

Universitat de Barcelona
Departament de Filologia Romànica
Secció de Lingüística



Tesis Doctoral

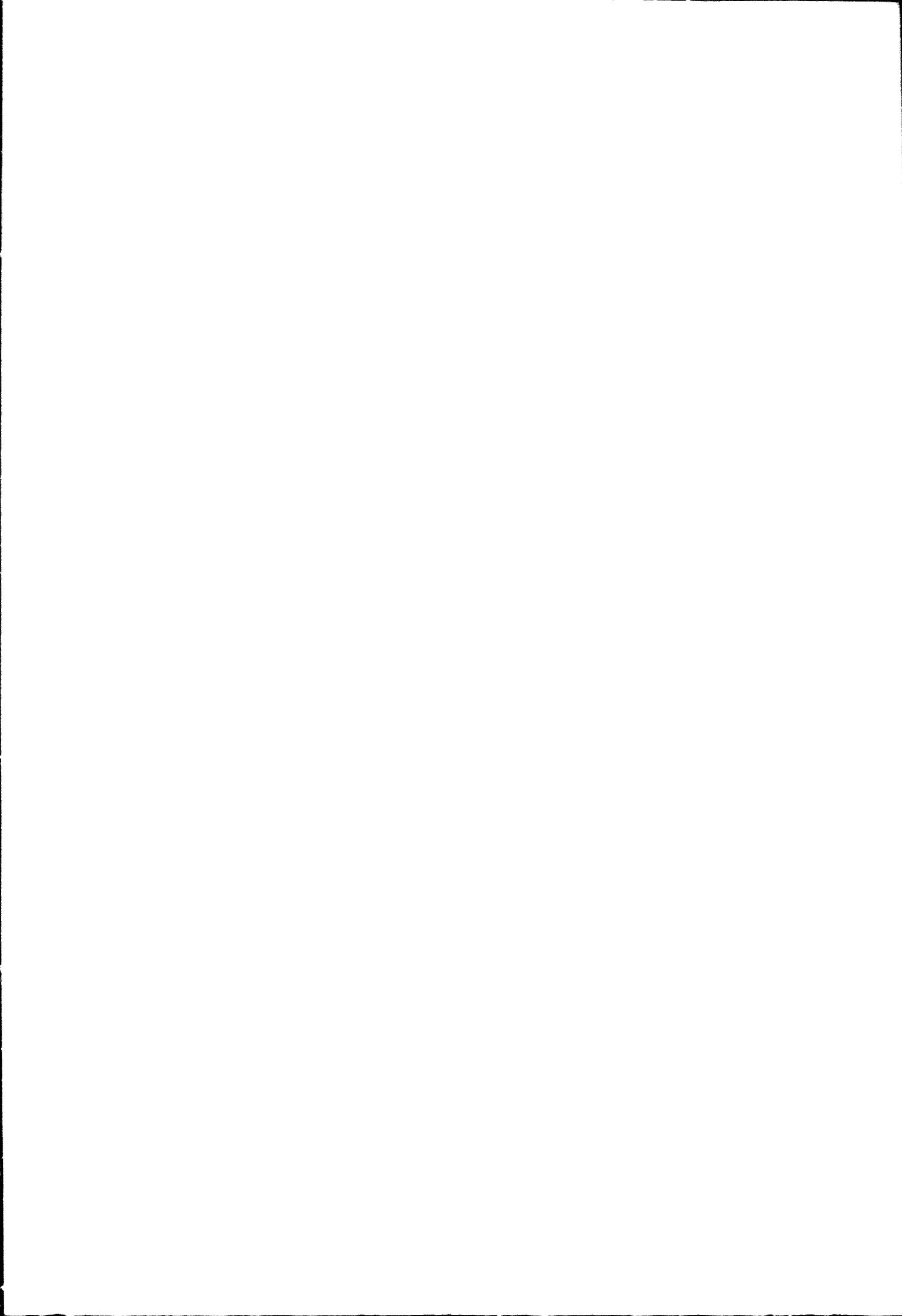
Representación de las entradas verbales en una
Base de Conocimiento Léxico:
Diátesis y Semántica Léxica

Octubre, 1995

Mariona Taulé Delor

Directora: Maria Antònia Martí Antonín

La elaboración de esta tesis ha sido posible gracias a la beca
de Formación de Personal Investigador concedida por el
Ministerio de Educación y Ciencia.



Agraïments

Vull expressar una gratitud especial a M^{re} Antònia Martí, directora d'aquesta tesi i la meva mestra durant tants anys, per l'atenció amb que ha seguit aquest treball i la justesa de les seves indicacions.

A la Irene, "amiga i companya", per l'estreta col·laboració i per tot el que hem après juntes.

A l'Horaci, per tot el que m'ha ensenyat i pel seu constant suport i amista.

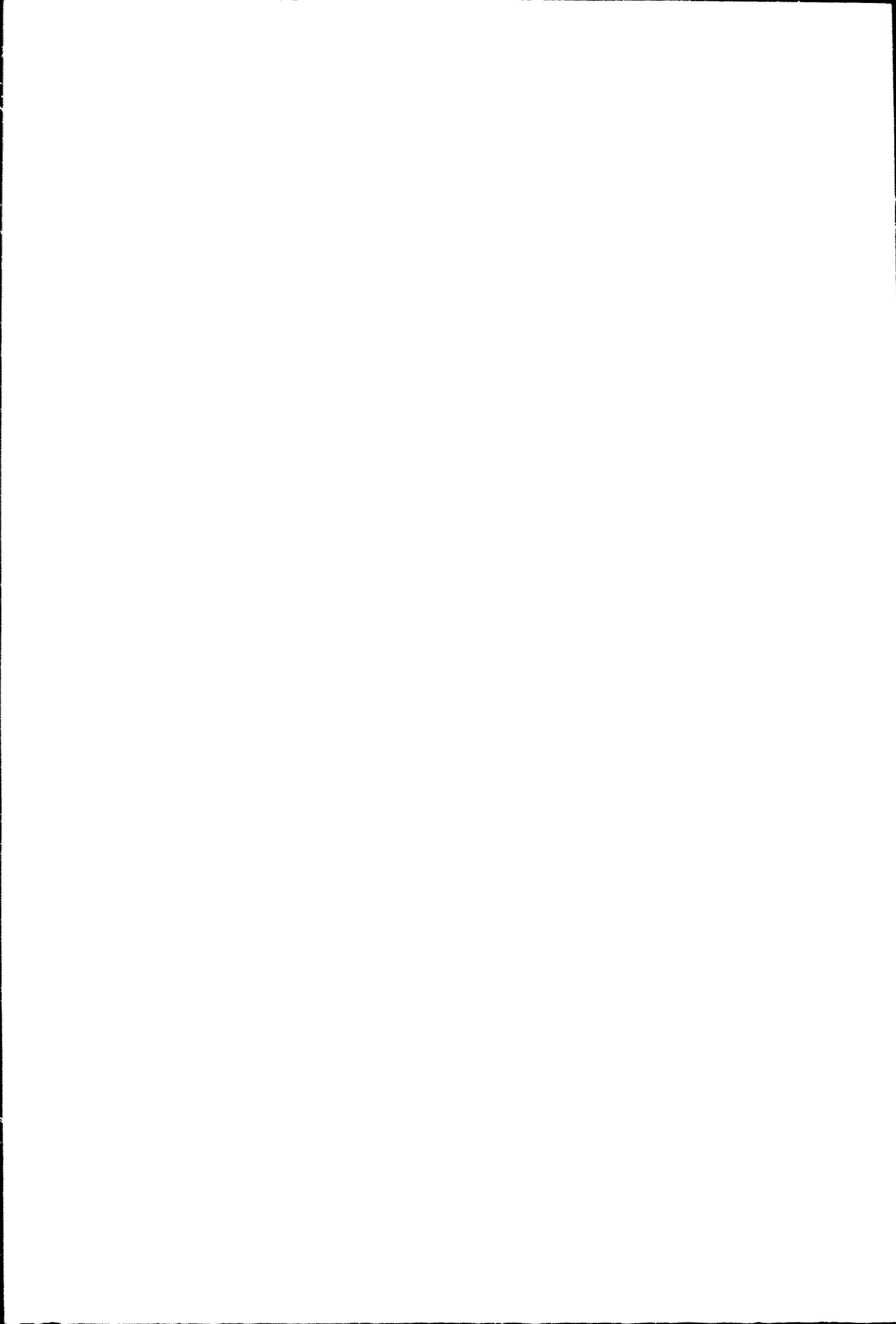
A la Felisa, per la confiança que sempre m'ha demostrat i per haver-me permès de col·laborar i formar part en el projecte Acquilex.

Als incondicionals German, Kiku i Alicia, sempre al peu del canó.

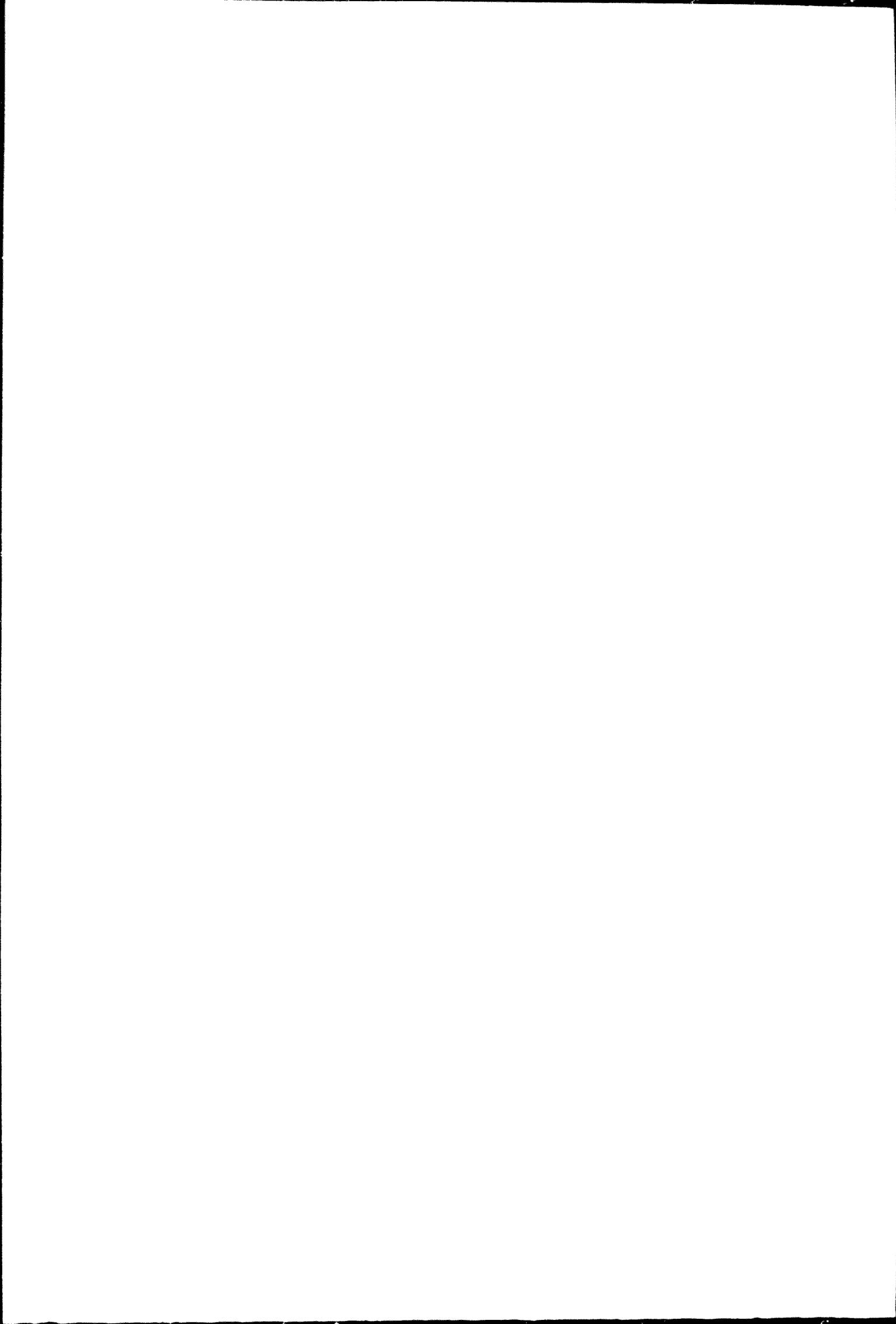
Gràcies, també, a tots els amics del despatx 566 i del laboratori de Lingüística Computacional.

L'agraïment més sincer a Sebastià Serrano, cap del departament de Filologia Romànica, per la seva disponibilitat a facilitar-me tota mena d'ajut, així com als companys de la secció de Lingüística General.

Vull acabar expressant el meu agraïment a la família, amics i mestres, que d'una manera o altra han contribuït a la meua formació en el pla científic i humà.



Ai Manel, pel seu afecte i paciència...



Indice

Capítulo 1: Las entradas verbales en el componente léxico

1. Introducción	1
1.1. Organización del trabajo	5
2. La representación del verbo en distintos modelos teóricos	7
2.1. La Teoría de la Rección y el Ligamiento (GB)	8
2.2. La Gramática Léxico-Funcional (LFG)	14
2.3. La Gramática de Estructura Sintagmática Generalizada (GPSG)	19
2.4. La Gramática de Estructura Sintagmática Regida por el Núcleo (HPSG)	25
2.5. Conclusiones	32

Capítulo 2: Adquisición y representación del Conocimiento Léxico

1. Introducción	37
2. El proyecto ACQUILEX	39
2.1. Del Diccionario (MRD) a la Base de Datos Léxica	40
2.2. Extracción de la información léxica contenida en la Base de Datos: Construcción de taxonomías	45
2.3. Proceso de conversión	49
3. Proceso de extracción de información léxica: Análisis de las definiciones verbiales	53
3.1. Definición del corpus	53
3.2. Analizadores	57
3.3. Análisis de las definiciones	61
3.3.1. La definición	62
3.3.1.1. "Flex"	62
3.3.1.2. Término Genérico	63
3.3.1.3. "Differentia"	64

<i>Indice</i>	ii
3.4. Gramáticas	65
3.4.1. Reglas de Análisis	67
3.4.2. Reglas de Interpretación	71
3.5. Resultados	79
4. La Base de Conocimiento Léxico	82
4.1. Lenguaje de Representación Léxica	82
4.1.1. Sistema de tipos	84
4.1.2. Unificación	86
4.1.3. Herencia	87

Capítulo 3: El signo léxico: el verbo en la Base de Conocimiento

1. Introducción	89
2. La información morfológico-sintáctica	92
3. La información semántica	98
4. Clasificación del signo verbal	104

Capítulo 4: La Estructura de Qualia Verbal (VRQS)

1. Introducción	109
2. Niveles de representación semántica	112
2.1. La saturación léxica en las entradas verbales	118
3. 'VRQS': Semántica Compositiva	121
3.1. 'VRQS' y descomposición	121
3.2. 'VRQS' y saturación léxica	125
3.2.1. Saturación Argumental	126
3.2.2. Saturación Morfológica	128
3.2.3. Saturación Sobreentendida	129
3.3. Saturación Léxica, Estructura Argumental y Diátesis	130

4. Descripción de las clases semánticas	135
4.1. Verbos de Cocción	135
4.1.1. Verbos básicos de cocción	136
4.1.2. Verbos que indican la Manera de cocción	137
4.1.3. Verbos que indican el Medio de cocción	137
4.1.4. Verbos que especifican un ingrediente	138
4.2. Verbos de Ingestión	139
4.2.1. Verbos básicos de ingestión	140
4.2.2. Verbos relacionados con la ingestión de una comida particular	140
4.2.3. Verbos que indican la Manera de Ingestión	141
4.2.4. Verbos que destacan un alimento concreto	142
4.3. Clase semántica de "Emborrachar"	143
4.4. Verbos de Movimiento	144
4.4.1. Verbos que indican de forma Inherente la Trayectoria	149
4.4.2. Verbos que indican un Cambio de Posición	152
4.4.3. Verbos que no indican de forma inherente la Trayectoria	153
4.4.4. Verbos que indican la Ruta del movimiento	154
4.4.5. Verbos que indican el vehículo o Medio de locomoción	155
4.4.6. Verbos que expresan la Manera del movimiento	156
4.4.7. Verbos que indican maneras de bailar	157
4.4.8. Verbos que expresan el movimiento corporal	158
4.5. Verbos de Meteorología	159
5. Representación de la 'VRQS' en la Base de Conocimiento Léxico	162
5.1. El tipo <vrqs>	162
5.2. Los componentes semánticos y el tipo de saturación léxica	164
5.2.1. El Agente	166
5.2.2. El Paciente	167
5.2.3. La Trayectoria	168
5.2.4. La Manera	169
5.2.5. El Medio	170
5.2.6. El Tiempo	170
5.3. Las clases semánticas en la BCL	171
5.4. Unos ejemplos de entradas léxicas	178

Capítulo 5: La Diátesis

1. Introducción	183
2. Diátesis: Definición e Hipótesis	186
2.1. Hipótesis de trabajo	188
2.2. Cambios morfológicos, sintácticos y semánticos	189
2.1.1. Diátesis y forma verbal	190
2.1.2. Diátesis y Subcategorización	191
2.1.3. Diátesis y Estructura Argumental	193
3. Clasificación de las alternancias de diátesis	197
3.1. Alternancias transitivas	197
3.1.1. Alternancias Causativas	198
3.1.1.1. Alternancia de Pasiva Refleja	199
3.1.1.2. Alternancia Causativa-Incoativa	200
3.1.2. Alternancias de ausencia de Objeto	202
3.1.2.1. Alternancia de Objeto Indefinido	202
3.1.3. Alternancia de Pasiva	203
3.1.4. Alternancias Reflexivas	203
3.1.4.1. Alternancia Reflexiva de Parte-del-cuerpo	204
3.1.4.2. Alternancia Reflexiva de Origen	205
3.2. Alternancias Intransitivas	205
3.2.1. Alternancia de Objeto Cognado	206
3.2.2. Alternancia de Complemento de Medida	207
3.2.3. Alternancia del componente Trayectoria	208
3.2.3.1. Alternancia del componente Destino	209
3.2.3.2. Alternancia del componente Origen	210
3.2.3.3. Alternancia del componente Dirección	210
3.2.3.4. Alternancia del componente Ruta	210
3.3. Alternancias Oblicuas	211
3.3.1. Alternancia de omisión del Oblicuo	212
3.3.2. Alternancia de 'Caída' de preposición	213
3.4. Alternancias Pronominales	214
4. Representación de la Diátesis en la Base de Conocimiento Léxico	217
4.1. Las reglas léxicas	219
4.1.1. Un ejemplo: la regla léxica de Pasiva Refleja	222

<i>Indice</i>	v
4.2. Las reglas de Diátesis	227
4.2.1. Reglas Transitivas	229
4.2.2. Reglas Intransitivas	231
4.2.3. Reglas Oblícuas	233
4.2.4. Reglas Pronominales	234

Capítulo 6: Diátesis y Semántica Léxica (VRQS)

1. Introducción	237
2. Diátesis y componentes de significado	240
3. Diátesis y saturación léxica	245
3.1. Diátesis y Saturación Argumental	246
3.2. Diátesis y Saturación Morfológica	247
3.3. Diátesis y Saturación Sobreentendida	251
4. Diátesis y clase semántica	254
4.1. Verbos de Cocción	256
4.2. Verbos de Ingestión	258
4.2.1. Clase semántica de "Emborrachar"	261
4.3. Verbos de Movimiento	263

<u>Conclusiones</u>	271
---------------------	-----

<u>Bibliografía</u>	275
---------------------	-----

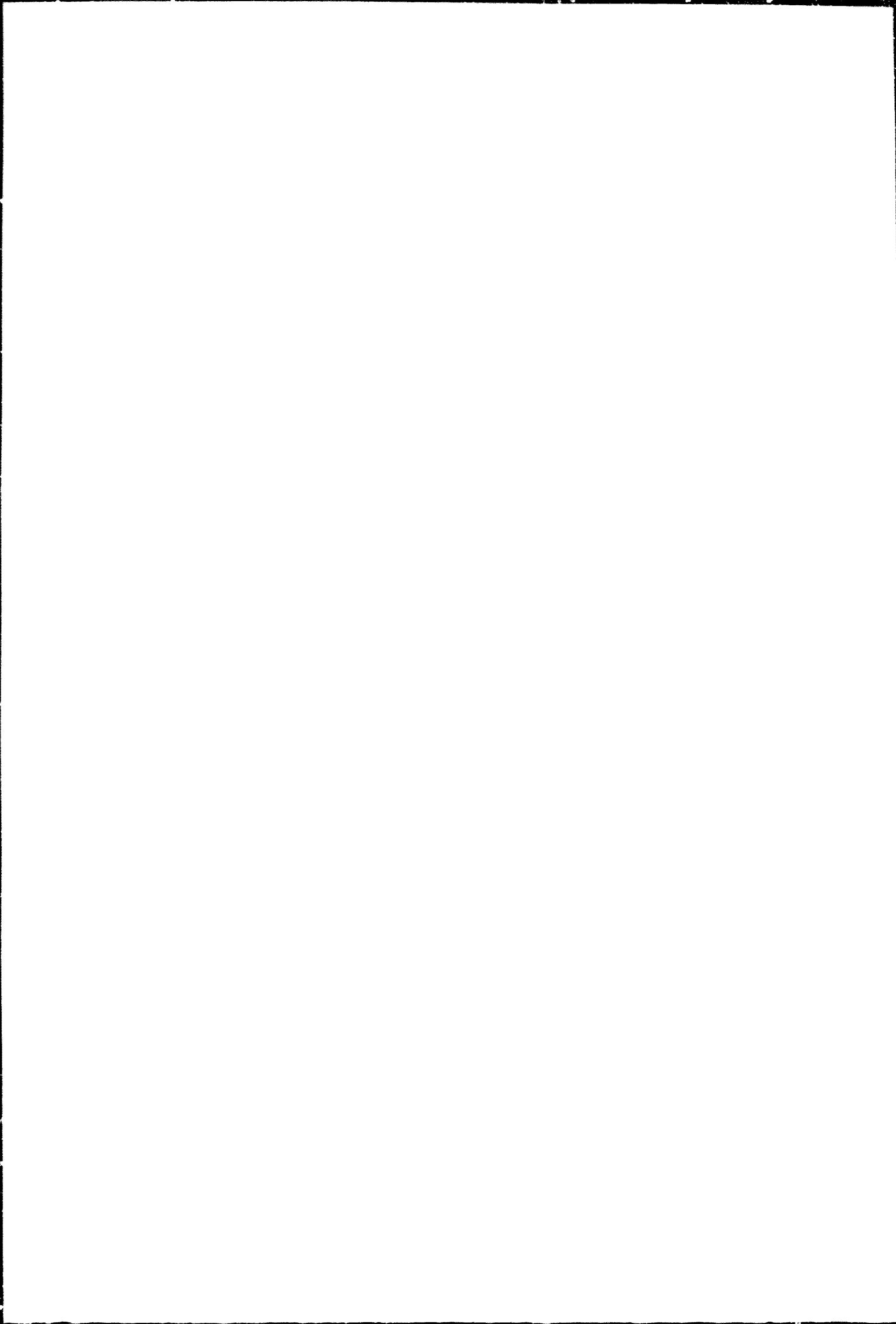
Apéndices

1. Gramáticas para el análisis de las definiciones verbales	1
1.1. Reglas de Análisis	1
1.2. Reglas de Interpretación	4
1.2.1. Reglas de Interpretación: verbos de cocción e ingestión	4

<i>Indice</i>	vi
1.2.2. Reglas de Interpretación: verbos de movimiento	7
2. Resultados del análisis: Taxonomías analizadas	11
2.1. Cocción	11
2.2. Ingestión	18
2.3. Movimiento	25
3. Sistema de Tipos	57
3.1. Tipos generales	57
3.2. Tipos sintácticos	58
3.2.1. Tipos sintácticos verbales	61
3.3. Tipos semánticos	63
3.3.1. Tipos semánticos verbales	68
3.4. El signo	70
3.4.1. Signo nominal	70
3.4.2. Signo verbal	70
3.4.3. Otros signos	74
3.5. Semántica léxica: Estructura de Qualia	76
3.5.1. NOMRQS	76
3.5.2. VRQS	81
4. Reglas Léxicas	85
4.1. Tipos generales	85
4.2. Reglas de diátesis transitivas	85
4.3. Reglas de diátesis intransitivas	90
4.4. Reglas de diátesis oblicuas	91
4.5. Reglas de diátesis pronominales	93
5. Entradas Léxicas verbales	95
5.1. Cocción	95
5.2. Ingestión	101
5.3. Movimiento	107

Lista de abreviaturas

BCL	Base de Conocimiento Léxico
BDL	Base de Datos Léxica
CG	Gramática Categorial
EA	Estructura Argumental
EC	Estructura Conceptual
ELC	Estructura Léxica Conceptual
EP	Estructura Profunda
ES	Estructura Superficial
FLEX	Flexión
FL	Forma Lógica
GB	Teoría de la Rección y el Ligamiento
GPSG	Gramática de Estructura Sintagmática Generalizada
GRZ	Il Nuovo Dizionario Italiano Garzanti
HPSG	Gramática de Estructura Sintagmática Regida por el Núcleo
LDOCE	Longman Dictionary of Contemporary English
LFG	Gramática Léxico-Funcional
LRL	Lenguaje de Representación Léxico
MRD	"Machine-Readable Dictionary" (Diccionario en cinta magnética)
Reglas DI	Reglas de Dominio Intermediato
Reglas PL	Reglas de Precedencia Lineal
UCG	Gramática Categorial de Unificación
VANDALE	Van Dale Groot Woordenboek Hedendaags Nederlands
VOX	Diccionario General Ilustrado de la Lengua Española VOX
VRQS	Estructura de Qualia Verbal



Capítulo 1: Las entradas verbales en el componente léxico

1. Introducción

El creciente interés en el contenido y en la organización de las entradas léxicas desde una perspectiva teórica, unido a la necesidad desde un punto de vista computacional (o del Procesamiento del Lenguaje Natural) de disponer de léxicos de tamaño 'real', ha situado el componente léxico en el centro de atención de muchas investigaciones en lingüística.

En las teorías lingüísticas actuales existe una tendencia generalizada a asignar un papel central dentro de la teoría al componente léxico, ya que muchas de las generalizaciones lingüísticas que antes se expresaban en el componente sintáctico, en la actualidad se consideran que tienen una dimensión léxica. El componente léxico contiene, en general, especificaciones sobre el conocimiento fonológico, morfológico, sintáctico y semántico de las unidades léxicas, convirtiéndose de este modo en un componente básico a partir del cual se proyecta la mayoría de información lingüística. En este sentido, el léxico ya no se concibe como una mera lista de entradas con información asociada, sino como un componente complejo y estructurado.

Este nuevo enfoque permite la reducción de las reglas sintagmáticas y, consecuentemente, la simplificación de la gramática, pero, por otro lado, la cantidad

y complejidad de información especificada en este nivel puede comportar, también, la aparición de información redundante así como crear, además, problemas de control. Se hace necesario, por lo tanto, no sólo decidir qué tipo de información deben contener las entradas léxicas, sino también cómo estructurar toda esta información con el fin de evitar redundancia y captar las similitudes entre clases de palabras que se comportan de igual forma.

En consecuencia, gran parte del interés en lingüística (computacional) se ha centrado, por un lado, en la definición y estructuración del contenido de las entradas léxicas y, por el otro, en el desarrollo de lenguajes para la representación del conocimiento léxico, que eviten la redundancia de la información y que resuelvan problemas de control y consistencia de los datos especificados a este nivel (bases de conocimiento, representación mediante estructuras de rasgos tipificadas, lenguajes de unificación, mecanismos de herencia, reglas léxicas, etc.).

Desde una perspectiva computacional, la necesidad de disponer de léxicos de tamaño 'real' para aplicaciones de Procesamiento del Lenguaje Natural ha convertido la adquisición automática o semi-automática de conocimiento léxico¹, y su posterior representación, en uno de los intereses fundamentales de la lingüística, y en especial de la lexicografía, computacional.

El presente trabajo se enmarca en esta línea de investigación, el objetivo fundamental del cual es la representación de las entradas léxicas verbales en una Base de Conocimiento Léxico. El proyecto ACQUILEX ha proporcionado el marco adecuado para la realización de esta investigación, facilitando la metodología y herramientas necesarias para la consecución de los objetivos trazados.

Dentro del componente léxico, las entradas verbales adquieren un protagonismo extraordinario por considerar el verbo como el núcleo léxico, ya sea de manera explícita o implícita, de la oración. Es decir, el verbo es el que proporciona o proyecta las propiedades, fundamentalmente, sintácticas y también semánticas de las oraciones en las que aparece. En nuestra metodología, la especificación de la información sintáctica y semántica está estrechamente relacionada y no se puede concebir de manera aislada.

La presente investigación se centra, por lo tanto, en la definición y posterior representación de las entradas verbales en una Base de Conocimiento Léxico y, en

¹ La adquisición de conocimiento se realiza, en general, mediante el uso de fuentes léxicas existentes, como son los diccionarios de uso común y los corpus analizados.

concreto, en la organización y formalización de la información sintáctica y semántica y la manera en que dicha información interactúa.

Se examina, de forma específica, el nivel de Semántica Compositiva o 'VRQS'² (Fustejovsky 1991), en el que se indican los componentes semánticos básicos para la descripción del significado léxico verbal y la manera en que dichos componentes se saturan o realizan léxicamente (de manera argumental, morfológica o sobreentendida). La 'VRQS' es, en definitiva, una propuesta de organización léxica, en la que los verbos se clasifican en distintas clases semánticas en función del tipo de componentes de significado y la saturación léxica que éstos presentan.

El tipo de información incluida en la 'VRQS', y especialmente aquella que hace referencia a la saturación de los componentes semánticos, se encuentra en clara correspondencia con las alternancias de diátesis del verbo. De esta manera, se observa que ciertos componentes semánticos se relacionan estrechamente con alternancias de diátesis concretas y que cada tipo de saturación léxica tiene consecuencias distintas a nivel sintáctico y de la Estructura Argumental.

La 'VRQS' permite explicar, por lo tanto, cómo un mismo predicado verbal puede expresar de diferentes maneras sus argumentos, es decir puede aparecer en Estructuras Argumentales alternativas.

En este sentido, se argumenta que cada clase o subclase semántica verbal participa en un conjunto determinado de alternancias de diátesis y, en consecuencia, dichas alternancias pueden ser útiles, también, como criterio metodológico para clasificar los verbos semánticamente (Levin 1991).

Se propone una clasificación de las alternancias de diátesis basada en los distintos cambios producidos, es decir según las modificaciones que puedan efectuarse en la forma del verbo, en la realización de sus argumentos o bien en la interpretación semántica de éstos. Básicamente, se establecen cuatro grandes clases de alternancias: transitivas, intransitivas, oblicuas y pronominales.

Dichas alternancias de diátesis se representan en la Base de Conocimiento mediante reglas léxicas.

El resultado final es una propuesta de clasificación verbal en términos de los componentes semánticos más destacados, del tipo de saturación léxica que éstos

² Acrónimo del inglés "Verbal Relativized Qualia Structure".

presentan y de las alternancias de diátesis que exhiben.

El trabajo se ha centrado, básicamente, en el estudio de tres clases semánticas concretas: verbos de cocción, ingestión y movimiento. La selección de estas clases verbales se ha realizado teniendo en cuenta que se trata de una muestra suficientemente representativa y amplia para poder ilustrar tanto la descomposición del significado en componentes semánticos, como el tratamiento de las alternancias de diátesis.

La presente investigación trata de ser una contribución al desarrollo de léxicos computacionales para sistemas de Procesamiento del Lenguaje Natural, pero también contiene un componente fundamental de investigación lingüística básica al tratar ampliamente y en profundidad cuestiones tan críticas como la interrelación entre la sintaxis y la semántica, el papel de cada una de ellas en el estudio del lenguaje y una propuesta clara de esta última como el componente determinante de la gramática de una lengua.

1.1. Organización del trabajo

La presente investigación se estructura en seis capítulos, de los cuales los dos primeros sirven para situar desde una perspectiva teórica y metodológica, respectivamente, los objetivos de esta tesis, y los restantes desarrollan estrictamente la propuesta de entrada léxica y representación de la información sintáctica y semántica contenida en ella.

En el primer capítulo, se expone de forma general el papel del componente léxico y la representación de la información contenida en él en distintos modelos teóricos: *Teoría de la Rección y el Ligamiento* (Véase apartado 2.1), *Gramática Léxico-Funcional* (Véase apartado 2.2), *Gramática de Estructura Sintagmática Generalizada* (Véase apartado 2.3) y *Gramática de Estructura Sintagmática Regida por el Núcleo* (Véase apartado 2.4). La exposición se centra en el contenido de las entradas verbales y, en concreto, en la información sintáctica y semántica. La revisión de estas distintas propuestas de entrada léxica verbal sirve de punto de partida teórico a partir del cual se propone la entrada verbal de esta investigación.

El segundo capítulo, organizado en tres partes básicas, se presenta la metodología adoptada. En primer lugar, se describe el proyecto ACQUILEX (Véase apartado 2), destacando las tres etapas principales en las que se realiza la adquisición y representación de la información extraída. A continuación, se describe detalladamente el proceso de extracción de la información léxica (Véase apartado 3), haciendo especial hincapié en el análisis de las definiciones verbales del diccionario fuente utilizado. Finalmente, se describen las características básicas de la Base de Conocimiento Léxico en la que se representa la información léxica (Véase apartado 4).

En el capítulo 3, se define la entrada léxica verbal propuesta en este trabajo, destacando, en especial, la información morfo-sintáctica (Véase apartado 2) y semántica (Véase apartado 3) contenida en ella.

El cuarto capítulo empieza con una revisión de diversas propuestas de representación de la semántica léxica (Véase apartado 2), en las que se perfila la tendencia a proponer distintos niveles de descripción semántica para poder dar cuenta del significado y comportamiento de los verbos (Estructura Argumental, Estructura Eventual, Restricciones de Selección y Semántica Compositiva).

Seguidamente, se define de forma exhaustiva el nivel de representación de Semántica Compositiva o 'VRQS' (Véase apartado 3) y se describen, detalladamente, las clases semánticas analizadas a lo largo de esta investigación (Véase apartado 4). El capítulo termina con la representación de la información contenida en la 'VRQS' en la Base de Conocimiento (Véase apartado 5).

El quinto capítulo se organiza en tres apartados. En los dos primeros, se revisan y definen los distintos cambios (morfológicos, sintácticos y semánticos), que producen las alternancias de diátesis (Véase apartado 2), y se establece una primera clasificación de éstas (Véase apartado 3). En el último apartado, se detallan las reglas léxicas desarrolladas en la Base de Conocimiento, encargadas de generar las distintas alternancias de diátesis analizadas (Véase apartado 4).

En el capítulo seis, se examina en qué medida interactúan la información contenida en la 'VRQS' y las alternancias de diátesis. Primero, se presenta la relación entre los componentes de significado y la diátesis (Véase apartado 2); en segundo lugar, las consecuencias sintáctico-semánticas de los distintos tipos de saturación léxica (Véase apartado 3) y, por último, se muestra el conjunto de alternancias que admite cada clase semántica (Véase apartado 4).

La tesis termina con la presentación de las conclusiones sintetizadas en el capítulo siete.

Finalmente, se incluye una serie de apéndices, en los que se muestran las distintas gramáticas realizadas para el análisis de la información contenida en las definiciones verbales (Véase Apéndice 1), las taxonomías realizadas (Véase Apéndice 2), el sistema completo de tipos desarrollado (Véase Apéndice 3), las reglas léxicas de diátesis (Véase Apéndice 4) y una muestra de entradas léxicas verbales de la Base de Conocimiento (Véase Apéndice 5).

2. La representación del verbo en distintos modelos teóricos

“El desarrollo de la Gramática Generativa Transformacional desde sus inicios a la actualidad puede concebirse, en cierto modo, como el refinamiento progresivo de la estructura del componente léxico” (Scalise, 1984). Efectivamente, la concepción inicial del componente léxico como una simple lista de piezas léxicas, con información idiosincrática, se ha visto substituida por un nuevo enfoque en el que dicho componente aparece dotado de una estructura interna compleja capaz de regir una amplia variedad de fenómenos. Como veremos a lo largo de este apartado la frase de S. Scalise puede hacerse también extensible a la *Teoría de la Rección y Ligamiento*³ (GB), desarrollada inicialmente por N. Chomsky (1981) que, en cierto modo, es la continuación inmediata de la Gramática Transformacional.

De la misma manera, en los modelos teóricos desarrollados posteriormente, fuera del ámbito estrictamente transformacional, es decir la *Gramática Léxico-Funcional*⁴ (Kaplan & Bresnan 1982), la *Gramática de Estructura Sintagmática Generalizada*⁵ (Gazdar et al. 1985), la *Gramática de Estructura Sintagmática Regida por el Núcleo*⁶ (Pollard & Sag 1987 y 1992), y las distintas versiones de la denominada *Gramática Categorial*⁷ (Uszkoreit 1986), (Zeevat et al. 1987), (Zeevat 1988) y otros, el componente léxico se ha convertido en una parte importante, por no decir central, de estas teorías gramaticales. Dichas teorías, pertenecientes a la familia de los formalismos gramaticales basados en unificación, otorgan una extraordinaria importancia al componente léxico, ya que es donde se especifica toda la información necesaria para el análisis gramatical. De hecho, este enfoque más lexicalista viene motivado, en gran parte, por la formulación, en el marco teórico de la GB, de teorías y principios, como por ejemplo la Teoría de la X', la Teoría Temática y el Principio de Proyección, en los que se postula la idea

³ “Government and Binding Theory”

⁴ “Lexical-Functional Grammar”

⁵ “Generalized Phrase Structure Grammar”

⁶ “Head-Driven Phrase Structure Grammar”

⁷ “Categorial Grammar”

⁸ La Gramática Categorial no se va a presentar en este apartado, en el capítulo 2 se hará una breve referencia a ella de forma específica.

básica que la sintaxis es una proyección del léxico y que las exigencias de dicho componente deben mantenerse en todos los niveles de análisis, es decir a nivel sintáctico, semántico y fonológico.

El objetivo fundamental de este apartado es una presentación general del lugar que ocupa el componente léxico en estas teorías gramaticales 'actuales', así como una revisión del tipo de información contenida en las entradas léxicas verbales de dicho componente y la manera en que ésta se representa. Para ello me he basado en las versiones más generalizadas de estas teorías, sin entrar demasiado en las distintas concepciones o análisis particulares que se puedan haber realizado dentro de una misma teoría. De la misma manera, sólo se hace referencia a aquellos principios o subteorías que tienen una relación directa con el componente léxico y la información especificada en éste.

Esta exposición se centra en la información y organización de las entradas verbales, destacando de forma especial la información de tipo sintáctico y semántico contenida en los verbos. En este sentido, no se definen de forma específica las otras piezas léxicas incluidas en el lexicon, es decir, nombres, adjetivos, preposiciones, adverbios, etc.

2.1. La Teoría de la Rección y el Ligamiento (GB)

En la *Teoría de la Rección y el Ligamiento* (de aquí en adelante GB), introducida por N. Chomsky (1981), la gramática se concibe como un sistema compuesto por un conjunto de principios —Teoría de la X', Teoría Temática, Teoría del Caso, Teoría de la Rección, Teoría del Ligamiento, Teoría del Límite y Teoría del Control— que restringen la capacidad generativa de las reglas que generan las representaciones sintácticas (convertidas, en esta teoría, en un mecanismo de índole muy general) y que controlan la buena formación del resultado.

En GB la gramática se organiza de forma modular, distinguiendo, básicamente, cuatro módulos (o componentes de representación) autónomos pero relacionados entre sí (Véase figura 1.1). a) el módulo léxico; b) el módulo sintáctico, en el que se distingue el nivel de Estructura-Profunda (EP) y el nivel de Estructura-Superficial (ES), ambos relacionados mediante la operación

* Evidentemente, los modelos teóricos que aquí se presentan no son los únicos que pueden ser etiquetados de actuales, pero sí que se trata probablemente de los más representativos.

transformacional de desplazamiento de 'alpha'; c) el módulo fonológico (o Forma Fonológica), encargado de dar una representación fonética de los sonidos de la oración, es el nivel de interfase entre la facultad del lenguaje y las facultades perceptuales y motoras y, por último, d) el módulo semántico (o Fórmula Lógica, FL.)¹⁰, que enlaza el sistema con el significado; es el nivel de interfase entre la facultad del lenguaje y las facultades conceptuales del cerebro.

La teoría GB es el único de los modelos que se van a exponer en este apartado que todavía recurre a las denominadas 'transformaciones', aunque se hayan reducido de forma drástica.

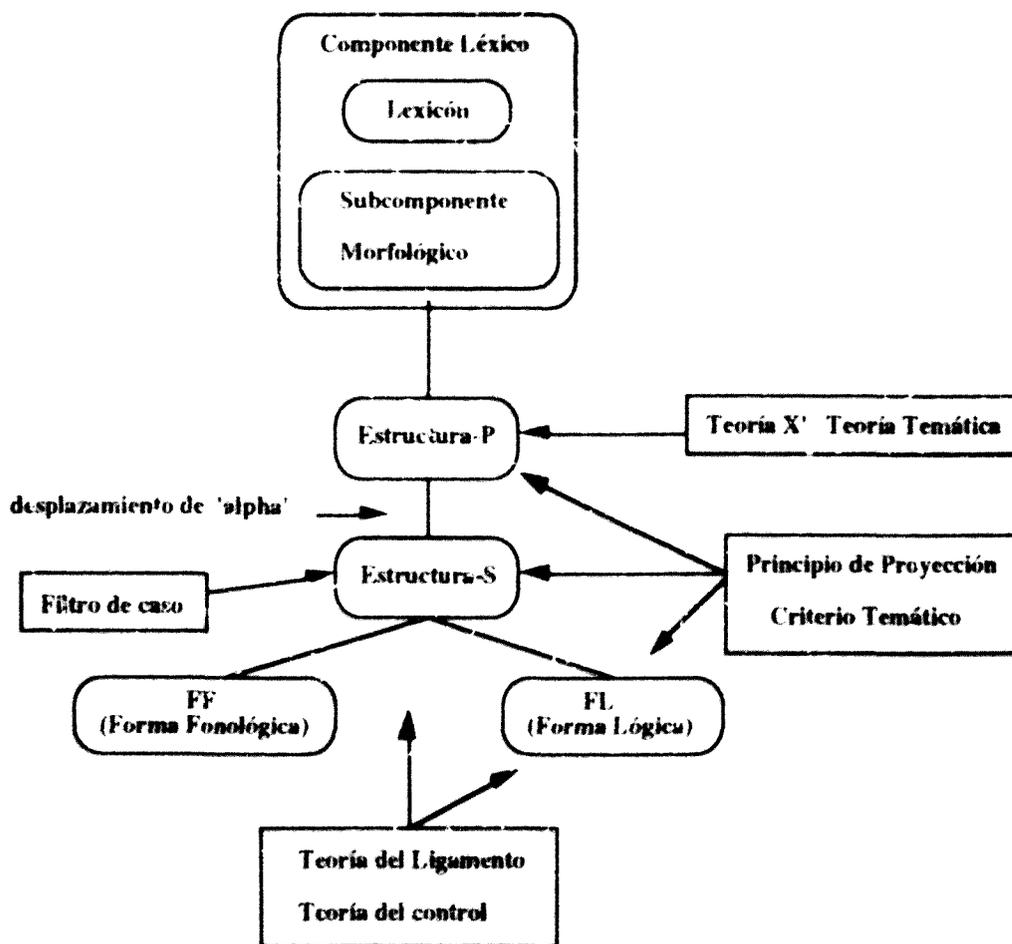


Figura 1.1: Organización de la GB¹¹

¹⁰ La Forma Lógica constituye un primer nivel de representación semántica, pero también se puede definir, de hecho, como el último nivel de la derivación sintáctica, en el que se especifican las propiedades sintácticas necesarias y pertinentes para la interpretación semántica.

¹¹ El esquema de organización de los distintos componentes de la Teoría GB es una adaptación tomada de P. Sells (1989).

En la organización de la gramática propuesta por la GB, se acepta de forma generalizada que el componente léxico está integrado, básicamente, por un lexicón o vocabulario de la lengua, y un subcomponente morfológico.

El lexicón incluye una lista de propiedades idiosincráticas de los ítems léxicos, con sus propiedades fonológicas, morfo-sintácticas y semánticas, que constituyen las unidades atómicas de la sintaxis. En este sentido, en el lexicón se especifica la información asociada a las piezas léxicas que no se puede predecir sobre la base de principios universales o restricciones paramétricas de las lenguas particulares, es decir la información idiosincrática del lenguaje.

En el subcomponente morfológico, se especifican las reglas de formación de palabras (derivación, composición, flexión, etc.), que dan cuenta de la estructura interna de las piezas léxicas, así como de la creación de nuevos ítems léxicos.

La naturaleza de las entradas léxicas y, en concreto, el tipo de fenómenos que las reglas morfológicas deben tratar son puntos de controversia dentro de la teoría. Se pueden destacar, básicamente, dos aproximaciones distintas para el tratamiento de estos fenómenos, que son comúnmente conocidas como la *Hipótesis Lexicalista Fuerte* y la *Hipótesis Lexicalista Débil*.

La *Hipótesis Lexicalista Fuerte* ((Halle 1973) (Jackendoff 1975), (Lieber 1980, 1990), (Williams 1981), (Scalise 1984), etc.) sostiene que la derivación, la composición, así como los procesos de flexión son fenómenos determinados por un conjunto de reglas y principios del componente morfológico y, por lo tanto, deben tratarse exclusivamente en el componente léxico². Estas reglas también pueden expresar alteraciones en la Estructura Argumental de los ítems léxicos, explicar la incorporación nominal, etc.

La *Hipótesis Lexicalista Fuerte* ha tenido una gran influencia en la mayoría de teorías gramaticales actuales, además de su evidente repercusión en el marco de la GB, también se aplica en la *Gramática Léxico-Funcional* (LFG), en la *Gramática de Estructura Sintagmática Generalizada* (GPSG), en la *Gramática de Estructura Sintagmática Regida por el Núcleo* (HPSG) y en algunas versiones de la *Gramática Categorical* (CG).

² Existen, también, posturas distintas sobre si la flexión debe tratarse de forma separada de las reglas de derivación (Scalise 1984) o, por el contrario, no ((Halle, 1973), (Jackendoff 1975), (Lieber 1990), (Williams 1981), etc.).

Por otro lado, los que adoptan la *Hipótesis Lexicalista Débil* ((Aronoff 1976), (Anderson 1982), (Chomsky 1989), (Baker 1988), etc.) postulan que los procesos de flexión se deben tratar en la sintaxis o en la fonología excluyéndolos, por lo tanto, del componente léxico. La flexión es considerada de naturaleza distinta a los procesos de derivación y composición. De hecho, en N. Chomsky (1989) se distingue entre morfología derivativa (donde también incluye la composición) y morfología flexiva.

Por lo que se refiere a la naturaleza de las piezas léxicas incluidas en el lexicon existen, también, diversas posiciones. Mientras que autores como ((Halle, 1973), (Lieber, 1990, 1992) y otros) consideran que las unidades básicas del lexicon son los morfemas, que combinados adecuadamente, mediante una serie de reglas, forman las palabras de la lengua, otros autores propugnan que dichas unidades son las palabras ((Jackendoff, 1975), (Aronoff 1976, 1992), (Anderson, 1982), (Scalise, 1984) y otros)¹³. En general, en el primer caso los autores ponen más énfasis en la función dinámica del léxico es decir en la capacidad del lenguaje de generar nuevas entradas léxicas, y en el segundo caso se enfatiza la descripción y análisis de la estructura interna de las palabras.

Las entradas léxicas, que se caracterizan en el lexicon, se definen por sus propiedades fonológicas, morfo-sintácticas y semánticas, así como por cualquier otra propiedad que no pueda expresarse a través de las reglas generales de los demás componentes de la gramática. A continuación, nos centraremos en la información sintáctica y semántica de las entradas léxicas verbales.

En las entradas verbales se incluye, básicamente, la información sobre la categoría, el marco de subcategorización, la Estructura Argumental, así como las restricciones de selección asociadas a los argumentos (Véase el ejemplo de (1)).

¹³ Existen, también, diferencias entre los autores que proponen, en general, la palabra como unidad básica del lexicon. En R. Jackendoff (1975), el lexicon contiene todas las entradas léxicas existentes de la lengua (palabras simples, compuestos, derivados, afijos, etc.), mientras que en M. Aronoff (1976, 1992) y en S. Anderson (1982b) se considera que la palabra, o la raíz, es la unidad mínima del lexicon y los morfemas (portadores de información relacional) se introducen mediante las reglas de formación morfológica.

(1) *Entrada Léxica:*

Kill , [+V, -N]	(Categoría)
[NP, V']	(Marco de Subcategorización)
(Agent, Patient)	(Estructura Argumental)
[+concret], [+anim]	(Restricciones de Selección)

La especificación categorial, junto con el marco o esquema de subcategorización, determinan en que contextos puede aparecer un verbo, es decir su distribución. Las categorías se definen mediante los rasgos binarios N y V. En consecuencia, los símbolos de las categorías atómicas N, V, A y P son una abreviación del conjunto de rasgos primitivos que las definen¹⁴. La Teoría de rasgos permite diferenciar las categorías mayores (N, V, A y P), distinguiéndolas a su vez entre terminales y no terminales, de la categorías menores (DET, COMP, CONJ, ADV, etc.).

Es importante señalar que en el marco de subcategorización sólo se especifican los complementos verbales, es decir, los constituyentes que son obligatorios dentro del SV. El dominio de la subcategorización se limita a la máxima proyección que contiene la cabecera o núcleo léxico. En virtud de la teoría de la X', el sujeto no entra dentro del dominio del verbo porque la proyección máxima del verbo es SV¹⁵. En consecuencia, el sujeto no se incluye en el marco de subcategorización. En GB se asume como propiedad general de las oraciones tener sujeto y es el Principio de Proyección Extendido el que asegura que esta propiedad se realice.

En la Estructura Argumental se especifica el número de argumentos requeridos por el verbo (es decir, los participantes implicados en la actividad o estado expresado por el predicado) y el tipo de relación semántica que dichos argumentos mantienen con el predicado. La relación semántica entre el verbo y sus argumentos es expresada mediante papeles temáticos (o "theta-grid" (Stowell 1981)) del tipo Agente, Paciente, Tema, Origen, etc. La Estructura Argumental se deriva del significado del verbo y, por lo tanto, también se incluye en ella el papel temático que el verbo asigna composicionalmente al sujeto.

¹⁴ Las categorías se definen como: [+N, -V]= N; [-N, +V]= V; [+N, +V]= A; [-N, -V]= P.

¹⁵ En GB la oración es una proyección de FLEX (es decir, de la flexión).

Siguiendo la propuesta de E. Williams (1981), la Estructura Argumental también incluye la distinción entre argumentos externo e internos, los cuales se distinguen porque el argumento externo (generado fuera del SV, es decir el sujeto) se representa subrayado. En este sentido, la Estructura Argumental del verbo contiene una cantidad de información más amplia que aquella contenida en el marco de subcategorización. De hecho, la distribución de los argumentos verbales podría, también, deducirse de la Estructura Argumental (Pesetsky 1982), (Grimshaw 1990) y, por lo tanto, convirtiéndose en redundante, e innecesaria, la información incluida en el marco de subcategorización. El problema principal con el que se encuentran estas propuestas es que no existe siempre una correspondencia sistemática entre categoría sintáctica y papel temático. En consecuencia, la diferencia fundamental entre estos dos niveles de representación es que en el marco de subcategorización se expresa el tipo de constituyentes requeridos por el verbo, mientras que en la Estructura Argumental se especifica el número de argumentos semánticos.

Por lo que se refiere al estatus de los papeles temáticos, al nivel de especificación de éstos y a la manera de establecer cuál es el papel temático que se utiliza en un contexto dado, no hay un acuerdo global dentro de la teoría (Véase apartado 2 del capítulo 4).

En cuanto a la representación de las alternancias en la Estructura Argumental (o diátesis) de los verbos, se realiza mediante transformaciones, es decir a través del mecanismo de desplazamiento de 'alpha', donde 'alpha' se interpreta como una variable de categoría sintáctica. Las transformaciones convierten las estructuras creadas por las reglas sintagmáticas en nuevas estructuras. En GB las oraciones pasivas, por ejemplo, se derivan por medio de un desplazamiento de 'alpha' (concretamente, un desplazamiento o movimiento del SN), en el que se desplaza el objeto de la estructura profunda para convertirlo en sujeto en la estructura superficial. En GB no se formula una regla de pasiva sino las restricciones sobre el desplazamiento. En consecuencia, la alternancia en los esquemas de subcategorización se tratan, en general, como un fenómeno sintáctico y, por lo tanto, en las entradas verbales sólo se especifica un único marco de subcategorización (por ejemplo, la construcción activa).

La relación básica entre el lexicon y los distintos niveles de representación sintáctica (es decir, en la EP, ES y en la FL) se establece mediante dos principios de la Teoría Temática: el Principio de Proyección Extendido y el Criterio Temático. En este sentido, la Teoría Temática asegura el cumplimiento de las exigencias léxicas. El Principio de Proyección Extendido establece que la sintaxis es una proyección del

léxico y, en consecuencia, que las exigencias léxicas deben ser respetadas en todos los niveles de representación sintáctica. Además, la formulación de este principio, también, garantiza que todas las oraciones tengan sujeto. La Teoría Temática se encarga, también, de regular la asignación de los papeles temáticos mediante la formulación del Criterio Temático, que prohíbe que un mismo papel temático se atribuya a más de un argumento y que un argumento reciba más de un papel temático del mismo predicado. La Teoría de la X' asegura que los elementos léxicos se proyecten de forma adecuada en la estructura sintagmática, de manera que define los sintagmas como proyecciones de los núcleos léxicos (o cabeceras).

2.2. La Gramática Léxico-Funcional (LFG)

La *Gramática Léxico-Funcional* (en adelante LFG) (Bresnan & Kaplan 1982) y (Kaplan & Bresnan 1982) es un formalismo declarativo perteneciente a la familia de las denominadas Gramáticas de Unificación¹⁶, en el que el componente léxico recibe un tratamiento prioritario, ya que proporciona toda la información necesaria para el análisis sintáctico y, como veremos más adelante, muchos de los fenómenos que en la teoría de la GB se tratan mediante transformaciones, es decir en el componente sintáctico de la gramática, en LFG se solucionan a nivel léxico.

Una gramática LFG asigna dos niveles de descripción sintáctica a toda oración de una lengua. El primero se denomina estructura de constituyentes (o estructura-c) y en él se representan las configuraciones sintagmáticas; la estructura-c es un árbol convencional de estructura sintagmática que indica, básicamente, las relaciones de dominio y precedencia de las palabras y sintagmas de una oración. Esta es la representación que determina la forma fonológica de la oración. En el segundo nivel de descripción sintáctica, denominado estructura funcional (o estructura-f), se representan las funciones gramaticales y especifica explícitamente la información interpretable semánticamente (Véase figura 1.2). La estructura-f representa esta información como un conjunto de pares ordenados atributo-valor

Es importante señalar que no existe ningún tipo de relación derivacional entre la estructura-c y la estructura-f, se trata de representaciones formalmente diferentes y con información distinta. La estructura-c se genera mediante una gramática independiente del contexto ligeramente modificada, junto con la

¹⁶ En el apartado 4.1 del capítulo 2, se exponen brevemente las características principales (y comunes) de las Gramáticas de Unificación, es decir la definición de estructura de rasgos, la unificación, etc.

información especificada en las entradas léxicas. Los nodos de la estructura-c sólo pueden derivarse mediante reglas de estructura sintagmática y, por lo tanto, no hay ni operaciones de movimiento ni de eliminación de constituyentes. De la misma manera, ninguna regla de la sintaxis puede manipular la información de las estructuras-f. La eliminación de reglas transformacionales implica, en consecuencia, que las reglas de estructura sintagmática deben generar directamente todas las oraciones de la lengua.

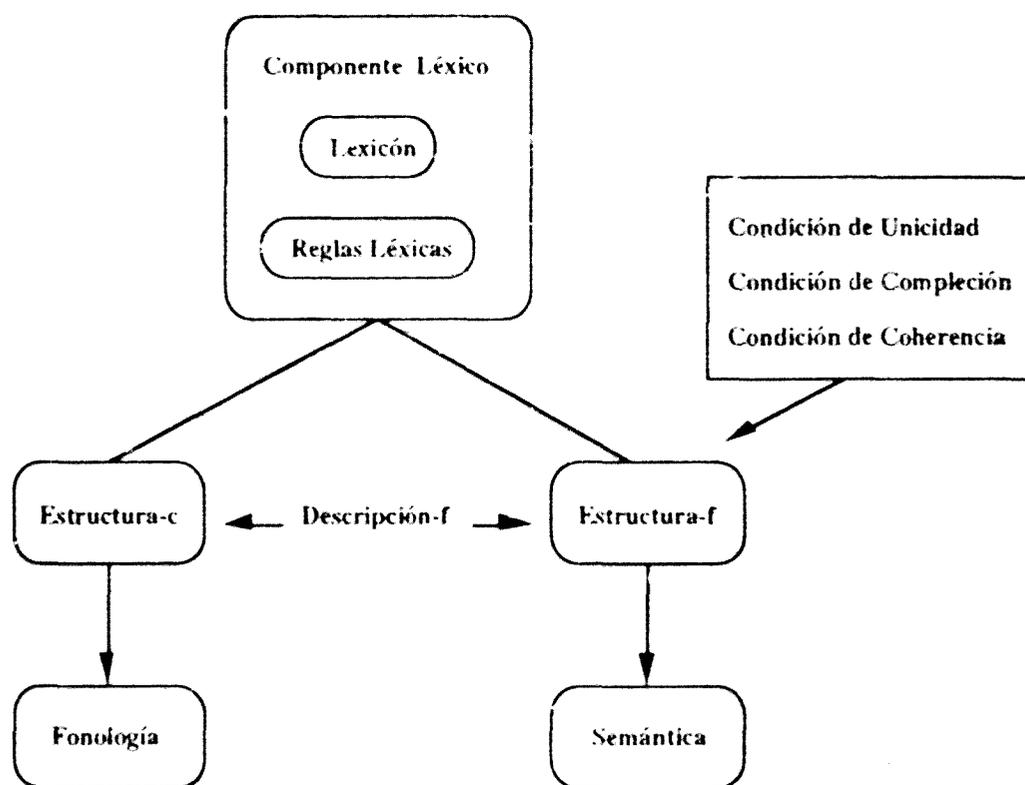


Figura 1.2: Organización de la LFG¹⁷

Las reglas de la gramática incorporan, además, unas anotaciones que especifican la función gramatical del nodo sintáctico, información que, también, se expresa en las entradas léxicas. Es a partir de esta información funcional especificada en la gramática y en el léxico que se producen las denominadas descripciones funcionales (descripciones-f), a partir de las cuales se construyen las estructuras-f. Las descripciones-f, entendidas como un conjunto de ecuaciones que

¹⁷ El esquema de organización de los distintos componentes de la Teoría LFG es una adaptación tomada de P. Sells (1989).

expresan propiedades, sirve de elemento de conexión entre la estructura-c y la estructura-f, estableciendo una relación estrecha y sistemática entre los dos niveles de descripción. Esto no significa, sin embargo, que los nodos de las estructuras-c sean isomórficos a las estructuras-f subsidiarias.

En LFG se asume que no siempre hay una correspondencia elemento a elemento entre las relaciones semánticas de los argumentos del predicado y las configuraciones sintagmáticas y palabras que los expresan. Es ésta la motivación principal de la postulación de los dos niveles de descripción sintáctica, es decir la estructura-c y la estructura-f.

En LFG se postulan, también, una serie de condiciones de buena formación —Condición de Unicidad, de Compleción y de Coherencia— que se aplican a las estructuras-f y aseguran que las cadenas de la estructura-c sean gramaticales. La condición de Unicidad garantiza que toda estructura-f sea una función, es decir que cada atributo tenga un valor. Mientras que la condición de Compleción y de Coherencia aseguran que las funciones gramaticales y los predicados léxicos aparezcan en configuraciones de estructura-f mutuamente compatibles.

Por lo que se refiere al componente léxico, éste incluye el conjunto de las entradas léxicas, es decir el lexicón, y una serie de reglas léxicas que captan las relaciones redundantes entre dos piezas léxicas. Las unidades que integran el lexicón son formas básicas y afijos, que combinados adecuadamente mediante procesos regulares de formación de palabras derivan las formas flexionadas. En LFG todos los procesos de formación de palabras (derivación, composición y flexión) se realizan enteramente en el componente léxico: son tratados como procesos estrictamente léxicos.

Los objetos lingüísticos del componente léxico, tanto las entradas como las reglas léxicas, se representan mediante estructuras de rasgos. Las entradas léxicas incluyen una especificación categorial (P.e.: N, V, A, P, V, AFFIX, COMP, etc.), que indica la categoría preterminal debajo la cual la pieza léxica debe ser introducida, sus rasgos (morfo-)sintácticos (es decir, género, número, tiempo, etc.), así como la forma de la estructura predicado-argumento.

A continuación, se muestran a modo de ejemplo las entradas léxicas de los verbos ingleses *hand* y *persuaded* (2).

(2) Entradas Léxicas:

Hand. V. (_iTENSE) = PRESENT
 ~ [_iSUBJ NUM) = SG \wedge (_iPERS SUBJ) = 3]
 (_iPRED) = 'HAND' <(_iSUBJ) (_iOBJ) (_iOBJ2)>¹⁹

Persuaded. V. (_iTENSE) = PAST
 (_iPRED) = 'PERSUADED' <(_iSUBJ) (_iOBJ) (_iVCOMP)>²⁰
 (_iVCOMP TO) = C +
 (_iVCOMP SUBJ) = (_iOBJ)

Las entradas léxicas verbales de (2)¹⁹ especifican además de la categoría léxica (V), la información flexiva, que en el caso del verbo *hand* establece que está en tiempo presente ((_iTENSE) = PRESENT) y que no puede tener un sujeto en tercera persona del singular (~ [_iSUBJ NUM) = SG \wedge (_iPERS SUBJ) = 3), y en el caso del verbo *persuaded* que se trata de un verbo en pasado ((_iTENSE) = PAST).

Las entradas también contienen la denominada forma semántica o Estructura Argumental, en la que se indica el nombre del predicado semántico (es decir, 'HAND' y 'PERSUADED', en los ejemplos), seguido de una lista de argumentos con funciones gramaticales especificadas, es decir los argumentos interpretables gramaticalmente (<(_iSUBJ) (_iOBJ) (_iOBJ2)> para el verbo *hand* y <(_iSUBJ) (_iOBJ) (_iVCOMP)> para *persuaded*)²⁰. En la entrada léxica de un verbo del tipo 'equi' como *persuaded* (P.e.: *The girl persuaded the baby to go*), también se especifica que la cláusula subcategorizada es infinitiva (indicado en la ecuación restrictiva por el atributo TO, (_iVCOMP TO) = C +) y que el sujeto de esta cláusula es el mismo que el objeto del predicado *persuaded* (indicado mediante un esquema de control (_iVCOMP SUBJ) = (_iOBJ)).

¹⁹ El signo ortográfico '~' equivale a la flecha que indica hacia arriba propuesta en LFG. Estas flechas indican que la información declarada en el léxico debe expandirse por la estructura sintáctica.

²⁰ Las entradas léxicas presentadas en los ejemplos de (2) y (3) ya se encuentran flexionadas, pero, en principio, el lexicon contiene las formas básicas y los afijos que combinados mediante reglas léxicas de flexión, derivación o composición generarán las palabras.

²⁰ El orden de las funciones gramaticales en la forma léxica no corresponde al orden de los constituyentes en una estructura sintagmática, sino al orden numérico de los argumentos lógicos o temáticos en la Estructura Argumental (Bresnan 1982)

Es importante señalar que el verbo subcategoriza funciones sintácticas —sujeto, objeto, vcomp, etc.—, las cuales permiten una cierta independencia categorial en los procesos gramaticales, a diferencia de la GB en que el esquema de subcategorización es un conjunto de categorías gramaticales (NP, SP, etc.). Nótese, además, que en la Estructura Argumental de las formas semánticas también se incluye el sujeto. En LFG la oración no es proyección de ninguna categoría, es considerada una proyección exocéntrica, que puede seleccionar como núcleo cualquier categoría funcional que sea necesaria. Sin embargo, esto no quiere decir que los rasgos de una oración no provengan del núcleo verbal.

En cuanto al tratamiento de la "polyadicity" variable (Bresnan 1982), o las alternancias en la Estructura Argumental del verbo (como por ejemplo la posibilidad de algunos verbos de alternar formas transitivas con intransitivas, construcciones activas con pasivas, etc.), se realiza mediante la formulación de reglas léxicas, es decir, como un proceso estrictamente léxico²¹ (Véase ejemplo (3)).

(3) *Regla de Pasiva para el Inglés:*

Cambio Funcional

(,SUBJ) → ∅ / (,BY OBJ)

(,OBJ) → (,SUBJ)

Cambio Morfológico

V → V_[Part]

(,PARTICIPLE) = PASSIVE

Esta regla indica las sustituciones funcionales que se deben realizar y especifica, también, que se debe añadir un esquema de participio ((,PARTICIPLE) = PASSIVE) en las entradas pasivas. En consecuencia, el resultado de la aplicación de la regla léxica de pasiva será una nueva entrada léxica en la que se señalarán todos los cambios especificados en dicha regla (Véase la entrada léxica pasiva (4) del verbo *persuaded*). Es decir, que la forma léxica de pasiva se deriva de la forma activa.

²¹ En este modelo, las relaciones sistemáticas entre formas léxicas se expresan, en general, mediante reglas léxicas.

(4) *Entrada Léxica:*

Persuaded. V. (_iPARTICIPLE) = PASSIVE
 (_iPRED) = 'PERSUADE <(_iBY OBJ) (_iSUBJ) (_iVCOMP)>
 (_iVCOMP TO) = C +
 (_iVCOMP SUBJ) = (_iSUBJ)

Las reglas que tratan la "poliadicity" variable pueden interactuar con otras reglas léxicas (es decir, con las reglas de formación de palabras), pero nunca con reglas no léxicas.

2.3. La Gramática de Estructura Sintagmática Generalizada (GPSG)

La *Gramática de Estructura Sintagmática Generalizada* (en adelante GPSG) (Gazdar et al. 1985), es una gramática generativa interesada en la descripción de las lenguas particulares, cuyo objetivo final es el desarrollo de una teoría general de la estructura de las lenguas naturales. La GPSG es una teoría perteneciente al conjunto de las denominadas Gramáticas de Unificación, en la que se otorga una atención primordial al concepto de formalización y a la precisión matemática.

La GPSG es una gramática de carácter declarativo e independiente del contexto, en la que sólo se postula un nivel de representación sintáctica y se elimina, por lo tanto, el componente transformacional y las reglas de movimiento.

En GPSG, se considera que una gramática tiene una estructura sintagmática, es decir que existe un solo tipo de objeto sintáctico: la regla de estructura sintagmática. Sin embargo, dichas reglas no se definen de forma directa, sino a través de la teoría de los rasgos sintácticos, de la factorización de las reglas sintagmáticas en reglas de dominio, reglas de precedencia lineal y metarreglas, y de una serie de principios de instanciación de rasgos (Véase figura 1.3).

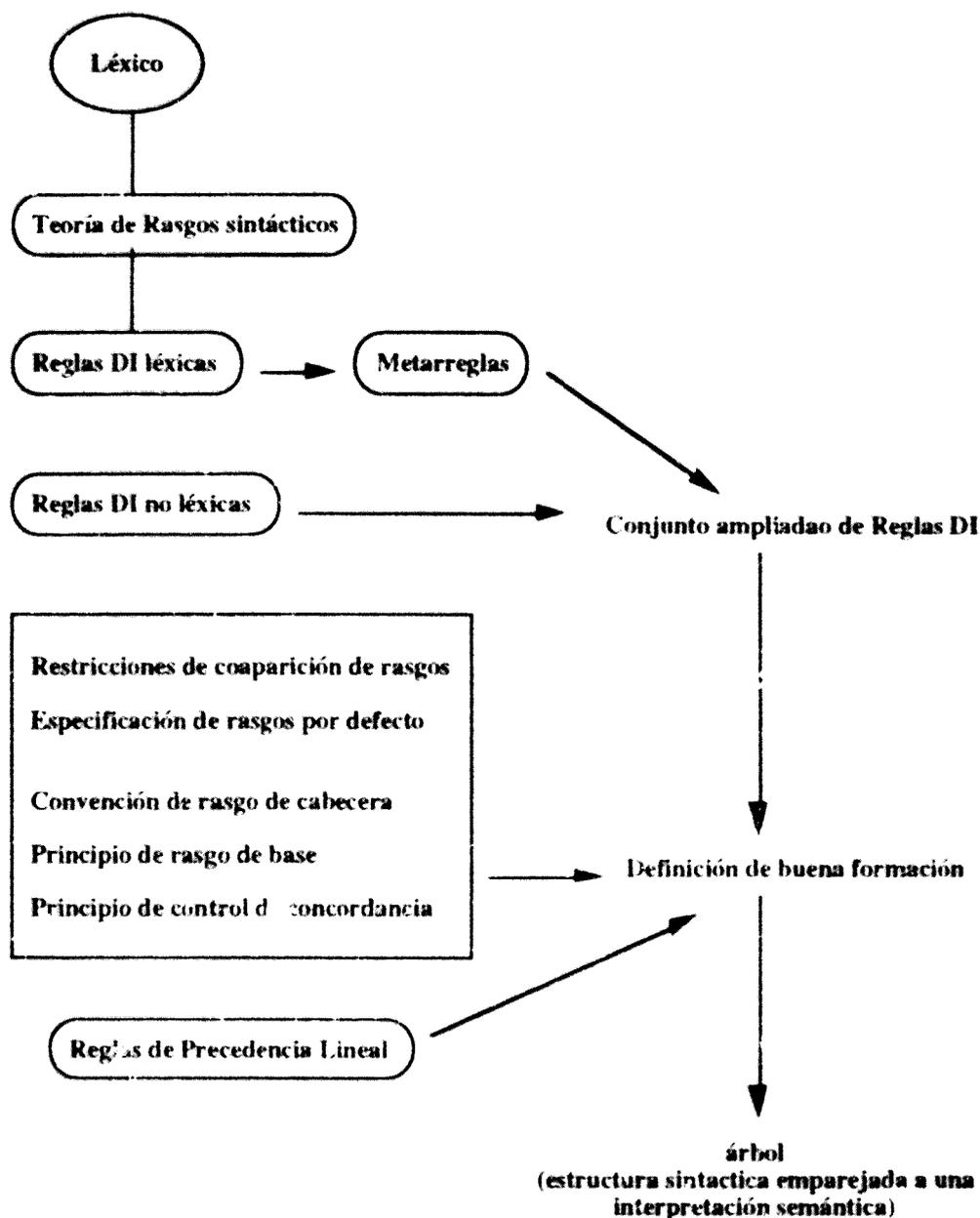


Figura 1.3: Organización de la GPSG²²

En la gramática, cada regla sintáctica está conectada de manera biunívoca a una regla semántica que define el significado del constituyente resultante de la operación sintáctica. El significado de una estructura se concibe como una función del significado de sus componentes. La parte semántica de la regla se define en un sistema lógico-intensional análogo al elaborado en R. Montague (1970). De esta

²² El esquema de organización de los distintos componentes de la Teoría GPSG es una adaptación tomada de P. Sells (1989)

forma, la interpretación semántica se construye al mismo tiempo que la estructura sintáctica.

En GPSG, se asume la Teoría de la X' (Chomsky 1970) que define la categoría sintáctica como una proyección de los rasgos del núcleo, es decir de la categoría léxica. Se parte, igualmente, de la idea de que las categorías gramaticales tienen una estructura interna y que ésta puede analizarse como un conjunto de rasgos ordenados de la forma atributo-valor. La especificación del conjunto de atributos y de valores es muy importante en esta gramática (y, de hecho, en cualquier gramática de unificación) porque definen aquellas propiedades que se consideran pertinentes para la caracterización de los objetos lingüísticos.

Los pares atributo-valor se asignan a las categorías por tres procedimientos distintos: por libre instanciación (es decir, asignados en las entradas léxicas o mediante los principios de Restricción de Coaparición de Rasgos y Especificación de Rasgos por Defecto), por atribución (es decir, mediante una regla de dominio inmediato) o bien por propagación, o herencia, (es decir, en virtud de la Convención de Rasgo de Cabeceera, del Principio de Rasgos Base o del Principio de Control de Concordancia).

Las reglas de Dominio Inmediato (o reglas DI) definen las relaciones de dominio entre la categoría madre y las categorías hijas, además de definir los requisitos de construcción de las categorías. Las reglas DI pueden ser léxicas, que son las que introducen los argumentos subcategorizados por el núcleo léxico, o bien no léxicas que no introducen ningún núcleo léxico y en las que no se refleja, por lo tanto, ningún tipo de subcategorización. Esto significa que en GPSG se definen tantas reglas como esquemas de subcategorización sean necesarios y en las categorías léxicas se especifica (mediante el atributo SUBCAT) en que estructuras pueden aparecer.

Las reglas de Precedencia Lineal (o reglas PL) son las que se encargan de determinar el orden relativo de los elementos hermanos de un árbol. Las reglas PL se aplican sobre todas las reglas DI de la gramática, se trata de condiciones que deben cumplir todas las reglas.

La separación entre la información referente al dominio y la información concerniente al orden sólo afecta a las reglas. Las estructuras sintácticas contienen ambas relaciones.

La GPSG dispone de otro tipo de reglas, las denominadas metarreglas, encargadas de captar las generalizaciones acerca de los constituyentes y, en concreto, acerca de la subcategorización de los núcleos léxicos. En este sentido, sólo pueden aplicarse sobre reglas DI léxicas, es decir que las metarreglas sirven para definir nuevas reglas DI léxicas a partir de las ya existentes.

Los principios de la GPSG pueden ser particulares (Restricción de Coaparición de Rasgos y Especificación de Rasgos por Defecto) o universales (Convención de Rasgo de Cabecera, Principio de Rasgos Base y Principio de Control de Concordancia) y son los que se encargan de la buena formación de las estructuras, es decir que actúan como restricciones en la distribución de los pares atributo-valor.

Por lo que se refiere al léxico está compuesto por una lista de entradas léxicas, en las que se incluye para cada una la información fonológica u ortográfica, la información acerca de la categoría léxica, así como la referente a la subcategorización, una indicación morfológica (en el caso en que tenga una irregularidad morfológica, en caso contrario será una especificación vacía) y la información semántica (el significado o denotación de la pieza léxica):

(5) *Entradas Léxicas:*

<DEVOUR.

[[-N], [+V], [BAR 0], [VFORM BSE], [SUBCAT 2]],

∅.

DEVOUR'>

<GIVE.

[[-N], [+V], [BAR 0], [VFORM BSE], [SUBCAT 3]],

GAVE,

GIVE'>

En la información sintáctica de la entrada se especifica la categoría de la pieza léxica, es decir que se trata de una categoría léxica mayor (especificado mediante el par atributo-valor [BAR 0]) y, en concreto, un verbo ([-N], [+V]) en forma básica ([VFORM BSE]). Además, se incluye, también, la información relativa a la subcategorización mediante el rasgo [SUBCAT], el valor del cual es un índice numérico que asocia la entrada léxica a un esquema de subcategorización definido

por una regla DI léxica ([SUBCAT 2] y [SUBCAT 3]). Las Restricciones de Coaparición de Rasgos aseguran que el rasgo [SUBCAT] sólo se defina en las categorías léxicas mayores, es decir las que funcionan como núcleos léxicos ([BAR 0]).

Las reglas DI léxicas que definen los esquemas de subcategorización de los verbos *devour* y *give* de los ejemplos son las siguientes:

(6) Reglas DI Léxicas:

VP --> H[2], NP

VP --> H[3], NP, PP_[TO]

Nótese que en el marco de subcategorización de los verbos no se incluye información acerca del sujeto. En GPSG, se considera que la oración es una proyección del V, pero la información relativa al sujeto no se especifica directamente en las reglas DI léxicas que introducen los distintos marcos de subcategorización. El hecho de tener sujeto es una propiedad general de todos los verbos y, por lo tanto, dicha información se especifica en una regla DI no léxica más general:

S --> X², H [-SUBCAT]

y mediante una Restricción de Coaparición de Rasgos, en la que se indica que la existencia de un SV implica la de un sujeto:

[+SUBJ] > [+V, -N, BAR 2]

El modelo asigna la misma categoría al SV y a la oración (es decir V²), que sólo se distinguen entre sí por la presencia del rasgo [SUBJ +] en la última.

La información referente a la alternancia en los esquemas de subcategorización se expresa, o bien mediante la postulación de dos entradas léxicas distintas (por ejemplo, en el caso del verbo inglés *give*, que puede aparecer en frases como *Terry gave the book to Lou* o *Terry gave Lou the book*, será considerado en el léxico como V[3] y V[5]), o bien mediante la formulación de una metarregla, como en el caso de la pasiva (Véase ejemplo (8)).

En el ejemplo (7), se muestra la entrada léxica de *give* con [SUBCAT 5] y la regla DI correspondiente.

(7) *Entrada Léxica:*

<GIVE,
 [[-N], [+V], [BAR 0], [VFORM BSE], [SUBCAT 5]],
 GAVE,
 GIVE'>

Regla DI Léxica:

VP --> H[5], NP, NP

La metarregla de pasiva es la que se encarga de captar la generalización lingüística que tanto el verbo *devour* con el rasgo [SUBCAT 2] como el verbo *give* con el rasgo [SUBCAT 3] y [SUBCAT 5], admiten la construcción de pasiva. En este sentido, las metarreglas sirven sólo para expresar generalizaciones sobre las posibilidades de subcategorización de los núcleos léxicos. Dichas reglas operan sobre estructuras sintácticas y no léxicas.

La metarregla de pasiva presenta el siguiente esquema:

(8) *Metarregla de Pasiva del Inglés:*

VP --> W, NP
 ||
 VP [PAS] --> W, (PP_[BY])

(Donde W es una variable que representa una secuencia)

Esta metarregla expresa que toda regla DI léxica de la gramática que tenga un VP que domina un NP, y otro material (W), puede expandirse en una nueva regla que permita la categoría pasiva ([VP [VFORM PAS]]), la cual dominará el material W de la regla originaria, opcionalmente con un PP que contiene la especificación

[PFORM BY]. En consecuencia, la aplicación de la metarregla de pasiva de (8) a las reglas DI léxicas especificadas en (6) y (7) da como resultado las reglas incluidas en (9).

(9) Reglas DI Léxicas:

VP[PAS] --> H[2], (PP[BY])

VP[PAS] --> H[3], PP[TO], (PP[BY])

VP[PAS] --> H[5], NP, (PP[BY])

Las alternancias en los esquemas de subcategorización de los verbos se tratan, básicamente, a nivel sintáctico mediante reglas DI léxicas o mediante la aplicación de una metarregla.

2.4. La Gramática de Estructura Sintagmática Regida por el Núcleo (HPSG)

La *Gramática de Estructura Sintagmática Regida por el Núcleo* (en adelante HPSG) (Pollard & Sag 1987 y 1992), desarrolla un modelo de tratamiento de la sintaxis y la semántica de las lenguas naturales basado en la idea básica que el lenguaje es un vehículo de información. Se trata de una teoría declarativa, basada en la unificación y en la que la especificación de los objetos lingüísticos se realiza de manera acumulativa mediante la interacción de restricciones especificadas en las entradas léxicas, las reglas (esquemas de combinatoria y reglas de precedencia lineal) y los principios. En consecuencia, en HPSG sólo se postula un nivel de representación gramatical, se trata de una teoría monoestratal, de la misma manera que la LFG y la GPSG.

Al igual que las otras teorías basadas en la unificación, los objetos lingüísticos también se analizan en términos de estructuras parciales de información, es decir estructuras de rasgos de la forma atributo-valor. Las entradas léxicas, las reglas, los árboles de estructura sintagmática, los contenidos semánticos, así como también los principios de la gramática se expresan mediante estructuras de rasgos. En este sentido, se considera la HPSG como el formalismo por excelencia de los basados en unificación.

Se trata de una teoría ecléctica ya que muchas de las hipótesis que en ella se desarrollan proceden de las teorías precedentes, en especial de la GPSG y de la *Gramática Categorial*, aunque también tiene influencias de la GB y la LFG, así como de otras disciplinas académicas como son la semántica (en concreto, la Semántica de Situaciones) y la informática (sobre todo, por lo que se refiere a la representación de la información).

El modelo gramatical de la HPSG se compone, básicamente, de un componente léxico, en el que se incluyen las entradas y las reglas léxicas, un conjunto finito de reglas gramaticales y una serie de principios. Las reglas sintagmáticas consisten en esquemas de combinatoria muy generales, que determinan los requisitos de construcción y distribución entre los signos, y en reglas de precedencia lineal, que restringen el orden temporal en que se realizan los constituyentes de un signo. Existen principios universales (Principio de Rasgos Núcleo, Principio de Herencia del Ligamen, Principio de Subcategorización y Principio Semántico) y particulares de cada lengua. Son estos principios los que garantizan la buena formación de los signos (Véase figura 1.4).

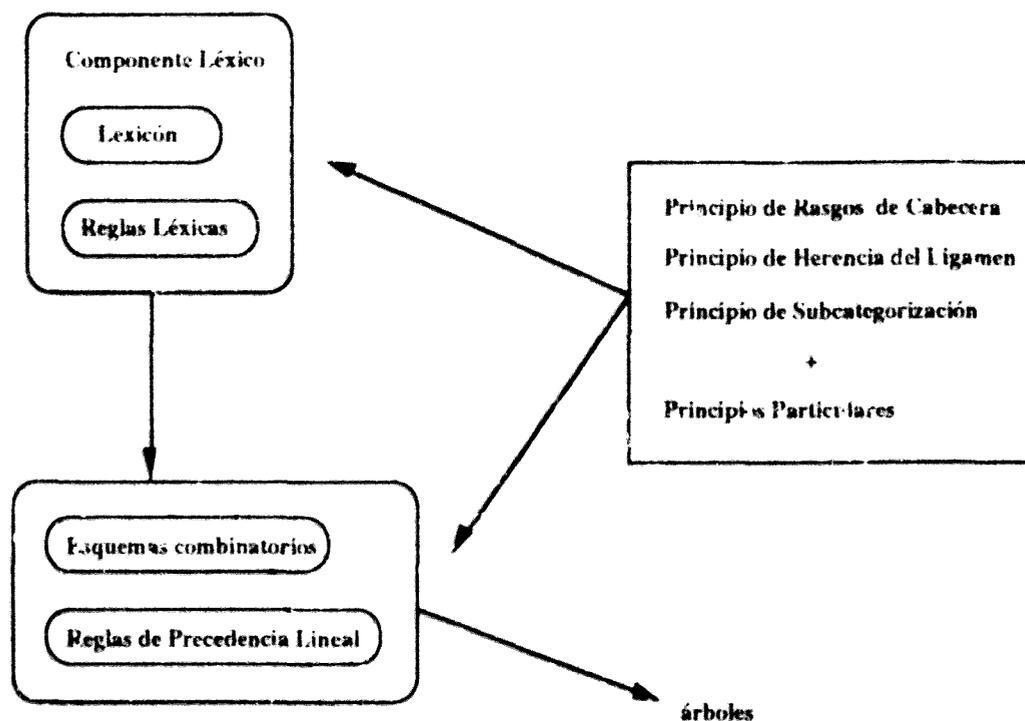


Figura 1.4: Organización de la HPSG

La HPSG es un modelo que no se limita al tratamiento sintáctico del lenguaje. En esta teoría se parte de la idea fundamental que los aspectos sintácticos y semánticos de una teoría gramatical deben tratarse de forma integrada, sin priorizar ninguno de ellos. Este es uno de los aspectos que distingue la HPSG del resto de teorías que hemos expuesto.

El componente léxico incluye las entradas léxicas de la lengua (tanto palabras como afijos) y un conjunto de reglas léxicas (de derivación, composición, flexión, de alternancias de diátesis, etc.).

Las entradas léxicas se definen como signos léxicos en los que la información fonológica (o en su defecto la ortográfica), a partir de la cual se realiza la interpretación fonética, la información (morfo-)sintáctica, la información semántica y la información pragmática se especifican al mismo nivel (Véase la entrada léxica del inglés *tried* de la figura 1.5).

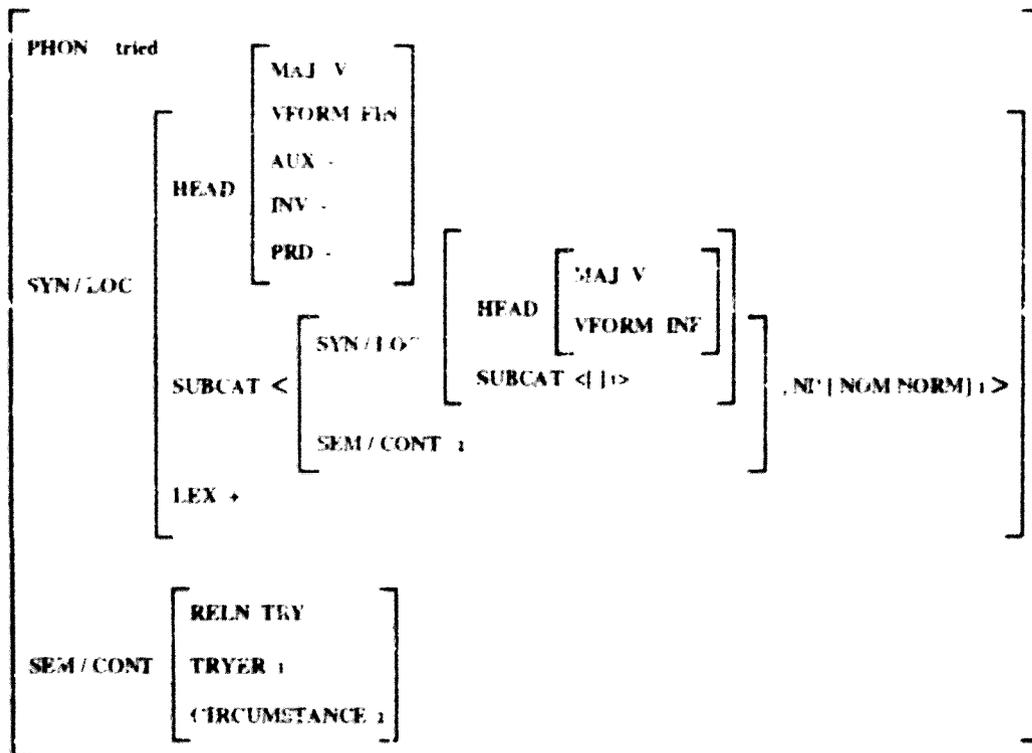


Figura 1.5: Entrada Léxica de 'tried'

Las propiedades sintácticas de los signos léxicos se especifican en el rasgo [LOCALS] del atributo [SYN], en el que se distinguen tres tipos de información:

a) En el rasgo [HEAD], se incluye información acerca de la categoría léxica o sintáctica (N, V, A, P, etc.), así como de las propiedades morfológicas inherentes (es decir, la forma léxica ([VFORM], [NFORM], [PFORM], etc.), el caso para las formas pronominales ([CASE]), si se trata de un verbo auxiliar ([AUX]), si se trata de un verbo predicativo ([PRD]), etc.). Es decir, las propiedades sintácticas que un signo léxico comparte con sus proyecciones.

b) En el rasgo [SUBCAT], se indica la información sobre la valencia del signo, es decir el número y tipo de signos sintagmáticos subcategorizados. Se distingue entre signos saturados, en los que la lista del rasgo [SUBCAT] se encuentra vacía (P.e.: O, SN, DET), y los signos no saturados, en los que dicha lista contiene información (P.e.: N, V, SV, etc.). Según esta distinción, las entradas léxicas verbales son signos no saturados.

El orden de aparición de los signos de la lista [SUBCAT] no refleja, necesariamente, el orden superficial de los complementos, sino más bien la noción de oblicuidad de las relaciones gramaticales, donde el argumento más a la izquierda es el menos oblicuo, es decir el sujeto.

Es importante señalar que el sujeto también se incluye en el marco de subcategorización. En HPSG, la oración es proyección del verbo y la diferencia entre las categorías sintagmáticas S y VP es, precisamente, que en VP el sujeto no se encuentra saturado.

El rasgo [SUBCAT] tiene un papel importante en diversos aspectos de la teoría sintáctica y semántica, porque una de las funciones esenciales de dicho rasgo es establecer la correspondencia entre las relaciones gramaticales y los roles semánticos (dicha relación se refleja en la entrada léxica *tried* (Véase figura 1.5), mediante los subíndices numéricos, que relacionan el SN sujeto₍₁₎ con el argumento semántico 'Tryer₍₁₎' y el objeto VP [INF]₍₂₎ con el argumento semántico 'circumstance₍₂₎').

c) El rasgo [LEX] se utiliza para distinguir los signos léxicos ([LEX +]) de los sintagmáticos ([LEX -]), es decir aquéllos que poseen una estructura constituyente interna. Otro rasgo fundamental que establece la distinción entre signos léxicos y sintagmáticos, es que estos últimos incluyen el atributo [DTRS]

('Daughters') en el que se especifica la información sobre los constituyentes inmediatos (léxicos o sintagmáticos) del signo en cuestión.

En cuanto a la información de tipo semántico, especificada en [SEM/CONT], se basa en la *Semántica de Situaciones* propuesta por (Barwise & Perry 1983), que consideran el significado como la relación entre las expresiones y los tipos de situaciones que éstas describen. Dichas situaciones se definen en términos de individuos, relaciones y propiedades, en una localización espacio-temporal. Asumiendo esta perspectiva, los verbos expresan relaciones (estados, actividades o procesos), que se distinguen según el número de argumentos que participen en dicha relación (es decir, la aridad), y según la manera en que participan estos argumentos en la situación descrita, es decir según sus roles semánticos. El contenido semántico de las oraciones y, en general, de los sintagmas es determinado por la información presente en las entradas léxicas juntamente con principios universales como el Principio Semántico. Dicho principio asegura que el contenido semántico de un signo sintagmático sea idéntico al de su núcleo léxico.

La cantidad de información especificada en el léxico supone una reducción importante del número y complejidad de las reglas de la gramática pero, por otro lado, puede comportar también redundancia e inconsistencia en la especificación de los signos léxicos. Con el fin de evitar redundancias innecesarias o incoherencias en la descripción de las piezas léxicas, la HPSG recurre a la noción de Tipos léxicos y a la formulación de reglas léxicas.

Los tipos léxicos son estructuras de rasgos que permiten organizar y clasificar jerárquicamente la información según las propiedades morfológicas, sintácticas y semánticas de las palabras.

Se propone una jerarquía de tipos léxicos²³, en la cual se describen todos los objetos lingüísticos y en la que cada tipo incluye explícita o implícitamente una serie de rasgos (Véase figura 1.6).

²³ Los tipos se ordenan parcialmente por subsunción (P.e.: signo-léxico \leq signo; signo-sintagmático \leq signo).

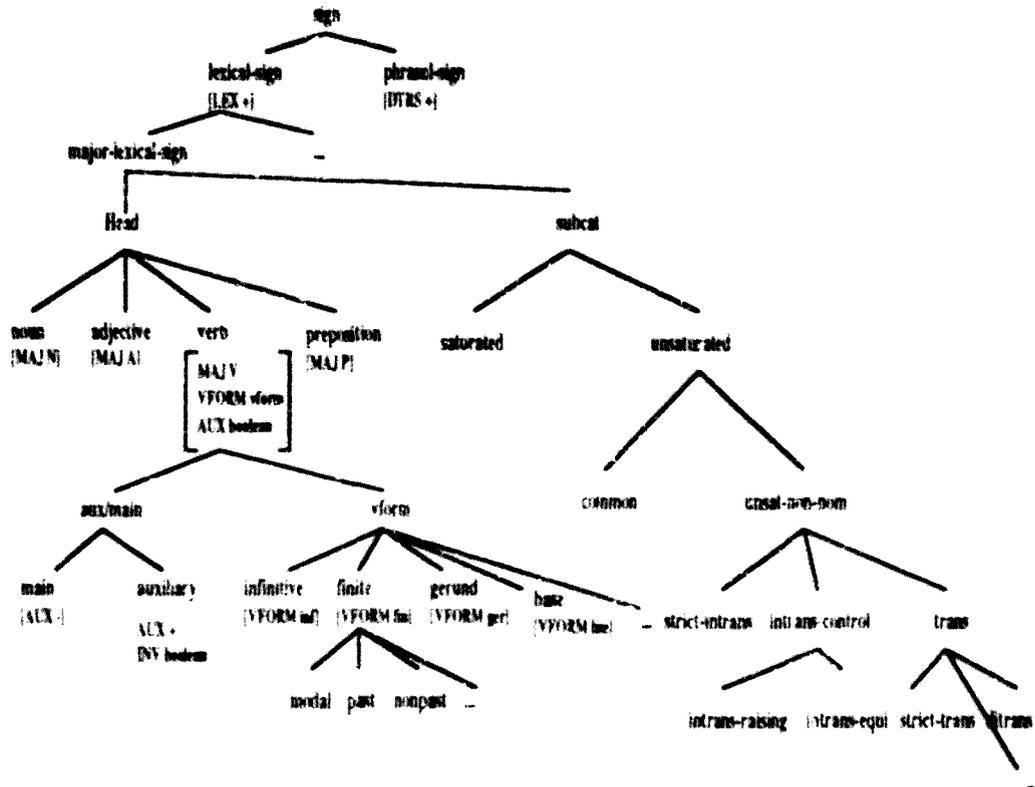


Figura 1.6: Muestra parcial de la jerarquía de tipos

Las reglas léxicas son un mecanismo que se utiliza para generar entradas léxicas nuevas a partir de las ya existentes. En este sentido, las reglas léxicas sirven para captar generalizaciones lingüísticas de tipo morfológico (como los procesos de derivación, composición o fenómenos flexivos) o bien de tipo sintáctico-semántico (como la relación entre forma activa y pasiva (Véase figura 1.7), transitiva e intransitiva, causaliva e incohativa, etc.).

Es decir, las reglas léxicas se pueden utilizar para dar cuenta de las alternancias de diátesis, así como de los procesos derivativos y flexivos. El tratamiento de todos estos fenómenos mediante dichas reglas supone la simplificación de las reglas gramaticales, la reducción del número de entradas léxicas y el control de la redundancia que pudiera darse en el léxico.

Nótese que la regla de pasiva propuesta (Véase figura 1.7), deja intacto el contenido semántico (SEM/CONT₅₃), mientras que cambia el marco de subcategorización (SYN/LOC/SUBCAT <(PP[BY]₁₄) ... []₅₃>) y modifica la fonología de la entrada, aplicando una función que calcula la forma de participio

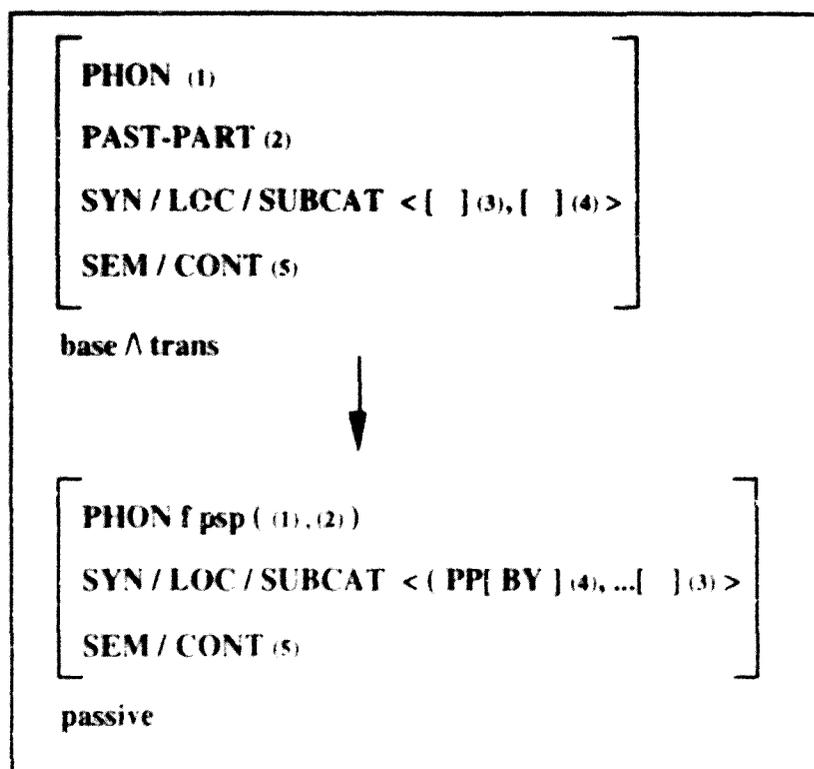
(f_{psp} ((1)(2))).

Figura 1.7: Regla Léxica de Pasiva para el inglés

Los principios generales de la teoría garantizan la buena formación de las estructuras lingüísticas. El Principio de Rasgos Núcleo asegura la unificación de los rasgos núcleo de las piezas léxicas y sus proyecciones. El Principio de Herencia de Ligamen controla la transmisión de la información de los rasgos de ligamen²⁴. El Principio de Subcategorización²⁵ asegura la transmisión de la información contenida en el rasgo [SUBCAT], es decir los requerimientos de subcategorización.

Los principios son, pues, los que controlan la proyección, a través de las estructuras lingüísticas, de la información codificada en los rasgos sintácticos de las

²⁴ El Principio de Rasgos Núcleo y el Principio de Herencia de Ligamen son, esencialmente, una reformulación de la Convención de Rasgos de Cabecera y del Principio de Rasgos Base respectivamente, formulados en la GPSG.

²⁵ El principio de Subcategorización es una generalización de la Cancelación de argumentos utilizada en la Gramática Categorial.

entradas léxicas. Los rasgos sintácticos se clasifican en <HEAD>, <SUBCAT> y <BINDING>²⁶ según los principios a que obedecen. El Principio Semántico es el que asegura que la información semántica especificada en las entradas léxicas se proyecte en los signos sintagmáticos.

2.5. Conclusiones

Una de las características comunes de todas las teorías gramaticales que se han expuesto en los apartados anteriores es la importancia que se da a la información especificada a nivel léxico y, en concreto, a la información de tipo sintáctico-semántico, es decir la Estructura Argumental de las unidades léxicas. En general, las cuatro teorías conciben el léxico como un componente complejo y estructurado, en el que se especifica gran parte de la información lingüística a partir de la cual se derivan las estructuras sintácticas. De ahí el gran interés en el estudio, tratamiento y definición de la Estructura Argumental de las piezas léxicas, sobre todo en GB, LFG y HPSG.

En la Estructura Argumental²⁷ se establece, en definitiva, la relación sistemática entre las propiedades semánticas de los predicados y las configuraciones sintácticas en las que aparecen, es decir la manera en que éstos se realizan sintagmáticamente. En este sentido, hay un descubrimiento de la semántica léxica: en principio, el significado léxico de un predicado determina en gran medida los argumentos que éste tendrá, es decir que la Estructura Argumental es predecible, en gran parte, a partir del significado de las piezas léxicas. De la misma manera, las alternancias de diátesis, es decir las alternancias en la expresión de los argumentos de un verbo, son igualmente determinadas por el significado léxico verbal y, por lo tanto, se tratan a nivel léxico, generalmente, mediante reglas léxicas²⁸.

Otra característica común en todas estas teorías es la tendencia a reducir las reglas gramaticales, convertidas en esquemas combinatorios muy generales, que se

²⁶ Los rasgos de Ligamen (<BINDING>) son los que proporcionan información sobre las dependencias a larga distancia, que no son léxicamente gobernadas (o regidas).

²⁷ La Estructura Argumental también se denomina "Theta-grid" o estructura predicado-argumento.

²⁸ En GB y en GPSG, se realizan a nivel sintáctico, mediante un desplazamiento o una metarregla respectivamente.

limitan a establecer ciertas generalizaciones de gran alcance lingüístico, sobre la manera en que se combinan los elementos de las oraciones y acerca del orden de aparición de éstos en dichas oraciones. La simplificación de las reglas es una consecuencia inmediata del enriquecimiento del componente léxico y, sobre todo, de la definición en las entradas léxicas de información que en modelos anteriores se especificaba a nivel sintáctico, como por ejemplo la especificación del marco de subcategorización o de la Estructura Argumental.

En general, los cuatro modelos gramaticales incluyen en las entradas léxicas la siguiente información: a) información fonológica (u ortográfica), que se utilizará para la interpretación fonológica o fonética; b) información morfológica; c) información sintáctica, en la que se especifica, normalmente, la categoría léxica y el marco de subcategorización e d) información semántica, es decir la Estructura Argumental de los predicados²⁹. Sin embargo, la manera de representar y organizar esta información es distinta en las diversas teorías gramaticales.

A continuación, se muestra de forma esquemática las características principales de la información contenida en el componente léxico de dichas teorías (Véase las figuras 1.8 y 1.9).

Componente Léxico	GB	LFG	GPSG	HPSG
Unidades Léxicas	Palabras/ F. Básicas + Morf.	F. Básicas + Morf.	F. Básicas + Morf.	F. Básicas + Morf.
Derivación y Composición	Léxico	Léxico	Léxico	Léxico
Flexión	Léxico/ Sintaxis	Léxico	Léxico	Léxico
Alternancias de Diátesis	Transformaciones sintácticas	Reglas Léxicas	Metarreglas	Reglas Léxicas

Figura 1.8: El Componente Léxico en GB, LFG, GPSG, HPSG

²⁹ El término predicado entendido en un sentido amplio, ciertos adjetivos y nombres también presentan Estructuras Argumentales, es decir también son predicados semánticos.

Las unidades principales del componente léxico son, en general, formas básicas y afijos, que mediante reglas léxicas (de derivación, composición y flexión) se combinan adecuadamente para formar las palabras o piezas léxicas de la lengua.

En cuanto al tratamiento de las alternancias de diátesis, mientras que en la GB y la GPSG se realizan en el componente sintáctico, mediante transformaciones y metarreglas respectivamente, en LFG y HPSG estos fenómenos se tratan mediante reglas léxicas, es decir en el componente léxico. En consecuencia, las reglas léxicas dan cuenta, por un lado, de la capacidad del lenguaje de generar o crear palabras y, por otro lado, dan cuenta de las relaciones, regularidades y generalizaciones que las unidades léxicas de la lengua guardan entre sí. Las reglas léxicas también reflejan la estructura interna de las palabras o unidades léxicas.

Análogamente, en estas cuatro teorías gramaticales, el verbo adquiere un papel importantísimo, al considerarse que es el predicado verbal el que determina la estructura del SV y, en definitiva, la estructura de la oración. El verbo es considerado, explícita o implícitamente (Véase figura 1.9), el núcleo léxico de las oraciones y, por lo tanto, el que determina las propiedades semánticas y también sintácticas de la oración. En consecuencia, las entradas verbales, al igual que las otras piezas léxicas, contienen una gran cantidad de información, especialmente de tipo sintáctico y semántico (Véase figura 1.9).

Entradas Léxicas	GB	LFG	GPSG	HPSG
Categoría Léxica	+	+	+	+
Rasgos Morfológicos	+	+	+	+
Esquema de Subcategorización	Categorías gramaticales [- Sujeto]	Funciones Gramaticales [+ Sujeto]	Categorías gramaticales (*) [- Sujeto]	Categorías y relaciones gramaticales [+ Sujeto]
Estructura Argumental	Roles Temáticos Arg. Externo e Internos	Predicado Semántico (**)	Predicado Semántico	Roles Semánticos (que describen una Situación)
Oración (Proyección)	FLEX	Exocéntrica	Verbo	Verbo

Figura 1.9: Las Entradas Léxicas en GB, LFG, GPSG, HPSG

(*) En GPSG, son las Reglas DI Léxicas las que introducen los esquemas de subcategorización. En las entradas léxicas se indica mediante el rasgo [SUBCAT] en qué esquemas pueden aparecer.

(**) En LFG, las entradas léxicas sólo especifican la Estructura Predicado-Argumento (es decir, la Estructura Argumental) en la que se indica el predicado semántico y la lista de argumentos con sus correspondientes funciones gramaticales.

Las diferencias fundamentales residen en la formalización de estas especificaciones en las entradas verbales: mientras que en GB y GPSG el esquema de subcategorización es un conjunto de categorías gramaticales en las que sólo se incluyen los complementos verbales, es decir los constituyentes regidos por el verbo³⁰, en LFG es una lista de argumentos con sus respectivas funciones gramaticales y en HPSG una lista de signos, en los que la correspondencia entre relaciones gramaticales y roles semánticos queda establecida. Esta lista de signos está ordenada según la oblicuidad de las relaciones gramaticales. Tanto en la LFG como en la HPSG, en las que la información semántica del predicado es la que

³⁰ En GB y en GPSG, se entiende que el sujeto es una propiedad general de las oraciones y, por lo tanto, su especificación está gobernada por principios generales.

determina básicamente las configuraciones sintácticas, el argumento sujeto forma parte del esquema de subcategorización. De hecho, la Estructura Argumental propuesta en la GB, también, incluye el papel temático del sujeto.

El tipo de entrada léxica que se propone en este trabajo, se basa en las propuestas teóricas de los modelos gramaticales que se han revisado, especialmente de la HPSG, del que se toma la idea de signo léxico, en el que se incluye la información fonológica (u ortográfica), morfo-sintáctica y semántica al mismo nivel (Véase el capítulo 3, donde se expone de forma detallada nuestra propuesta de entrada léxica).

Capítulo 2: Adquisición y representación del Conocimiento Léxico

1. Introducción

La gran cantidad de información contenida en las entradas léxicas y, especialmente, en las entradas verbales, puede plantear problemas en la estructuración y organización de los datos léxicos. En consecuencia, no sólo es importante definir el tipo de información que se ha de especificar en las entradas verbales, sino también determinar el modo en que ésta debe representarse.

Por otro lado, la necesidad en aplicaciones de Procesamiento del Lenguaje Natural de disponer de léxicos computacionales de gran volumen ha supuesto la orientación de muchas investigaciones a la reutilización de la información contenida en los diccionarios de uso común, así como en corpus textuales, como fuente potencial para la adquisición automática (o semi-automática) de conocimiento léxico.

En este capítulo, se presenta la metodología propuesta en el proyecto ACQUILEX, en el marco del cual se ha desarrollado la presente investigación, en el que se propone una solución a los problemas que puede plantear el volumen y la complejidad de los datos especificados en el componente léxico: a) la reutilización de la información de fuentes existentes y b) la representación de la información léxica en una Base de Conocimiento.

El objetivo principal de ACQUILEX¹ es, precisamente, la adquisición de conocimiento léxico a partir de diccionarios de uso común (y de corpus analizados) y su posterior representación en una Base de Conocimiento Léxico multilingüe, a partir de la cual se pueden derivar componentes léxicos para el Procesamiento del Lenguaje Natural.

La exposición se inicia con una breve introducción de dicho proyecto (Véase apartado 2), en la que se describen, de forma general, las distintas etapas que han sido necesarias para llevarlo a cabo, así como también se citan las distintas herramientas y entornos informáticos utilizados, que reflejan, en definitiva, la metodología seguida.

Se pueden destacar, fundamentalmente, tres fases: a) el proceso de volcado del diccionario en cinta magnética (MRD²) a la Base de Datos Léxica (BDL) (Véase sección 2.1), b) el proceso de extracción de la información contenida en los diferentes campos de la BDL y, en concreto, el análisis de las definiciones de las entradas léxicas (Véase sección 2.2) y c) el proceso de conversión, en el que se traduce la información extraída en términos de la Base de Conocimiento Léxico (BCL) (Véase sección 2.3).

El presente capítulo se centra, de forma específica, en las fases de extracción y de conversión de la información, mientras que la fase referente al volcado de MRD a la BDL sólo se hace una ligera referencia.

Por lo que se refiere al proceso de extracción (Véase apartado 3), se expone de forma exhaustiva la manera en que se ha realizado el análisis de las definiciones verbales (Véase sección 3.3), cuyo objetivo principal es la extracción de la información semántica, en especial de los componentes de significado, contenida en ellas.

En cuanto al proceso de conversión (Véase apartado 4), se resaltan más las características básicas de la Base de Conocimiento Léxico (Véase apartado 4.1) que el proceso en sí de conversión o traducción de esta información a la BCL.

¹ Véase (Acquilex-I, 1992) para una visión general del proyecto.

² MRD es el acrónimo del término inglés "Machine-Readable Dictionary".

2. El proyecto ACQUILEX

La presente investigación se ha realizado en el marco del proyecto ACQUILEX³, el objetivo fundamental del cual es el desarrollo de técnicas y metodologías que permiten la adquisición de conocimiento léxico a partir del análisis de la información contenida en diversos diccionarios de uso común, disponibles en cinta magnética (o MRD), y su posterior representación en una Base de Conocimiento Léxico multilingüe, a partir de la cual se pueden derivar componentes léxicos para aplicaciones de Procesamiento del Lenguaje Natural.

"The goal of ACQUILEX is to demonstrate that information can be usefully extracted from multiple MRD sources in a resource efficient fashion in order to (eventually) construct a multilingual Lexical Knowledge Base for use in Natural Language Processing applications" (Briscoe 1991).

En consecuencia, se parte de la hipótesis inicial que los diccionarios de uso común, en los que se incluye una gran cantidad de datos léxicos, constituyen una fuente potencial y, por lo tanto, reutilizable (Calzolari 1991) para la adquisición de conocimiento léxico (Boguraev & Briscoe 1989).

El desarrollo del proyecto se estructura, básicamente, en las siguientes etapas que se describen a continuación (Véase figura 2.1).

1) El proceso de volcado⁴ (o carga) del diccionario en soporte magnético (MRD) a la Base de Datos Léxica.

2) El proceso de extracción de la información contenida en la BDL, mediante la construcción de taxonomías y el análisis de éstas.

3) El proceso de conversión de la información de la taxonomía a la Base de

³ ACQUILEX (1989-1992) y ACQUILEX-II (1992-1995) son proyectos integrados financiados por la CEE a través del programa Esprit, en el que participan las universidades de Amsterdam, Cambridge, Dublín (sólo en el primer proyecto), la Universitat Politècnica de Catalunya, el Istituto di Linguistica Computazionale de Pisa y las empresas editoriales (sólo en Acquilex-II) Bibliograf, Cambridge University Press y Van Dale.

⁴ Traducción del término inglés "loading".

Conocimiento Léxico.

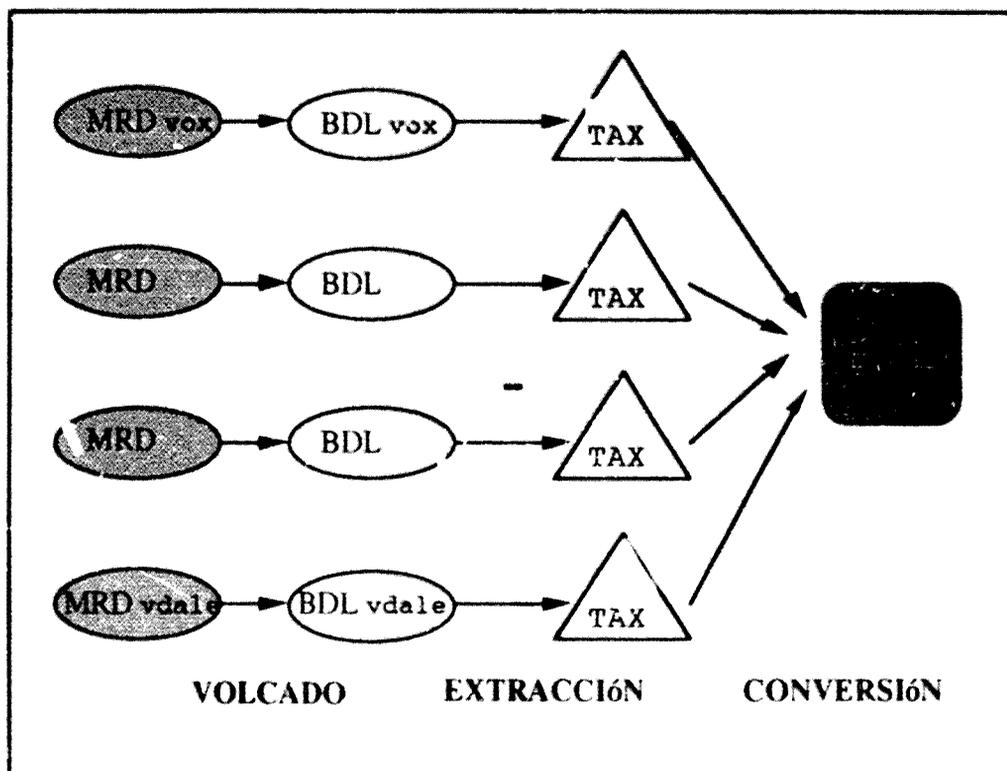


Figura 2.1: Etapas del proyecto ACQUILEX

2.1. Del Diccionario (MRD) a la Base de Datos Léxica

En el proceso de volcado, se transforma la información contenida en el MRD, consistente en una secuencia continuada de texto y códigos tipográficos (Véase figura 2.2), en un formato adecuado para la carga a la Base de Datos Léxica —una estructura lispificada o parentizada (Véase figura 2.3)—, que se obtiene a partir del análisis del MRD (Boguraev & Briscoe 1989).

[e1] freír [r1] (1. [c1] frigere) tr. [r1] Cocer {un manjar} en aceite o grasa hirviendo: ~ con, oen, aceite.[c1] 2. [r1] Mortificar, encocarar: me tiene frito.[c1] 3. [r1] pop.Matar {a alguien} a balazos.- [c1] 4. pml. [r1]fam. pasar mucho calor.[c1] 5. [r1] fr.fig.fam. mandar, ir, etc., a ~ monas, espárragos, etc., despedir a alguien a quien no se desea ver más. [e2][v1] Conjug.[v2] {9} como [n1] reír. [r1]pp.reg.: [n1] freído; [r1]irreg.: [n1] frito.[v1] Sin.[v2] [n1] 1 Sofreír,[r1] freír ligeramente.[f1]

Figura 2.2: *Entrada de diccionario en cinta magnética*

Para el análisis del MRD es necesaria la realización de una gramática ((Castellón & Martí 1990), (Castellón et al. 1992)) que, a grandes rasgos, segmenta la información del MRD por entradas y, cada entrada, por los diferentes campos que contiene. El resultado final es una estructura parentizada que contiene la misma información del diccionario fuente pero etiquetada por campos (Véase figura 2.3).

((freír)
 (ETIM 1. frigere)
 (Sense 1)
 (CA tr.)
 (DEF Cocer {un manjar} en aceite o grasa hirviendo: ~ con. oen. aceite.)
 (Sense 2)
 (CA tr.)
 (DEF Mortificar, encocarar: me tiene frito.)
 (Sense 3)
 (CA tr.)
 (DEF pop.Matar {a alguien} a balazos.-)
 (Sense 4)
 (CA prnl.)
 (SEM fig.)
 (REG fam.)
 (DEF Pasar mucho calor.)
 (Sense 5)
 (CA prnl.)
 (USO fr.)
 (SEM fig.)
 (REG fam.)
 (DEF Mandar, ir, etc., a ~ monas, espárragos, etc., despedir a alguien a quien no se desea ver más.)
 (FORMA 1)
 (TIPOF conjug)
 (TXF .{9} como reír.Partic.reg.: freído; irreg.: frito.)
 (RELA 1)
 (TIPOR Sin.)
 (TXR 1 Sofreír, freír ligeramente.))

Figura 2.3: Estructura parentizada de la entrada verbal 'freír'

Una Base de Datos Léxica (BDL) es una estructura en la que se incluye información léxica organizada (consistente, normalmente, de entradas con campos etiquetados), que permite el acceso fácil y selectivo a los datos que tiene almacenados, es decir a las entradas léxicas y, en principio, a cualquier información que éstas contengan. El sistema de BDL utilizado en el proyecto ACQUILEX se describe extensamente en el trabajo de J. Carroll (1990).

La BDL, dotada con una interfase gráfica, no sólo hace posible la realización de preguntas que permiten el acceso a las entradas léxicas mediante las palabras o formas básicas, es decir de la misma manera que se consulta un diccionario de uso común (p.e: preguntar por las palabras *hervir*, *tragar*, *nadar*, etc.) (Véase figura 2.4), sino que también permite recuperar y preguntar sobre la información específica de cada entrada y sobre las relaciones con otras entradas léxicas (p.e: solicitar todas las palabras que tengan una categoría verbal, o sólo aquéllas que sean verbos transitivos, o bien las que contengan en la definición la palabra *freír*, etc.) (Véase figura 2.5 y figura 2.6).

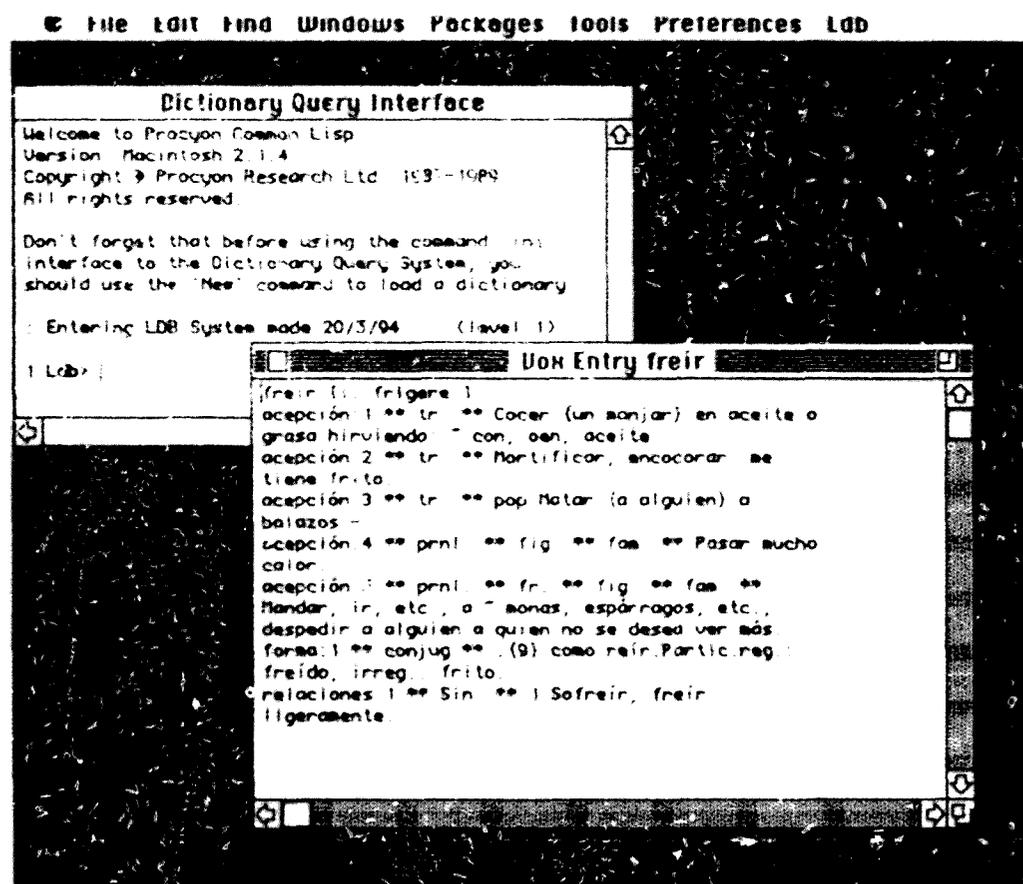


Figura 2.4: Entrada verbal *freír* en la BDL

En la figura 2.4, se muestra la entrada léxica verbal *freír* de la BDL. Dicha entrada contiene la misma información especificada en el diccionario fuente, es decir la etimología, la categoría y las distintas acepciones con sus respectivas definiciones y ejemplos.

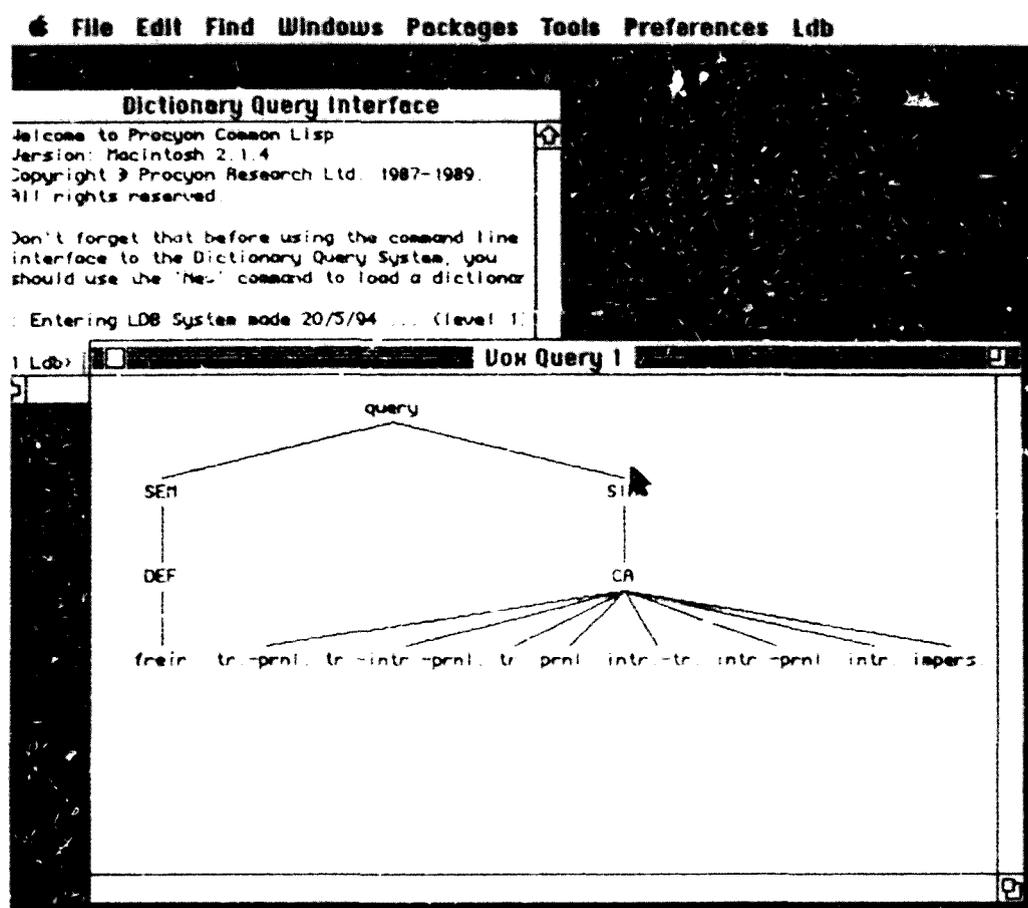


Figura 2.5: Consulta a la BDL

En la figura 2.5, se ejemplifica el tipo de pregunta efectuada a la BDL, mediante la formulación de forma gráfica de la información referente a la categoría ('tr-prnl', 'tr-intr-prnl', 'tr.', 'rec.', 'prnl.', 'intr-tr.', 'intr.', etc.) y sobre la aparición en el campo definición de la palabra 'freír'.

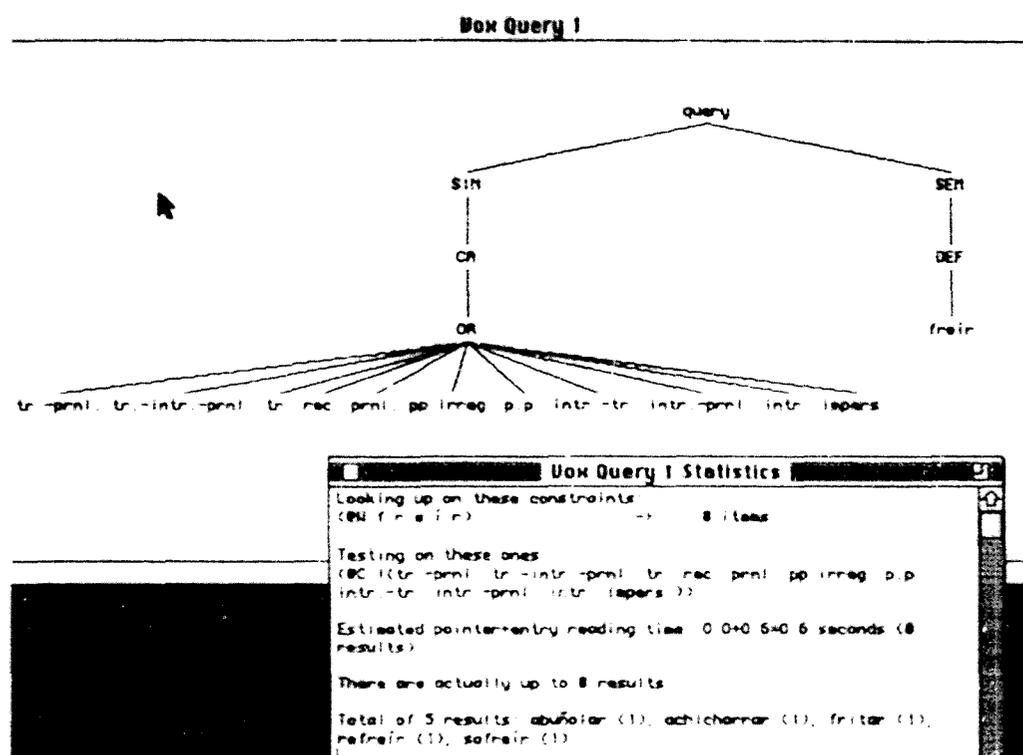


Figura 2.6: Resultado de la consulta a la BDL

En la figura 2.6, se presenta el resultado de entradas léxicas del diccionario que contienen la información requerida en la pregunta realizada a la BDL, es decir aquellas entradas que son verbos y que en su definición aparece la palabra 'freír' (un total de cinco: *abuñolar (1)*, *achicharrar (1)*, *fritar (1)*, *refreír (1)* y *sofreír (1)*).

2.2. Extracción de la información léxica contenida en la Base de Datos: Construcción de taxonomías

El proceso de extracción de la información léxica se lleva a cabo, básicamente, en dos etapas⁵:

En primer lugar, se realiza la extracción de la estructura taxonómica implícita en las definiciones del diccionario (es decir la identificación del "genus" o término

⁵ En los trabajos de A. Ageno et al. (1992b) y de I. Castellón (1992), se presenta extensamente el proceso de construcción de las taxonomías y el posterior análisis de éstas.

genérico⁶) y, a continuación, se extrae mediante el análisis de estas definiciones la información concreta de las propiedades sintácticas y semánticas de las entradas léxicas (es decir la extracción de la "differentia" o elementos que modifican el término genérico) (Véase apartado 3).

Se entiende por estructura taxonómica la clasificación jerárquica de los sentidos del diccionario conectados entre sí por la relación semántica ES-UN (Amsler 1981) y (Cruse 1987). La relación de hiponimia ES-UN, aunque no es la única que subyace en las definiciones del diccionario (en las que se puede encontrar también, por ejemplo, la relación semántica meronímica de PARTE-TODO, CONJUNTO-MIEMBRO, etc. ((Vossen 1990), (Climent & Martí 1995)), es la que se establece de manera más general entre las distintas acepciones (o sentidos) de las entradas léxicas.

La unidad básica de la estructura taxonómica son las acepciones o sentidos de las entradas léxicas y, en la presente investigación, la única relación semántica entre estos sentidos que se ha tratado es la de hiponimia (es decir la relación ES-UN).

Las taxonomías se construyen de modo incremental a partir de subconjuntos temáticos y mediante un proceso descendente, que se inicia a partir de un concepto inicial que funciona como cabeza (nodo o raíz⁷) de la taxonomía. A partir de esta cabeza taxonómica empieza la búsqueda de los respectivos hipónimos, hasta llegar a sentidos que ya no son parte nuclear de ninguna definición. Este proceso se realiza de forma semi-automática, es decir que el lingüista controla y valida, de manera directa, la construcción de la taxonomía y, en consecuencia, el resultado final es una estructura taxonómica totalmente desambiguada (Véase figura 2.7)⁸.

⁶ El término genérico es la palabra o término que funciona como núcleo de la definición de una entrada léxica, normalmente, un nombre en las taxonomías nominales y un verbo en las taxonomías verbales.

⁷ Traducción del término inglés "Top" de la taxonomía.

⁸ En el Apéndice 2, se dan las taxonomías verbales realizadas que han sido necesarias para esta investigación.

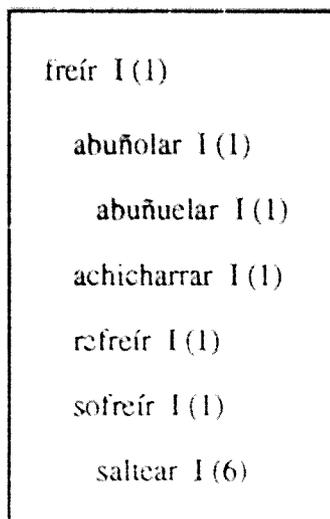


Figura 2.7: Taxonomía de 'freír_I_(1)'

La taxonomía de la figura 2.7 incluye los verbos que se definen mediante el término genérico 'freír' (elegido como cabeza taxonómica) y sólo aquéllos que pertenecen a la acepción primera de dicho verbo, es decir los verbos que se incluyen en la clase semántica de cocción⁹. Es en este sentido que se trata de una taxonomía desambiguada.

Por lo tanto, en los verbos del primer nivel de profundidad de la taxonomía, es decir *abuñolar*, *achicharrar*, *refreír* y *sofreír*, el término genérico es 'freír', mientras que en los verbos de niveles inferiores de la taxonomía, es decir *abuñuelar* y *saltear*, tienen como término genérico *abuñolar* y *sofreír* respectivamente. No hay ningún verbo en el diccionario que se defina con los verbos *abuñuelar* y *saltear*.

Una vez elaborada la taxonomía se procede a la extracción de las propiedades sintácticas y semánticas contenidas en las definiciones de las entradas léxicas, que consiste exactamente en la obtención de la información especificada en la "diferencia" o modificadores del término genérico (Véase figura 2.8)¹⁰.

⁹ La taxonomía de 'freír_I_(1)' forma parte de la estructura taxonómica más general de 'cocer_I_(1 2)' (Véase el Apéndice 2.1).

¹⁰ En el apartado 3 de este mismo capítulo, se expone de forma exhaustiva el análisis de la información especificada en las entradas verbales así como las gramáticas y los analizadores utilizados para llevarlo a cabo.

freír I (1)
 ((CLASS COCER) (AG-TEM MANJAR) (SPREP (EN (OBJECT ACEITE GRASA)) (PROPERTIES HIRVIENDO)) (R210))

abuñolar I (1)
 ((CLASS FREÍR) (AG-TEM HUEVOS OTROS MANJARES) (MODO DEJÁNDOLOS) (R410))

abuñuelar I (1)
 ((CLASS ABUÑOLAR) (R110))

achicharrar I (1)
 ((CLASS FREÍR) (RELATED-TO ASAR TOSTAR) (AG-TEM MANJAR) (R210))

refreír I (1)
 ((CLASS FREÍR) (MODO MUCHO BIEN) (AG-TEM COSA) (R220))

sofreír I (1)
 ((CLASS FREÍR) (AG-TEM COSA) (R210))

saltear I (6)
 ((CLASS SOFREÍR) (AG-TEM MANJAR) (SPREP (A (OBJECT FUEGO)) (PROPERTIES VIVO)) (SPREP (EN (OBJECT MANTECA ACEITE)) (PROPERTIES HIRVIENDO)) (R210))

Figura 2.8: *Taxonomía analizada de 'freír_1_1)*

En este sentido, la información que se puede extraer de forma automática (o semi-automática) de los diccionarios incluye tanto aquella que se considera más explícita, como puede ser la lista de palabras (la cadena ortográfica y, según el diccionario, la descripción fonética¹¹) y la información especificada mediante códigos tipográficos (p.e: categoría léxica, información sobre la etimología, áreas temáticas, registros, etc.), como la información que se presenta de forma implícita (p.e: la descripción semántica contenida en las definiciones y, en particular los componentes de significado, las taxonomías u otras relaciones semánticas, etc.) y que su acceso, por lo tanto, no es tan directo.

¹¹ La descripción fonética aparece, normalmente, en diccionarios bilingües.

2.3. **Proceso de conversión**

El proceso de conversión consiste en la traducción o conversión de las taxonomías, y de la información extraída de las definiciones de las entradas del diccionario, al lenguaje de representación (LRL) de la Base de Conocimiento Léxico multilingüe (BCL).

Una Base de Conocimiento Léxico permite la representación y codificación de forma estructurada del conocimiento léxico, permitiendo, por lo tanto, expresar de manera relacionada la información morfo-sintáctica y semántica a gran escala (Véase apartado 4).

El proceso de conversión exige, en primer lugar, la definición del lenguaje de representación con el cual se define la BCL y, en segundo lugar, las reglas de conversión que permiten incorporar automáticamente la información extraída (en términos del LRL) a la BCL.

El lenguaje de representación utiliza un sistema de estructuras de rasgos tipificadas¹², ordenadas jerárquicamente, basado en la unificación (Carpenter 1990, 1991) y aumentado con mecanismos de herencia (Copestake 1992). La BCL también admite, además de las operaciones de herencia y unificación, la implementación de reglas —léxicas y gramaticales— y la creación de enlaces de traducción¹³ entre entradas léxicas de distintas lenguas.

Las reglas de conversión (Véase figura 2.9), encargadas de traducir la información contenida en las taxonomías al lenguaje de la BCL, forman parte del entorno Convert desarrollado en los trabajos de A. Ageno et al. (1992a y 1992b).

Se trata de reglas declarativas, del tipo condición-acción, que se aplican de forma automática sobre las taxonomías y que producen como resultado las unidades o entradas léxicas de la BCL (Véase figura 2.11).

¹² Traducción del término inglés "typed".

¹³ Traducción del término inglés "link".

```
(EXTRACTION
(CA3 ?CAT *PARES)
((en_lista ?cat '(tr\ tr\ -prnl \ intr\ -tr\))
(intersection_var '((asar I (1))) *PARES)
(unify_tree ?$tot '(*) (sprep (**) (object *x)) *) *)
(intersection_var '*x' (brasas llama parrillas magosto)))
((fill_feature 'etiqueta 'strict-trans-sign)
(fill_feature 'arg1' (** (* qua-eve *)
(** (* p-ag-cause * e-human) (* p-pat-affected * e-inanimate))))
(fill_feature 'arg2' (cat \ result \ result \ m-feats \ vform))
(fill_feature 'tra2' 'base)
(fill_feature 'arg3' (cat \ result \ result \ m-feats \ comp-form))
(fill_feature 'tra3' 'no-comp)
(fill_feature 'arg4' (cat \ result \ result \ m-feats \ diatheses \ trans alt))
(fill_feature 'tra4' 'pass)
(fill_feature 'arg6' (rqs)
(fill_feature 'tra6' 'cook-means))
R_ext_001)
```

Figura 2.9: Regla de conversión

En la regla de conversión de la figura 2.9, se pueden destacar, básicamente, un conjunto de condiciones, que si se cumplen, darán paso a una serie de acciones.

Las condiciones de esta regla concreta (R_ext_001) especifican que los sentidos pertenecientes a la taxonomía de 'asar I (1)' —(intersection_var '((asar I (1))) *PARES)—, cuya categoría sea transitiva, pronominal-transitiva o intransitiva-transitiva ((en_lista ?cat '(tr\ tr\ -prnl \ intr\ -tr\)) y en las definiciones de los cuales se incluya un SP que contenga los nombres 'brasas', 'llama', 'parrilla' o 'magosto' —(unify_tree ?\$tot '(*) (sprep (**) (object *x)) *) *) (intersection_var '*x' (brasas llama parrillas magosto))—, se representarán en la BCL como signos verbales transitivos ((fill_feature 'etiqueta 'strict-trans-sign)) con unos rasgos morfo-sintácticos y semánticos determinados.

En cuanto a los rasgos de tipo morfo-sintáctico, se establece que los atributos 'vform'¹⁴, 'comp-form'¹⁵ y 'diatheses'¹⁶ de estos signos tendrán como

¹⁴ El atributo 'vform' indica la forma verbal, es decir si se trata de un verbo en forma base, infinitiva, pronominal, etc.

¹⁵ El atributo 'comp-form' sirve para indicar el tipo de completiva que rige el verbo.

¹⁶ En el atributo 'diatheses', se especifican el tipo de alternancias de diátesis que admite un verbo, es decir Pasiva, Pasiva Refleja, Objeto Indefinido, etc.

valores 'base', 'no-comp' y 'pass' respectivamente. Referente a los rasgos semánticos, se especifican los argumentos semánticos —(* * (* * qua-eve *) (* * (* p-ag-cause * e-human) (* p-pat-affected * e-inanimate)))) y la clase verbal a la que pertenecen (fill_feature 'tra6 'cook-means), es decir verbos que indican el medio de cocción¹⁷.

La aplicación de esta regla sobre el sentido 'magostar_I_1' (Véase figura 2.10), incluido en la taxonomía de 'asar_I_1', da como resultado la entrada léxica *magostar_X_I_1* de la BCL (Véase figura 2.11)

```
magostar I(1)
((CLASS ASAR)(AG-TEM CASTAÑAS)(SPREP (EN (OBJECT
MAGOSTO))))(R210))
```

Figura 2.10: Análisis del sentido *magostar_I_1*

```
magostar X_I_1
STRICT-TRANS-SIGN
< lex-sign sense-id : sense-id dictionary > = ("VOX")
< lex-sign sense-id : sense-id word > = ("magostar")
< lex-sign sense-id : sense-id homonym-no > = ("I")
< lex-sign sense-id : sense-id sense-no > = ("1")
< cat : result : result : m-feats : vform > = base
< cat : result : result : m-feats : comp-form > = no-comp
< cat : result : result : m-feats : diatheses : trans-alt > = pass
< sem : arg1 : arg1 > = qua-eve
< sem : arg2 : arg1 : pred > = p-agt-cause
< sem : arg2 : arg1 : arg2 > = e-human
< sem : arg2 : arg2 : pred > = p-pat-affected
< sem : arg2 : arg2 : arg2 > = e-inanimate
< rqs > = cook-means.
```

Figura 2.11: Entrada léxica de *magostar_X_I_1* en la BCL

En la figura 2.11, se pueden observar los distintos atributos con sus respectivos valores (en negrita) instanciados. Los distintos valores del rasgo <lex-sign : sense-id> se obtienen de forma automática de la BDL.

¹⁷ En el capítulo 3, se detalla toda esta información incluida en las entradas verbales.

El resultado final es que cada una de los sentidos o acepciones incluidos en las taxonomías se corresponde con una entrada léxica de la BCL.

En el marco del proyecto ACQUILEX, se ha diseñado el entorno SEISD (Sistema de Extracción de Información Semántica de Diccionarios) (Ageo et al. 1992b)¹⁸, en el que se integran todas las herramientas y sistemas o aplicaciones informáticas necesarias para llevar a cabo las distintas etapas mencionadas previamente, especialmente el proceso de extracción de información y la posterior conversión de la misma.

¹⁸ El entorno SEISD ha sido diseñado por el equipo ACQUILEX de Barcelona. Dicho equipo es un grupo interdisciplinar integrado por lingüistas e informáticos de la Universitat Politècnica de Catalunya, la Universitat de Barcelona y la Universidad Nacional de Educación a Distancia.

3. Proceso de extracción de información léxica: Análisis de las definiciones verbales

El objetivo principal de este apartado es la descripción del proceso de extracción de información léxica que contienen las entradas verbales^{19,20} del diccionario fuente y, en concreto, del tipo de información que se obtiene, cómo está organizada en el diccionario y las gramáticas que se han elaborado para extraerla.

Teniendo en cuenta estas consideraciones, la exposición se ha organizado en cuatro puntos claramente diferenciados que reflejan básicamente la metodología que se ha seguido. En primer lugar, se presenta el corpus con el cual se trabaja y las estrategias adoptadas para su selección (Véase punto 3.1). Seguidamente, se describen las herramientas utilizadas para llevar a cabo la extracción de información, se trata sobre todo de los analizadores usados, tanto del morfológico como del sintáctico (Véase punto 3.2). A continuación, se exponen las gramáticas realizadas con una presentación exhaustiva de las reglas desarrolladas (Véase punto 3.3). Finalmente, se muestran los resultados del análisis, donde se ve la correspondencia entre la información que contienen las entradas del diccionario y aquella información que se obtiene, y una valoración general del proceso que se ha seguido (Véase punto 3.4).

3.1. Definición del Corpus

Trabajar con diversas fuentes de conocimiento y con la totalidad de la información que éstas contienen sería la manera ideal de realizar nuestra investigación y, de hecho, es la estrategia seguida dentro del proyecto ACQUILEX, que dispone de diferentes diccionarios monolingües y bilingües para diversas

¹⁹ Los trabajos de Ageno et al. (1991) y Castellón (1992) dan una descripción exhaustiva del análisis de las definiciones de las entradas nominales.

²⁰ El estudio se realiza a nivel de acepciones.

lenguas (Español, Holandés, Inglés e Italiano)²¹.

En el caso particular de esta investigación se ha utilizado el diccionario monolingüe del español Diccionario General Ilustrado de la Lengua Española, VOX (de ahora en adelante VOX). Evidentemente, esta elección influirá de forma clara en el trabajo porque en cada diccionario la información que se incluye se presenta de una forma y con unos criterios concretos que no tienen porque coincidir con los de otros diccionarios. En este sentido la organización de los datos y la manera en que se describen condiciona las tareas que se realizan en el proceso de extracción.

La elaboración del proceso de extracción de información se ha realizado a partir de un subconjunto del diccionario. Se ha considerado oportuno diseñar dicho proceso a partir de una muestra reducida, pero representativa, de entradas verbales del diccionario, y una vez conseguido este primer objetivo, extender el proceso a la totalidad de los verbos.

Se ha trabajado con tres subconjuntos específicos de entradas verbales: verbos de cocción (p.e: *preparar, asar, tostar, cocer, sazonar*, etc.), verbos de ingestión (p.e: *ingerir, comer, beber, tragar*, etc.) y verbos de movimiento sobre todo aquellos que expresan acción (p.e: *mover, andar, correr, pasar, volver, ir*, etc.). Aunque se trabaja con un subconjunto de verbos, el tipo de información que se extrae y la manera en que se representa es, en general, extensible al resto de verbos que se encuentran en el diccionario. Por lo tanto, esta muestra se puede calificar de suficientemente representativa para empezar la investigación.

La elección de estos tres subconjuntos se ha realizado teniendo en cuenta que se trata de conjuntos bastante homogéneos y ricos. Los tres subconjuntos verbales han estado muy estudiados a lo largo de la historia de la lingüística, y esto es consecuencia de la riqueza sintáctica y semántica que presentan, sobre todo por lo que se refiere a las alternancias de diátesis, es decir las diferentes maneras en que se pueden expresar los argumentos de un verbo, y a los componentes de significado o propiedades semánticas que los definen.

²¹ Las fuentes de conocimiento disponibles en el proyecto ACQUILEX son: los diccionarios monolingües, Longman Dictionary of Contemporary English (LDOCE) (Procter 1978) del inglés, Il nuovo dizionario Garzanti del italiano (Garzanti 1984), el Van Dale Groot Woordenboek Hedendaags Nederlands del holandés (Van Dale 1984) y el Diccionario General Ilustrado de la Lengua Española VOX (VOX 1987); y los diccionarios bilingües, Italian-English Collins Bilingual (COLLINS 1987) y el Diccionario Manual VOX: Inglés-Español, Español-Inglés (VOX 1990).

Como ya se ha esbozado previamente, se han adoptado dos restricciones básicas para la elección del corpus lingüístico que se analiza: una, de tipo sintáctico, determinada por la categoría gramatical de la entrada (sólo se tratan entradas verbales) y la otra, de tipo semántico, determinada por el tipo de información que hay en la definición de cada verbo teniendo en cuenta el dominio semántico, es decir los verbos de cocción, ingestión y movimiento.

La selección del corpus se realiza de manera automática mediante preguntas efectuadas a la Base de Datos Léxica (BDL) y con la posterior creación de taxonomías verbales.

Las preguntas a la BDL contienen las dos restricciones mencionadas, es decir que contengan 'verbo' como valor en el campo categoría y, además, que contengan en el campo de la definición una de las formas 'preparar', 'tostar', 'cocer' o 'sazonar' para los verbos de cocción; 'ingerir', 'comer' o 'tragar' para los verbos de ingestión; y 'mover' o 'moverse', 'pasar', 'volver' o 'irse' en el caso de los verbos de movimiento. Precisamente, son estas formas las que se han elegido como nodos iniciales de cada taxonomía. La construcción de taxonomías verbales permite tener los datos de forma estructurada y es la estrategia para obtener de forma interactiva²² el corpus verbal que se quiere analizar.

A continuación, presentamos una muestra de las preguntas que se formulan a la Base de Datos Léxica con el fin de ejemplificar la manera en que se selecciona el corpus de verbos.

Muestra de preguntas a la Base de Datos Léxica

```
[[SEM
 [DEF preparar ]]
 [SIN
 [CA (OR tr.-prnl. tr.-intr.-prnl. tr. rec. prnl. intr.-tr. intr.-
 prnl. intr. impers.)]]]
```

²²El lingüista es el que decide si un determinado verbo tiene que formar parte o no de una taxonomía, esto significa que en las taxonomías resultantes los problemas de desambiguación de sentidos ya han sido resueltos.

[[SEM
 [DEF ingerir]]
 [SIN
 [CA (OR tr.-prnl. tr.-iatr.-prnl. tr. rec. prnl. intr.-tr. intr.-
 prnl. intr. impers.)]]]

[[SEM
 [DEF moverse]]
 [SIN
 [CA (OR tr.-prnl. tr.-intr.-prnl. tr. rec. prnl. intr.-tr. intr.-
 prnl. intr. impers.)]]]

La lectura de estas preguntas es la siguiente: "Cuáles son las entradas del diccionario que contienen en su definición la forma 'preparar', 'ingerir' o 'moverse' (respectivamente en cada pregunta) y, de éstas, cuáles son las que solo tienen una categoría verbal, es decir que son verbos?".

Por lo que se refiere al número de taxonomías realizadas es de 12, en las que se incluyen un total de 479 acepciones verbales (Véase figura 2.12): 4 para los verbos de cocción con 71 acepciones, 3 para los verbos de ingestión con 72 acepciones y 5 para los verbos de movimiento con 336 acepciones²³.

CLASE SEMÁNTICA	Nº Acep.
COCCIÓN	71
INGESTIÓN	72
MOVIMIENTO	336

Figura 2.12: *Relación de Taxonomías*

Una vez realizada la elección y seleccionado el corpus podemos empezar a definir las gramáticas que usaremos con el fin de analizarlo.

²³ En el Apéndice 2, se incluyen las taxonomías realizadas.

3.2. Analizadores

El proceso de extracción de información de las definiciones verbales se ha realizado de forma automática mediante la aplicación del analizador morfológico *Seg-Word* de A. Sanfilippo (1990) y del analizador sintáctico-semántico *FPar* de H. Alshawi (1989).

El analizador morfológico *Seg-Word* está pensado para ser fácilmente integrable en una Base de Datos Léxica que contenga información de diccionario, y esto lo convierte en una herramienta muy interesante para nuestro propósito por el hecho que trabajamos, precisamente, con una Base de Datos Léxica obtenida a partir de un diccionario en cinta magnética.

La característica fundamental de este analizador es que utiliza el propio diccionario (BDL) como la fuente básica de información para el análisis de las formas flexivas y derivadas.

FPar es un analizador sintáctico-semántico basado en esquemas ("pattern-based") organizados jerárquicamente, que realiza un análisis descendente aplicando patrones de definición cada vez más específicos a medida que los niveles superiores (más generales) se van verificando. Los esquemas más específicos son dominados por los esquemas más generales.

El algoritmo de análisis empieza con el nodo inicial de la jerarquía de esquemas y hace que coincida con la definición que se analiza. Si este esquema se verifica, es decir que tiene éxito, entonces lo intenta con cada uno de sus esquemas hijos ("daughter patterns"). El procedimiento se va repitiendo recursivamente hasta que el análisis es completo.

La ventaja de utilizar este procedimiento es que permite proporcionar un análisis parcial cuando no es posible aplicar esquemas más detallados y, por lo tanto, siempre se obtiene un análisis (como mínimo se extrae el elemento nuclear o término genérico de la definición, en este caso, el verbo). Por otro lado, la jerarquía de esquemas nos permite especificar los componentes más importantes.

El analizador contempla dos conjuntos de reglas.

a) las reglas de análisis encargadas de analizar y reconocer los diferentes elementos que presentan las definiciones y,

b) las reglas de interpretación (o de construcción de la estructura semántica, término utilizado por H. Alshawi), encargadas de dar una salida semántica (o bien sintáctica, en caso de ambigüedad) a estos elementos.

Cada una de las reglas de análisis tiene asociada su regla de interpretación correspondiente. En las reglas de interpretación se etiqueta o codifica semántica o sintácticamente la información obtenida en el análisis.

Toda regla de análisis consta de un identificador de regla, de un esquema léxico o sintáctico ("phrasal pattern") y de una lista de identificadores de regla para los esquemas hijos ("daughter patterns"), es decir una llamada a los esquemas más específicos (Véase figura 2.13).

Una regla de análisis se escribe de la siguiente manera:

```
(<rule identifier>  
  <phrasal pattern> <daughter identifiers>)
```

Figura 2.13: Regla de análisis

Las reglas de interpretación también tienen un identificador de regla que se hace coincidir con el de su regla de análisis correspondiente. Las reglas de interpretación asociadas a las reglas de análisis se aplican cuando ninguno de los descendientes inmediatos de la regla de análisis tiene éxito.

En la figura 2.14, se ejemplifica una regla de análisis, concretamente la regla verbal v-110, con su respectiva regla de interpretación. También se pueden observar las tres partes que integran una regla de análisis.

<i>Regla de análisis</i>	<i>Regla de interpretación</i>
(v-110 (v && +v *0sadv &&) v-111))	(v-110 ((compound-class +v) (modo *0sadv) (r110)))
(v-111 (v && +v *0sadv *pp-modi *0sn &&))	(v-111 ((compound-class +v) (modo *0sadv) (ag-tem *0sn) (pr.1l *pp-mod) (r111)))

Figura 2.14: *Regla de análisis y regla de interpretación*

Siguiendo con el ejemplo de la figura 2.14, el identificador de regla es **v-110**, al cual le sigue el esquema léxico o sintáctico (**v && +v *0sadv &&**) y finalmente la lista de identificadores de regla para los esquemas hijos, en este caso sólo uno, **v-111**. La regla **v-111**, por lo tanto, es más específica que la regla **v-110** y sólo se realizará si se cumple primero la más general.

En cuanto a las reglas de interpretación, se etiqueta semántica o sintácticamente la información extraída.

La posibilidad de separar la información analizada y la salida (sintáctico-) semántica en dos tipos distintos de reglas permite que un mismo grupo de reglas de análisis pueda generar estructuras semánticas o sintácticas diferentes. Este hecho es particularmente interesante en el caso de las definiciones verbales porque sólo ha sido necesaria la creación de un único conjunto de reglas de análisis válido para todos los verbos, mientras que se han creado diversos conjuntos de reglas de interpretación según el dominio temático, es decir, tantas como subconjuntos semánticos verbales. De hecho, lo que realmente diferencia un subconjunto semántico de otro son los diferentes esquemas semánticos.

A continuación, se revisan las unidades con las cuales trabaja el analizador y la anotación en que se escriben los esquemas léxicos y sintácticos.

La lista de operadores utilizados es la siguiente:

- + indica que la categoría aparece obligatoriamente una vez
- +0 indica que la categoría aparece opcionalmente una o ninguna vez
- & indica que la categoría aparece obligatoriamente una o más veces
- &() indica que la categoría aparece opcionalmente una, más de una o ninguna vez
- * indica que el nodo de la gramática es una regla independiente del contexto
- *0 indica que el nodo de la gramática es una regla independiente del contexto opcional
- && indica un segmento arbitrario de palabras

En la regla independiente del contexto de la figura 2.15 podemos ver, a modo de ejemplo, estos elementos:

Regla de análisis

```
(sub-pats (QUOTE (*pp-mod))
  (QUOTE
    (**prep-pp (+p +0det +noun &0adj))
    (**prep-pp6 (como +v +noun))
    (**locp-9 (por medio de +0det +noun &0adj))))
```

Regla de interpretación

```
(**prep-pp (sprep (+p (object +noun)) (properties &0adj)))
(**prep-pp6 (simil (+v (object +noun))))
(**locp-9 (instrument +noun))
```

Figura 2.15: Regla independiente del contexto

En la regla independiente del contexto de la figura 2.15, podemos observar que hay elementos terminales (+noun, +0det, &0ad, +p, +v) y no terminales (*pp-mod). Reglas independientes del contexto (*pp-mod) y esquemas léxicos expresados de forma literal ((como +v +noun); (por medio de +0det +noun

&Oadj): (**simil** (+v (object +noun))) y (**instrument** +noun)).

3.3. Análisis de las definiciones

Antes de presentar las gramáticas correspondientes a las definiciones verbales se expone brevemente cuál es la información de la que se parte. A continuación, se presentan las dos gramáticas que se han construido, una para los verbos de cocción e ingestión y otra para los verbos de movimiento y, por último, se revisa la información sintáctico-semántica que se extrae de ellas.

Una entrada verbal está compuesta, básicamente, de la palabra que se define (en infinitivo) y del número de acepciones diferentes que tiene introducidas por un número. La información léxica que contienen es de tipo etimológico, morfológico, gramatical, semántico y de uso. Exceptuando la información semántica que se encuentra en la definición de la entrada, el resto de información, por norma general, se presenta codificada de forma clara con etiquetas específicas, o bien con códigos tipográficos.

De toda la información contenida en las entradas verbales, se ha tratado, de momento, sólo la referente a la categoría gramatical y, sobre todo, la contenida en la definición de la entrada. La información que se extrae de las entradas verbales es, en principio, de tipo morfo-sintáctico y semántico. No se contempla la información de uso, porque el objetivo de la Base de Conocimiento Léxico (BCU) es la representación general de las entradas verbales y no nos interesa, por ahora, el uso específico de un verbo en un determinado registro o área geográfica.

El único dato sintáctico que nos proporcionan las definiciones de los verbos es el referente a la categoría verbal. Las categorías que se dan pueden ser de dos tipos, simples o complejas. Las categorías verbales simples informan sobre si se trata de un verbo transitivo, intransitivo, pronominal, recíproco o impersonal (p.e: '*tr.*', '*intr.*', '*prnl.*', '*rec.*', '*impers.*'), y las categorías complejas indican las posibles alternancias sintácticas que presenta aquel verbo (p.e: '*tr.-prnl.*', '*intr.-prnl.*', '*intr.-tr.*', '*tr.-intr.-prnl.*').

Hay verbos que a veces se comportan, por ejemplo, de forma transitiva y otras veces de forma intransitiva (Véase capítulo 5), como es el caso del verbo *conducir* (etiquetado como '*intr.-tr.*'), que se usa de forma transitiva en la frase

Conduce un deportivo rojo y, en cambio, en la frase *Conduce de forma temeraria* se comporta como un intransitivo.

Por lo que se refiere a la extracción de la información semántica se parte de la definición de la entrada verbal. En este sentido, se puede considerar la definición de un verbo como una descomposición del significado y, por lo tanto, considerarla como una fuente a partir de la cual extraer los componentes semánticos que configuran el significado léxico de los verbos (Véase apartado 3.3 del capítulo 4).

3.3.1. La definición

En la definición de una entrada verbal podemos distinguir básicamente tres partes fundamentales: en primer lugar, lo que se ha denominado "ilex" (o ámbito temático), a continuación, el "genus" o término genérico y, por último, la "differentia" o complementos del término genérico. Así como el campo "ilex" y la "differentia" pueden o no encontrarse en la definición, son opcionales, la presencia del término genérico es siempre obligatoria.

3.3.1.1. "Ilex"

En el código "ilex", que es la abreviación de 'Información Léxica', se indica el ámbito temático en que se utiliza una determinada acepción verbal. En este sentido, "ilex" es equiparable a las etiquetas de ámbito temático del diccionario (es decir, 'Agr.', 'Quim.', etc.), en ningún momento se la tiene que considerar como un complemento del término genérico.

"Ilex" se realiza, normalmente, como un sintagma preposicional o una frase de gerundio a la izquierda del término genérico y encabezando la definición. En el caso de tratarse de un sintagma preposicional casi siempre es introducido por la preposición *en* (1), y si se trata de una frase de gerundio, mayoritariamente es introducida mediante el gerundio *'hablando de...'* (2).

- (1) **ENROCAR I (1):**
En el juego de ajedrez, mover en una misma jugada el rey y un roque bajo condiciones prescritas.
- (2) **ARREMETER I (1):**
Hablando del caballo, partir con furia.

En los ejemplos (1) y (2), tanto el sintagma preposicional *En el juego de ajedrez*, como la frase de gerundio *hablando del caballo*, son etiquetados como "ilex" porque aportan información sobre el ámbito temático en el que se usan estos verbos.

De todas formas, este tipo de información es reconocida por la gramática, pero ignorada en el momento de extraerla, porque como ya se ha señalado anteriormente no nos interesa representar el uso de un verbo en un registro concreto o en una área geográfica específica sino su uso general, por este motivo no se encuentra ningún nodo en la gramática que refleje esta información.

3.3.1.2. Término Genérico

La característica fundamental de las definiciones verbales es que el verbo principal definidor de la entrada se encuentra en infinitivo y, normalmente, es este infinitivo el que inicia la definición. Precisamente, este infinitivo es el que denominamos "genus" o término genérico, el cual se comporta como el elemento nuclear de la definición, en consecuencia, su presencia es obligatoria. Por lo tanto, la definición de una entrada léxica verbal consiste en una oración de infinitivo o simplemente en un infinitivo.

El término genérico puede ser simple, o bien complejo si se trata de una serie de verbos en infinitivo (normalmente no superior a cuatro infinitivos coordinados). Esta lista de infinitivos se presenta concatenada o bien por conjunciones, en particular por la copulativa *y* (3) y por la distributiva *o* (4), o bien por la yuxtaposición mediante comas (5), o por la combinación de conjunciones y comas (6).

- (3) FREZAR I (2):
Tronchar y comer las hojas los gusanos de seda.
- (4) DEAMBULAR I (1):
Andar o pasear sin objeto determinado, por pasatiempo.
- (5) COCINAR I (1):
Guisar, preparar manjares.
- (6) ZIGZAGUEAR I (1):
Andar, moverse o extenderse en zigzag.

También se puede encontrar coordinación o yuxtaposición de dos o más oraciones de infinitivo, es decir de dos términos genéricos con sus respectivos complementos (7).

- (7) SOASAR I (1):
Medio asar o asar ligeramente.
- EMBOCAR I (3):
Comer mucho y de prisa, **exgullir**.

En estos casos sólo se considera la primera frase de infinitivo, es decir las frases *medio asar* y *comer mucho y de prisa*, respectivamente de (7).

3.3.1.3. "Differentia"

Los modificadores o complementos del término genérico, que denominamos "differentia" siguiendo la terminología latina, siempre se encuentran a la derecha del término genérico y, como ya se ha indicado, no son obligatorios.

La "differentia" aporta información semántica sobre las características del Agente, el Paciente, la Trayectoria, la Manera, etc. del verbo que se está definiendo. Estos complementos se realizan estructuralmente en sintagmas nominales, sintagmas preposicionales, adjetivos o participios, adverbios y frases de

gerundio.

El motivo por el cual se habla de componentes o propiedades semánticas, y no de funciones gramaticales, es que el tipo de información que se extrae de las definiciones verbales es, fundamentalmente, semántica y no sintáctica. Los componentes semánticos también son diferentes según el subconjunto temático que se analice. Por ejemplo, en el caso de los verbos de movimiento a menudo encontramos componentes semánticos que indican la Trayectoria, en cambio, en los verbos de cocción o de ingestión no nos encontramos nunca con un componente semántico de este tipo. Es precisamente este tipo de información la que diferencia un subconjunto temático de otro y es, también, uno de los motivos por los cuales se han creado reglas de interpretación distintas para subconjuntos temáticos diferentes.

Por otro lado, la información contenida en la "diferencia" es la que distingue y caracteriza, también, los verbos incluidos en una misma taxonomía.

3.2. Gramáticas

En este apartado se exponen las dos gramáticas que se han construido con el fin de analizar las definiciones de las entradas verbales, la gramática de los verbos de cocción y de ingestión y la gramática de los verbos de movimiento (Véase Apéndice 1). También se ha realizado una tercera gramática, la del término genérico, la finalidad de la cual es la de obtener exclusivamente el término genérico para la creación de taxonomías.

Es importante puntualizar que el análisis de la información se realiza en dos fases: en la primera se obtiene el término genérico y la correspondiente taxonomía, y en la segunda se obtienen las propiedades que modifican el término genérico, es decir la "diferencia". Por este motivo es necesaria esta tercera gramática que sólo contiene las dos reglas que se presentan a continuación (Véase figura 2.16).

Reglas de análisis

(v- (v &&) v-100)

(v-100 (v && +v &&))

Reglas de interpretación(v- ((class thing) (r₀ 0)))

(v-100 ((compound-class +v) (r100)))

Figura 2.16: Gramática del Término Genérico

La primera regla de la figura 2.16 selecciona únicamente aquellas definiciones que están categorizadas como verbos y la segunda es la responsable de la obtención del término genérico.

'Compound-class' no es una etiqueta semántica, sino una función que da una etiqueta diferente en el caso que los genéricos sean simples o coordinados. Si es un término genérico simple la salida del análisis es 'class' (8) y en el caso de los términos genéricos coordinados la salida del análisis es 'class' y 'related-to' (9).

(8) ABUÑUELAR [paras .]
acepción:1 ** tr. ** **Abuñolar.**

abuñuelar I(1)
((CLASS ABUñOLAR) (R110))

(9) RECULAR [probl . del fr. reculer]
acepción:1 ** intr. ** **Cejar o retroceder.**

recular I(1)
((CLASS CEJAR) (RELATED-TO RETROCEDER) (R100))

Como se ha indicado previamente (Véase sección 3.2), las gramáticas utilizadas por el analizador *FPw* soportan dos tipos de reglas: las reglas de análisis, encargadas de reconocer los diferentes elementos que se presentan en las definiciones, y las reglas de interpretación, encargadas de dar una salida semántica (o bien sintáctica, en caso de ambigüedad) a estos elementos.

Como las reglas de análisis de los verbos de cocción y de ingestión son muy similares a las de los verbos de movimiento (la diferencia radica, precisamente, en las reglas de interpretación, y en concreto en sus 'esquemas semánticos') se ha elaborado un único conjunto de reglas de análisis común para los tres subconjuntos semánticos (Véase Apéndice 1.1) y dos conjuntos de reglas de interpretación diferentes (Véase Apéndice 1.2): uno para los verbos de movimiento y otra para los verbos de cocción e ingestión, cada uno de los cuales refleja los diferentes 'esquemas semánticos' de cada subconjunto.

3.4.1. Reglas de Análisis:

La gramática tiene un total de 28 reglas de análisis: 17 reglas dependientes del contexto y 11 independientes del contexto. Trataremos, en primer lugar las reglas dependientes del contexto y, a continuación, las independientes del contexto.

Las dos primeras reglas de la gramática se encargan de la extracción del término genérico (Véase figura 2.17).

(v- (v &&) v-100)

(v-100)

(v && +v &&) v-110 v-200 v-300 v-400)

Figura 2.17: Reglas de análisis: Reglas dependientes del contexto

Las reglas restantes (Véase figura 2.18) se ocupan del análisis de los modificadores y de las diferentes combinaciones en que éstos aparecen en las definiciones verbales. Las reglas más generales son las que analizan los cinco tipos de modificadores de los verbos: adverbio (v-110), sintagma preposicional (v-112)²⁴, sintagma nominal (v-200), adjetivo (v-300) y frase de gerundio (v-400).

²⁴ La regla V-112 sería de tipo complejo, pero por razones de exposición y dado que todos los otros nodos son opcionales, se ha incluido también como una regla general.

(v-110
 (v && +v *0sadv &&) v-111)

(v-112
 (v && +v *pp-mod *0sadv *0pp-mod1 *0sn *0pp-mod1&&))

(v-200
 (v && +v && *sn1 &&) v-210)

(v-300
 (v && +v && *adj &&) v-310 v-320)

(v-400
 (v && +v && *sger &&) v-410 v-420 v-430)

Figura 2.18: Reglas de análisis: Reglas dependientes del contexto

A partir de aquéllas más generales, se aplica el resto de reglas (v-111, (v-112), v-210, v-220, v-310, v-320, v-410, v-420, v-421, v-430 i v-440), que son las que presentan una estructura sintáctica más específica y compleja. Éstas son las que muestran las diferentes combinaciones posibles entre los modificadores, es decir la combinación de sintagmas preposicionales y sintagmas nominales, la combinación de adverbios con sintagmas nominales y preposicionales, etc. (Véase figura 2.19).

En estas reglas se refleja el orden en que se presentan los diferentes modificadores en las definiciones. Las reglas más específicas se llaman desde las reglas más generales, de manera que si no se cumplen las reglas más generales no se aplicarán aquéllas más específicas.

(v-111
 (v && +v *0sadv *pp-mod *0sn *0pp-mod1 *0pp-mod *0sn1 &&))

(v-210
 (v && +v && *sn1 *0sadv *0sn *0pp-mod1 *0pp-mod &&)
 v-220 v-421)

(v-310
 (v && +v &adj &0adv *sn1))

(v-410
 (v && +v &0adj *0pp-mod1 *(0sn *sger &&))

(v-430
 (v && +v *sger *sn1 &&))

(v-440
 (v && +v *0sn1 *pp-mod *sger &&))

Figura 2.19: Reglas de análisis dependientes del contexto

La regla (v-210), que es una de las más productivas, es un buen ejemplo para mostrar las diferentes combinaciones posibles de los modificadores. Los únicos elementos obligatorios de la regla son el verbo (+v) y un sintagma nominal (*sn1), el resto de elementos son opcionales (*0sadv, *0sn, *0pp-mod1, *0pp-mod). La gran cantidad de combinaciones de estructuras posibles, un total de 12, es lo que la convierte en una de las reglas más utilizadas de la gramática, juntamente con la regla (v-111).

Las reglas independientes del contexto desarrollan los diferentes nodos no terminales de la gramática, es decir el sintagma nominal (*sn), el sintagma preposicional (*pp-mod), el sintagma adverbial (*sadv) y la frase de gerundio (*sger). Es en estas reglas donde se ven más las diferencias entre los subconjuntos semánticos, diferencias que quedan claramente reflejadas en sus reglas de salida correspondientes²⁵.

²⁵ Hay esquemas que no tienen su regla de salida correspondiente porque no se dan en el subconjunto semántico.

En la figura 2.20, se da una muestra de las cuatro reglas independientes del contexto mencionadas que contienen los diferentes esquemas en los que se realizan estos nodos en la gramática.

```
(sub-pats (QUOTE (*sn))
(QUOTE
(**sub10 (+0det +noun &0adj)
(**sub12 (+pron))))))

(sub-pats (QUOTE (*sadv))
(QUOTE
(**sadv1 (+adv))))

(sub-pats (QUOTE (*sger))
(QUOTE
(**ger1 (+ger &0adv +0det +noun &0adj
*0pp-mod1)
(**ger2 (+ger *0sadv *pp-mod))
(**ger6 (+ger))))))

(sub-pats(QUOTE (*0pp-mod))
(QUOTE
(**prep-pp (+p +0det +noun &0adj))
(**prep-pp2 (+p +num +noun))
(**prep-pp3 (+p +pron))
(**prep-pp6 (como +v +noun))
(**prep-pp8 (simil +noun))
(**alargo-pp (a lo largo de +0det +noun &0adj))
(**loc-1 (de +0det +noun +p +0det +noun1))
(**locp-6 (por encima de +0det +noun &0adj))
(**locp-9 (por medio de +0det +noun &0adj))))))
```

Figura 2.20: Reglas de análisis independientes del contexto

La regla del sintagma preposicional (***0pp-mod**) presenta diversos esquemas que se han diferenciado entre sintagmas preposicionales (etiquetados en la gramática como ****prep-pp**, con un total de once) y locuciones preposicionales (etiquetadas como ****loc** o ****locp**, con un total de diez). Esta variante está

condicionada por las definiciones de los verbos de movimiento, en las que se usan muchas construcciones preposicionales con el fin de expresar nociones como las de Trayectoria, Origen, Destino, Dirección, etc. Muchos de estos esquemas preposicionales no tienen estructura de salida en la gramática de los verbos de cocción e ingestión, porque simplemente no se encuentran en las definiciones de estos subconjuntos verbales.

La diferencia entre las dos gramáticas, como ya se ha mencionado anteriormente, radica en las reglas de interpretación y, en concreto, en la regla del sintagma preposicional. Las reglas de interpretación de los verbos de cocción e ingestión no contienen locuciones preposicionales, y en cuanto a los sintagmas preposicionales, sólo encontramos ocho de los once que se pueden encontrar en la gramática de los verbos de movimiento. Las reglas de interpretación se ejemplifican con las de los verbos de movimiento.

3.4.2. Reglas de Interpretación

A continuación, se revisan las diferentes etiquetas semánticas y sintácticas que se han asignado a todos estos modificadores del término genérico en las reglas de interpretación de los verbos de cocción, ingestión y movimiento.

En el momento de determinar de forma automática el valor semántico de estos modificadores se plantea el problema de su desambiguación. Los casos más problemáticos, posiblemente porque son los más frecuentes, corresponden a los sintagmas nominales y a los sintagmas preposicionales.

Un sintagma nominal puede funcionar tanto como un Agente o como un Paciente (o Tema) y, por ahora, no hay forma de determinar en qué casos se comporta como un Agente y en cuáles como un Paciente. El orden de aparición de los elementos no es un criterio suficiente para su desambiguación, porque el orden en las definiciones no es fijo. Por este motivo en las reglas de interpretación se ha introducido la etiqueta AG-TEM (Agente-Tema) que indica esta ambigüedad.

A continuación, se muestran unos ejemplos de definiciones de entradas del diccionario (10), en los que se puede observar que dicho orden no es fijo. (En los ejemplos de (11), se dan los análisis correspondientes de estas definiciones.)

- (10) ALETEAR [frecuent.]
 acepción:1 ** intr. ** Mover las **aves** frecuentemente las **alas**,
 sin echar a volar.

OREJEAR

acepción:1 ** intr. ** Mover las **orejas** un **animal**.

- (11) aletear 1(1)
 ((CLASS MOVER) (AG-TEM AVES) (MODO
 FRECUENTEMENTE) (AG-TEM ALAS) (R210))

orejear 1(1)
 ((CLASS MOVER) (AG-TEM OREJAS) (AG-TEM ANIMAL)
 (R210))

En las definiciones de estos dos verbos —*aletear* y *orejear* (10)—, con el mismo término genérico '*mover*', aparecen dos sintagmas nominales, pero así como en el caso del verbo *aletear 1(1)* el primer sintagma nominal es el Agente y el segundo el Paciente o Tema, en el caso del verbo *orejear 1(1)* el primer sintagma nominal se interpreta semánticamente como Paciente, mientras que el segundo se interpreta como Agente.

En el caso de los sintagmas preposicionales la diversidad semántica es todavía mayor, porque pueden expresar Trayectoria (Destino, Origen, Dirección o Ruta), Manera, Finalidad, Localización, etc. Sin embargo, el grado de ambigüedad es menor, porque las distintas preposiciones se interpretan de manera distinta según el término genérico de la definición en la que aparecen. A modo de ejemplo, en las definiciones en las que el término genérico es *andar* los sintagmas preposicionales introducidos por la preposición *por* indican, en general, la Ruta de la Trayectoria (12), mientras que la preposición *hacia* en verbos cuyo genérico es *irse* introduce Trayectorias que expresan la Dirección (13). Las preposiciones *sin* y *con*, en cambio, en las definiciones que contienen el verbo *comer* como término genérico indican, normalmente, Manera (14) y (15), etc.

(12) LADEAR [de lado]

acepción:3 **intr.** Andar o caminar **por** las laderas.

ladear I (3)

((CLASS ANDAR) (RELATED-TO CAMINAR) (SPREP (**POR**
(OBJECT LADERAS)))) (R111))

(13) LEVANTAR [l. levantare < levare, alzar]

acepción:1 **tr.** Mover de abajo **hacia** arriba: ~el brazo.

levantar I (1)

((CLASS MOVER) (SPREP (DE ABAJO)) (SPREP (**HACIA**
ARRIBA)) (R231))

(14) EMBUCIAR

acepción:3 **tr.** Comer **con** exceso.

embuchar I (3)

((CLASS COMER) (SPREP (**CON** (OBJECT EXCESO)))) (R111))

(15) PAPAR [l. pappare]

acepción: 1 **tr.** Comer [cosas blandas] **sin** mascar.

papar I (1)

((CLASS COMER) (AG-TEM COSAS) (SPREP (**SIN** (OBJECT
MASCAR)))) (R210))

El problema se plantea en el momento de especificar en la gramática la interpretación semántica concreta de cada sintagma preposicional, porque los verbos que tienen como término genérico *andar* también forman parte de la taxonomía de *moverse*, que contiene muchos otros verbos en las definiciones de los cuales pueden aparecer sintagmas preposicionales introducidos por la misma preposición *por* que no indiquen Ruta.

Esta estrategia de desambiguación, teniendo en cuenta el término genérico y la preposición, se realiza en las reglas de conversión, donde sí que se dispone de mecanismos para poder expresarlo. Las reglas de conversión son las encargadas de

traducir (y enriquecer) la información obtenida en el proceso de extracción a la Base de Conocimiento Léxico²³ (Véase apartado 2, sección 2.3).

Como los sintagmas preposicionales son ambiguos, la etiqueta semántica asignada es de tipo formal PP-MOD (Modificadores Preposicionales). Un claro ejemplo de esta ambigüedad o polisemia de las preposiciones se presenta en los diferentes significados de la preposición *en* en los verbos de cocción e ingestión, en los que tanto puede funcionar como un complemento que indica el Instrumento o el Medio en el que se cocina *en almíbar* (16), como el Tiempo *en el almuerzo* (17), la Manera *en ligeras porciones* (18) o bien el Locativo *en prado o dehesa* (19).

(16) CONFITAR

acepción:2 **tr.** Cocer [las frutas] **en almíbar**.

confitar I (2)

((CLASS COCER) (AG-TEM FRUTAS) (SPREP (EN (OBJECT ALMÍBAR)))) (R210))

(17) ALMORZAR [l. admordere, morder]

acepción:2 **tr.** Comer **en el almuerzo** {algún manjar}:
- chuletas.

almorzar I (2)

((CLASS COMER) (AG-TEM MANJAR) (SPREP (EN (OBJECT ALMUERZO)))) (R111))

(18) PICOTEAR [frequent]

acepción:4 **intr.** Comer de diversas cosas y **en ligeras porciones**.-

picotear I (4)

((CLASS COMER) (SPREP (DE (OBJECT COSAS)) (PROPERTIES DIVERSAS)) (SPREP (EN (OBJECT PORCIONES)) (PROPERTIES LIGERAS))) (R111))

²³ En el trabajo de A. Agno et al. (1992a y 1992b), se trata el proceso de conversión y se presentan de forma completa las reglas de conversión realizadas.

(19) HERBAJAR

acepción:1 ** tr. ** Apacentar {el ganado} en prado o dehesa.-

herbajar I(1)

((CLASS APACENTAR) (AG-TEM GANADO) (SPREP (EN
(OBJECT PRADO DEHESA))) (R210))

Por lo que se refiere a las locuciones preposicionales, etiquetadas en las reglas de interpretación como LOC-PP (LOC-PP1, LOCP-PP), responden a esquemas del tipo *de un lado a otro*, *a un lado* y *a otro*, *por encima de +n*, *por delante de +n*, etc. Estas construcciones se tratan en el proceso de conversión como indicadores del componente semántico Trayectoria ("Path") (20) y (21).

(20) VENIR [l. -ire]

acepción:1 ** intr. ** Caminar una persona o moverse una cosa **de allá para acá**;

venir I(1)

((CLASS MOVERSE) (AG-TEM COSA) (LOC-PP1 (DE ALLÁ)
(PARA ACÁ)) (R210))

(21) BARBEAR

acepción:4 ** intr. ** Andar el toro **a lo largo de las tablas** rozándolas con el hocico, como buscando la salida del ruedo.

barbear I(4)

((CLASS ANDAR) (AG-TEM TORO) (LOC-PP TABLAS)
(R210))

El esquema *por medio de +n* tiene una etiqueta semántica propia INSTRUMENT (22).

- (22) **BOMBEAR**
 acepción: 3 ** tr. ** Elevar {agua u otro líquido} **por medio de una bomba.**

bombear I (3)
 ((CLASS ELEAR) (AG-TEM AGUA OTRO LÍQUIDO)
 (INSTRUMENT **BOMBA**) (R210))

El esquema *como +n* y *como +v +n* también tienen una etiqueta semántica particular, la de SIMIL (Similar) (23), así como la etiqueta NUMERAL que indica las veces que se repite una misma acción (24).

- (23) **REPTAR** [l. reptare, der. de repere]
 acepción: 1 ** intr. ** Andar **arrastrándose como los reptiles.**

reptar I (1)
 ((CLASS ANDAR) (MODO ARRASTRÁNDOSE) (**SIMIL REPTILES**) (R420))

- (24) **REBINAR**
 acepción: 1 ** tr. ** Cavar **por tercera vez {las viñas}.**

rebinar I (1)
 ((CLASS CAVAR) (AG-TEM VIÑAS) (**NUMERAL TERCERA VEZ**) (R111))

Estos esquemas, pueden parecer, en principio, muy "ad hoc", pero de hecho se trata de un tipo de estructura que se encuentra más de una vez en las definiciones verbales de estos subconjuntos semánticos.

A continuación, se da una muestra de las reglas de interpretación correspondientes a los diferentes esquemas incluidos en la regla independiente del contexto del sintagma preposicional (****pp-mod**) (Véase figura 2.21).

(**prep-pp (**sprep** (+p (object +noun)) (properties &Adj)))
 (**prep-pp2 (**numeral** +num +noun))
 (**prep-pp3 (**sprep** (+p (object +pron))))
 (**prep-pp6 (**simil1** (+v (object +noun))))
 (**prep-pp8 (**simil** +noun))
 (**alargo-pp (**loc-pp** +noun))
 (**loc-1 (**loc-pp1** (de +noun) (+p +noun1)))
 (**locp-6 (**locp-pp** +noun))
 (**locp-9 (**instrument** +noun))

Figura 2.21: Reglas de interpretación independientes del contexto²⁷

El resto de modificadores, es decir frases de gerundio (25), adjetivos (26) y adverbios (27), juegan el mismo rol semántico en los tres subconjuntos semánticos verbales. Semánticamente se trata de complementos que denotan *Manera*, los cuales se han etiquetado en la gramática como MODO.

(25) SORBER [l. -ere]

acepción:1 ** tr. ** Beber **aspirando**: ~ una horchata.

sorber I (1)

((CLASS BEBER) (**MODO ASPIRANDO**) (R410))

(26) TRAVESEAR [de travieso]

acepción:1 ** intr. ** Andar **inquieto** y **revoltoso** de una parte a otra.

travesear I (1)

((CLASS ANDAR) (**MODO INQUIETO REVOLTOSO**)
 (LOC-PP1 (DE PARTE) (A OTRA)) (R320))

(27) GLOTONEAR

acepción:1 ** intr. ** Comer **glotonamente**

glotonear I (1)

((CLASS COMER) (**MODO GLOTONAMENTE**) (R110))

²⁷ En el Apéndice 1 se da la regla del **pp-mod completa.

Para terminar con este apartado se presentan, a continuación, una muestra de reglas de interpretación con el fin de ilustrar lo que se ha expuesto (Véase figura 2.22 y 2.23).

```
(v-110
((compound-class +v)
(pm1 *0sadv)
(r110)))

(v-200
((compound-class +v)
(ag-tem *sn1)
(r200)))

(v-300
((compound-class +v)
(modo &adj)
(r300)))

(v-400
((compound-class +v)
(modo *sger)
(r400)))
```

Figura 2.22: *Muestra de Reglas de interpretación dependientes del contexto*

```
(**prep-pp8 (simil +noun))
(**loc-1 (loc-pp1 (de +noun) (+p +noun1)))
(**locp-9 (instrument +noun))
(**sub10 (+noun (properties &0adj)))
(**sadv1 (modo (+adv)))
(**ger6 (+ger))
```

Figura 2.23: *Muestra de Reglas de interpretación independientes del contexto*

3.5. Resultados

La gramática se ha aplicado sobre un total de 479 definiciones de entradas verbales correspondientes a los subconjuntos semánticos de cocción, ingestión y movimiento. El análisis de las definiciones ha sido correcto en el 95% de entradas analizadas. El 5% restante se debe a errores tipográficos del diccionario que dan lugar a palabras inexistentes y, por lo tanto, imposibles de categorizar y reconocer por la gramática, o bien a errores producidos por ambigüedad categorial, es decir palabras que tienen asignada más de una categoría. El método de análisis garantiza que en el peor de los casos se obtenga como mínimo el término genérico.

Las reglas v-210 y v-111 son las más utilizadas de la gramática seguidas por las reglas v-220, v-320, v-420 y v-440. Se trata de las reglas que contienen mayor número de elementos opcionales y, por lo tanto, son también las que resuelven casi todas las combinaciones posibles entre los diferentes modificadores. El resto de reglas se utiliza pocas veces porque se trata de estructuras o combinaciones muy poco usadas en las definiciones. Por lo que se refiere a las reglas v-100, v-200, v-300 y v-400, son tan generales (desde éstas se llaman el resto de reglas) que no aparecen nunca como resultado de un análisis, aunque casi siempre se apliquen. En cuanto a la regla v-110 es la tercera más utilizada porque es la que cubre el análisis del término genérico y es muy frecuente que los verbos se definan mediante otro verbo o mediante la concatenación de más de uno.

El hecho de que el analizador *FPar* permita diferenciar las reglas de análisis de las reglas de interpretación facilita el proceso de elaboración de la gramática ya que las reglas de análisis son válidas para cualquier verbo del diccionario y sólo las reglas de interpretación deberán elaborarse para cada subconjunto semántico.

A continuación (Véase figura 2.24), se muestran los resultados del análisis sintáctico-semántico, en los que se puede observar la relación de modificadores extraídos de las definiciones de los verbos pertenecientes a cada una de las taxonomías analizadas.

RESULTADOS DEL ANÁLISIS SINTÁCTICO-SEMÁNTICO									
CLASS	RELATED-TO	AG-TEM	MODO	SPREP	LOC-PP1	NUM	INSTR.	CANTIDAD	SIMIL.
COCCIÓN	10	56	9	28					
Cocer	1	19	3	18					
Preparar	4	22	4	18					
Sazonar	2	7		7					
Testar	3	8	2	5					
INGESTIÓN	10	47	9	41					
Comer	6	33	5	32		1			
lamer	1	9	1	6					
Tragar	3	5	3	3					
MOVIMIENTO	85	179	87	178	19		3	1	4
Irse	10	4		9	1				
Mover	16	74	22	38	1			1	
Moverse	24	40	43	55	9				4
Pasar	36	45	19	61	7				
Volver	11	16	3	15	1				

Figura 2.24: Resultados del análisis de las definiciones

La etiqueta AG-TEM, que se corresponde con los SNs interpretables desde un punto de vista semántico como Agente o Paciente, es aquella que aparece más veces en las definiciones verbales, especialmente en las de los verbos de cocción e ingestión.

En las taxonomías de movimiento el número de modificadores de tipo preposicional (SPREP y LOC-PP1) es ligeramente superior a las ocurrencias de AG-TEM. Las etiquetas SPREP y LOC-PP1 expresan generalmente, en dichas taxonomías, la Trayectoria, componente semántico característico de la clase semántica de movimiento. Nótese que la etiqueta LOC-PP1 no aparece ni una sola vez en las otras taxonomías.

En las taxonomías de cocción e ingestión, la etiqueta SPREP suele interpretarse semánticamente de diversas maneras, a veces indica el Medio (o Instrumento) en el que se realiza la acción, otras el Lugar o el Tiempo y, a menudo, la Manera. En este sentido, en dichas taxonomías la presencia de modificadores que expresan Manera es considerable, porque también hay que añadir las ocurrencias

(desambiguadas) etiquetadas como MODO.

El resto de etiquetas —NUMERAL, INSTRUMENT, SIMIL y CANTIDAD— ocurren de manera puntual en las taxonomías analizadas.

4. La Base de Conocimiento Léxico

Una vez se han extraído las taxonomías y los componentes de significado de las definiciones, esta información se traduce, en términos del Lenguaje de Representación Léxica (LRL), a la Base de Conocimiento.

La cantidad y complejidad de la información contenida en las entradas léxicas puede significar la presencia de información redundante, así como también puede crear problemas de control de ésta. En consecuencia, el modo en que se estructura y organiza toda esta información es fundamental, tanto para evitar la posible redundancia e inconsistencia en los datos, como para reflejar las similitudes entre clases de palabras que se comportan de la misma manera.

Las Bases de Conocimiento Léxico (BCL) permiten abordar este problema de forma eficaz mediante la codificación estructurada del conocimiento. Las BCLs incorporan mecanismos de herencia y reglas léxicas que evitan tanto la redundancia como la inconsistencia de la información.

A continuación se describe, de forma más detallada, la BCL desarrollada en el marco del proyecto ACQUILEX y, en concreto, el lenguaje de representación que la define (Véase sección 4.1).

4.1. El Lenguaje de Representación Léxica (LRL)

La Base de Conocimiento Léxico desarrollada en el marco del Proyecto ACQUILEX (Copestake 1992) está diseñada para representar la información extraída de diversas fuentes de conocimiento multilingües y con el objetivo final de proporcionar un lexicón estructurado, que contenga información morfo-sintáctica y semántica a gran escala.

La BCL utiliza un lenguaje de representación (LRL) basado en la unificación de tipos, o estructuras de rasgos tipificadas²⁰ ordenadas jerárquicamente, que incorpora la herencia por defecto. El LRL está basado en el formalismo gráfico de

²⁰ Traducción del término inglés "typed".

unificación de tipos²⁹ propuesto en (Carpenter 1990, 1991), ampliado con mecanismos de herencia y reglas — léxicas y gramaticales—, además de permitir también la creación de enlaces de traducción^{30,31}.

Las reglas léxicas³² (Véase capítulo 5 sección 4.1) se encargan de captar las generalizaciones lingüísticas entre palabras pertenecientes a la misma clase léxica³³, tanto las de tipo morfológico (como son los procesos de derivación, composición y fenómenos flexivos), como las de tipo sintáctico-semántico (como por ejemplo las alternancias de diátesis).

Las reglas gramaticales³⁴, en cambio, son las que permiten la combinación de signos (tanto léxicos como sintagmáticos) para crear un nuevo y único signo sintagmático (Véase Apéndice 3).

Los enlaces de traducción³⁵ se utilizan, fundamentalmente, para expresar la equivalencia en la traducción entre expresiones de lenguas distintas. Esto es perfectamente posible en una BCL multilingüe, que utiliza el mismo LRL y un sistema de tipos común, en la que los objetos o entradas léxicas se describen en el mismo 'metalenguaje' y, en consecuencia, es posible relacionar entradas de lenguas distintas mediante los denominados enlaces de traducción (Véase Apéndice 3).

La BCL se compone, básicamente, de un sistema de tipos, que describe el universo de elementos con los que se representan los objetos lingüísticos, y de un conjunto de entradas léxicas, representadas también mediante tipos (signos léxicos).

²⁹ Traducción del inglés "Typed graph-based unification formalism".

³⁰ Traducción del término inglés "Tlink".

³¹ Las reglas gramaticales y los enlaces de traducción no se exponen de forma detallada porque no son objeto de estudio en la presente investigación.

³² Véase Apéndice 4.

³³ En el apartado 4 del capítulo 5, se presentan detalladamente las reglas léxicas.

³⁴ Los trabajos de A. Copestake (1992) e I. Castellón (1992) muestran de forma exhaustiva las reglas gramaticales desarrolladas en la BCL.

³⁵ En los trabajos de A. Sanfilippo et al. (1992), I. Castellón (1992), C. Soler & M.ª Martí (1994), A. Agero et al. (1994) se exponen de forma detallada los enlaces de traducción.

4.1.1. El Sistema de tipos

Un tipo es una estructura de rasgos (de la forma atributo-valor) sujeta a restricciones, las cuales definen la buena formación de dicho tipo. Por lo tanto, los tipos léxicos son estructuras de rasgos que permiten organizar y clasificar jerárquicamente la información según las propiedades morfológicas, sintácticas y semánticas de las palabras (Véase figura 2.25).

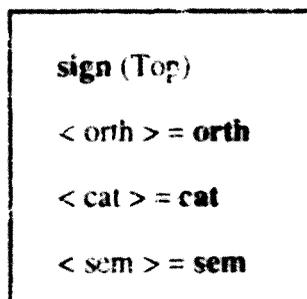


Figura 2.25: Estructura de Rasgos Tipificada

En la figura 2.25, se muestra el tipo 'sign' definido mediante una estructura de rasgos, en la que se destacan tres atributos (<orth>, <cat> y <sem>) con sus respectivos valores ('orth', 'cat' y 'sem')³⁶ que, a su vez, son también tipos. En cada tipo, también, se especifica el tipo superior que lo subsume, en este caso 'Top' (es decir el más general de la jerarquía).

Siguiendo el planteamiento teórico expuesto en la *Gramática de Estructura Sintagmática Regida por el Núcleo* (HPSG), el sistema de tipos de la BCL ordena jerárquicamente³⁷ las estructuras de rasgos tipificadas (o tipos), especificando el tipo más general (o raíz³⁸) en el nodo más alto del diagrama y los tipos restantes en los niveles siguientes (Véase figura 2.26).

³⁶ En el capítulo 3, se explican con más detalle todos estos rasgos.

³⁷ Véase Apéndice 3

³⁸ Traducción del término inglés "Top".

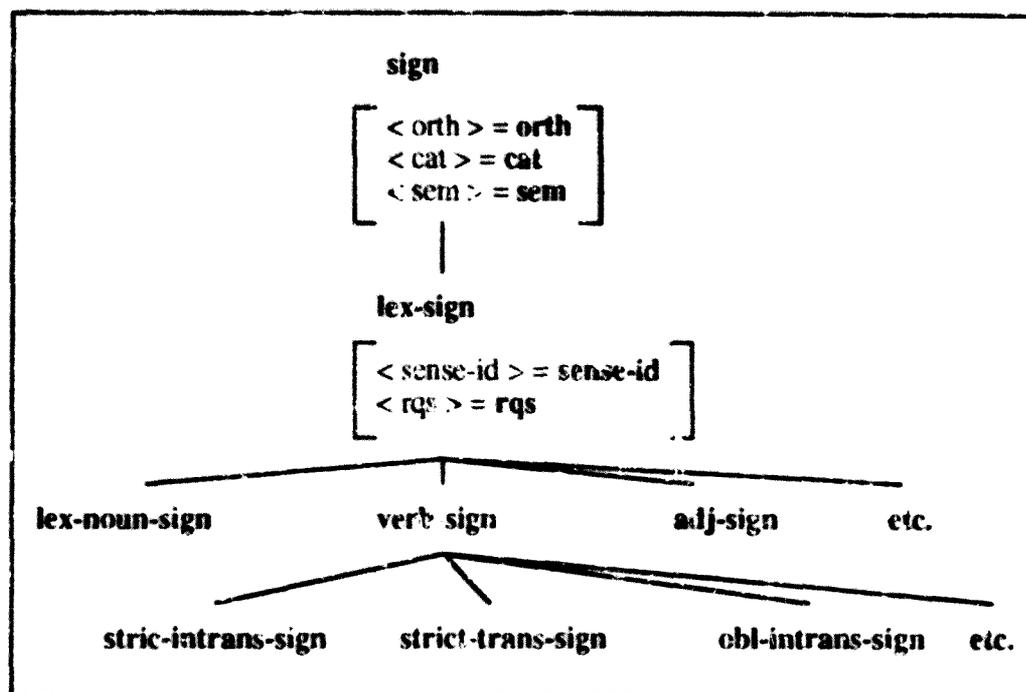


Figura 2.26: Jerarquía de tipos

En la organización jerárquica de tipos de la figura 2.26, se puede observar que los tipos 'lex-noun-sign', 'verb-sign' y 'adj-sign' son subtipos del tipo más general 'lex-sign' que, a su vez, es subtipo de 'sign'. El tipo 'lex-sign' tiene los mismos rasgos que 'sign'²⁰ (es decir <orth>, <cat> y <sem>), además de los rasgos propios (es decir <sense-id> y <rqs>).

El sistema de tipos se ordena parcialmente mediante la relación de subsunción. Se entiende que "una expresión A subsume una expresión B, si A y B no contienen información incompatible, y B es más informativa que A." (Copestake et al. 1991).

En definitiva, todos los objetos de la BCL se describen mediante estructuras de rasgos tipificadas: los tipos funcionales (p.e: 'boolean', 'string', 'list', etc.); los operadores lógicos (p.e: la conectiva 'and'); los descriptores lingüísticos (tanto morfológicos (p.e: número, persona, forma verbal, etc.), sintácticos (p.e: categoría, marco de subcategorización, diátesis, etc.) semántico-lógicos (p.e:

²⁰ La transferencia de la información especificada en el tipo superior al tipo inferior es posible gracias a la herencia monotónica múltiple ortogonal.

fórmula lógica, índice ontológico, papeles temáticos, etc.), como ontológicos (p.e: 'vrqs', clase semántica, componentes de significado, etc.); las reglas léxicas y gramaticales; los enlaces de traducción y las entradas léxicas.

4.1.2. Unificación

La unificación es una operación básica que se aplica sobre las estructuras de rasgos. La unificación permite combinar la información contenida en dos (o más) estructuras de rasgos, siempre que la información que contengan sea compatible, con el fin de crear una nueva estructura de rasgos o bien para completar y enriquecer su información.

Por otro lado, la unificación no autoriza la combinación de estructuras incompatibles. Se considera que dos estructuras de rasgos A y B se pueden unificar si sus atributos y valores son compatibles (Véase figura 2.27).

"Unification is the operation that determines whether two types are compatible by building the most general type compatible with both" (Pereira 1984).

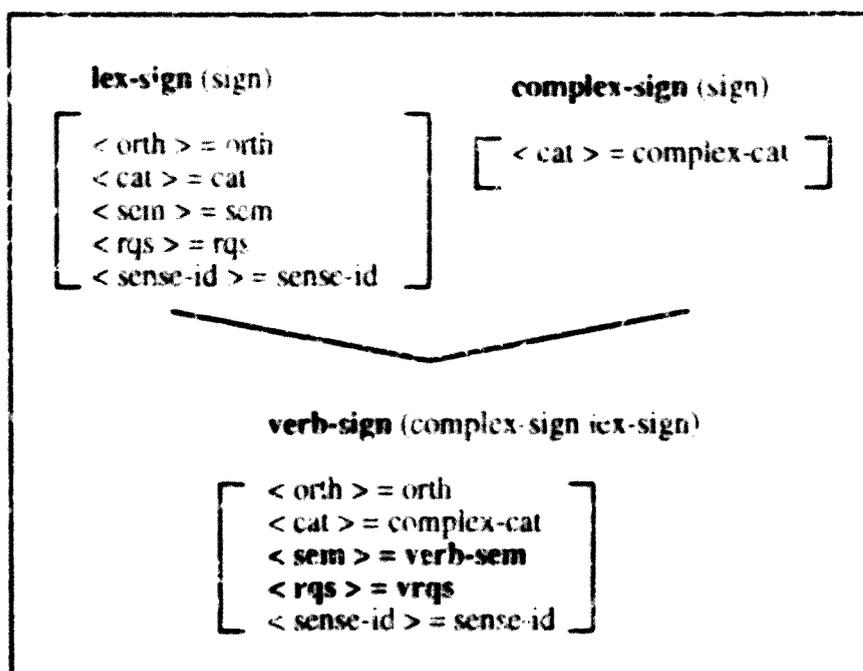


Figura 2.27: Unificación de dos estructuras de rasgos

La unificación de la información contenida en 'lex-sign' y 'complex-sign' en el tipo más específico (o informativo) 'verb-sign' es posible porque sus rasgos no son contradictorios. El tipo 'verb-sign' hereda los rasgos de 'lex-sign' y los de 'complex-sign'. Nótese que el valor de <cat> en 'verb-sign' es el heredado de 'complex-sign' y no de 'lex-sign', esto se debe a que el valor 'complex-cat' es más específico que 'cat'. En cuanto a los valores <verb-sem> y <vrqs> son los propios de 'verb-sign'.

4.1.3. Herencia

La herencia es una operación de inferencia, que permite obtener conocimiento que no se ha declarado de forma explícita y que, además, posibilita la propagación de la información por la jerarquía de tipos.

La Base de Conocimiento Léxico admite dos tipos concretos de herencia: la herencia monotónica múltiple ortogonal y la herencia por defecto.

La herencia monotónica múltiple ortogonal es la que permite la transferencia, de manera global, de la información especificada en un (o más) tipo(s) superior(es) de la jerarquía a un subtipo (es decir a un tipo inferior subsumido por aquél (o aquéllos)). Cuando la herencia de la información procede de dos o más tipos, éstos no pueden contener información incompatible (Véase figura 2.28).

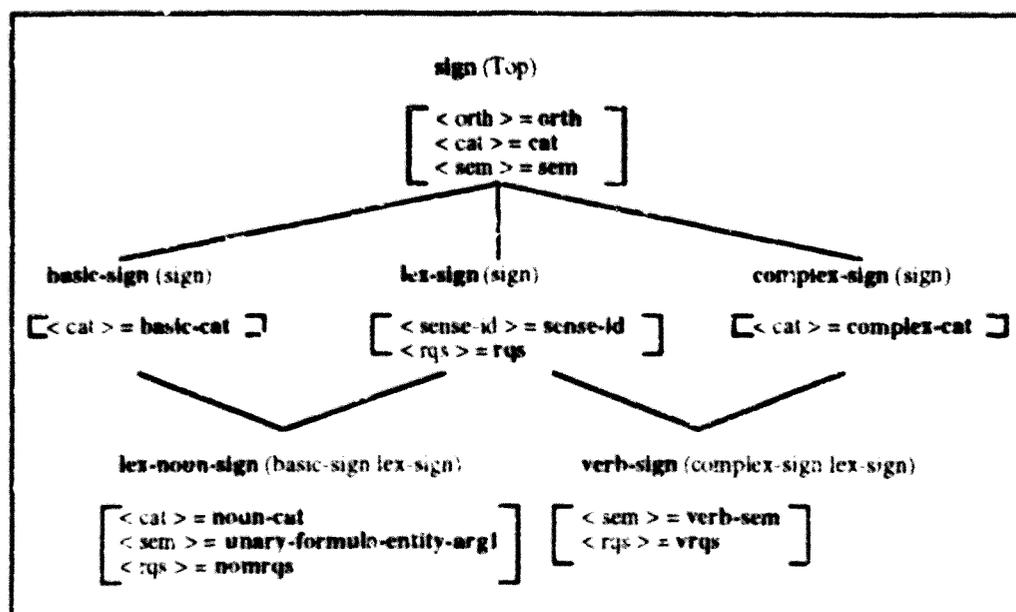


Figura 2.28: Herencia monoiónica múltiple ortogonal

En la figura 2.28, los tipos 'basic-sign', 'lex-sign' y 'complex-sign' heredan monotónicamente los rasgos de un único tipo, en el ejemplo los incluidos en 'sign', mientras que los tipos 'lex-noun-sign' y 'verb-sign' heredan monotónicamente sus rasgos de múltiples tipos, es decir de 'lex-sign' y 'basic-sign' en el caso de 'lex-noun-sign', y de 'lex-sign' y 'complex-sign' en el caso de 'verb-sign'.

En cuanto a la herencia por defecto, es la que permite, dado un valor en un tipo, variar dicho valor en el subtipo subsumido. Este tipo de herencia sólo se aplica en las entradas léxicas.

"La herencia por defecto se formaliza en términos de unificación por defecto de estructuras de rasgos ordenadas por una jerarquía de herencia" (Copestake et al. 1991a).

Los distintos tipos de herencia permiten representar la información léxica de una manera muy estructurada, en la que la información compartida por más de un tipo sólo es necesaria expresarla una única vez y, por lo tanto, evita la presencia de información redundante.

Capítulo 3: El signo léxico: el verbo en la Base de Conocimiento

1. Introducción

En este capítulo se expone la entrada léxica verbal propuesta en esta investigación y, en especial, la manera en que la información contenida en ella se representa en la Base de Conocimiento, es decir la manera en que se organiza toda la información léxica: la descripción, básicamente, de los tipos lingüísticos morfo-sintácticos (Véase apartado 2) y semánticos (Véase apartado 3) desarrollados en la jerarquía de la BCL y la clasificación verbal que se propone (Véase apartado 4).

Siguiendo las propuestas teóricas de los modelos gramaticales presentados en el primer capítulo¹, y en particular la *Gramática de Estructura Sintagmática Regida por el Núcleo* (Pollard & Sag 1987 y 1992), las entradas léxicas se representan como signos léxicos en los que la información fonológica², morfo-sintáctica y semántica se especifica al mismo nivel y de forma integrada.

¹ Véase capítulo 1, apartado 2.

² En esta investigación no se trata la información de tipo fonológico, que es substituida por la ortografía.

En la BCL, se distingue entre signos léxicos (palabras) y signos sintagmáticos (sintagmas y oraciones), pero en la presente investigación sólo se hace referencia a la representación de los signos léxicos y, en concreto, a los signos verbales^{3,4} (Véase figura 3.1).

lex-sign (sign)	verb-sign (lex-sign complex-sign)
<orth> = orth	<orth> = orth
<cat> = cat	<cat> = complex-cat
<sem> = sem	<sem> = verb-sem
<rqs> = rqs	<rqs> = vrqs
<sense-id> = sense-id.	<sense-id> = sense-id.

Figura 3.1: *Signo léxico y Signo verbal*⁵.

En este sentido, el tipo de información que incluyen los signos léxicos es, básicamente, de tipo morfo-sintáctico y semántico. La información de tipo ortográfico (representada mediante el rasgo <orth>) se reduce a una secuencia ortográfica de la palabra (o sintagma).

La información morfo-sintáctica (codificada en <cat>) (Sanfilippo 1990 y 1991) indica el tipo de categoría léxica o sintáctica ("Part of Speech"), las propiedades morfológicas inherentes y, en el caso de las entradas verbales, el número de argumentos subcategorizados, así como las alternancias de diátesis que admite el verbo en cuestión (véase capítulo 5).

En cuanto a la información de tipo semántico incluida en las entradas léxicas verbales, se ha diferenciado entre el nivel de semántica lógica (representada en <sem>) y el nivel de semántica compositiva (codificada en <vrqs>).

³ Sanfilippo, A. (1991)

⁴ Los trabajos de I. Castellón (1992) y A. Copestake et al. (1992), presentan un tratamiento completo de los signos nominales.

⁵ En los signos de la figura 3.1, se incluyen todos los rasgos ya sean propios o heredados de tipos superiores.

En el nivel de representación de la semántica lógica (Sanfilippo 1990 y 1991), se especifica la naturaleza de los argumentos subcategorizados mediante proto-roles temáticos, es decir la Estructura Argumental del verbo, las Restricciones de Selección así como el tipo de verbo según su comportamiento aspectual. En el nivel de representación de la semántica compositiva, en cambio, se especifica la información sobre la clase semántica a la que pertenece el verbo e información relativa a los componentes semánticos que integran su significado léxico, así como el tipo de saturación léxica que éstos presentan (Véase capítulo 4).

El rasgo <sense-id> informa sobre el diccionario fuente (p.e: VOX, GRZ, LDOCE, VANDALE, etc.), la lengua (p.e: español, holandés, inglés, italiano, etc.), la forma gráfica, el número de homógrafo, el número de acepción y el número de entrada en la BDL, de una palabra determinada.

En esta aproximación —al igual que en la *Gramática de Estructura Sintagmática Regida por el Núcleo* (Pollard & Sag 1987 y 1992)—, los verbos se consideran elementos nucleares (o "Head") de sus oraciones. En este sentido, los verbos son los que proporcionan la estructura relacional y semántica a sus oraciones, es decir que las oraciones son proyección de la información léxica especificada en las entradas verbales.

A continuación, se detalla la información especificada en los rasgos <cat> y <sem> de las entradas léxicas verbales, basada en la propuesta de A. Sanfilippo (1990, 1991). En cuanto a la información incluida en la <vrqs>, propuesta en esta investigación, se expone de forma exhaustiva en el siguiente capítulo*.

* Véase capítulo 4.

2. La información morfológico-sintáctica

En el rasgo <cat> de los signos léxicos verbales, se incluye tanto la información morfológica (codificada en el rasgo <m-feats>) como la propiamente sintáctica.

En el rasgo <sent-m-feats>, que es un subtipo de <m-feats>, se describen las propiedades morfológicas específicas de los verbos (Véase figura 3.2).

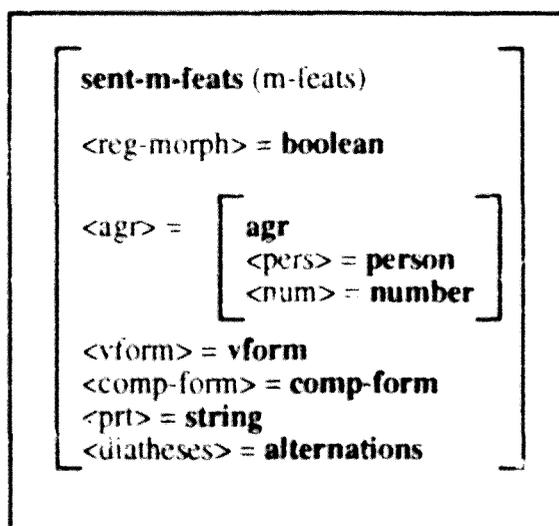


Figura 3.2: '*sent-m-feats*'.

El rasgo <reg-morph> indica si se trata de un verbo irregular o regular. En el rasgo de concordancia <agr>, se especifica el número y la persona. En <v-form>, se codifica la información relativa al tiempo y modo verbal, así como también la forma verbal (infinitivo, gerundio, pasiva, pronominal, etc.). El rasgo <comp-form> indica si se trata de un verbo que rige o no una oración y, en el caso en que la admita, especifica el tipo de complementador que requiere. Las preposiciones regidas se indican en el rasgo <prt> y, por último, las posibles alternancias de diátesis (p.e: Pasiva, Pasiva Refleja, Objeto Indefinido, etc.) se especifican en el rasgo

<diatheses>⁷.

El verbo como categoría léxica nuclear ("Head") impone restricciones sobre el número y tipo de complementos (o signos) con los cuales se combina de forma característica con el fin de ser completo. Este sería el tipo de información codificada en <cat>, la información relativa a los argumentos subcategorizados.

Para la representación de dicha información sintáctica se ha utilizado una aproximación basada en la *Gramática Categorial de Unificación* ("Unification Categorical Grammar" (UCG)) desarrollada en H. Zeevat et al. (1987); y H. Zeevat (1988).

En la UCG, las categorías se definen de la siguiente manera:

- a) Si A es una categoría básica, A es una categoría.
- b) Si A y B son categorías, A/B y A\B son categorías.

Siguiendo la propuesta de estas gramáticas, se define un conjunto de categorías básicas —nombre ('n'), sintagma nominal ('sn') y frase ('sent')— a partir de las cuales se describe el conjunto de categorías complejas —verbo, adjetivo, preposición, etc.— mediante una aplicación funcional (Véase figura 3.3).

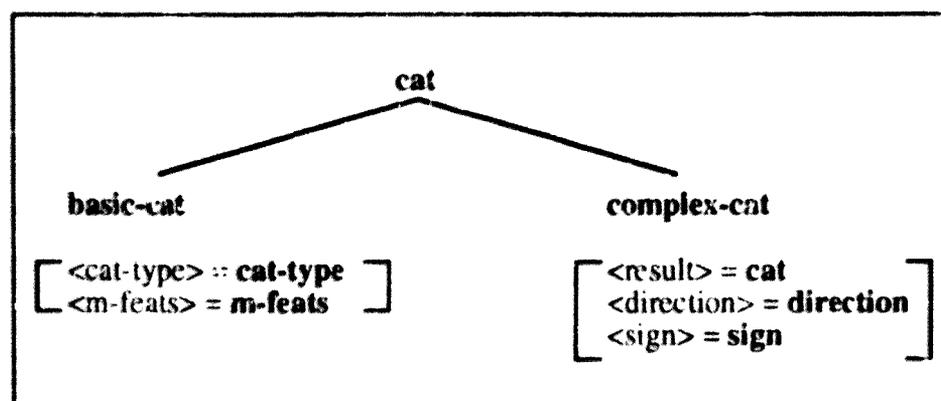


Figura 3.3: 'cat'

⁷ En el capítulo 5, se expone exhaustivamente la información referente a las alternancias de diátesis desarrolladas y la manera en que se han representado en la BCL.

En las categorías básicas (<basic-cat>), se introduce el tipo de categoría (<cat-type>), es decir 'n', 'np' y 'sent', y las propiedades morfológicas (<m-feats>).

Las categorías complejas se definen recursivamente mediante tres rasgos: 'result', 'direction' y 'active'. Esta es la representación en estructura de rasgos de la anotación más familiar mediante el operador "/" o "\" adoptado en la UCG.

El valor de 'active' es un signo, que indica el elemento que se activa para dar como resultado la categoría expresada en 'result' (B en la anotación de UCG, A/B o B\A).

El rasgo 'direction' codifica el orden de combinación relativa de la parte activa, es decir hacia adelante ('forward') (el operador "/" en la anotación de la UCG, A/B) o hacia atrás ('backward') (el operador "\" en la anotación de la UCG, A\B).

El valor de 'result', que puede ser una categoría básica o compleja, es la categoría resultante de la aplicación (A en la anotación de UCG, A/B o B\A).

Según esta propuesta, un verbo intransitivo, por ejemplo, se define del siguiente modo:

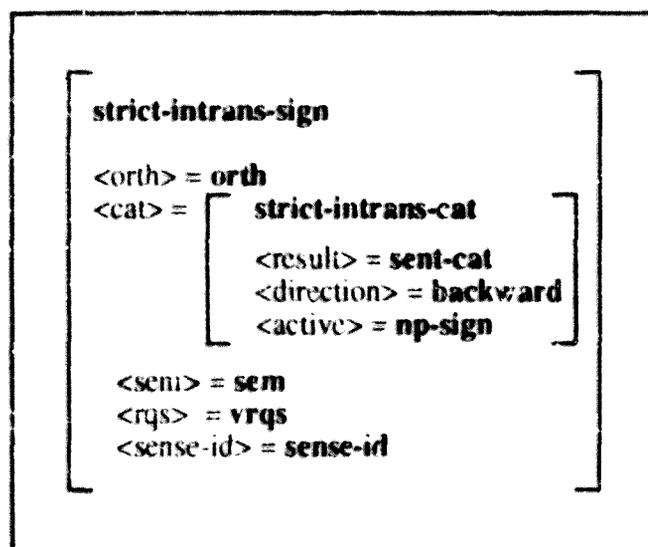


Figura 3.4: 'strict-intrans-cat'

La categoría intransitiva 'strict-intrans-cat' indica que se trata de un verbo que busca un SN (<active> = np-sign), para dar como resultado una oración (<result> = sent), mediante la aplicación funcional hacia atrás (<direction> = backward).

Todos los verbos de la Base de Conocimiento se representan como categorías complejas, porque se considera que como mínimo subcategorizan el argumento sujeto.

Los criterios que han determinado la clasificación sintáctica (y como veremos más adelante también la semántica⁸) de los verbos en la BCL han sido el número y tipo de argumentos que subcategorizan.

Según el número de argumentos, los verbos se clasifican, fundamentalmente, en dos clases: los verbos con un argumento ('strict-intrans-cat'), y los verbos con dos o más argumentos ('x-sign-x-cat').

El tipo 'x-sign-x-cat' se subclasifica en diversas clases según el tipo de argumento que subcategoriza el verbo —'strict-trans-cat', 'ditrans-cat', 'obl-intrans-cat', 'obl-trans-cat', 'comp-cat', 'xcomp-cat', etc.—, es decir según si el argumento subcategorizado es un sintagma nominal, un complemento preposicional, una completiva, una frase de infinitivo, una frase de gerundio, un adjetivo, etc. (Véase la figura 3.5).

⁸ Véase apartado 3 de este capítulo.

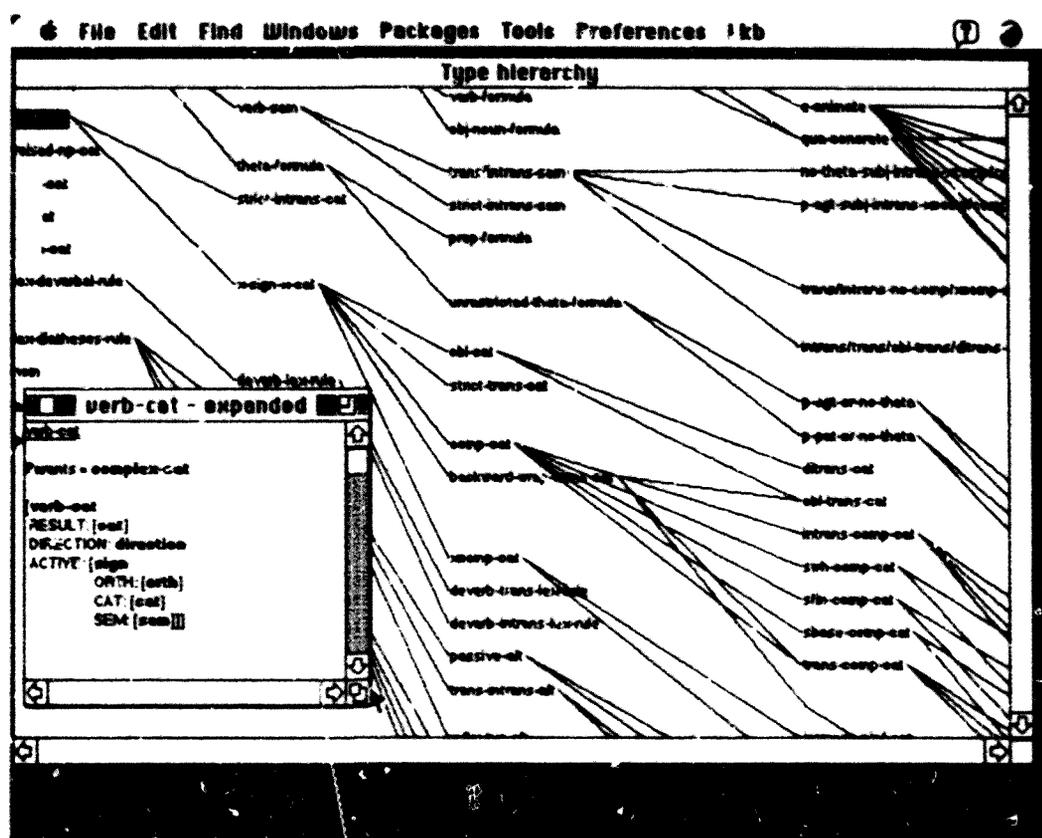


Figura 3.5: Clasificación verbal según la categoría

En consecuencia, la categoría de un verbo transitivo, cuyos argumentos se realizan sintagmáticamente mediante dos SNs (p.e: *Laura ha comprado chocolate*), es 'strict-trans-cat', mientras que un verbo transitivo que subcategoriza una oración (p.e: *Quiere que vengas*) se codifica como 'comp-cat'. Según el tipo de oración subcategorizada, se pueden definir distintos subtipos de 'comp-cat': 'sfin-comp-cat' si se trata de una completiva (p.e: *Dice que Laura no vendrá*); 'swh-comp-cat' si se trata de una interrogativa indirecta (p.e: *Pregunta si el tren llega con retraso*), etc.

La categoría 'strict-intrans-cat' corresponde a los verbos intransitivos (p.e: *Laura duerme*) y 'obl-cat' a los verbos que subcategorizan un SP (p.e: *Se peleó con Laura*), etc.

En la categoría 'xcomp-cat', se incluyen los verbos con complementos predicativos: 'vpinf-control-cat' si se trata de verbos con oraciones infinitivas (p.e:

Laura quiere salir); 'vping-control-cat' si se trata de oraciones de gerundio (p.e.: *Salieron cantando*), etc.

3. La información semántica

A continuación, se presenta la manera en que se ha codificado la información semántica correspondiente a la semántica lógica, es decir el nivel de representación semántica que describe la Estructura Argumental, la Estructura Eventual y las Restricciones de Selección.

Este tipo de información se representa en el rasgo <sem> de la BCL (Sanfilippo 1990, 1991). La semántica compositiva, codificada en <vrqs>, se presenta detalladamente en el próximo capítulo².

La estrategia básica adoptada en la BCL es la propuesta por A. Sanfilippo (1990, 1991), la cual consiste en la combinación de un sistema de interpretación semántica neo-Davidsoniana ((Parsons 1980), (Carlson 1984), (Dowty 1989)), aumentada con la teoría de relaciones gramaticales y de la selección argumental propuesta en (Dowty 1988).

Siguiendo la aproximación neo-Davidsoniana de la semántica de los verbos, éstos son tratados como propiedades de eventualidades y los roles temáticos (es decir la interpretación semántica de los argumentos del verbo) como relaciones entre eventualidades y individuos. Las relaciones temáticas, por lo tanto, proporcionan un estado indispensable de interpretación semántica para combinar los significados de los verbos y de los sintagmas nominales en significados de frase.

Las relaciones temáticas se definen mediante los denominados proto-roles temáticos (Dowty 1988). Los proto-roles se conciben como "clusters" o conjuntos de propiedades semánticas que permiten definir un Agente prototipo (1) y un Paciente prototipo (2), en función de las propiedades características predominantes. Los proto-roles están, además, relacionados con la selección de argumentos mediante el Principio de Selección Argumental (3).

² Véase capítulo 4.

(1) Contributing Properties for the Agent Proto-Role:

- a. volition
- b. sentience (and/or perception)
- c. causes event
- d. movement

(2) Contributing Properties for the Patient Proto-Role:

- a. change of state (including coming-to-being, going-out-of-being)
- b. causally affected by event
- c. stationary (relative to movement of Proto-Agent)

(3) Argument Selection Principle:

The argument of a predicate having the greatest number of Proto-Agent properties entailed by the meaning of the predicate will be lexicalized as the subject of the predicate. The argument of a predicate having the greatest number of Proto-Patient properties entailed by the meaning of the predicate will be lexicalized as the direct object of the predicate.

Corollary 1:

If two arguments of a relation have (approximately) equal numbers of entailed proto-agent and proto-patient properties, then either may be lexicalized as the subject (and similarly for objects).

Corollary 2:

With a three-place predicate, the non-subject argument having a greater number of entailed Proto-Patient properties will be lexicalized as the direct object. The non-subject argument having fewer entailed Proto-Patient properties will be lexicalized as an oblique or prepositional object (and if two non-subject arguments have approximately equal entailed Proto-Patient properties, either may be lexicalized as direct object).

En definitiva, el Agente proto-típico de un verbo ('p-ag') es la relación temática asociada al argumento que tiene el mayor número de propiedades de proto-Agente implicadas en el significado del verbo, que se asocia sintácticamente con el sujeto; mientras que el Paciente proto-típico, en cambio, es la relación temática asociada con el

argumento de un verbo transitivo que tiene el mayor número de propiedades de proto-Paciente y que se corresponde, normalmente, con el objeto sintáctico. El resto de argumentos verbales se asocian con el proto-rol Oblícuo (o preposicional).

El nivel de semántica lógica se representa en la BCL mediante el rasgo <verb-sem>, que es una fórmula lógica binaria (Véase figura 3.6).

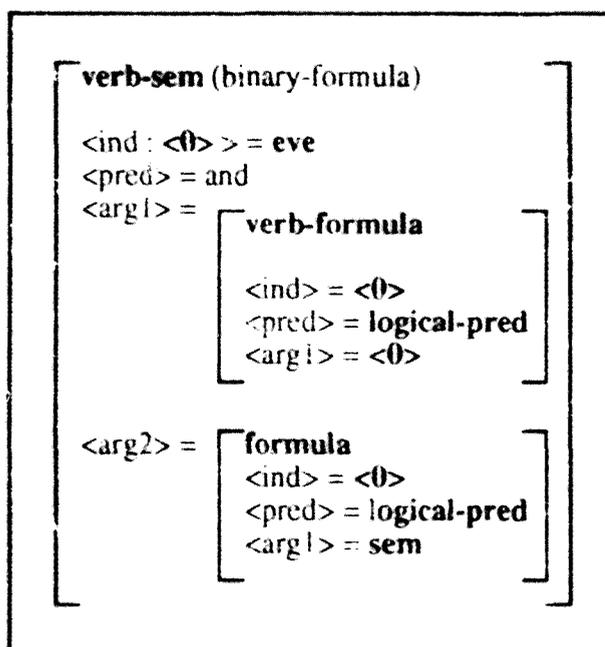


Figura 3.6: 'verb-sem'

El primer argumento (<arg1> = verb-formula) es una fórmula unaria, que describe el predicado como un evento. Esta es la información relativa a la Estructura Eventual.

El segundo argumento (<arg2> = formula) es una fórmula, en la que se describen los argumentos verbales mediante los proto-roles temáticos (p.e: 'p-ag', 'p-pat', 'p-obl', 'p-ag-cause', 'p-pat-affected', etc.) y las Restricciones de Selección, que especifican las propiedades semánticas que deben cumplir dichos proto-roles ('e-animado', 'e-inanimado', 'e-humano', 'e-animal', 'e-concreto', etc.). En consecuencia, esta fórmula refleja la Estructura Argumental del verbo.

El índice ontológico (<ind>) caracteriza el predicado verbal como un evento o eventualidad¹⁰ ('eve'). Siguiendo la clasificación de Z. Vendler (1967), D. Dowty (1988) y otros, los predicados pueden denotar estados ('state'), acciones ('dyn-eve') o procesos ('proc') (Véase figura 3.7).

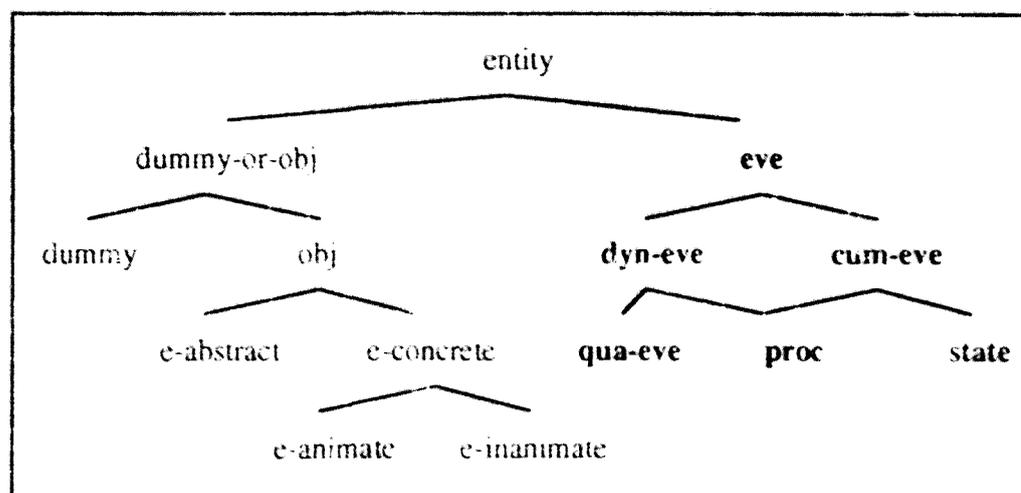


Figura 3.7: Jerarquía de tipos: objetos y eventos

Los estados denotan eventos en los que no se produce ningún cambio, se trata de una acción homogénea en la que no se hace referencia al inicio o final de un período (p.e: *La niña está enferma*). Los procesos denotan, en general, una actividad de duración indefinida, que ocurre de forma homogénea y en la que cada parte del proceso es de la misma naturaleza que el total (p.e: *La niña anda*). En el caso de las acciones, se distingue entre las de tipo "accomplishment", que denotan una actividad que tiene una duración o culminación lógica (p.e: *La niña anduvo hasta el colegio*), y las de tipo "achievement", que son las que ocurren de forma instantánea, en un momento determinado (p.e: *Murió a las tres*). Los "accomplishment" y "achievement" se codifican con el mismo rasgo ('qua-eve') en la BCL.

La clasificación semántica se ha realizado teniendo en cuenta, en primer lugar, el número de argumentos requeridos por el verbo y, en segundo lugar, el tipo de argumento.

¹⁰ Traducción del término inglés "event".

La primera distinción se establece entre los verbos con un argumento, cuyo rol semántico es un proto-Agente ('strict-intrans-sem') (p.e: *Laura duerme*) y los verbos de más de un argumento ('trans/intrans-sem'). Los verbos, que además del argumento proto-Agente, toman un segundo argumento que no es una cláusula u oración se definen mediante el tipo 'trans/intrans-no-comp/xcomp-sem'. Este tipo subsume, por ejemplo, los tipos semánticos que caracterizan los verbos estrictamente transitivos ('strict-trans-sem') (p.e: *Laura corta la carne*), en los que el segundo argumento es un proto-Paciente, y los intransitivos con un objeto oblicuo, es decir con un proto-Oblicuo ('intrans-obl-sem') (p.e: *Laura piensa en él constantemente*).

Otro subtipo de 'trans/intrans-sem', por ejemplo, es el tipo semántico 'intrans/trans/obl-trans/ditrans-sem', que describe los verbos que toman tres argumentos. Este tipo se subclasifica, a su vez, en nuevos tipos semánticos según los argumentos requeridos por el verbo, por ejemplo, 'obl-trans/ditrans-sem' que caracteriza verbos con un argumento proto-Paciente y un proto-Oblicuo, además del proto-Agente (p.e: *Laura da las pastillas al enfermo*), etc. (Véase figura 3.8)^{11,12}.

¹¹ En los trabajos de A. Saufilippo (1990 y 1991) se presenta de forma detallada la clasificación semántica representada en <sem>.

¹² En el Apéndice 3, se incluye toda la jerarquía de tipos desarrollada en la BCL.

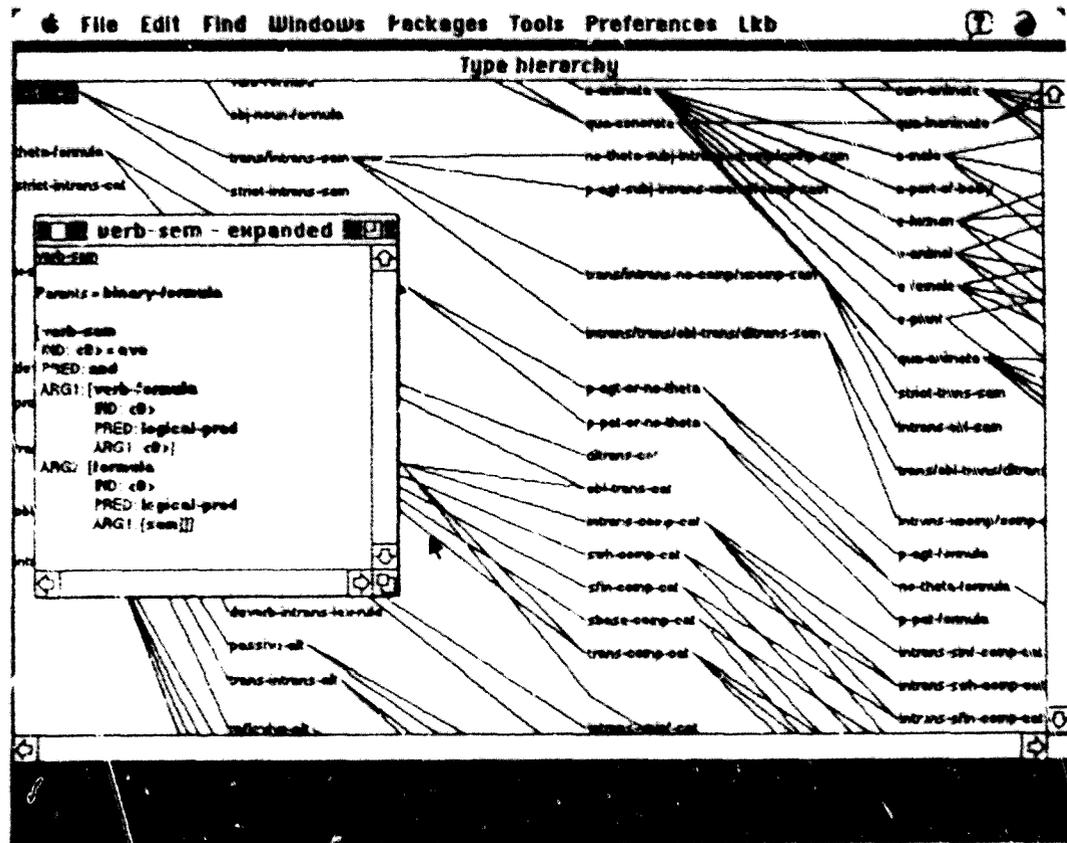


Figura 3.8: Clasificación verbal según la semántica

4. Clasificación del signo verbal

Una vez descritos los tipos sintácticos y semánticos verbales, a continuación se presenta la clasificación de los signos verbales en función de la clase de información morfo-sintáctica y semántica especificada.

En los signos verbales¹³ ('verb-sign'), se integra la información sintáctica y semántica mediante la coindexación de los argumentos subcategorizados que se especifican en <cat> y los argumentos de la fórmula lógica descritos en <sem>. De esta manera, se asegura que todo argumento verbal tiene asociado un (y solamente uno) proto-rol semántico.

Teniendo en cuenta todas estas especificaciones morfo-sintácticas y semánticas de los argumentos verbales, expuestas en los apartados anteriores, la organización de los verbos en la BCL establece, básicamente, tres grandes clases de signos verbales: a) los signos intransitivos ('strict-intrans-sign'), los signos de dos complementos ('2-complements-verb-sign') y c) los signos de tres complementos ('3-complements-verb-sign') (Véase figura 3.9).

¹³ En el Apéndice 5, se da una muestra de entradas léxicas verbales.

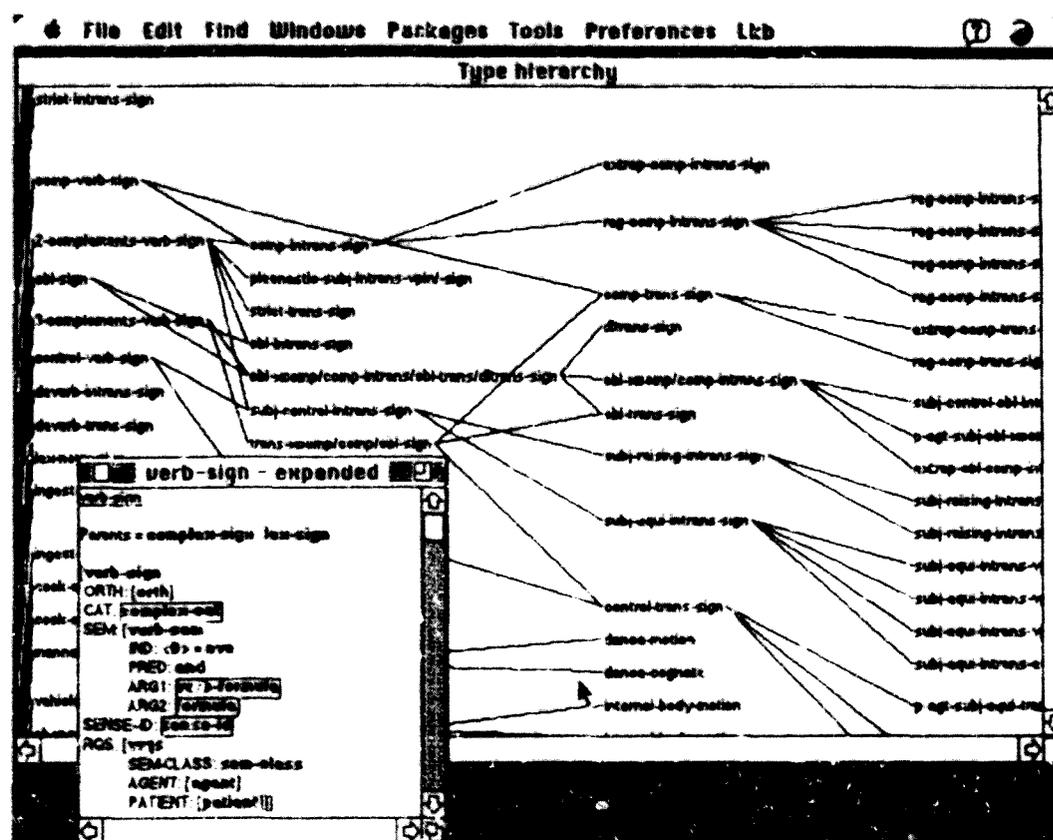


Figura 3.9: Clasificación de los signos verbales

Los tipos '2-complements-verb-sign' y '3-complements-verb-sign' se subclasifican en tipos posteriores según la categoría sintáctica (<cat>) y la fórmula semántica (<sem>) del verbo.

A continuación, y a modo de ejemplo, se explican brevemente aquellos signos verbales que se utilizan más en la presente investigación, el resto se exponen de forma exhaustiva en los trabajos de A. Sanfilippo (1990 y 1991).

El signo intransitivo 'strict-intrans-sign' se caracteriza por tener una categoría del tipo 'strict-intrans-cat' y una fórmula lógica 'strict-intrans-sem', en la que se especifica que el argumento proto-Agente se corresponde con el signo nominal activo en <cat> (Véase figura 3.10).

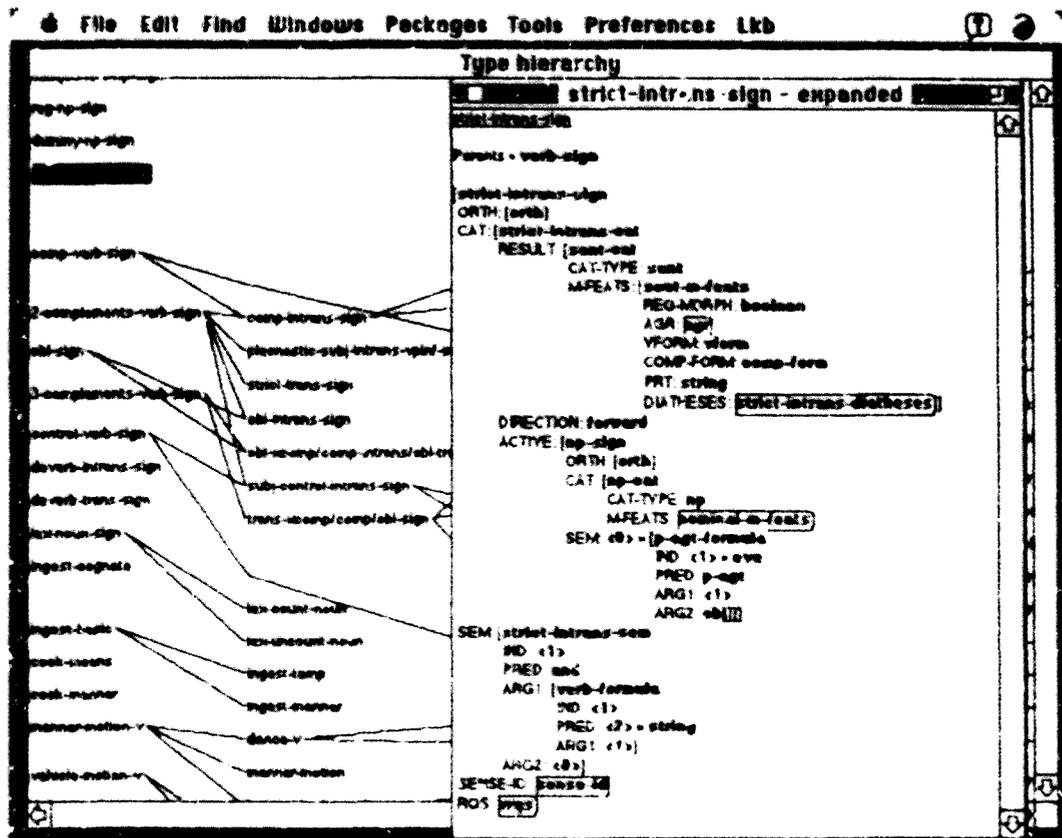


Figura 3.10: 'strict-intr-ns-sign'

El signo transitivo 'strict-trans-sign' es un subtipo de '2-complements-verb-sign', en el que la categoría es 'strict-trans-cat' y la fórmula lógica es 'strict-trans-sem' (Véase figura 3.11).

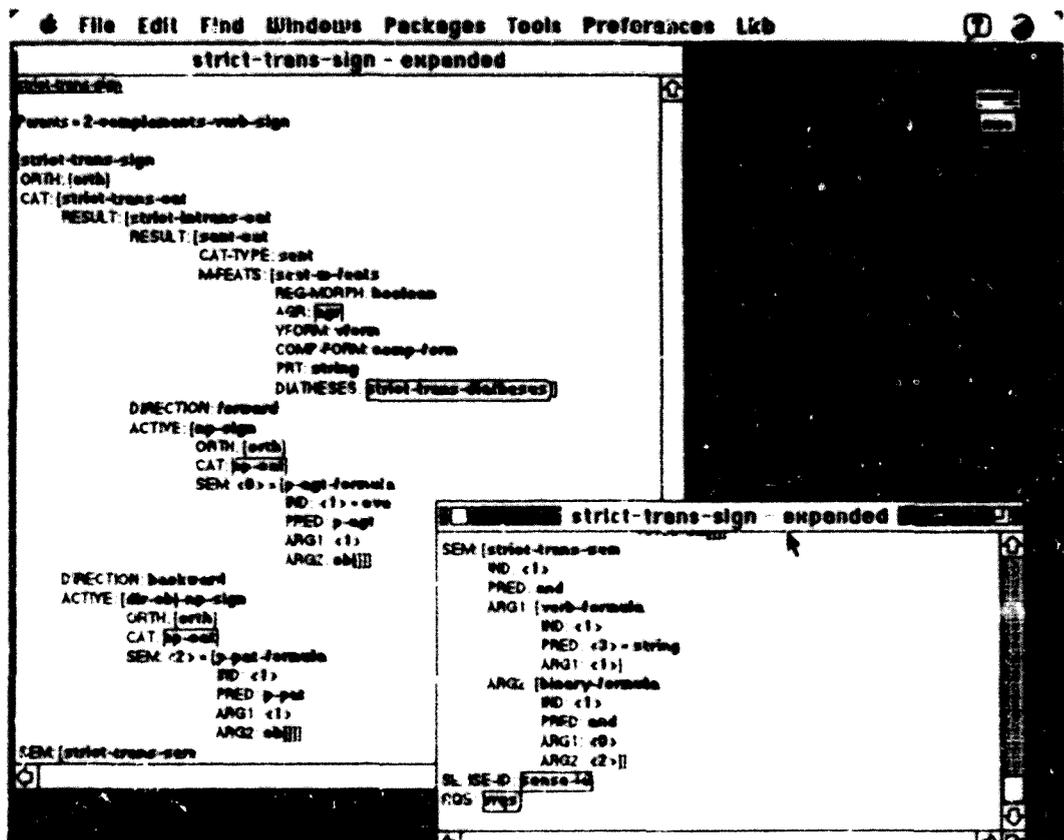


Figura 3.11: 'strict-trans-sign'

En la figura 3.11, se puede observar también la coindexación de los argumentos especificados en <sem> con los signos nominales expresados en <cat>. El proto-Agente se asocia con el signo nominal que se realiza como sujeto ('np-sign') y el proto-Paciente con el signo nominal que funciona como objeto directo ('dir-obj-np-sign').

El signo intransitivo oblicuo 'obl-intrans-sign', en cambio, es un subtipo de '2-complements-verb-sign' y de 'obl-sign'. La categoría es 'obl-cat' y la fórmula lógica 'intrans-obl-sem' (Véase figura 3.12). En la presente investigación, se caracterizan los verbos que rigen preposición (p.e. *dirigirse a*, *pensar en*, etc.) como signos intransitivos oblicuos.

The screenshot displays a software application window titled "oblique-intrans-sign - expanded". The main area contains the following text:

```

parents = obl-sign 2-complements-verb-sign
obl-intrans-sign
ORTH: [orth]
CAT: [obl-ent]
RESULT: [strict-locus-ent]
RESULT: [past-ent]
CAT-TYPE: past
MFEATS: [past-a-features]
REG-MORPH: be-locus
AGR: [p]
YFORM: [verb]
COMP-FORM: comp-form
PRT: string
DIATHESES: [oblique-intrans-sign]

DIRECTION: forward
ACTIVE: [sp-sign]
ORTH: [orth]
CAT: [sp-obl]
SEM: <0> = [p-agt-formula]
      ID: <1> = eve
      PRED: p-agt
      ARG1: <1>
      ARG2: ob[]

DIRECTION: backward
ACTIVE: [oblique-ent-sp-sign]
ORTH: [orth]
CAT: [sp-ent]
CAT-TYPE: sp
MFEATS: [past-a-features]
SEM: <2> = [prep-formula]
      ID: <1>
      PRED: string
      ARG1: <1>
      ARG2: ob[]

SEM: [intrans-obl-sp]
    
```

The "Type hierarchy" window on the right shows a network of relationships between various linguistic types, with "obl-intrans-sign" highlighted in the center.

The smaller window at the bottom right, titled "oblique-intrans-sign - expanded", shows a semantic structure:

```

SEM: [intrans-obl-sem]
ID: <1>
PRED: ent
ARG1: [verb-formula]
      ID: <1>
      PRED: <2> = string
      ARG1: <1>
ARG2: [biary-formula]
      ID: <1>
      PRED: ent
      ARG1: <0>
      ARG2: <2>
SENSE-D: sense-10
RGS: [sp]
    
```

Figura 3.12: 'obl-intrans-sign'

En el Apéndice 3, se muestra la jerarquía de tipos completa desarrollada en la Base de Conocimiento Léxico.