves in a certain manner Y; X show a certain behavior Y; X conduct or comport oneself on Y way.

X = I	Y = II
1. SN	1. Adv

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.C_{II,1}: Adv denota un adverbio de modo.

Combinaciones posibles	Ejemplos
C _{I,1} + C _{II,1}	{ <i>Jorge</i> } <i>se comportó</i> { <i>mal</i> } [<i>con su novia</i>]

(4)

00009090

descansar₁

(Tourism) X is at rest.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{ <i>Mi hijo</i> } <i>descansó</i> [<i>toda la noche</i>]



(5)

00009090

reposar₁

(Tourism) X is at rest.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{ <i>Mi hijo</i> } <i>reposó</i> [<i>toda la noche</i>]



(6)

00009222

moverse₁

(Factotum) X is in a state of action.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{ <i>El venado</i> } <i>se mueve</i> [<i>muy rápido</i>]

(7)

00011259

dormirse₁

(Factotum) X changes from a waking to a sleeping state.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{Luis} se durmió [temprano]



(8)

00011259

adormecerse₁

(Factotum) X changes from a waking to a sleeping state.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{Luis} se adormeció [temprano]



(9)

00011671

acostarse₁

(Factotum) X goes to bed in order to sleep.

X = I
1. SN

C_{L,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{L,1}	{Los niños} se acostaron [en mi recámara]

(10)

00014589

sedar₁

(Medicine, Psychology) X causes Y to be calm or quiet.

X = I	Y = II
1. SN	1. a SN 2. Pro

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.2}: Pro denota un pronombre clítico acusativo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{La enfermera} sedó {al paciente} [porque estaba inquieto]
C _{I.1} + C _{II.2}	[El doctor ordeno que] {ella} {lo} sedara.



(11)

00019853

derrumbarse₁

(Factotum) X collapses due to fatigue, an illness, or a sudden attack.

X = I	Y = II
1. SN	1. por SN 2. debido a SN

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo inanimado.C_{II.2}: SN denota un sintagma nominal con un sustantivo inanimado.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{Laura} se derrumbó {por el cansancio.}
C _{I.1} + C _{II.2}	{El ladrón} se derrumbó {debido a el golpe recibido.}



(12)

00021823

parecer₁

(Factotum) X has a certain outward Y or facial expression Y.

X = I	Y = II
1. SN	1. Adj

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: Adj denota un adjetivo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{Ismael} parece {triste.}

(13)

00033668

vestir₅

(Factotum) X is dressed in Y.

X = I	Y = II
1. SN	1. de Adj

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: Adj denota un adjetivo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{Manuel} vestirá {de azul} [en la ceremonia]



(14)

00033668

llevar₃

(Factotum) X is dressed in Y.

X = I	Y = II
1. SN	1. SN

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo inanimado.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{María} lleva {una sudadera nueva.}



(15)

00048767

cuidar₂

(Medicine) X provides treatment for Y.

X = I	Y = II
1. SN	1. a SN 2. Pro

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.2}: Pro denota pronombre clítico acusativo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{Ese doctor} cuida [muy bien] {a Don Luis.}
C _{I.1} + C _{II.2}	{La enfermera Lilia} {los} cuidaba [con cariño].

(16)

00048767

tratar₁

(Medicine) X provides treatment for Y.

X = I	Y = II
1. SN	1. a SN 2. Pro

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.C_{II,1}: SN denota un sintagma nominal con un sustantivo animado.C_{II,2}: Pro denota pronombre clítico acusativo.

Combinaciones posibles	Ejemplos
C _{I,1} + C _{II,1}	{ <i>Ese doctor</i> } trata [<i>muy bien</i>] { <i>a Don Luis.</i> }
C _{I,1} + C _{II,2}	{ <i>La enfermera Lilia</i> } { <i>los</i> } trataba [<i>con cariño</i>].



(17)

00059503

ejercitarse₁

(Sport) X do physical exercise.

X = I
1. SN

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.

Combinaciones posibles	Ejemplos
C _{I,1}	{ <i>El maestro</i> } se ejercita [<i>todas las mañanas.</i>]



(18)

00059503

ejercitar₁

(Sport) X do physical exercise with some part of the body Y.

X = I	Y = II
1. SN	1. SN

C_{I,1}: SN denota un sintagma nominal con un sustantivo animado.C_{II,1}: SN denota un sintagma nominal con un sustantivo inanimado.

Combinaciones posibles	Ejemplos
C _{I,1} + C _{II,1}	{ <i>El maestro</i> } ejercita { <i>sus piernas</i> } [<i>todas las mañanas.</i>]

(19)

00059503

trabajar₁

(Sport) X do physical exercise with some part of the body Y.

X = I	Y = II
1. SN	1. SN

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo inanimado.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{ <i>El maestro</i> } trabaja { <i>sus piernas</i> } [<i>en el gimnasio.</i>]



(20)

00062229

expeler₁

(Biology) X eliminate, as of bodily substances Y.

X = I	Y = II
1. SN	1. SN
	2. Pro

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo inanimado.C_{II.2}: Pro denota pronombre clítico acusativo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{ <i>Su cuerpo</i> } expelió { <i>las mucosidades</i> } [<i>por la boca.</i>]
C _{I.1} + C _{II.2}	{ <i>El niño enfermo</i> } { <i>lo</i> } expelió [<i>a tiempo.</i>]



(21)

00062229

expulsar₁

(Biology) X eliminate, as of bodily substances Y.

X = I	Y = II
1. SN	1. SN
	2. Pro

C_{I.1}: SN denota un sintagma nominal con un sustantivo animado.C_{II.1}: SN denota un sintagma nominal con un sustantivo inanimado.C_{II.2}: Pro denota pronombre clítico acusativo.

Combinaciones posibles	Ejemplos
C _{I.1} + C _{II.1}	{ <i>Su cuerpo</i> } expulsó { <i>las mucosidades</i> } [<i>por la boca.</i>]
C _{I.1} + C _{II.2}	{ <i>El niño enfermo</i> } { <i>lo</i> } expulsó [<i>a tiempo.</i>]

